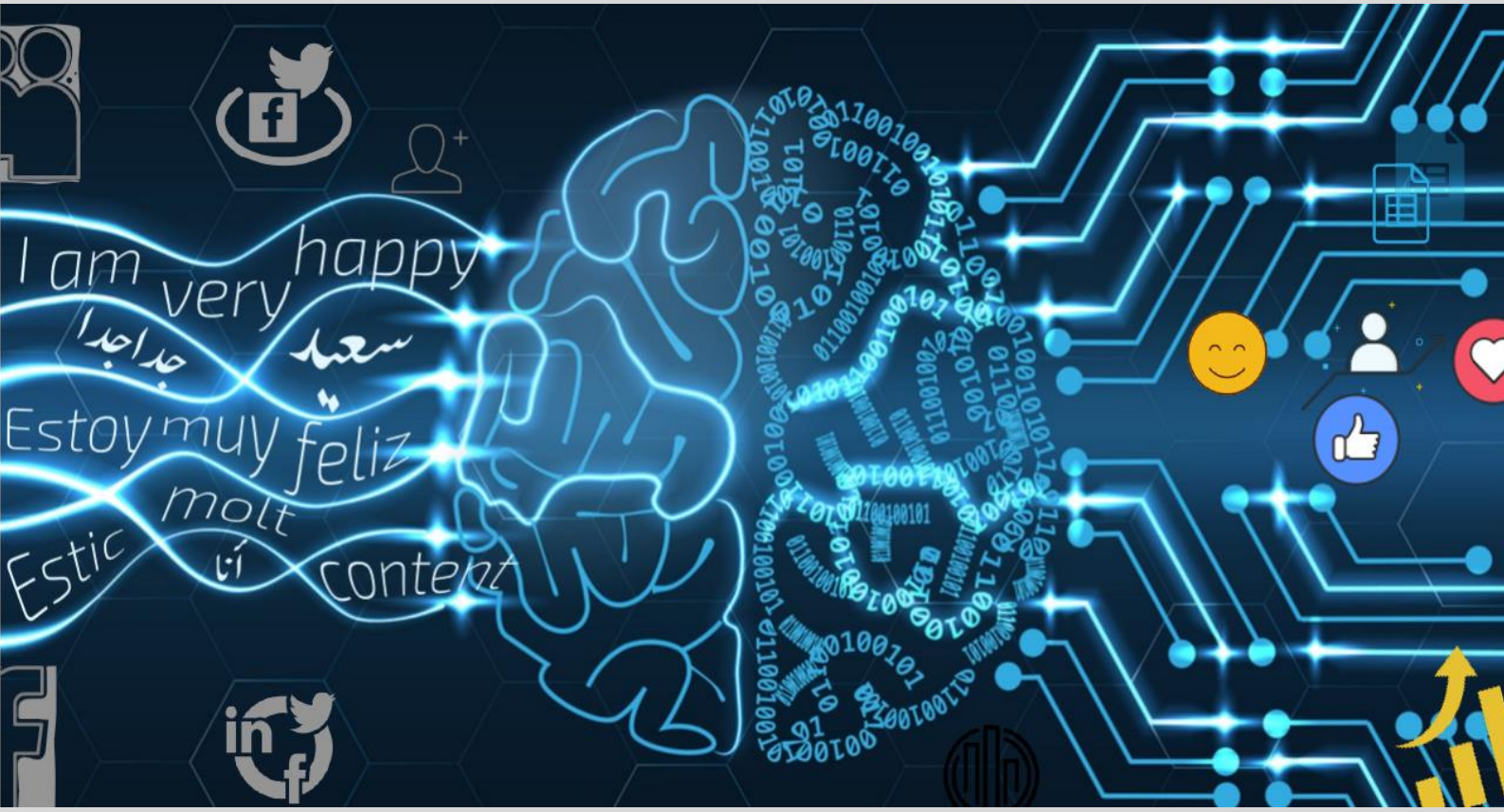
UNIVERSITAT
ROVIRA i VIRGILI

Sentiment Analysis of Textual Content in Social Networks

From Hand-Crafted to Deep Learning-Based Models

Mohammed Jabreel



DOCTORAL THESIS
2020

Sentiment Analysis of Textual Content in Social Networks

From Hand-Crafted to Deep Learning-Based Models

UNIVERSITAT ROVIRA I VIRGILI

Mohammed Jabreel

Supervised by Prof. Dr. Antonio Moreno

Universitat Rovira i Virgili

This dissertation is submitted for the degree of

Doctor of Philosophy

Tarragona 2020

UNIVERSITAT
ROVIRA I VIRGILI
Departament d'Enginyeria Informàtica i Matemàtiques

Av. Paisos Catalans, 27
43007 Tarragona
Tel. +34 977 55 95 95
Fax. +34 977 55 95 97

I STATE that the present study, entitled "Sentiment Analysis of Textual Content in Social Networks, From Hand-Crafted to Deep Learning-Based Models", presented by Mohammed Hamood Abdullah Jabreel, for the award of the degree of Doctor, has been carried out under my supervision at the Departament d'Enginyeria Informàtica i Matemàtiques.

Tarragona, February 2020.

Doctoral Thesis Supervisor,

Prof. Dr. Antonio Moreno Ribas

I would like to dedicate this thesis to
my daughter Lyan and my son Loay.
You are my dream come true, the greatest gift of my life I have ever received.
my parents.
For their endless love, support and encouragement.
my wife Najlaa.
You made my life so much better in so many ways that it is hard to imagine doing this without you.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Mohammed Jabreel

Tarragona 2020

Acknowledgements

This dissertation would not have been possible without the support of many people.

First and foremost, I would like to thank my advisor and role model Antonio Moreno. You provided me with the perfect balance of guidance and freedom. Toni, you helped me see the pros and cons of so many decisions, small and large. I admire your ability to see the nuances in everything. Thank you also for reading countless drafts of (often last-minute) papers. I also want to thank you for your perspective and for helping me pursue and define works with more impact.

I am also thankful to Mohamed Abdel-Nasser for his helpful comments, support and encouragement.

I am grateful for my dear friend Adel Saleh, who I hope will always stay in my life. I would be remiss if I did not record a special 'thank you' for this friend who showed up in the middle of the night to help with various serious situations.

For three years, I was supported by the by URV Research Support Funds and Marti Franques PhD grant for which I want to express my gratitude sincerely.

Abstract

Sentiment analysis, known as opinion mining, is the computational study of people's opinions, sentiments, attitudes, feelings, and emotions. It is considered to be one of the most active research fields in natural language processing, data mining, information retrieval, and Web mining. Sentiment analysis is not just the problem of classifying whether a piece of text expresses a positive or negative sentiment or opinion. It is indeed a much more complex problem than that. It is a generic name and involves many facets and multiple related tasks, each of which has its challenges.

This thesis proposes several advanced methods to automatically analyse textual content shared on social networks and identify people' opinions, emotions and feelings at a different level of analysis and in different languages.

We start by proposing a sentiment analysis system, called SentiRich, based on a set of rich features, including the information extracted from sentiment lexicons and pre-trained word embedding models. Then, we propose an ensemble system based on Convolutional Neural Networks and XGboost regressors to solve an array of sentiment and emotion analysis tasks on Twitter. These tasks range from the typical sentiment analysis tasks, to automatically determining the intensity of an emotion (such as joy, fear, anger, etc.) and the intensity of sentiment (aka valence) of the authors from their tweets. We also propose a novel Deep Learning-based system to address the multiple emotion classification problem on Twitter.

Moreover, we considered the problem of target-dependent sentiment analysis. For this purpose, we propose a Deep Learning-based system that identifies and extracts the target of the tweets.

While some languages, such as English, have a vast array of resources to enable sentiment analysis, most low-resource languages lack them. So, we utilise the Cross-lingual Sentiment Analysis technique to develop a novel, multi-lingual and Deep Learning-based system for low resource languages.

We propose to combine Multi-Criteria Decision Aid and sentiment analysis to develop a system that gives users the ability to exploit reviews alongside their preferences in the process of alternatives ranking.

Finally, we applied the developed systems to the field of communication of destination brands through social networks. To this end, we collected tweets of local people, visitors, and official brand destination offices from different tourist destinations and analysed the opinions and the emotions shared in these tweets.

Overall, the methods proposed in this thesis improve the performance of the state-of-the-art approaches and show exciting findings.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In the last decade, the World Wide Web (WWW) has become one of the essential sources of information. People share their views, opinions, feelings and experiences online by writing blogs, posting comments on a microblogging service (e.g., Twitter), using a social networking service (e.g., Facebook), publishing a product review, or commenting in discussion forums and other types of social media. Such social media have adapted the role of users to be not only content consumers but also producers. Consequently, the web offers a large amount of public discourse and accessible content. Figure 1 shows a screenshot from the Social Media Statistics website, which shows some real-time statistics about the volume of information generated on the web. In about 60 seconds it is notable the vast volume of data generated through the most popular social platforms and applications (e.g. 4M Facebook posts, 400K tweets, 4K reviews on TripAdvisor and 34K posts on Tumblr). Thus, we can figure out the massive amount of content published by people on the web.

The content shared by users comes in different modalities, i.e., textual content, images, and video. The analysis of user-generated content on the web has attracted researchers in different fields, including, among others, the fields of Natural Language Processing (NLP) [105], computer vision [99, 167], tourism, multi-criteria decision aid, smart health, complex networks, and location-based services analysis.

One of the hottest topics in the field of NLP that targets the problem of analysing user-generated content is Sentiment Analysis (SA) or Opinion Mining (OM). It has a wide range of applications in commerce, public health, social welfare, etc. For instance, it can be used in public health for detecting depression, identifying cases of cyber-bullying and tracking of well-being [26, 32]. SA can also be used in public opinion detection about political tendencies [92, 105], brand management [72], stock market monitoring [91] and education

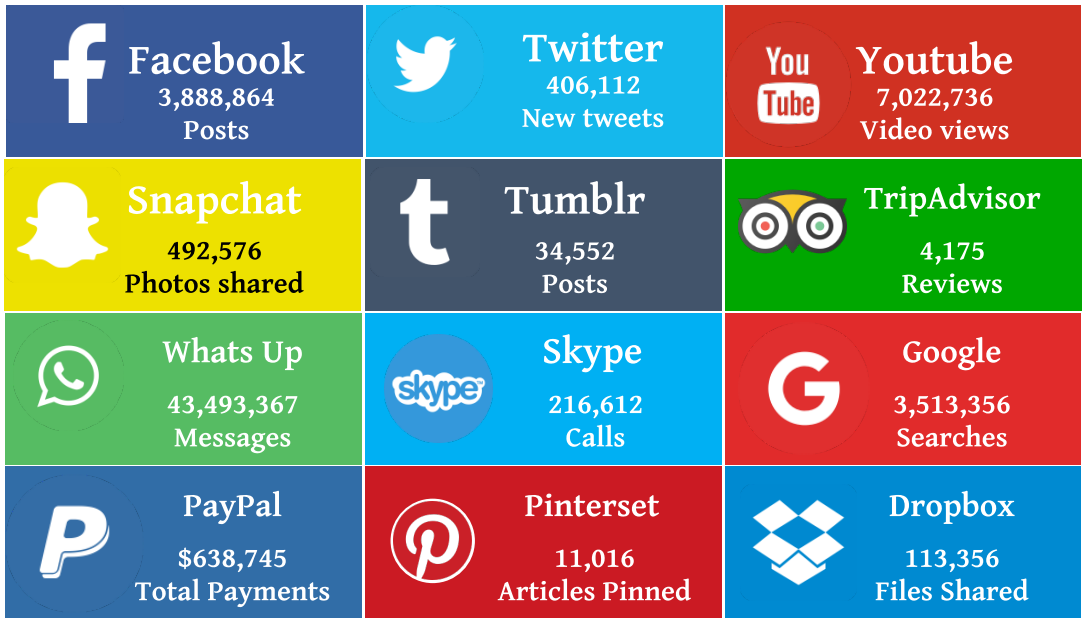


Figure 1.1: The Amount of data generated in Internet in real time.

[153]. Nowadays, several companies make massive investments towards the automatic analysis of the opinions of the customers about their products on the Web, seeking to detect trends that boost sales, consumer satisfaction and corporate profits [74]. Governments also show interest in the analysis of the public opinion on different social and economic issues on the web [164].

SA can be used to refer to many diverse but related problems. Most commonly, it is used to refer to the problem of automatically identifying the polarity of a piece of text, i.e., whether it is positive, negative, or neutral. However, more generally, it refers to determining one's attitude towards a particular target or topic. Here, attitude can mean an evaluative judgment, such as positive or negative, or an emotional or effectual attitude such as frustration, joy, anger, sadness or excitement.

As shown in Figure 1.2 (the review of a restaurant), opinions can come in varying degrees of granularity. We can see the overall polarity of the review in the document, but we can also analyse in more depth each of the sentences. The review is composed of neutral sentences that contain only factual information and do not give any opinion (e.g., the first two sentences) and sentences that contain a mixture of polarities regarding specific entities or characteristics of those entities.

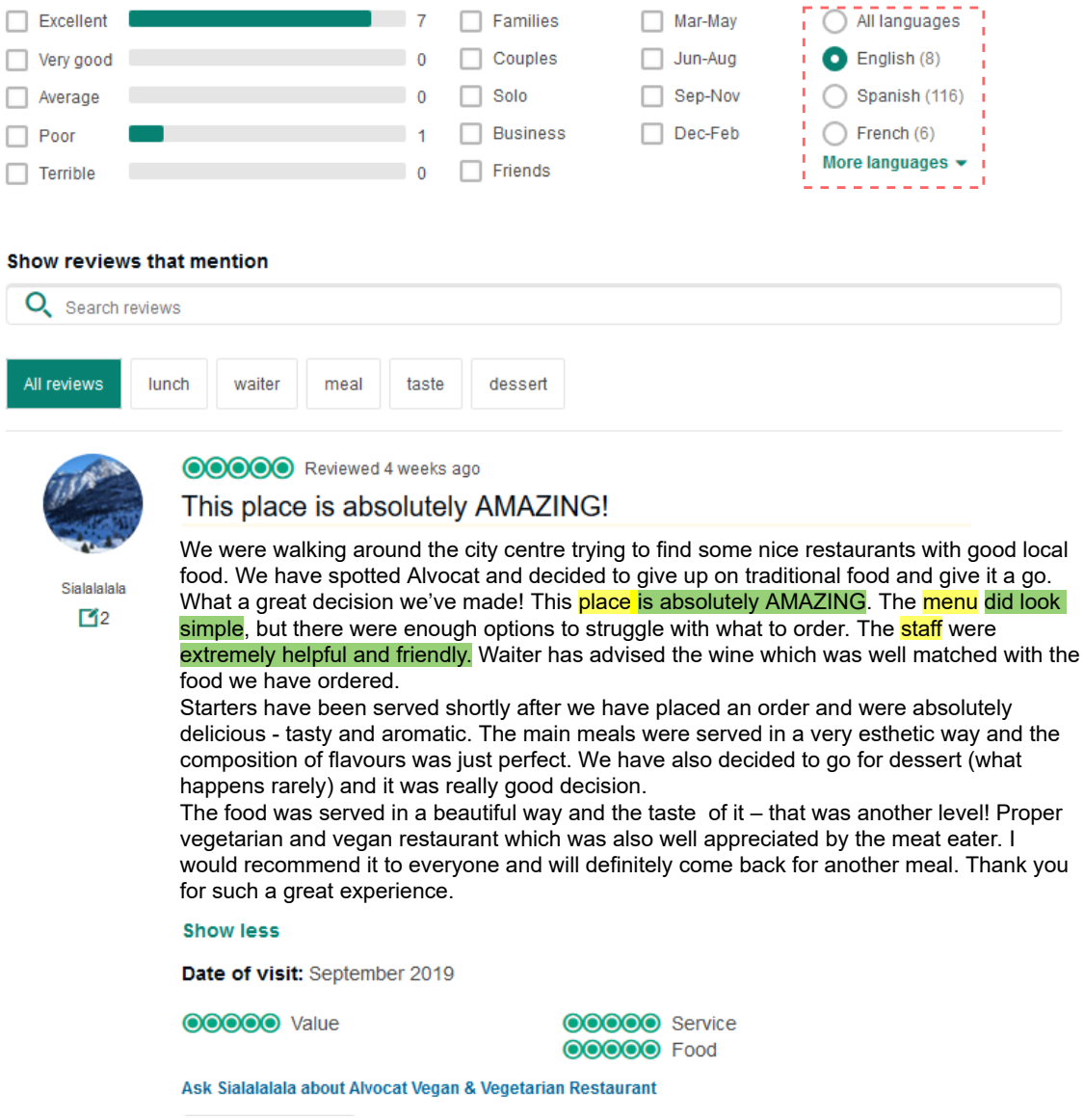


Figure 1.2: The layout of a typical online review.

Hence, sentiment analysis can be performed at different levels (coarse-grained and fine-grained). The former attempts to extract the overall polarity on a document or sentence level. The problem in the fine-grained level of analysis is to identify the sentiment polarity towards a specific aspect in a given text; this level of analysis is called Aspect-Based Sentiment Analysis (ABSA) or targeted sentiment analysis [69]. For example, we can see in the review shown in Figure 1.2 some sentences refer to the *service*, while others talk about the *food*.

Moreover, we can see in the information on the top right of Figure 1.2 that online reviews or textual content may be expressed in different languages. So, sentiment analysis systems needed in all languages. However, building an independent sentiment analysis system for each language is a challenging task as much as building a universal sentiment analysis system for all languages. Most of the existing systems on sentiment analysis rely heavily on a rich set of sentiment resources (such as text corpora with manually annotated sentiment polarity, sentiment lexicons, and word embeddings). However, the distribution of sentiment resources is very imbalanced among languages. Thus, building a sentiment analysis system in low-resource languages requires tremendous human effort to construct such resources, which is a time-consuming and expensive task.

Although opinions are the most common shared content in online social networks and forums, people also tend to share their emotions which are the keys to their feelings and thoughts. Emotion analysis is the task of determining the attitude towards a target or topic. The attitude can be the polarity (positive or negative), or an emotional state such as joy, anger or sadness [101–103].

1.2 Objectives

In this thesis, we aim to develop methods to automatically analyse textual content shared on social networks and identify people' opinions, emotions and feelings at different level of analysis and in different languages. As such, we introduce the following set of goals:

- **To develop efficient sentiment analysis systems based on new features that can be used with traditional Machine Learning (Machine Learning) methods.**

  Towards this objective, we aim to use pre-built resources such as sentiment lexicons and word embeddings to extract new features.

- **To move sentiment analysis beyond sentence-level and polarity-based analysis.**

  Towards this objective, we are interested in:

  - Proposing ensemble systems that combine classical Machine Learning models with Deep Learning to analyse emotions expressed on Twitter.
  - Developing a Deep Learning based model to solve the problem of multi-label emotions classification.
  - Utilising Deep Learning based models to analyse opinions at the aspect level.

- **To move sentiment analysis beyond a single language.**

  We aim to utilise the concept of transfer learning to develop a system that can transfer sentiment knowledge from high resources languages to low resources languages. The final goal of this objective is to obtain a universal sentiment analysis system that works with low resource languages and does not require machine translation.

- **To test the developed systems on real cases of analysis.**

- The idea of this objective is to test our developed systems in real applications and different domains. Hence, towards this objective, we define the following sub-goals:

  - To collect tweets of local people, visitors and official brand destination offices from different tourist destinations and analyse the opinions shared in these tweets.
  - To combine the aspect-based sentiment analysis with multi-criteria decision aid systems to improve the decision-making process.

1.3 Contributions and Thesis Structure

In response to the objectives of this thesis, the main contributions of this PhD thesis are the following:

1. We have developed a sentiment analysis system, called SentiRich, based on a set of rich features, including the information extracted from sentiment lexicons and pre-trained word embedding models. We have used the system to contribute to the international challenge Sentiment Analysis in Twitter task of SemEval-2017. The system was ranked seventh out of 38 systems in the English language. We have developed another version of SentiRich system for the Arabic language, and the system was ranked second out of eight systems. We describe and explain these systems in **Chapter 3**. The results of this study have been published in the following papers:

- Mohammed Jabreel and Antonio Moreno, "SentiRich: Sentiment Analysis of Tweets Based on a Rich Set of Features", *Artificial Intelligence Research and Development*, vol. 288, pp. 137 - 146, IOS Press, 2016.
- Mohammed Jabreel and Antonio Moreno, "SiTAKA at SemEval-2017 Task 4: Sentiment Analysis in Twitter Based on a Rich Set of Features", *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Association for Computational Linguistics, pp. 694–699, 2017.

2. We have proposed a system based on Convolutional Neural Networks and XGboost regressors to solve an array of sentiment and emotion analysis tasks on Twitter. These tasks range from the typical sentiment analysis tasks, to automatically determining the intensity of an emotion (such as joy, fear, anger, etc.) and the intensity of sentiment (aka valence) of the authors from their tweets. We have used the system to contribute to the international challenge SemEval-2018 Task 1: Affect in Tweets. In the analysis of emotions in tweets written in Arabic, the system obtained the first or second position –out of 14 participants- in four subtasks (numerical/categorical emotion intensity and numerical/categorical valence). In the analysis of English tweets, the proposed system was ranked between the fifth and the eleventh position –out of 39 participants- in the same tasks. The proposed system is explained in **Chapter 4**. The results of this work were published in the following paper:

   - Mohammed Jabreel and Antonio Moreno, "EiTAKA at SemEval-2018 Task 1: An Ensemble of N-Channels ConvNet and XGboost Regressors for Emotion Analysis of Tweets", *Proceedings of the 12th International Workshop on Semantic Evaluation*, Association for Computational Linguistics, pp. 193-199, 2018.

3. We have proposed a novel Deep Learning based system that addresses the multiple emotion classification problems on Twitter. We proposed a novel method to transform it into a binary classification problem and exploit a deep learning approach to solve the transformed problem. **Chapter 5** explains this system. The results of this work have been published in the following journal paper:

   - Mohammed Jabreel and Antonio Moreno, "A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets", Applied Sciences, vol. 9(6), March 2019.

4. We have proposed Deep Learning based, target-dependent sentiment analysis systems that identify and extract the target of the tweets. The proposed systems are composed of

two main steps. First, the targets of the tweet to be analysed are extracted. Afterwards, the polarities of the tweet towards each extracted target are identified. **Chapter 6** is based on the following papers and book chapters:

- Mohammed Jabreel, Fadi Hassan, and Antonio Moreno, "Target-dependent sentiment analysis of tweets using bi-directional gated recurrent neural networks", *Advances in Hybridization of Intelligent Methods. Models, Systems and Applications*, Series on Smart Innovations, Systems and Technologies, vol. 85, pp. 39-55. Springer, 2017. ISBN 978-3-319-66789-8.
- Mohammed Jabreel, Fadi Hassan, Saddam Abdulwahab and Antonio Moreno, "Recurrent Neural Conditional Random Fields for Target Identification of Tweets", *Artificial Intelligence Research and Development*, vol. 300, pp. 66 - 75, IOS Press, 2017.

5. We have proposed a novel, multi-lingual and Deep Learning-based system called Universal Sentiment Analysis System for Low-Resource Languages (UniSent). The system was designed to address the problem of Cross-lingual Sentiment Analysis. It aims to transfer knowledge extracted from annotated sentiment resources in a rich-resource language (e.g., English) to low-resource languages. The proposed system was presented at the 22nd edition of the International Conference of the Catalan Association for Artificial Intelligence (CCIA-2019) and it was awarded the best paper prize. **Chapter 7** describes this work and is based on the following papers:

   - Mohammed Jabreel, Najlaa Maaroof, Aida Valls and Antonio Moreno, "UniSent: Universal Sentiment Analysis System for Low-Resource Languages", *Artificial Intelligence Research and Development*, vol. 319, pp. 387 - 396, IOS Press, 2019.
   - Mohammed Jabreel, Najlaa Maaroof, Aida Valls and Antonio Moreno, "Towards a Universal and Multi-Lingual Sentiment Analysis System", *Engineering Applications of Artificial Intelligence*, **under review**.

6. As a novel contribution to the fields of Multi-Criteria Decision Aid (MCDA) and SA, we have developed a system called SentiRank. It is a system based on MCDA techniques and sentiment analysis that give users the ability to exploit reviews alongside their preferences in the process of alternatives ranking. To achieve that, we primarily examine the task of aspect-based sentiment analysis to automatically detect and analyse all expressions of sentiment towards a set of aspects in users' reviews. After that, we exploit an MCDA method, named ELECTRE-III, to develop a ranking system

based on the decision maker's preferences and the users' reviews. The description of the proposed system is presented in **Chapter 8**. It will be submitted to the following conference:

- Mohammed Jabreel, Najlaa Maaroof, Aida Valls and Antonio Moreno, SentiRank: Combining Aspect-Based Semantic Analysis with Multi-Criteria Decision Support Systems, *The 17th International Conference on Modeling Decisions for Artificial Intelligence*, 2020.

7. As a case study, and by applying the developed systems, we have collected tweets of local people, visitors, and official brand destination offices from different tourist destinations and analysed the opinions and the emotions shared in these tweets. This work has been carried out in collaboration with national and international researchers specialised in destination brand communication. The results of these studies, presented in **Chapter 9**, are based on the following papers:

   - Mohammed Jabreel, Antonio Moreno, and Assumpció Huertas, "Do local residents and visitors express the same sentiments on destinations through social media?." *Information and Communication Technologies in Tourism 2017*. Springer, pp. 655-668, 2017.
   - Mohammed Jabreel, Antonio Moreno, and Assumpcio Huertas, "Semantic comparison of the emotional values communicated by destinations and tourists on social media", *Journal of Destination Marketing & Management*, vol 6(3), pp. 170-183, 2017.
   - Mohammed Jabreel, Antonio Moreno, and Assumpció Huertas, "Semantic analysis and the evolution towards participative branding: Do locals communicate the same destination brand values as DMOs?", *Public Library of Science (PloS one)*, 2018.
   - Lidija Lalicic, Assumpció Huertas, Antonio Moreno, Stefan Gindl, and Mohammed Jabreel. "Do DMOs communicate their emotional brand values? A comparison between Twitter and Facebook." In *Information and Communication Technologies in Tourism*, pp. 159-171. Springer, Cham, 2018.
   - Antonio Moreno, Mohammed Jabreel, and Assumpció Huertas, "Automatic analysis of the communication of tourist destination brands through social networks", *10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 546-553, 2015

The next chapter, i.e. **Chapter 2**, is an introductory chapter that presents an overview of the sentiment analysis field and the main concepts used in this thesis. **Chapter 10** distils the findings and discusses shortcomings, advantages and potential future research directions. **Appendix A** gives an overview about the concepts of Multi-Criteria Decision Aid. **Appendix B** lists some of our publications that describe the case studies of our systems to the field of communication of destination brands through social networks.

Chapter 2

Background

2.1 Introduction

This chapter places the contributions of this thesis into a broader context. We provide a broad overview of the relevant concepts. In section 2.2, we introduce the problem of sentiment analysis, the related terminology, and point out some subproblems. Section 2.3 provides an overview of Machine Learning and Deep Learning. In section 2.4, we provide a brief overview of the most relevant Natural Language Processing concepts. Subsequently, section 2.5 overviews the word representation techniques. In section 2.6, we present the description of the main evaluation metrics used in this thesis. Finally, Section 2.4 summarizes the most relevant information provided in this chapter.

2.2 Sentiment Analysis

Sentiment analysis, known as opinion mining, is the computational study of people's opinions, sentiments, attitudes, feelings, and emotions. It is considered to be one of the most active research fields in Natural Language Processing, data mining, information retrieval, and Web mining. Sentiment analysis is not just the problem of classifying whether a piece of text expresses a positive or negative sentiment or opinion. It is indeed a much more complex problem than that. It is a generic name and involves many facets and multiple related tasks, each of which has its challenges. The next sub-sections provide an overview of these tasks.

2.2.1 Sentiment Polarity Classification

Sentiment polarity classification is the most well-studied problem in the field of sentiment analysis. Typically, the task is considered as a multi-class classification problem: Given

a subjective text, the goal is to determine whether the general tone of the text is positive, negative or neutral. This task can be conducted at various levels of granularity: from the sentiment polarity associations of words and phrases, to the sentiment of sentences, SMS messages, chat messages, and tweets, to the analysis of sentiment in product reviews, blog posts, and whole documents.

Word Level: The idea of this level of analysis is to assign a sentiment value to a word. A collection of word-sentiment associations, called entries, forms a sentiment lexicon. Each entry is composed by a key, which is the word, and a value, which can be positive or negative, or a real value that indicates the strength of the association between the word and the sentiment category. Such lexicons can be created either by manual annotation or automatically [85]. Manually created lexicons usually contain a few thousand entries. In contrast, lexicons that are automatically generated can capture sentiment associations for hundreds of thousands of words or sequences of words.

Sentence Level: Sentiment analysis systems at the sentence level aim to assign labels such as positive, negative, or neutral to whole sentences. It is worth noting that the sentiment level of a sentence can not be obtained by simply aggregating the polarities of its words. Hence, it is required to develop Machine Learning-based systems that understand the context and infer the correct sentiment expressed in sentences. Such systems learn a mapping from labelled training data (sentences that are already marked as positive, negative, or neutral) using a large number of features extracted from the text and/or external resources, e.g., sentiment lexicons.

Document Level: Sentiment analysis at the document level is usually decomposed into the sentiment analysis of the component sentences. However, there are some works which propose to summarise the sentiment in whole documents [91].

2.2.2 Detecting Sentiment Towards a Target

People express their sentiments or opinions in texts towards entities or their aspects. We usually call these entities the targets of the opinions. For example, a restaurant review that communicates a positive opinion about the food may express a negative attitude towards the quality of service. Thus, it is necessary to have systems that can automatically identify the targets in an opinionated text and the polarity towards these extracted targets. This problem is called *target-dependent sentiment analysis*. In this problem, it is necessary to determine the target and its context, which can be defined as follows:

Target A *target* is an entity (person, organisation, product, object, etc.) referred to in a text, about which an opinion is expressed.

Context The *context* of the target is the text surrounding it, that provides information about the polarity of the sentiment towards it.

It is quite usual to give several opinions on different aspects of an object in a single sentence. For example, the text *"I have got a new mobile. Its camera is wonderful, but the battery life is too short."*, gives both positive and negative remarks about a mobile phone. It may be seen that the example contains three targets (*"mobile"*, *"camera"* and *"battery life"*) and the sentiment polarities towards them can be seen as "neutral", "positive" and "negative", respectively.

2.2.3 Automatically Detecting and Analyzing Affect and Emotions

The task of detecting the expression of emotion in a text can be considered as a refinement of the sentiment polarity classification task. The goal is to assign to a piece of text one (in the case of multi-class classification problem) or more (in the case of multi-label classification problem) of a predefined set of basic emotions. Unlike the sentiment polarity classification systems, emotion classification systems try to identify more fine-grained differences in the expression of sentiment. Most commonly, Ekman's six "basic" emotions -anger, disgust, fear, happiness, sadness, and surprise- are used as class labels for this task [4, 7, 44, 46]. Other theories, such as Plutchik's wheel which considered eight primary emotions [121], Scherer's effect categories [143], or Ekman's extended model [45] may also be used as a basis.

Beside deriving emotion categories from psychological theories of emotion, we may also define an ad-hoc set of emotion categories to suit the needs of a specific application [20, 61, 162, 188]. Applications for emotion classification are manifold, ranging from the observation of trends in public social networks [20] to the analysis of clinical records [119]. In general, emotion classification is closely related to the research area of affective computing [143]. Although emotions are closely related to sentiments/opinions, there are some differences between emotions and opinions. Rational opinions express no emotions, e.g., "the service of this restaurant is really good", and emotional sentences do not necessarily express sentiment or opinion, e.g., "I am so glad to see you here". Additionally, emotions may not have targets, but just people's feelings, e.g., "I am so sad today."

2.2.4 Multilingual Sentiment Analysis

Most of the research in sentiment analysis has focused on the English language. Thus, the sentiment resources (sentiment lexicons, annotated corpora, etc.) in other languages are fewer than in English. As a result, automatic sentiment analysis systems in other languages

tend to be less accurate than their English counterparts. Consequently, work on multilingual sentiment analysis has mainly focused on developing methods and systems to map sentiment resources from English into morphologically complex languages.

2.3 Machine Learning

Machine Learning is the sub-field of computer science that, according to Arthur Samuel in 1959, gives "computers the ability to learn without being explicitly programmed" [140]. It evolved from the study of pattern recognition and computational learning theory in Artificial Intelligence.

Machine Learning models can solve various types of problems such as classification, regression, and clustering problems. Those models can be categorized into supervised and unsupervised approaches. Supervised Machine Learning models need labelled datasets in which experts annotate manually a set of examples and usually split them into two or sometimes three data sets: training set, development set and testing set.

2.3.1 Supervised Learning Algorithms

Supervised learning algorithms require all the instances of the training set to be labelled or annotated. Usually, the annotation process is done by experts of the field, which produces a "knowledge-laden" data. From this previous knowledge, the algorithm can learn and generalize, being able to predict new and unseen samples.

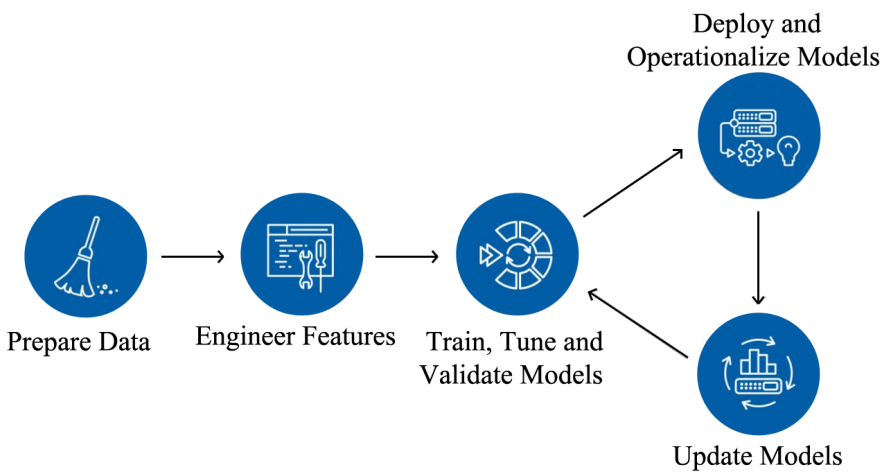


Figure 2.1: Supervised Machine Learning life cycle.

Figure 2.1 illustrates the life cycle of building a supervised Machine Learning model. First, a set of features must be designed and used to represent the examples as vectors. Afterwards, the computed vectors are used to train a model (classifier or regressor). The most popular supervised Machine Learning models that have been used recently, especially for the sentiment analysis problem, are Support Vector Machines (SVMs), Naive Bayes, Decision Trees, Random Forest, Logistic Regression and XGBoost [47, 70, 104, 115]. Usually, the model has a set of parameters that must be tuned. Thus, some model selection techniques like grid-search and k-fold cross-validation can be used to select the values of the parameters that give the best performance based on what is called a development set. Once the model is obtained, its effectiveness is evaluated by applying it on testing sets.

It is worth mentioning that all of our works presented in this thesis are kinds of supervised learning-based algorithms. Below, we give a brief overview of the most common supervised Machine Learning models.

Support Vector Machines

Recently, Support Vector Machines (SVMs) became well-known for solving problems in classification, regression and anomaly detection. One of its essential properties is that the resolution of the parameters of the model corresponds to a convex optimization problem; thus, any local solution is a global one.

Let us describe the concept of SVM by analysing the two class classification problem. Given a labeled training set of the form $(x_i, y_i), i = 1, 2, \ldots k$, where $x_i \in \mathbb{R}^n$ are the feature values, $y_i \in \{1, -1\}$ is the binary classification, n is the number of features and k is the number of samples, the SVM classifier solves the following optimization problem:

$$\begin{aligned} &\|\omega\|^2_{\omega,\xi} + C\sum_{i=1}^{k} \xi_i \\ &\qquad s.t. \quad y_i(\omega^T \phi(x_i) + b) \geq 1 - \xi_i, \\ &\qquad\qquad \xi_i \geq 0. \end{aligned} \tag{2.1}$$

In this expression, the soft margin parameter C tells the SVM optimization process how much is needed to avoid misclassifying each training instance. The weight vector ω is normal to the separating hyperplane. The parameter ξ is used to give a degree of flexibility to the algorithm when fitting the data and b represents the bias.

During the optimization process of the SVM the training data x_i are mapped into a higher dimensional space using a kernel function, $K(x_i, x_j) = (\phi^T(x_i).\phi(x_j))$. SVM uses the kernel trick, by which the data become linearly separable in the new space. The SVM classifier

finds the hyperplane with a maximum margin of separation between the classes in the new higher dimensional space. In the case of a *linear SVM (LSVM)* classifier, ϕ refers to a dot product. In a *non-linear SVM (NLSVM)* the classifier function is formed by non-linearly projecting the training data of the input space to a feature space of a higher dimension by using a kernel function. The radial basis function (RBF) is widely used as a mapping kernel. The RBF can be defined as follows:

$$K(x_i, x_j) = exp\left(-\gamma \|x_i - x_j\|_2^2\right) \tag{2.2}$$

In this expression $\gamma = 1/2\sigma^2$, $\|x_i - x_j\|_2^2$ is the squared Euclidean distance between the two feature vectors x_i and x_j, and σ is a free parameter.

Decision Trees

A decision tree is a predictive model which is a mapping from observations about an item to conclusions about its target value. In the tree structures, leaves represent decisions (e.g., classifications) (also referred to as labels), non-leaf nodes are attributes, and branches represent conjunctions of attributes that lead to the classifications [123]. The process of building a decision tree is a supervised Machine Learning method for producing a decision tree from training data.

Constructing a decision tree that is compatible with a given data set is easy. The difficulty is in building a good decision tree, which typically means a small decision tree. A popular method for building the smallest decision trees is ID3 proposed by Quinlan, which is based on information gain. An improved version of ID3 developed by Ross Quinlan is known as C4.5. It builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The improvements of C4.5 include handling both continuous and discrete attributes, working with training data with missing attribute values and pruning trees after creation. The pruning process is useful to prevent a tree from being overfitted just for the training set. This technique makes the tree general for unlabeled data and can allow some mistakenly labeled training data.

Model Ensembling

Ensemble learning is the process of combining multiple learning algorithms, such as classifiers, to solve a specific problem. It is mainly used to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Other applications of ensemble learning involve assigning confidence to the decision made by the model,

selecting optimal (or near-optimal) features, data fusion, incremental learning, non-stationary learning and error-correcting. Popular types of ensembles include the following:

- Bayes optimal classifier that is an ensemble of all the hypotheses in the hypothesis space.
- Bayesian parameter averaging. It approximates the Bayes Optimal Classifier by sampling hypotheses from the hypothesis space and combining them using Bayes' rule.
- Bootstrap aggregating (bagging). The idea of this method is to build multiple models, typically of the same kind, from different subsamples of the training dataset. An example of this method is a Random Forest, that combines random decision trees to achieve very high classification accuracy.
- Boosting. In this kind of ensembling approach, we build multiple models, typically of the same type. Each model learns to fix the prediction errors of a prior model in the chain. One of the most popular methods of this type of ensembling that is used in this work is XGBoost.
- Stacking. The idea is to build multiple models, typically of differing types, and then use a supervisor model to learn to combine the predictions of the best primary models.
- Bucket of models. In this kind of ensembling, a model selection algorithm is used to choose the best model for each problem, and the model selection is based on cross-validation.

2.3.2 Deep Learning

Deep Learning (DL) is part of a broader family of Machine Learning techniques based on learning data representations rather than task-specific approaches. Researchers in the field of DL attempt to create effective systems that learn representations from large-scale, unlabeled data sets. DL architectures such as deep neural networks, deep belief networks and recurrent neural networks have been applied to many fields including computer vision, speech recognition, Natural Language Processing (NLP), and bioinformatics. In such fields, DL-based systems have proven to be able to produce results comparable to (and in some cases, superior) traditional, hand-crafted ML-based systems and human experts [56]. Unlike the traditional supervised Machine Learning models, Deep Learning models do not require to design and define manually the set of features that will be used to train the models. They

can automatically learn high-level features from the row data by stacking many neural layers; the lowest layers learn low-level features while the highest layers learn high level ones. The next sections provide a short overview of the Deep Learning concepts used in this work.

Neural Networks

Neural Networks (NNs) are a class of Machine Learning algorithms that have become ubiquitous in sentiment analysis in the last five years [183]. Systems based on NNs obtain the state-of-the-art results on almost all NLP tasks and datasets which are commonly used for testing. A full overview of NNs is out of the scope of this thesis. So, the reader may refer to [56] for more details. Therefore, we will focus only on the models used in this thesis. NNs, like any Machine Learning algorithm, require massive amounts of training data to generalize well. One critical issue of NNs is the large number of free parameters. So, they are more likely to overfitting than linear models such as support vector machines or maximum entropy models. Hence, regularization techniques such as dropout and early stopping are often used to prevent the problem of overfitting.

Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are the family of neural networks specifically designed for processing sequential data of varying lengths such as text.

As shown in Figure 2.2, an RNN receives as an input a sequence of vectors $\{x_1, x_2, ..., x_n\}$, at each time step t, it takes the input vector $x_t \in \mathbb{R}^d$ and the hidden state vector $h_{t-1} \in \mathbb{R}^{d_h}$ and produces the next hidden state h_t by applying the following equation:

$$h_t = \phi\left(x_t, h_{t-1}\right) \tag{2.3}$$

Usually, h_0 is initialized to a zero vector in order to calculate the first hidden state. The most

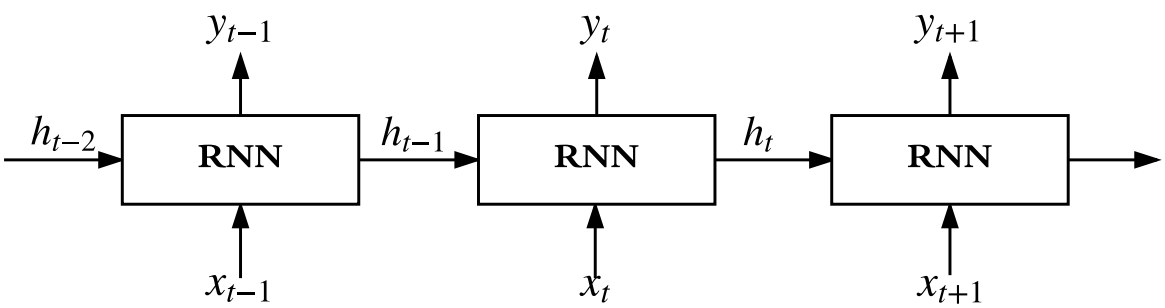


Figure 2.2: Recurrent Neural Network.

common approach is to use the affine transformation operation followed by an element-wise

non-linearity, e.g. Rectified Linear Unit (ReLU), as the function ϕ that produces the next hidden state vector h_t.

$$\phi(x_t, h_{t-1}) = f(W^x x_t + W^h h_{t-1} + b) \tag{2.4}$$

In this formula, $W^x \in \mathbb{R}^{d \times d_h}$, $W^h \in \mathbb{R}^{d_h \times d_h}$ and $b \in \mathbb{R}^{d_h}$ are the parameters of the model, and f is an element-wise non-linearity. The result of the recurrence process is a sequence of hidden states $\{h_1, h_2, ..., h_n\}$. We handle this sequence of states based on the type of the problem we are interesed to solve.

In practice, the major issue of RNNs using these transition functions is the difficulty of learning long-term dependencies due to vanishing/exploding gradients [17]. Long short-term memory (LSTM) units [66] and Gated Recurrent Units (GRU) [33] have been specifically designed to address this problem.

Bidirectional RNNs

The standard RNN reads an input sequence $X = (x_1, ..., x_n)$ in a forward direction (left-to-right) starting from the first symbol x_1 and ending in the last one x_n. Thus, it processes sequences in temporal order, ignoring the future context. For many tasks on sequences it is beneficial to have access to future as well as to past information. For example, in text processing, decisions are usually made after the whole sentence is known. The Bidirectional BiRNN architecture [59] proposed a solution for making predictions based on both past and future information.

Figure 2.3 illustrates the architecture of a BiRNN, which consists of forward $\overrightarrow{\phi}$ and backward $\overleftarrow{\phi}$ RNNs. The first one reads the input sequence in a forward direction $(x_1, ..., x_n)$ and produces a sequence of forward hidden states $(\overrightarrow{h_1}, ..., \overrightarrow{h_n})$, whereas the former reads the sequence in the reverse order $(x_n, ..., x_1)$ resulting in a sequence of backward hidden states $(\overleftarrow{h_n}, ..., \overleftarrow{h_1})$. We obtain a representation for each word x_t by concatenating the corresponding forward hidden state $\overrightarrow{h_t}$ and the backward one $\overleftarrow{h_t}$. The following equations illustrate the main ideas:

$$\overrightarrow{h_t} = \overrightarrow{\phi}(x_t, \overrightarrow{h_{t-1}}) \tag{2.5}$$

$$\overleftarrow{h_t} = \overleftarrow{\phi}(x_t, \overleftarrow{h_{t-1}}) \tag{2.6}$$

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{2.7}$$

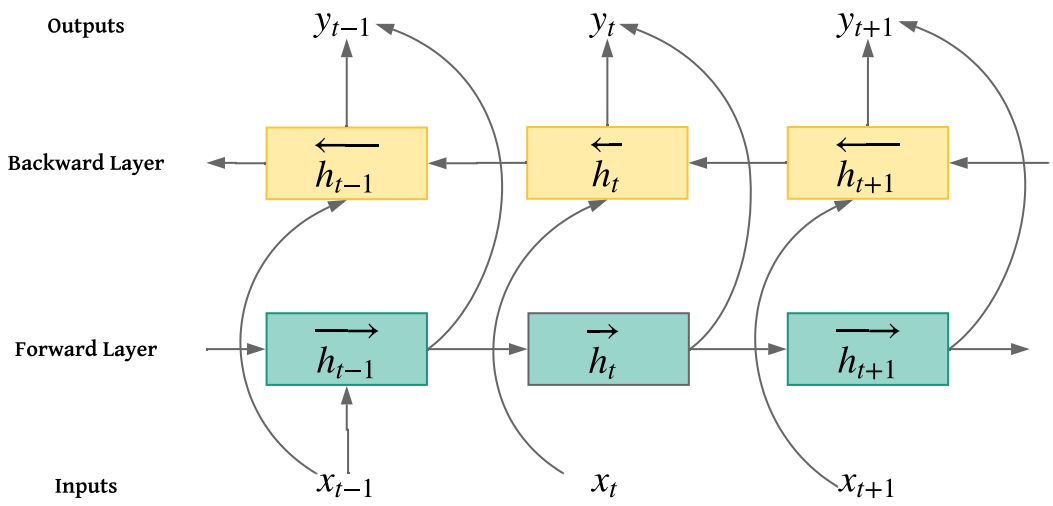


Figure 2.3: Bidirectional Recurrent Neural Network.

In practice, it is quite challenging to train RNNs for tasks that require a network to make use of all of the information distant from the current location of processing. Despite having access to the entire previous sequence, the information encoded in the hidden states of an RNN tends to be reasonably local, more relevant to the most recent pieces of the input sequence and recent decisions. However, it is often the case that distant information is significant to many language applications. One reason for the failure of RNNs to carry forward important information is that the hidden layers are asked to perform two tasks simultaneously: provide information useful for the current decision, and updating and carrying forward the information required for future decisions. A second difficulty with training RNNs emerges from the need to backpropagate the error signal back through time. To overcome these issues, more complex network architectures have been designed to manage the task of maintaining relevant context over time explicitly. More specifically, the network needs to learn to forget information that is no longer needed and to remember information required for decisions still to come. Below, we briefly describe the most popular types of RNNs that are used in NLP tasks.

Long Short-Term Memory

Long short-term memory (LSTM) networks [66] solve the context management problem by dividing it into two sub-problems. The problems of eliminating information no longer needed from the context and adding information likely to be needed for later decision making. The

clue to solving both problems is to learn how to handle this context instead of hard-coding a strategy into the architecture. LSTMs achieve this by first adding an explicit context layer to the architecture (in addition to the usual recurrent hidden layer), and by the use of specific neural units that make use of gates to control the flow of information into and out of the units. These gates are implemented by the use of additional weights that operate sequentially on the input and previous hidden layer.

The gates in an LSTM share a typical design pattern; each consists of a feedforward layer, followed by a sigmoid activation function and a pointwise multiplication with the layer to be gated. The reason behind choosing the sigmoid activation function stems from its tendency to clip its outputs to [0, 1]. The gates in LSTM can be seen as binary masks. Values in the layer being gated that align with values close to 1 in the mask are kept nearly unchanged; values corresponding to lower values are substantially omitted.

As shown in Figure 2.4, an LSTM is composed by three gates. The first one is called the **forget gate**. The objective of this gate to omit information from the context that is no longer required.

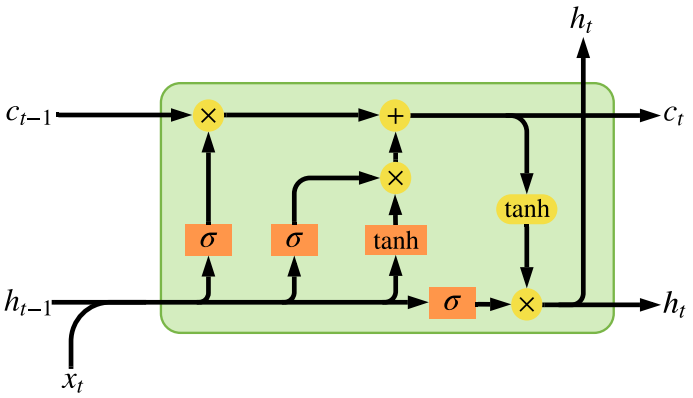


Figure 2.4: Long Short-Term Memory Architecture.

It computes a weighted sum of the previous state of the hidden layer and the current input and passes that into a sigmoid to produce a (pseudo) mask. This mask is then multiplied by the context vector to discard the information from the context that is no longer needed. Then, we compute the actual information we need from the previous hidden state and the current input. This step is similar to the basic operation in the typical RNNs. Next, we generate the mask for the **add gate** to select the information to be added to the context vector. By adding this to the modified context vector, we get the next context vector. Finally, we use the **output gate** to decide what information is required for the current hidden state. The following equations illustrate the complete computation for a single LSTM unit.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2.8}$$
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{2.9}$$
$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{2.10}$$
$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{2.11}$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{2.12}$$
$$h_t = o_t * \tanh\left(C_t\right) \tag{2.13}$$

In these formulas, x_t is the input at time t, h_{t-1} is the previous hidden state, C_{t-1} is the previous context vector, h_t is the current hidden state, C_t is the current context vector, and $\tilde{C}_t$ is candidate hidden state. The terms f_t, i_t and o_t refer to the forget, input and output gates respectively.

Gated Recurrent Unit

Gated recurrent units (GRUs) [34] were designed to have more persistent memory, making them very useful to capture long-term dependencies between the elements of a sequence.

Figure 2.5 shows a graphical depiction of a gated recurrent unit.

This kind of units have *reset* (r_t) and *update* (z_t) gates. The former has the ability to completely reduce the past hidden state h_{t-1} if it considers that it is irrelevant to the computation of the new state, whereas the later is responsible for determining how much of h_{t-1} should be carried forward to the next state h_t.

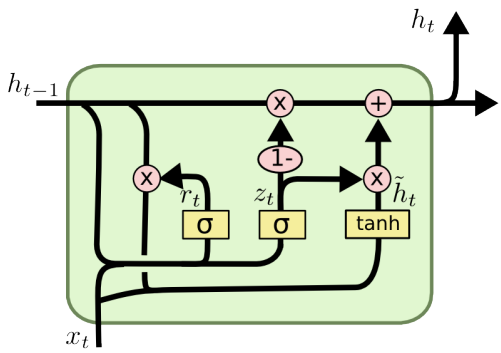


Figure 2.5: Gated Recurrent Unit (GRU).

The output h_t of a GRU depends on the input x_t and the previous state h_{t-1}, and it is computed as follows:

$$r_t = \sigma\left(W_r \cdot [h_{t-1}; x_t] + b_r\right) \tag{2.14}$$

$$z_t = \sigma\left(W_z \cdot [h_{t-1}; x_t] + b_z\right) \tag{2.15}$$

$$\widetilde{h}_t = tanh\left(W_h \cdot [(r_t \odot h_{t-1}); x_t] + b_h\right) \tag{2.16}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h}_t \tag{2.17}$$

In these expressions r_t and z_t denote the *reset* and *update* gates, $\widetilde{h}_t$ is the candidate output state and h_t is the actual output state at time t. The symbol $\odot$ stands for element-wise multiplication, σ is a sigmoid function and ; stands for the vector-concatenation operation. $W_r, W_z, W_h \in \mathbb{R}^{d_h \times (d+d_h)}$ and $b_r, b_z, b_h \in \mathbb{R}^{d_h}$ are the parameters of the *reset* and *update* gates, where d_h is the dimension of the hidden state.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs), also known as ConvNets, were initially designed to apply Deep Learning to Computer Vision tasks and have proven to be highly effective in solving problems such as image classification. They use the concept of a "convolution", where a sliding window or a "filter" passes over an image identifying important features and analyzing them one at a time and reducing them down to their essential characteristics. Although CNNs were originally invented for computer vision, they have subsequently shown to be effective for NLP and have achieved outstanding results in semantic parsing [176], information retrieval [146, 147], sentence modeling [78], and other traditional NLP tasks [36]. They have also been shown to achieve effective performance on text classification tasks [11, 168, 187], even when considering relatively simple one-layer architecture [82]. Hence, we focus on the usage of CNNs in the task of text classification. We show in Figure 2.6 the most common architecture of CNNs that was proposed by Kim, 2014 for sentence classification [82].

Let $x_i \in \mathbb{R}^d$ be the d-dimensional word vector corresponding to the i-th word in the input sentence. Usually, the word vector is obtained from a pre-trained word embedding model, see section 2.5. A sentence of length n, padded or truncated whenever is necessary, is represented as:

$$X_{1:n} = x_1 \oplus x_2 \oplus \ldots \oplus x_n \tag{2.18}$$

Here $\oplus$ refers to the concatenation operator. Generally, let $X_{i:i+h}$ refer to the concatenation of word vectors $x_i, x_{i+1}, \ldots, x_{i+h}$. A convolution operation, which involves a filter $w \in \mathbb{R}^{hk}$, is

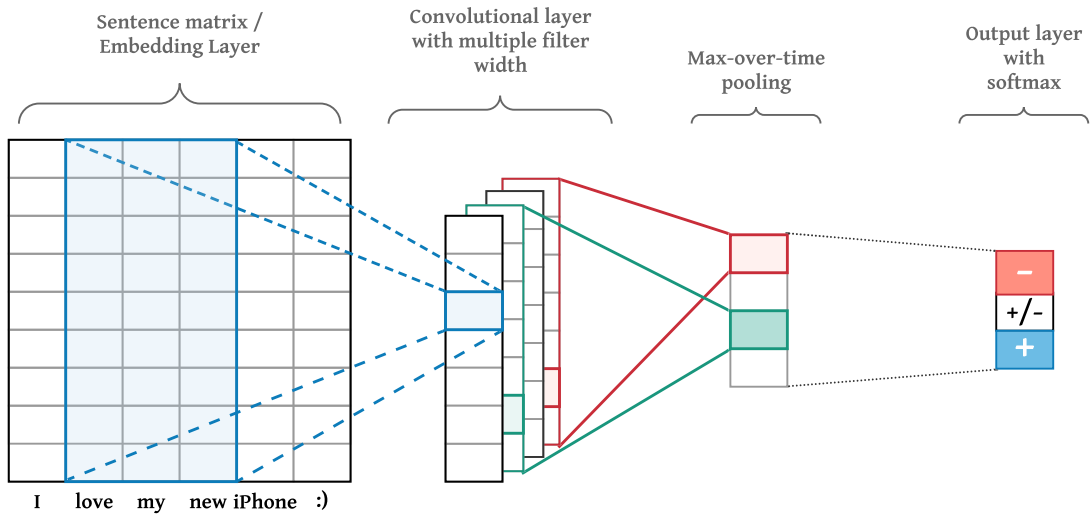


Figure 2.6: Convolutional Neural Networks for text classification.

applied to a window of h words to produce a new feature c_i as follows:

$$c_i = f(W \cdot X_{i:i+h} + b) \tag{2.19}$$

In this formula $b \in \mathbb{R}$ is a bias term and f is a non-linear activation function such as the hyperbolic tangent. This filter is slided through the sentence and applied to each possible window of words to produce a feature map $c \in \mathbb{R}^{n-h+1}$:

$$c = [c_1, c_2, \ldots, c_{n-h+1}]. \tag{2.20}$$

After that, we apply a max-over-time pooling operation [36] over the feature map and take the maximum value $\hat{c} = max\{c\}$ as the feature corresponding to this particular filter. We have described the process by which one feature is extracted from one filter. The model uses multiple filters (with varying window sizes) to obtain multiple features. The features from the penultimate layer are passed to a fully connected softmax layer, described below, whose output is the probability distribution over labels.

Softmax Classifier

The softmax classifier is a feed-forward neural network followed by the softmax function, which is used for multi-class classification (under the assumption that the classes are mutually

exclusive). It takes as input a vector $v \in \mathbb{R}^m$ and produces the probabilities for each class as follows:

$$p(y = i|v; W, b) = \frac{exp(w_i^T v + b_i)}{\sum_{j=1}^{C} exp(w_j^T v + b_j)}, i = 1, 2, ..., C \tag{2.21}$$

This can be interpreted as the (normalized) probability assigned to each class i given the input vector v, and parameterized by $W \in \mathbb{R}^{m \times C}$ and $b \in \mathbb{R}^C$, where C is the number of classes, w_i is the i-th column of W and b_i is a bias term.

2.4 Natural Language Processing

Natural Language Processing (NLP) includes the study of mathematical and computational models of different aspects of language and the development of a broad range of systems. Research in NLP is highly interdisciplinary, including concepts in computer science, linguistics, logic, and psychology. NLP has a unique role in computer science, particularly in the sub-field of Artificial Intelligence. Many aspects of the field deal with linguistic features of computation and NLP seeks to model language computationally. There are two main tasks in the field of NLP: language understanding (or text analysis) and language generation. Considering the text analysis, where the work in this thesis is focused on, there are some pre-processing steps used for that purpose, namely:

- Text Normalization. It is well known that text normalization is an important step for most NLP problems in general and especially for the sentiment analysis problem. The text in reviews and social posts (like tweets) is usually noisier than the regular text in articles and blogs. People tend to write their posts, messages and reviews in an informal way. Thus, in this thesis, we used several text normalization steps in order to utilize most of the information in text, performing tokenization, spell correction, word normalization, word segmentation hashtag splitting and word annotation.

- Part-of-Speech (POS) tagging. It is normally a sentence-based NLP task; given a sentence formed of a sequence of words, POS tagging tries to label (tag) each word with its correct part of speech (also named word category, word class, or lexical category). POS tags are useful because they expose a lot of information about a word and its neighbours. For instance, knowing whether a word is a noun or a verb tells us about likely neighbouring words (nouns are preceded by determiners and adjectives, verbs by nouns) and syntactic structure (nouns are generally part of noun phrases). Hence, in this thesis, we used POS tags to extract features.

- Words Clustering. It is the technique of partitioning sets of words into subsets of semantically similar words. It is increasingly becoming a major technique used in several NLP tasks ranging from word sense or structural disambiguation to information retrieval and filtering. One of the most common algorithms for words clustering is Brown clustering. It is a kind of hierarchical clustering of words based on the contexts in which they occur. The intuition behind the algorithm is that a class-based language model is used to address the data sparsity problem inherent in language modeling.

2.5 Word Representation

Word embeddings mostly follow the distributional assumption, according to which words with similar meanings tend to occur in a similar context. Thus, these vectors attempt to capture the features of the neighbours of a word. The main benefit of distributional vectors is that they capture the similarity between words. Measuring similarity between vectors is possible, using any similarity measure such as cosine similarity. We often use word embeddings as the first layer in Deep Learning-based models for NLP. When a text has to be analysed, the first step is to map each word into a continuous, low dimensional and real-valued vector, which can later be processed by a neural network model. All the word vectors are stacked into a matrix $E \in \mathbb{R}^{d \times N}$, where N is the vocabulary size and d is the vector dimension. This matrix is called the *embedding layer* or the *lookup table layer*. The embedding matrix can be initialized using a pre-trained model like *word2vec* or *Glove* [98, 117]. Typically, word embeddings are pre-trained to optimize an auxiliary objective based on large unlabeled corpora. For example, the work presented in [98] obtained the word embeddings by solving the problem of predicting a word based on its context. It has been shown that the learned word vectors can capture general syntactical and semantic information. Therefore, these word embeddings, due to its smaller dimensionality, have proven to be efficient, fast and beneficial in many NLP tasks.

Below, we introduce the word embedding models used in this thesis. Specifically, we provide a brief description of the Word2Vec and Global Vectors (GLove) models.

Word2Vec

Mikolov et al., 2013 [98] proposed two log-linear models for word embeddings: the continuous bag of words model (CBOW) and the Skip-gram model. CBOW estimates the conditional likelihood of a word given its surrounding context words with a window of size k. The skip-gram model does the opposite of the CBOW model. It tries to predict the

surrounding context words of the central word. The two models were trained in unsupervised settings. The word embedding dimension is determined by the accuracy of prediction. As the embedding dimension increases, the accuracy of prediction also increases until it converges at some point, which is considered the optimal embedding dimension.

It was found that the similarity of word representations goes beyond simple syntactic regularities. For example, by performing simple algebraic operations on the word vectors (as illustrated in Figure 2.7), it was shown that vector("King") - vector("Man") + vector(" Woman") results in a vector that is very close to the vector representation of the word Queen.

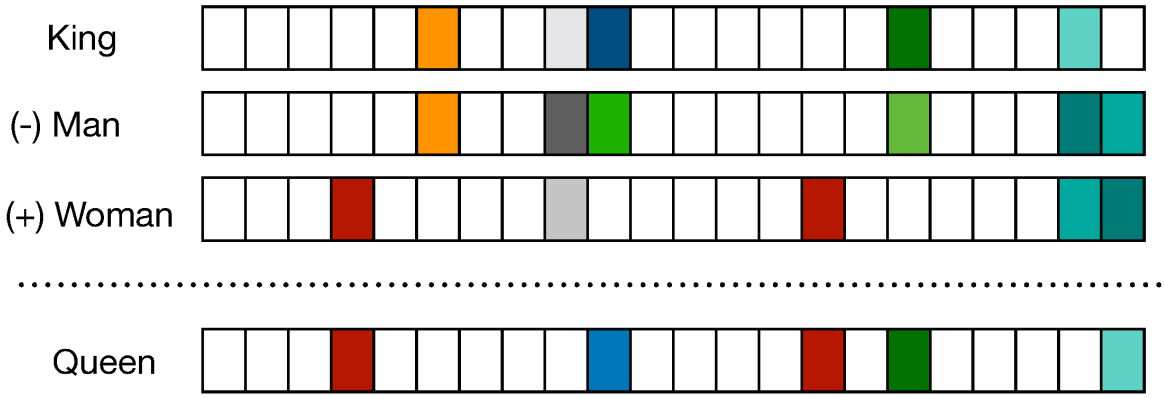


Figure 2.7: Example of algebraic operations on word vectors.

Moreover, it has been demonstrated that the Skip-gram embeddings perform better in case they were trained on enough data [98], so we will focus on using this model in this thesis.

Global Vectors for Word Representations (Glove)

Although the word2vec embedding models can make good use of local statistics, they do not utilize the overall statistics like the methods that rely on the co-occurrence matrix. Pennington et al. [117] attempt to overcome this shortcoming by combining the advantages of global matrix factorization and local context windows. They found that much of the relevant information can be determined by ratios of co-occurrence, rather than the raw counts and proposed a log bilinear regression model to take advantage of this fact. The word embeddings obtained by Glove have been empirically shown to perform well, as much as these obtained by Skip-gram, for sentimentn classification and semantic relatedness (Tai et al., 2015). Hence, in this thesis, we mainly use SkipGram and Glove embeddings.

2.6 Evaluation Metrics

Binary Classification

To introduce the evaluation metrics for sentiment analysis tasks, we first consider the binary sentiment classification problem. As we stated, our goal, in this case, is to assign a given piece of text to one of the two labels "positive" or "negative". We also need to know whether that piece of text is actually "positive" or "negative". The actual label is known as the human-annotated label, and we refer to it as the **gold label**. The evaluation process starts by building a **confusion matrix** like the one illustrated in Figure 2.8.

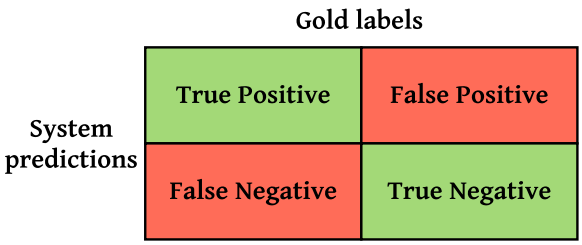


Figure 2.8: Confusion matrix.

The **true positive (tp)** cell in the matrix refers to the number of actually positive examples that the system being evaluated as such. The **true negative (tn)** is the number of examples that are negative and were correctly classified by the system. The **false positive (fp)** is the number of examples that are negative and were misclassified by the system as positive. The **false negative (fn)** is the number of examples that are positive and were misclassified by the system as negative.

Based on that, we can define the following evaluation metrics:

Accuracy: the number of samples correctly classified divided by the total number of samples.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{2.22}$$

Precision (P) measures the fraction of the number of positive examples that were correctly classified among the examples that were classified as positive.

$$P = \frac{tp}{tp + fp} \tag{2.23}$$

Recall (R) is the fraction of the number of positive examples that were correctly classified among the examples that are positive.

$$R = \frac{tp}{tp + fn} \tag{2.24}$$

The F_1-measure is the harmonic mean of the precision and recall.

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{2.25}$$

Multi-Class Classification

In the case of a multi-class classification problem for n classes we build an $n \times n$ confusion matrix and consider two cases: **macroaveraging** and **microaveraging**. In the first case, we compute the evaluation metric value, e.g., precision, for each class, and then average over classes. In the second case, we collect the decisions for all classes into a single confusion matrix, and then compute the precision and recall for that matrix. Figures 2.9 and 2.10 illustrate the confusion matrix of thre classes and calculation process for the macroaveraging and microaveraging cases.

		Actual Labels			
		Positive	*Neutral*	*Negative*	
Predicted Labels	*Positive*	8	10	1	$P_{pos} = \frac{8}{8+10+1}$
	Neutral	5	60	50	$P_{neu} = \frac{60}{5+60+50}$
	Negative	3	30	200	$P_{neg} = \frac{200}{3+20+200}$
		$R_{pos} = \frac{8}{8+5+3}$	$R_{neu} = \frac{60}{10+60+30}$	$R_{neg} = \frac{200}{1+50+200}$	

Figure 2.9: Example of multi-class confusion matrix.

Regression

We define the most popular statistical metrics that can be used to evaluate the performance of regression systems.

Positive vs Others

	True Pos	True Not
System Pos	8	11
System Not	8	340

Neutral vs Others

	True Neu	True Not
System Neu	60	55
System Not	40	212

Negative vs Others

	True Neg	True Not
System Neg	200	33
System Not	51	83

Pooled

	True Yes	True No
System Yes	268	99
System No	99	635

$P = \frac{8}{8+11} = 0.42$ $P = \frac{60}{60+55} = 0.52$ $P = \frac{200}{200+33} = 0.86$

$Macroaverage\ P = \frac{0.42+0.52+0.86}{3} = 0.60$ $Microaverage\ P = \frac{268}{268+99} = 0.73$

Figure 2.10: Example of multi-class evaluation.

The Pearson correlation coefficient, also referred to as **Pearson's** r is a measure of the linear correlation between the system outputs and the gold labels. Let X and Y be two variables that refer to the gold labels and system outputs respectively, the Pearson's r measure is defined as follows:

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)}\sqrt{(\sum y_i^2 - n\bar{y}^2)}} \tag{2.26}$$

Here, n refers to the number of examples, x_i, y_i are the individual sample points indexed with i, and $\bar{x}$ and $\bar{y}$ are the sample means for X and Y respectively.

Cohen's **kappa (k)** measures the agreement between two raters who each classify n items into C mutually exclusive classes.

$$k = \frac{p_o - p_e}{1 - p_e} \tag{2.27}$$

Here, p_o is the relative observed agreement among raters, and p_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly seeing each category.

2.7 Summary

This chapter presented an overview of some of the most relevant concepts to understand this thesis. We introduced the problem of sentiment analysis, the related terminology, and pointed out some subproblems. Moreover, we provided a brief description of Machine Learning, Deep Learning, Natural Language Processing, word embedding and evaluation

metrics. The following chapters utilize the concepts introduced in this chapter and explain the main contributions of this thesis.

Chapter 3

Sentiment Analysis in Twitter Based on a Rich Set of Features

3.1 Introduction

As we introduced in chapters 1 and 2, automated sentiment analysis is the problem of identifying people's opinions expressed in text. It normally involves the classification of text into categories such as "positive", "negative" and in some cases "neutral".

Most of the existing sentiment analysis systems are inspired in the work presented in [115]. Supervised classifiers are trained on a set of annotated corpora using a different set of hand-crafted features. The success of such models is based on two main factors: a large amount of labeled data and the intelligent design of a set of features that can distinguish between the samples. With this approach, most studies have focused on engineering a set of efficient features to obtain a good classification performance [47, 92, 114]. The idea is to find a collection of informative features to reflect the sentiments expressed in the text. Bag-of-Words (BoW) and its variation, i.e., n-grams, is the representation method used in most text classification problems and emotion analysis. Different studies have combined the BoW features with other features such as the Parts of Speech (POS) tags, the sentiment and the emotion information extracted from lexicons, statistical information, and word shapes to enrich the text representation.

Although BoW is a popular method in most text classification systems, it has some drawbacks. Firstly, it ignores the word order. That means that two documents may have the same or a very close representation as far as they have the same words, even though they carry a different meaning. The n-gram method resolves this disadvantage of BoW by considering the word order in a context of length n. However, it suffers from sparsity and

high dimensionality. Secondly, BoW is scarcely able to model the semantics of words. For example, the words *beautiful*, *wonderful* and *view* have an equal distance in BoW, where the word *beautiful* is closer to the word *wonderful* than the word *view* in the semantic space.

Word embeddings, as explained in Chapter 2, overcome the shortcomings of BoW. Some works have confirmed the effectiveness of utilizing such resources to represent the text. For example, the work in [160] proposed a model to learn sentiment-specific word embeddings, combining these embedded words with a set of state-of-the-art hand-crafted features to learn a deep model system.

It has been shown in several studies that sentiment and emotion lexicons play an essential role in developing efficient sentiment and emotion analysis systems. For instance, the authors in [104] used several sentiment lexicons and a variety of hand-crafted features to represent the tweets. They showed that the most outstanding features in their system were the lexicon-based ones. By using such features, the authors were able to improve the performance of the system with gains of more than 8.5%.

This chapter, following the ideas presented above, proposes the representation of tweets using a novel set of features including the information provided by seven lexicons and a bag of negated words (BonW). The concatenation of these features with a set of basic features improves the classification performance. The polarity of the tweets is determined by a classifier based on a Support Vector Machine. The system has been evaluated on the standard tweet sets used in the SemEval 2015 competition, obtaining results that, in four out of five cases, outperform those of the state-of-the-art sentiment analysis systems.

The rest of the chapter is structured as follows. Section 3.2 presents some related work found in the literature of sentiment analysis. In Section 3.3 we explain the methodology and the tools used to identify the sentiments of the tweets. The experiments and results are presented and discussed in Section 3.4. Section 3.5 provides the description of an extension of the proposed system, that includes an additional type of features based on pre-trained word embedding models, and the evaluation results of this extended version in English and Arabic texts. Finally, in the last section the conclusion is described.

3.2 Related work

Previous sentiment analysis systems (both of general text and social network messages) have used a basic set of features, including BoW, POS tags, clusters mapping and some features extracted from a sentiment lexicon. They have tried to improve the performance of the classifiers by adding new features to this basic set. For instance, in the NRC system [104], ranked in the first position in the SemEval 2013 competition, the following features

were added to the basic ones and used to train a SVM classifier: *non-contiguous ngrams* (n-grams in which one token is replaced with a place holder such as *), *all-caps* (the number of words with all characters in uppercase), the number of *hashtags*, the number of *elongated* words (i.e. the words with at least one character repeated more than two times, for example, 'happppyyyyyy'), the number of contiguous sequences of *punctuation*, a boolean feature indicating whether the last token contains an exclamation or question mark, the number of *emoticons*, a boolean feature that indicates whether the last token is a positive or negative emoticon, the number of *negated contexts* and a set of features extracted from some lexicons.

Similarly, TeamX [100] used the basic features plus the word sense (the result of the UKB word sense disambiguator). They also categorized the lexicons into formal and informal and used two POS taggers (the Stanford POS tagger was used with the formal lexicons and the CMU Ark Tweet NLP tagger was used with the informal lexicons). This system won the SemEval 2014 competition. Similarly to NRC and TeamX we use the set of basic features and we propose the addition of a new set of features including the bag of negated words (BonW) and a new lexicon-based polarity measure to improve the classifier's performance as described in the following section.

Another interesting system is Webis [62], the winner of the SemEval 2015 competition. Its authors used the ensemble learning technique. Four state of the art systems, including NRC and TeamX, were reproduced and combined in ensemble and the final decision was based on averaging the individual classifier's confidence scores for the classes. In the evaluation section we compare our approach with NRC, TeamX and Webis.

3.3 System description

This section explains the main steps of the SentiRich system, the tools and the resources that have been used in this work, the features used to describe a tweet and the classification method. Figure 3.1 shows a graphical depiction of the system.

3.3.1 Pre-Processing

Some standard pre-processing methods are applied on the tweets:

- *Normalization*: Each tweet is converted to the lowercase. URLs and usernames are omitted.

- *Tokenization and POS tagging*: All tweets are tokenized and tagged using Ark Tweet NLP [54].

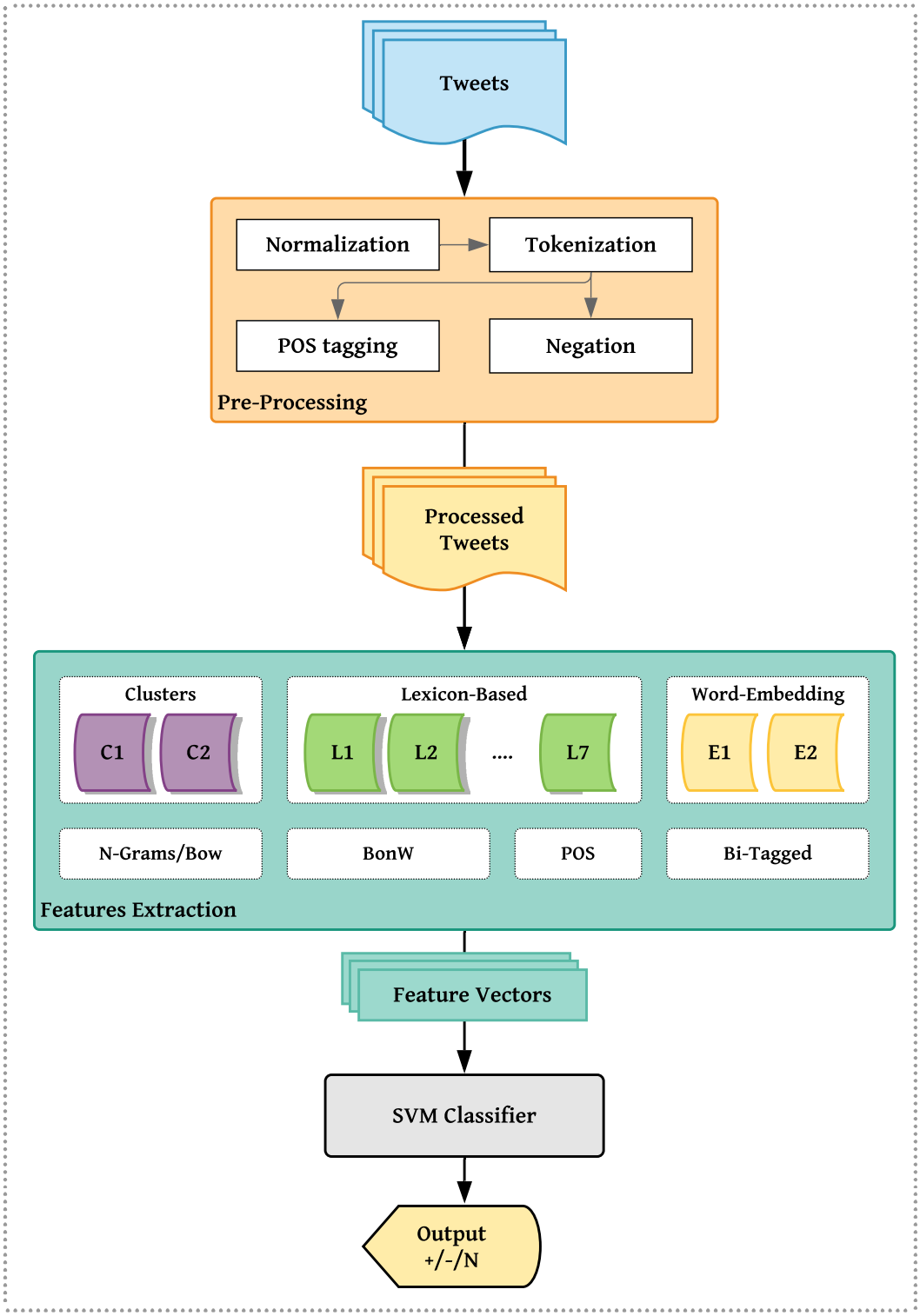


Figure 3.1: Architecture of SentiRich

- *Negation*: A negated context can be defined as a segment of the tweet that starts with a negation word (e.g. *no*, *don't*) and ends with a punctuation mark [115]. Each tweet is negated by adding the suffix "_NEG" to each word in the negated context.

As shown in [136], stopwords tend to carry sentiment information; thus, note that they were not removed from the tweets.

3.3.2 Sentiment Lexicons

We used seven lexicons in this work. They were chosen among a set of lexicons using a grid search cross validation on Twitter's test set of SemEval2013. These lexicons can be categorized based on their construction method into two types: manually constructed or automatically built.

Manual lexicons

- **General Inquirer**: General Inquirer [152] was one of the first sentiment lexicons. It contains 1195 positive and 2291 negative words.

- **MPQA**: MPQA or Opinion Finder [170] is a subjectivity lexicon, generated from the MPQA opinion corpus, which contains a large amount of manually annotated news articles. Each word in this lexicon has a polarity value (positive, negative or neutral) and a subjective value (strong subjective or weak subjective). As we consider only the polarity value on the message (i.e tweet) level, not on the word level, we only made use of the polarity value. As a polarity score for each word in the lexicon we assigned +1 to the positive words, -1 to the negative words and 0 to the neutral words.

- **AFINN**: the AFINN lexicon [112] consists of 2,477 entries (2,462 unigrams + 15 bigrams), each assigning to a word (or pair of words) an integer score between -5 "very negative" and +5 "very positive".

Automatic lexicons

- **Hu-Liu opinion lexicon (HL)**: This lexicon was designed in [67]. It was generated automatically from social media content. It contains a list of positive words (about 2000) and a list of negative words (about 4700). The polarity score for each word is assigned in the same way as with the MPQA.

- **NRC hashtags lexicon**: This lexicon was created automatically [104] from a set of 775,000 pseudo-labeled tweets which were collected via the Twitter API from April

to December 2012 based on 78 seed positive and negative hashtags such as *#good*, *#excellent*, *#bad* and *#terrible*. The sentiment score of each term (i.e. word or bigram), t, is computed using the point-wise mutual information (PMI) measure as follows:

$$SentScore(t) = PMI(t, pos) - PMI(t, neg) \tag{3.1}$$

Where *pos* and *neg* refer to the positive and negative corpora respectively. The *SentScore* is between $[-\infty, \infty]$, and it is normalized to the range [-5, +5]. The lexicon consists of 54,129 unigrams, 316,531 bigrams and 308,808 non-contiguous pairs. We used only the unigrams and bigrams from this lexicon.

- **SenticNet**: SenticNet[24] is a lexical resource that was constructed by clustering the vector space model of affective commonsense knowledge extracted from ConceptNet. This lexicon contains more than 14,000 terms along with their polarity scores which are in the range [-1, +1].

- **TS-Lex**: TS-Lex is a large-scale sentiment lexicon built using a representation learning approach from a set of tweets collected with the Twitter API [159]. It contains a list of 178,782 positive terms scored in the range (0, 1] and a list of 168,846 negative terms scored in the range [-1, 0).

The negation process also affects the lexicons. We added the suffix "_NEG" to each term in each lexicon, negated its score and appended it to the lexicon.

3.3.3 Features

SentiRich uses four types of features: *basic text*, *syntactic*, *lexicon* and *cluster*. These features are described below:

Basic text features (BTF)

These basic features are extracted from the text. They are the following:

- **Bag of words (BoW)**: Bag of words or n-grams features introduce some contextual information. The presence or absence of contiguous sequences of 1, 2, 3, and 4 tokens are used to represent the tweets.

- **Bag of negated words (BonW)**: Negated contexts are important keys in the sentiment analysis problem. Thus, we used the presence or absence of contiguous sequences of 1, 2, 3 and 4 tokens in the negated contexts as a set of features to represent the tweets.

Syntactic features (SF)

Syntactic features are useful to discriminate between neutral and non-neutral texts.

- **Part of speech (POS)**: Subjective and objective texts have different POS tags [113]. According to [191], non-neutral terms are more likely to exhibit the following POS tags in Twitter: nouns, adjectives, adverbs, abbreviations and interjections. The number of occurrences of each part of speech tag is used to represent each tweet.

- **Bi-tagged**: Bi-tagged features are extracted by combining the tokens of the bi-grams with their POS tag e.g. "*feel*_VBP *good*_JJ". It has been shown in the literature that adjectives and adverbs are subjective in nature and they help to increase the degree of expressiveness[1, 115].

Lexicon features (LF)

Opinion lexicons play an important role in sentiment analysis systems, and the majority of the existing systems rely heavily on them [129]. For each of the seven chosen lexicons, a tweet is represented by calculating the following features: (1) *tweet polarity*, (2) *the average polarity of the positive terms*, (3) *the average polarity of the negative terms*, (4) *the score of the last positive term*, (5) *the score of the last negative term*, (6) *the maximum positive score* and (7) *the minimum negative score*.

The polarity of a tweet T given a lexicon L is calculated using the function 3.2. First, the tweet is tokenized using the Ark Tweet NLP tokenizer and tagger tool. Then, the number of positive tokens (*PosCount*) and the number of negative tokens (*NegCount*) found in the lexicon are counted. Finally, the polarity measure is calculated.

$$polarity = \begin{cases} 1 - \frac{NegCount}{PosCount} & \text{if, } PosCount > NegCount \\ 0 & \text{if, } PosCount = NegCount \\ \frac{PosCount}{NegCount} - 1 & \text{if, } PosCount < NegCount \end{cases} \tag{3.2}$$

Cluster features (CF)

Representing the tweets by mapping each tweet to a set of clusters has proved to be an efficient method in text classification tasks in general and in sentiment identification in particular. We used two set of clusters in this work. The first one is the well known set of clusters provided by the Ark Tweet NLP tool which contains 1000 clusters produced with the Brown clustering algorithm from 56M English-language tweets. These 1000 clusters are used to represent each tweet by mapping each word in the tweet to its cluster. The second

one is *Word2vec cluster ngrams*, which is provided by [41]. They used the word2vec tool to learn 40-dimensional word embeddings of 255,657 words from a Twitter dataset and the K-means algorithm to cluster them into 4960 clusters.

3.3.4 Classifier

Up to now, support vector machines (SVM) [38] have been used widely and reported as the best classifier in the sentiment analysis problem. Thus, we trained a SVM classifier on the Twitter2013 train and dev sets of SemEval2013. We used the linear kernel with the value 0.5 for the cost parameter C.

3.4 Experiments and results

3.4.1 Datasets

We evaluated the effectiveness of our method by using it in the supervised task of Twitter sentiment classification (message-level) presented in SemEval 2015 [128]. The numerical description of the datasets used in SemEval 2015 is shown in table 3.1. The training and development datasets from Twitter 2013 were used to train the model whereas the Twitter's 2013 test set was used as a validation set.

Corpus	#Positive	#Negative	#Neutral	Total
Twitter2013-train	3662	1466	4600	9728
Twitter2013-dev	575	340	739	1654
Twitter2013-test [T13]	1572	601	1640	3813
SMS 2013	492	394	1207	2093
Twitter2014-test [T14]	982	202	669	1853
LiveJournal2014 [LJ14]	427	304	411	1142
Twitter2015-test [T15]	1040	365	987	2392

Table 3.1: Numerical description of the set of tweets.

3.4.2 Evaluation metric

We used the $F1_{pn}$ score as an evaluation metric in all the experiments. If $F1_p$ and $F1_n$ are the F1[1] scores for the *positive* and *negative* classes respectively, then the $F1_{pn}$ score is calculated as follows: $F1_{pn} = \frac{F1_p + F1_n}{2}$.

[1] F1 is the harmonic mean of precision and recall.

	T13	SMS 2013	T14	LJ14	T15
A. SentiRich					
All features	**72.82**	63.55	**71.23**	**73.02**	**68.00**
B. All features -					
All - BTF	70.75	58.97	70.71	70.69	66.63
All- SF	72.18	61.49	71.01	70.93	66.63
All - LF	65.97	56.20	65.85	65.06	61.66
All - CF	72.63	62.89	69.95	70.42	66.42
C. Single features set					
BTF	63.74	55.09	64.26	61.94	58.74
SF	55.68	43.27	55.17	52.12	51.45
LF	67.96	59.62	68.40	67.41	63.27
CF	62.37	52.77	62.21	63.18	57.55
D. State of the art systems					
NRC (all features)	69.02	**68.46**			
TeamX	72.12	57.63	70.96	69.44	
Webis	68.49	63.92	70.86	71.64	64.84

Table 3.2: The experiments results.

3.4.3 Results and discussions

We compared our system with the top three systems of SemEval 2013, 2014 and 2015. The rows under "A" in Table 3.2 show the results obtained by applying the proposed method on all the features, whereas the rows under "B" and "C" report the $F1_{pn}$ score obtained after removing one feature set and the results obtained when considering only one feature set respectively with the purpose of studying and analyzing the effectiveness of each feature set separately. Finally, we listed in the rows under "D" the results reported in the papers of the compared systems. The best results for each dataset are shown in bold.

In Table 3.2 it can be seen that SentiRich yields the bets performance in four cases: Twitter's 2013, 2014, 2015 and Live journal's 2014 tweets. NRC obtains the best performance with the SMS 2013 dataset.

An interesting observation is the consistency of the sets of features used in this work. This property is concluded from the results reported under "B" as it is clearly shown that there is not any improvement in any case when a feature set is removed.

Given the results in section B, the lexicon features seem to be the most relevant ones (their removal is the one that produces a bigger fall in the performance). This fact is confirmed by the results in section C, that show that these features are the ones which give a best individual score. The best results of the individual features are underlined in Table 3.2.

Another experiment in which we remove the new features proposed in this work ("BonW", the new "Polarity measure" shown in function 3.2, and the "Bi-tagged grams") has been conducted to evaluate their effectiveness. The results reported in Table 3.3 show that, in general, those features helped to improve the classifier's performance. The cells marked in bold indicate the stronger decrease of the classifier's performance when we removed one of those features.

	T13	SMS 2013	T14	LJ14	T15
All features - BonW	**72.38**	**62.32**	70.72	73.01	67.03
All features - Bi_tagged	72.44	63.00	**69.90**	**72.35**	**66.85**
All features - Polarity measure	72.69	63.22	70.99	72.73	67.60

Table 3.3: The results obtained after removing BonW, BiTagged or the polarity measure.

3.5 SiTAKA: An Extension of SentiRich

This section describes SiTAKA, an extended version of SentiRich, that has been used in the task 4A Sentiment Analysis in Twitter of SemEval2017 (English and Arabic languages). In addition to the four types of features used in SentiRich, SiTAKA proposes the representation of tweets using a set of features extracted from pre-trained word embedding models. As described in Chapter 2, an embedding space/model can be defined as a lookup-table or a matrix, where the key is a word, and the value is a vector that represents that word. In this version of the system, we used summation, standard deviation, minimum and maximum pooling functions [36] to obtain the tweet representation in the embedding space. The pooling function is an element-wise function, and it converts texts with various lengths into a fixed-length vector allowing to capture the information throughout the entire text. The final vector that represents the tweet is the concatenation of vectors derived from different pooling functions. More formally, let us consider an embedding matrix $E \in \mathbb{R}^{d \times N}$ and an input text $T = w_1, w_2, ..., w_n$, where d is the dimension size, N is the size of the vocabulary (i.e. the number of words in the embedding model), w_i is the ith word in the text and n is the number of words. First, each word w_i is substituted by the corresponding vector v_i^j in the matrix E where j is the index of the word w_i in the vocabulary. This step ends with the matrix $W \in \mathbb{R}^{d \times n}$. The vector $V_{T,E}$ is computed using the following formula:

$$V_{T,E} = \bigcup_{pool \in Funcs} pool_i^n v_i \tag{3.3}$$

where $\bigcup$ denotes the concatenation operation, and $Funcs = \{sum, std, max, min\}$ is the set of pooling functions.

The following subsections describe the word embedding models used in the English (*En*-SiTAKA) and Arabic (*Ar*-SiTAKA) versions of the proposed system, the preprocessing and normalization steps of Ar-SiTAKA and the results obtained with these two versions.

3.5.1 *En*-SiTAKA

We used two pre-trained embedding models in En-SiTAKA. The first one is word2vec which is provided by Google. It was trained on part of the Google News dataset (about 100 billion words). It contains 300-dimensional vectors for 3M words and phrases [98]. The second one is SSWEu, which was trained to capture the sentiment information of sentences as well as the syntactic contexts of words [160]. The SSWEu model contains 50-dimensional vectors for 100K words.

3.5.2 *Ar*-SiTAKA

Sentiment Lexicons

In this version of the SiTAKA system we used four lexicons [137]: Arabic Hashtag Lexicon, Dialectal Arabic Hashtag Lexicon, Arabic Bing Liu Lexicon and Arabic Sentiment140 Lexicon. The first two were created manually, whereas the rest were translated to Arabic from the English version using Google Translator.

Word Embeddings

In *Ar*-SiTAKA we used the model Arabic-SKIP-G300 [179]. Arabic-SKIP-G300 has been trained on a large corpus of Arabic text collected from different sources such as Arabic Wikipedia, Arabic Giga-word Corpus, Ksucorpus, King Saud University Corpus, Microsoft crawled Arabic Corpus, etc. It contains 300-dimensional vectors for 6M words and phrases.

Preprocessing and Normalization

The preprocessing and normalization steps used in *Ar*-SiTAKA, similar to those used in En-SiTAKA, are the following:

1. *Normalization*: Non-Arabic letters are removed from each tweet in the Arabic-language sets. Words with repeated letters (i.e. elongated) are corrected. URLs and usernames are omitted.

2. *Tokenization and POS tagging*: All Arabic-language tweets are tokenized and tagged using the Stanford Tagger (Green and Manning, 2010).

3. *Negation*: Each tweet is negated by adding a suffix #منفي to each word in the negated context. It is necessary to mention that in Ar-SiTAKA we did not use all the Arabic negation words due to the ambiguity of some of them. For example, the first word ما, is a question mark in the following " ما رأيك فيما حدث؟ - What do you think about what happened?" and it means "which/that" in the following example " إن ما حدث اليوم سيء جداً - The matter that happened today was very bad".

3.5.3 Results

Table 3.4 shows the numerical description of the datasets used in this work. We combined the English-language training sets of SemEval 13-16 and testing sets of SemEval 13-15 and used them to train the En-SiTAKA system. We used the development sets provided by the organizers of SemEval2017 to tune the systems' parameters.

System	Training set	Dev set
En-SiTAKA	27,700	20,632
Ar-SiTaka	2684	671

Table 3.4: Numerical description of the sets of tweets.

The evaluation metrics used by the task organizers were the macro averaged recall (ρ), the F1 averaged across the positives and the negatives $F1^{PN}$ and the accuracy (Acc) (Rosenthal et al., 2017). The systems have been tested on 12,284 English-language tweets and 6100 Arabic-language tweets provided by the organizers. The organizers omitted the golden answers of all the test tweets. We show the official evaluation results of our system along with the top 10 systems in Tables 3.5 and 3.6. Our system ranked 7th among 38 systems in the English-language tweets and 2nd among 9 systems in the Arabic language tweets.

3.6 Conclusion

We have proposed a new set of rich sentimental features for the sentiment analysis of the messages posted on Twitter. A Support Vector Machine classifier has been trained using a set of basic features, information extracted from seven useful and publicly available opinion lexicons, syntactic features and clusters. The proposed system, i.e., SentiRich, has been

#	System	ρ	$F1^{PN}$	Acc
1	DataStories	0.681_1	0.677	0.651
2	BB_twtr	0.681_1	0.685	0.658
3	LIA	0.676_2	0.674	0.661
4	Senti17	0.674_3	0.665	0.652
5	NNEMBs	0.669_4	0.658	0.664
6	Tweester	0.659_5	0.648	0.648
7	INGEOTEC	0.649_6	0.645	0.633
8	**En-SiTAKA**	0.645_7	0.628	0.643
9	TSA-INF	0.643_8	0.620	0.616
10	UCSC-NLP	0.642_9	0.624	0.565

Table 3.5: Results for SemEval-2017 Task 4, sub-task A, English.

#	System	ρ	$F1^{PN}$	Acc
1	NileTMRG	0.583_1	0.610	0.581
2	**Ar-SiTAKA**	0.550_2	0.571	0.563
3	ELiRF-UPV	0.478_3	0.467	0.508
4	INGEOTEC	0.477_4	0.455	0.499
5	OMAM	0.438_5	0.422	0.430
6	LSIS	0.435_5	0.469	0.445
7	1w-StAR	0.431_6	0.416	0.454
9	HLP@UPENN	0.415_7	0.320	0.443

Table 3.6: Results for SemEval-2017 Task 4, sub-task A, Arabic.

evaluated on five sets of tweets used in SemEval 2015. SentiRich outperformed the state-of-the-art systems in four out of those five sets. Extensive analysis has been conducted to evaluate the effectiveness of each kind of feature set used and also to evaluate the new proposed features (i.e. bag of negated words "BonW", the polarity measure and the Bi-tagged). The obtained results confirm the effectiveness of the sentiment lexicons as they played an important role in the improvement of the performance of the classifier.

We have developed an extended version of SentiRich that is called SiTAKA. The extended version includes the representation of the tweets using features extracted from pre-trained embedding models. Two versions of SiTAKA were developed, one for the Arabic language tweets and another one for the English language tweets. We used SiTAKA to participate in the SemEval2017 international competition obtaining the 2nd position in the Arabic version and the 8th position in the English version.

Deep Learning approaches have recently been used to build supervised, unsupervised or even semi-supervised methods to analyze the sentiment of texts and to build efficient opinion lexicons[145, 158, 160].

In the next chapter we describe how we have combined traditional Machine Learning techniques and new Deep Learning methods to build a system that can analyze emotions at different levels of analysis.

Chapter 4

An Ensemble of N-Channels ConvNet and XGboost Regressors for Emotion Analysis of Tweets

4.1 Introduction

Developing systems to automatically determine the intensity of emotions that are expressed in a text has a wide range of applications in several areas such as commerce, public health, and social welfare. For instance, detecting the intensity of emotions from customer's reviews could be useful to businesses to understand the customers' satisfaction with their services and products. Moreover, the recognition of the intensity of emotions could help dialogue systems like tutoring systems to detect how a student is happy, joy, unhappy, bored, hesitant, confident, etc. It could also be useful for detecting whether a student is confused, engaged, or sure when interacting with a tutorial system. One critical role that the intensity of emotions can play in the Medical Informatics tasks is detecting depression or suicidal intent.

However, most of the previous works in the field of Sentiment Analysis have focused on developing systems that only determine the polarity of a given text or the categorical classification of emotions (whether a given piece of text communicates anger, joy, sadness, etc.) [101, 118]. The absence of systems to automatically identify the intensity of the emotions can be attributed to the lack of suitable annotated data and resources for the emotion analysis tasks [101].

In this chapter we propose a system that works on an array of sentiment and emotion analysis tasks in Twitter including the typical sentiment analysis tasks, automatically determining the intensity of emotions and the intensity of sentiment (aka valence) of the users from their

tweets. These tasks were defined by the organizers of SemEval-2018, task1: Affect in Tweets. The organizers also provided annotated datasets for each task in English, Arabic, and Spanish (Mohammad et al., 2018). We define the tasks below:

EI-reg (Emotion Intensity Regression Task): *Given a tweet and an emotion* [1] *E, determine the intensity of E that best represents the mental state of the tweeter with a real-valued score between 0 (least E) and 1 (most E).*

EI-oc (Emotion Intensity Ordinal Classification Task): *Given a tweet and an emotion E, classify the tweet into one of four ordinal classes of intensity of E that best represents the mental state of the tweeter. The ordinal classes range from 0 (low amount) to 3 (high amount).*

V-reg (Sentiment Intensity Regression Task): *Given a tweet, determine the intensity of sentiment or valence V that best represents the mental state of the tweeter with a real-valued score between 0 (most negative) and 1 (most positive).*

V-oc (Sentiment Analysis, Ordinal Classification Task): *Given a tweet, classify it into one of seven ordinal classes, corresponding to various levels of positive and negative sentiment intensity, that best represents the mental state of the tweeter.*

Our system is an ensemble of two different approaches. The first one, called N-Channels ConvNet, is a DL based model, whereas the second one is an XGBoost regressor based on a set of embedding and lexicons-based features. We have developed two versions of the proposed system, one for the English language and the other for the Arabic language tweets.

The rest of the chapter is structured as follows. Section 4.2 gives a brief overview of the tools and the resources used in this work. In section 4.3, we present our proposed system. In Section 4.4, we report the experimental results. Finally, in Section 4.5, the conclusions are presented.

4.2 Resources

The proposed system is based on two main resources: sentiment lexicons and word embedding models. The following subsections describe these resources.

4.2.1 Sentiment Lexicons

As explained in Chapter 3, lexicons play an important role in the development of sentiment analysis systems, and most of the existing systems rely heavily on them [71, 110].

[1] The set of emotions used in this work are anger, fear, sadness and joy.

We used the following lexicons for the English version of our system: AFINN [112], General Inquirer [152], Bing-Liu opinion lexicon (HL) [67], MPQA [170], NRC hashtag sentiment lexicon [104], NRC emotion lexicon (EmoLex) [107], NRC affect intensity lexicon [106], NRC hashtag emotion lexicon and Vader lexicon [107]. More details about each lexicon, such as how it was created, the polarity score for each term, and the statistical distribution of the lexicon, can be found in Chapter 3 and in these papaers [70, 107].

For the Arabic version we used the following lexicons: Arabic Hashtag lexicon, Dialectal Arabic Hashtag lexicon, Arabic Bing Liu lexicon, Arabic Sentiment140 lexicon, and the Arabic translation of the NRC emotion lexicon. Those lexicons were created by [138]. The first two were created manually, whereas the rest were translated into Arabic from the English version using Google Translator.

4.2.2 Word Embeddings

We used two publicly available pre-trained embedding models in the English version of our system. The first one was used in [130]. It was trained using word2vec (skip-gram model) on an unannotated corpus of 20 million English tweets containing at least one emoticon. The second one was provided by Baziotis et al. [15]. It was trained on a big dataset of 330M English Twitter messages, gathered from 12/2012 to 07/2016 and a vocabulary size of 660K words using the Glove algorithm. Additionally, we trained two embedding models on 60M English tweets (30M contain positive emoticons, 30M negative ones). The first one was trained by applying word2vec skip-gram of window size five and filtering words that occur less than four times. The dimensionality of the vector was set to 300. The second one was trained using fastText [19]. The dimensionality of the vector was set to 300.

Similarly, we used two publicly available pre-trained embedding models in the Arabic version of our system and trained another two. The first one is the model Arabic-SKIP-G300, provided by Zahran et al., 2015 [179]. Arabic-SKIP-G300 was trained on a large corpus of Arabic text collected from different sources such as Arabic Wikipedia, Arabic Gigaword Corpus, Ksucorpus, King Saud University-Corpus, Microsoft crawled Arabic Corpus, etc. It contains 300-dimensional vectors for 6M words and phrases. The second one is Twitter-SGAraVec [**?**], which was trained using the word2vec skip-gram algorithm on 66M Arabic tweets and 1B tokens. The dimensionality of the vector was set to 300. Our embedding models were trained on the distant supervision corpus (about 16M Arabic tweets) provided by the organizers of SemEval-2018, task1. We were able to find about 12M tweets. Similar to our English embeddings, we trained these two Arabic embedding models.

Each pre-trained embedding model is used to initialize an embedding layer in our N-Channels ConvNet model. The pre-trained embedding models are also used to extract

features to represent the tweets in combination with features extracted from the sentiment lexicons to train XGBoost regressors, as explained in section 4.3.2.

4.3 System Description

This section explains the proposed system, whose architecture is shown in Figure 4.1. First, we preprocess the tweets as explained in Chapter 3. Afterwards, we pass them to the N-Channels ConvNet and the XGBoost regressors models (Subsections 4.3.1 and 4.3.2). Each model produces a value in $[0, 1]$ that represents the intensity of the emotion that can be inferred fr om the tweet. Finally, we ensemble the output of the two models to get the final result as described in subsection 4.3.3. The proposed system is also used as a feature extractor to train an ordinal Decision Tree classifier as described in subsection 4.3.4.

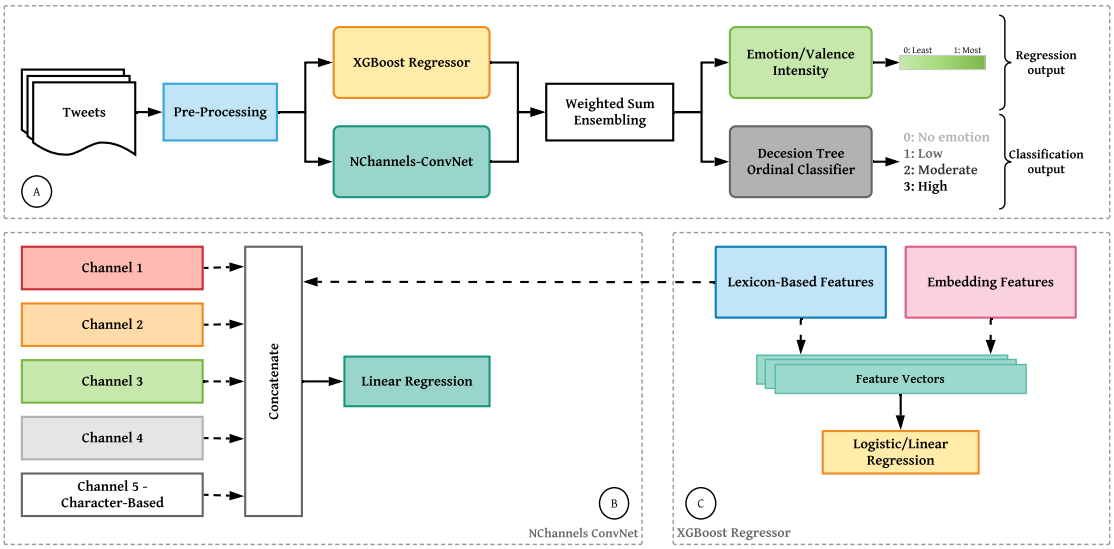


Figure 4.1: EiTAKA Architecture. (A) The overall architecture of the system. (B) The NChannels ConvNet Model. (C) The architecture of the XGBoost model.

4.3.1 N-Channels ConvNet

Convolutional Neural Networks (ConvNets) have achieved remarkable results in computer vision and speech recognition tasks in recent years. This subsection explains the architecture of our proposed ConvNet.

Architecture

The N-Channels ConvNet model architecture, shown in the bottom box in the part (B) of figure 4.1, is inspired by Inception-Net [155] and the CNN proposed by [82]. It is composed of multiple channels followed by a logistic regressor. Figure 4.2 shows the channel architecture. The input to each channel is a sequence of words $w_1, w_2, ...w_n$ where n is the number of words. To deal with the variation in the length of the tweets, we padded the tweet that has a length smaller than n, and truncated the one that has a length greater than n. We pass the input through an embedding layer to map each word w_i into a real-valued vector. Each channel has its own embedding layer which is initialized by a specific pre-trained embedding model. We used five channels with the four pre-trained embedding models described in subsection 2.2 and a character-based one.

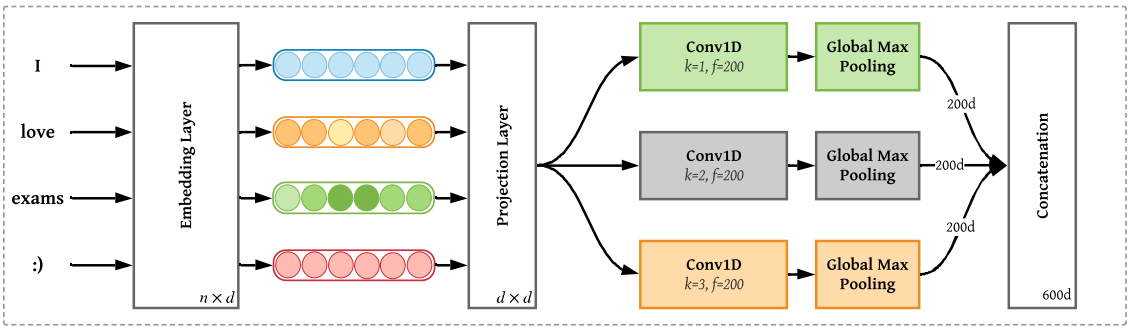


Figure 4.2: Channel Architecture.

The result from the embedding layer is a matrix $n \times d_c$ where d_c is the vector dimension. This matrix is passed to a projection followed by an activation layer. The projection layer is nothing but a fully-connected or dense layer whereas the activation layer can be any non-linear function such as a rectified linear unit (ReLU). Afterward, we feed the projected matrix to three Conv1D. Each one has a different kernel (1, 2, and 3) and 200 filters. To get more details about the architecture of this Conv1D please check [82]. We pass the output of each Conv1D through a global max-pooling layer which produces a vector with dimensionality of 200. Finally, the three vectors are concatenated. This yields a vector with dimensionality of 600 that represents the tweet (i.e. the input sequence of words). All this process is shown in Figure 4.2.

Finally, the outputs of all channels are concatenated with a lexicon-based vector (see next section) and fed to a single sigmoid neuron which gives the intensity of the emotion/valence.

Training

The proposed model was trained by minimizing the mean squared error between the real and predicted intensities. The optimization was done by applying back-propagation through layers via minibatch gradient descent. The training parameters were the following: batch size of 32, 100 epochs and Adam optimization method with learning rate of 0.001, $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and $\varepsilon = 10^{-}9$. To prevent over-fitting, we used dropout and early stopping methods.

4.3.2 XGBoost Regressor

XGBoost [27] has become a widely used and really popular tool among Data Scientists in industry, as it shows great performance on large-scale problems. It is a highly flexible and versatile tool that can work through most regression, classification and ranking problems as well as user-built objective functions.

We trained an XGBoost regressor to give the intensity of the emotion/valence based on lexicon features and embedding features.

Lexicon Features: For each lexicon, we used the sum of the scores provided by the lexicon for each word in the tweet. Let L denote the set of lexicons and $f_i^l(w)$ the score of the word w based on the feature i in the lexicon l (note that some lexicons have only one feature like the sentiment score and some of them have multiple features like anger emotion score, positive score, etc). Then, the set of features that represent a tweet T given a lexicon $l \in L$ can be obtained as follows:

$$V_{T,l} = \forall_{f_i^l \in F_l} \sum_{w \in T} f_i^l(w) \tag{4.1}$$

Here, F_l denotes the set of features in lexicon l.

Embedding Features: We extracted features from the word embedding models to represent the tweets as explained in Chapter 3. Given a tweet T and an embedding model E, the vector $V_{T,E}$ is obtained by applying the sum pooling function.

The final representation of the tweets is obtained by concatenating the two vectors $V_{T,l}$ and $V_{T,E}$.

Training

The XGBoost regressor has some parameters that need to be tuned. Table 4.1 shows the values of each parameter we chose for the different emotions. All those values were chosen using the grid-search on the development sets.

	P	# Est.	S	M	O
Eng.	Anger	300	0.75	5	Logistic
	Fear	300	0.75	5	Linear
	Sadness	300	0.75	5	Logistic
	Joy	300	0.75	7	Linear
	Valence	300	0.75	5	Linear
Ara.	Anger	200	0.9	9	Logistic
	Fear	200	0.9	5	Logistic
	Sadness	200	0.9	5	Logistic
	Joy	200	0.9	5	Logistic
	Valence	200	0.9	9	Logistic

Table 4.1: The XGBoost regressors parameters. #Est. refers to the number of estimators, S is the subsample, M is the maximum depth and O refers to the objective function.

4.3.3 Ensemble

We combined the results of the two systems described above with the intention of improving the performance and increasing the generalizability of the final system. We used the weighted average method to achieve that. Let r_1 and r_2 respectively denote the output of the XGBoost regressor and the N-Channels ConvNet system. The final output r was obtained as follows:

$$r = \alpha * r_1 + (1 - \alpha) * r_2; \quad \alpha \in [0, 1] \tag{4.2}$$

	Emotion	α
Eng.	Anger	0.3
	Fear	0.5
	Sadness	0.6
	Joy	0.2
	Valence	0.6
Ara.	Anger	0.5
	Fear	0.0
	Sadness	0.5
	Joy	0.4
	Valence	0.7

Table 4.2: The value of α for each individual model.

Table 4.2 shows the value of α for each individual model. All these values were obtained by grid search on the development set.

4.3.4 Decision Tree for Ordinal Classification Tasks

To solve the problem of ordinal classification we simply used the proposed model as feature extractor and trained a Decision Tree. The idea is to use the emotion/intensity as input feature and use rules generated from the Decision Tree to get the appropriate class. Figure 4.3 shows as an example of the Decision Tree classifier of the $fear$ emotion.

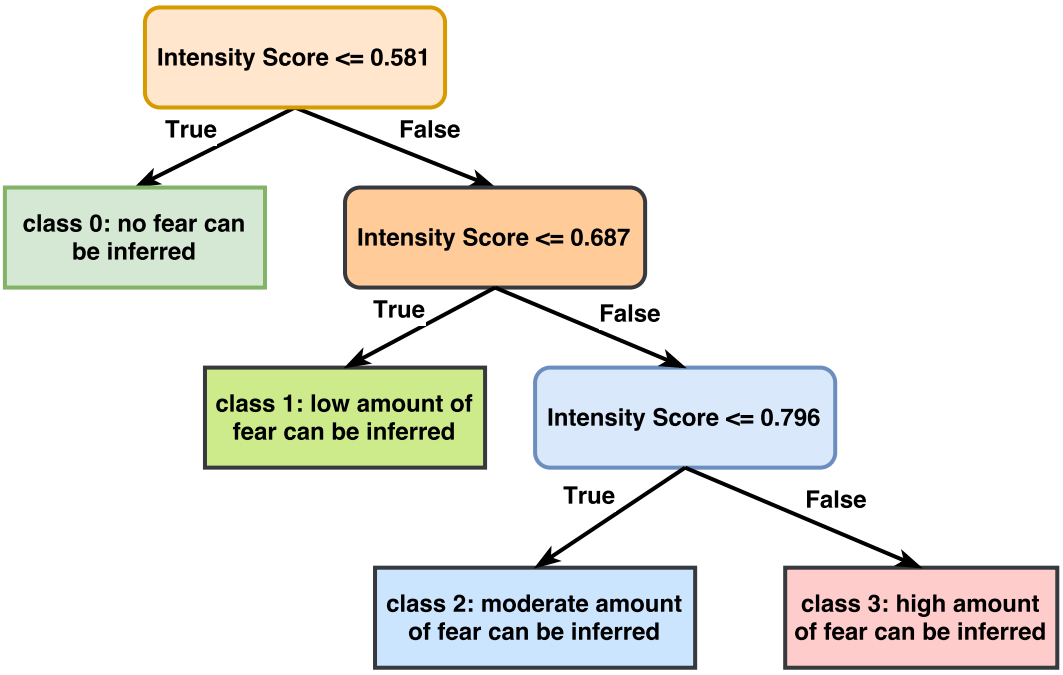


Figure 4.3: An example of a decision tree classifier.

4.4 Results

We trained and validated the proposed system on the training and validation sets of the SemEval-2018 Task 1: Affect in Tweets (AIT). To evaluate the performance of the systems, we applied it to the test sets of AIT. The organizers of the competition omitted the golden labels of the test sets, and we had no access to them. Table 4.3 describes the number of tweets in each dataset of AIT for both languages. The evaluation metrics used were the following:

- The Pearson Correlation Coefficient with the gold labels/scores as the primary metric for all of the tasks. We refer to this metric as Pearson in this chapter.
- The secondary metric used for the regression tasks (i.e. EI-reg and V-reg) was the Pearson correlation for a subset of the test set that includes only those tweets with

intensity score greater or equal to 0.5. We refer to this metric as Pearson5 in this chapter.

- The secondary metric used for the ordinal classification tasks (i.e. EI-oc and V-oc) was Kappa.

More details about the data and the evaluation metrics can be found in [102, 106].

Dataset	Train	Dev	Test	Total
English				
EI-reg, EI-oc				
Anger	1,701	388	1,002	3,091
Fear	2,252	389	986	3,627
Joy	1,616	290	1,105	3,011
Sadness	1,533	397	975	2,905
V-reg, V-oc	1,181	449	937	2,567
Arabic				
EI-reg, EI-oc				
Anger	877	150	373	1,400
Fear	882	146	372	1,400
Joy	728	224	448	1,400
Sadness	889	141	370	1,400
V-reg, V-oc	932	138	730	1,800

Table 4.3: The number of tweets in the SemEval-2018 Affect in Tweets Dataset.

Tables 4.4 and 4.5 show the results of the emotion and valence intensity regression tasks of our two systems and their combination (the ensemble model) in addition to the baseline results which were reported by the organizers of SemEval-2018. We can note from the results that both of our individual systems (i.e., the N-Channels ConvNet and the XGBoost regressor) outperformed the baseline system. The values in the tables also show the superiority of the N-Channels ConvNet over the XGBoost regressor. For instance, the results of the English version of the emotion intensity task show that the N-Channels ConvNet outperforms the XGBoost regressor by 5.9% with respect to macro-avg measure. Moreover, the performance of N-Channels Convnet is very close to the ensemble model. The gain in the case of using the ensemble system is only 1.2% for the English version and is very small (0.3%) in the case of the Arabic version. The results of the Pearson correlation of samples whose intensity score is greater or equal to 0.5 reveal that our system can be used as a classifier. This conclusion is confirmed by the results of the ordinal classification tasks, shown in Tables 4.6 and 4.7.

As we described in subsection 4.3.4, our approach to design a system to solve the ordinal classification tasks was to use the intensity score as input feature to train a Decision Tree.

	Model	Pearson					Pearson5				
		Avg.	A	F	J	S	Avg.	A	F	J	S
English											
	ConvNet	71.2	71.3	72.5	71.8	69.2	53.8	57.5	50.2	51.9	55.5
	XGBoost	65.3	67.4	64.4	62.5	66.8	50.3	56.3	45.5	43.7	55.5
	Ensemble	**72.4**	**73.1**	**73.3**	**72.2**	**71.1**	**56**	**60.6**	**52.2**	**52.5**	**58.7**
	SVM	52	52.6	52.5	57.5	45.3	39.6	45.5	30.2	47.6	35
Arabic											
	ConvNet	65.5	63.9	62.8	70.5	64.8	51.6	47.3	60.5	46.5	52
	XGBoost	59.6	49.4	54	71.3	63.7	46.4	37.6	49.2	44.9	54
	Ensemble	**66.7**	**62.7**	**62.7**	**73.8**	**67.5**	**53.3**	**47.9**	**60.4**	**49**	**56**
	SVM	45.5	40.6	43.5	53	45	35.3	34.4	36.6	33.2	36.7

Table 4.4: EI-reg task results. A = Anger, F = Fear, J = Joy and S = Sadness.

Language	Model	Pearson	Pearson5
English			
	ConvNet	82.5	64.5
	XGBoost	76.8	59.8
	Ensemble	**82.8**	**65.8**
	SVM	58.5	44.9
Arabic			
	ConvNet	81.7	55
	XGBoost	77.4	57.1
	Ensemble	**82.8**	**57.8**
	SVM	57.1	42.3

Table 4.5: V-reg task results.

	Model	Pearson					Kappa				
		Avg.	A	F	J	S	Avg.	A	F	J	S
English											
	Ours	63.3	65.1	59.5	65.1	63.6	60.8	61.9	57.4	60.7	63.2
	SVM	39.4	38.2	35.5	46.9	37	38.5	37.5	33.1	46.5	37
Arabic											
	Ours	57.4	57.2	52.9	63.4	56.3	54.2	54.7	51.6	58.8	51.8
	SVM	31.5	28.1	28.1	28.1	39.6	29.9	27.6	24.9	38.5	28.7

Table 4.6: EI-oc task results.

Language	Model	Pearson	Kappa
English			
	Ours	**79.6**	**79.1**
	SVM	50.9	50.4
Arabic			
	Ours	**80.9**	**78.3**
	SVM	47.1	47

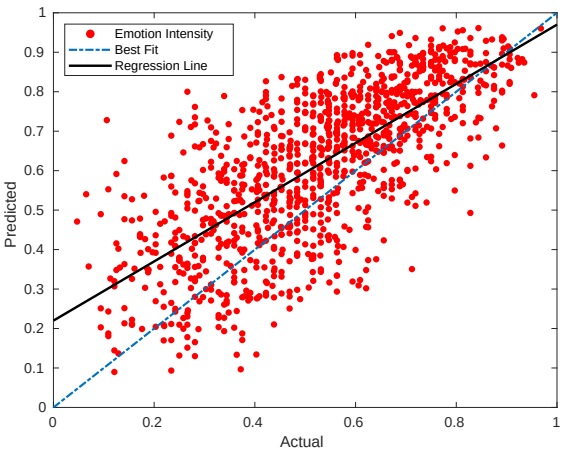
Table 4.7: V-oc task results.

During the inference phase we used our system to produce the intensity score for the new (unseen) samples (i.e. use it as feature extractor). Thus, the performance in this phase heavily relies on the performance of the proposed system. This is clearly shown in the results reported in tables 5 and 6. For example, our system gives very good results in the valence intensity regression task for both the English and Arabic versions (the Pearson correlation is 0.828 for both). This affects positively the performance of our system for the valence ordinal classification tasks (the Pearson correlation is about 0.80 for both).

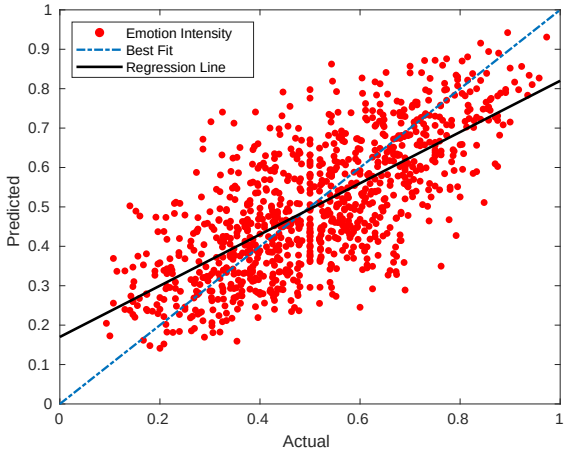
For further analysis, we created scatter plots to visualize the correlation between the actual values and those produced by our systems. We also plot the line of best fit. The scatter plots are shown in Figures 4.4, 4.5 and 4.6. Looking at the scatter plots, we can notice that, in general, our system consistently overestimates the intensity when the gold label score is low and underestimates the intensity when the gold label score is high. As exceptions to this observation, the scatter plots of the Joy emotion intensities in the English test set and the Valance intensities in the Arabic set show that as the gold score goes to 1, the variance between those scores and the predicted ones is reduced. The variance in the actual emotions intensities and the predicted ones in the Arabic set is larger than in the English one. The performance of the system in the valence regression task is markedly better in comparison to the performance in the emotion intensity regression tasks, as seen from the Pearson correlation, in Table 4.5, as well as the line of best fit, in Figure 4.6.

4.5 Conclusion

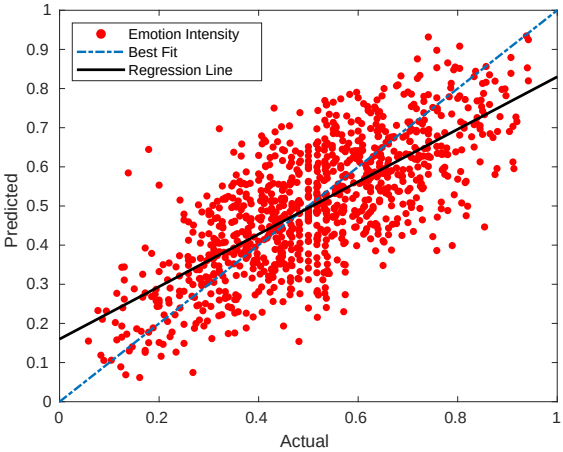
We have proposed a system that works on an array of sentiment and emotion analysis tasks in Twitter including the typical sentiment analysis tasks, automatically determining the intensity of emotions and the intensity of sentiment (aka valence) of the users from their tweets. The proposed system is an ensemble of two different approaches. The first one, called N-Channels ConvNet, is a Deep Learning approach whereas the second one is an



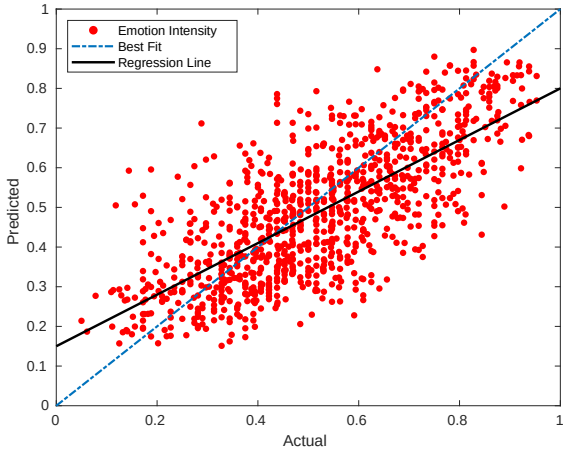
(a) Joy



(b) Sad

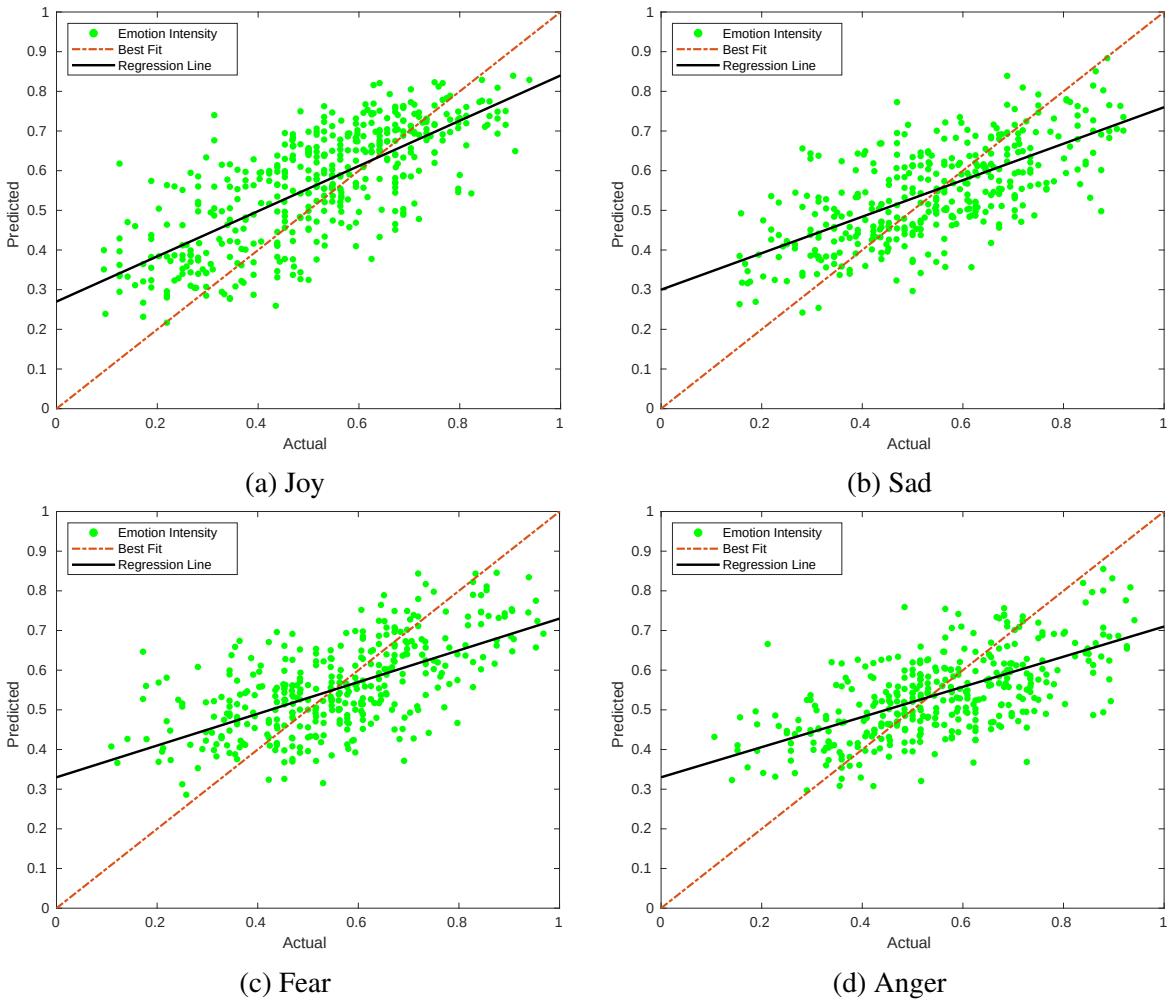


(c) Fear



(d) Anger

Figure 4.4: The correlation between the actual intensities of the emotions and the output of our system for the English test set.



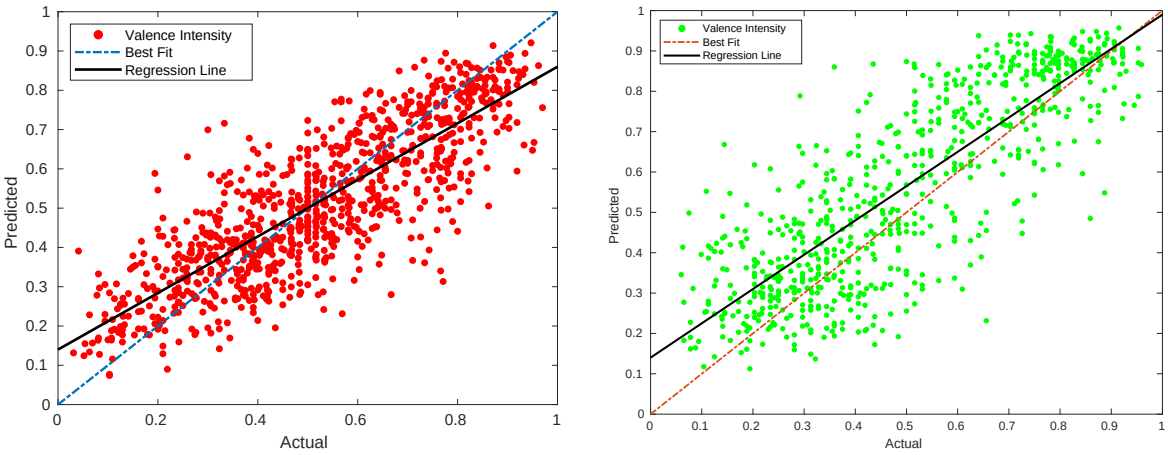
(a) Joy

(b) Sad

(c) Fear

(d) Anger

Figure 4.5: The correlation between the actual intensities of the emotions and the output of our system for the Arabic test set.



(a) English (b) Arabic

Figure 4.6: The correlation between the actual valance intensities and the output of our system for the English and the Arabic test sets.

XGBoost regressor based on a set of embedding and lexicons-based features. The ensemble technique helped to improve the performance of the final model in all subtasks. We have realized that the N-Channels ConvNet gives a performance very close to the ensemble model. This observation confirms the fact that Deep Learning models, and especially ConvNets, have achieved remarkable results in many fields such as computer vision, speech recognition and natural language processing. This chapter focused on developing systems to analyse the emotions in tweets independently (i.e., one seperate model for each emotion). However, in many real cases, a tweet might be assigned to multiple emotions jointly, espicially when there is a strong correlation between emotions. The next chapter proposes a novel method to solve the problem of multi-label emotion classification.

Chapter 5

A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets

5.1 Introduction

We showed in the previous chapters the need of analysing the emotions expressed in text in many fields such as commerce, public health, social welfare and tutoring systems. Emotions are the key to people's feelings and thoughts. However, most of the previous works on sentiment and emotion analysis have only focused on analysing sentiments and emotions independently. Systems or methods have been developed to solve the problem of sentiment and emotion analysis as single-label classification neglecting the correlation between different emotions.

Recently, the multi-label classification problem has attracted considerable interest due to its applicability to a wide range of domains, including text classification, scene and video classification, and bioinformatics [127]. Unlike the traditional single-label classification problem, where an instance is associated with only one label from a finite set of labels, in the multi-label classification problem, an instance is associated with a subset of labels.

Hence, in this chapter, we focus on the multi-label emotion classification task, which aims to develop an automatic system to determine the existence in a text of none, one, or more out of eleven emotions: the eight Plutchik [120] categories (joy, sadness, anger, fear, trust, disgust, surprise, and anticipation) that are shown in Figure 5.1, plus optimism, pessimism, and love.

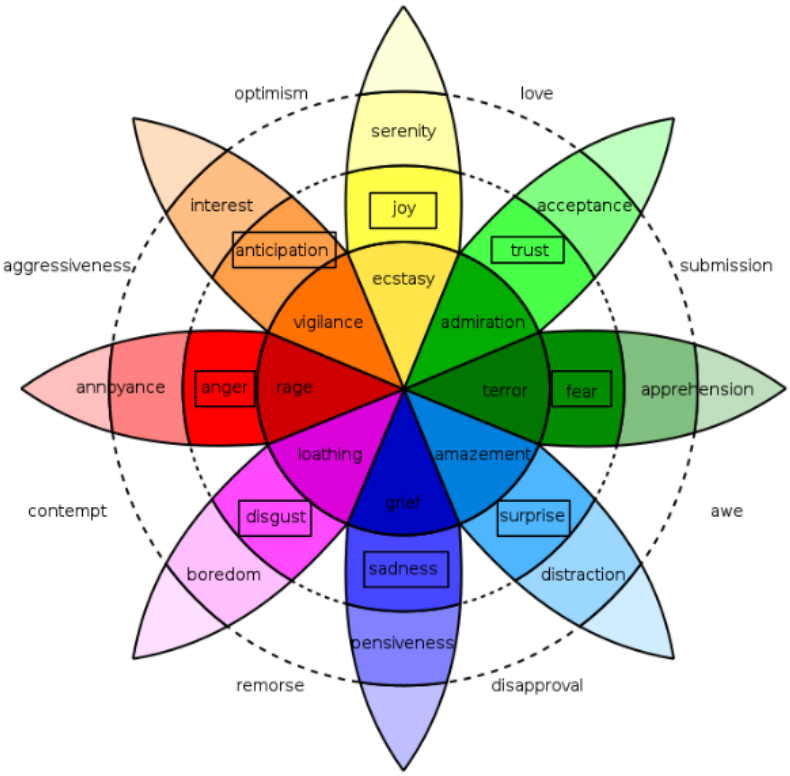


Figure 5.1: The set of the eight basic emotions proposed by Plutchik [120].

One of the most common approaches to addressing the problem of multi-label classification is the *problem transformation*. With this approach, a multi-label problem is transformed into one or more single-label (i.e., binary or multi-class) problems. Specifically, single-label classifiers are learned and employed; after that, the classifiers' predictions are transformed into multi-label predictions.

Different transformation methods have been proposed in the multi-label literature. The most common method is called *binary relevance* [126, 163]. The idea of the binary relevance method is simple and intuitive. A multi-label problem is transformed into multiple binary problems, one problem for each label. Then, an independent binary classifier is trained to predict the relevance of each individual label. Although binary relevance is popular in the literature, due to its simplicity, it suffers from directly modeling correlations that may exist between labels. However, it is highly resistant to overfitting label combinations, since it does not expect examples to be associated with previously-observed combinations of labels.

In this chapter, we propose a novel transformation method, called *xy*-pair-set, for the multi-label classification problem. Unlike binary relevance methods, our method transforms the problem into only one binary classification problem as described in Section 5.3. Additionally,

we exploit the successes of Deep learning models, especially the word2vec methods' family [98] and the recurrent neural networks [34, 144] and attention models [3, 165], to develop a system that solves the transformed binary classification problem. The critical component of our system is the embedding module, which uses three embedding models and an attention function to model the relationship between the input and the label.

To summarize, the contribution of this chapter is four-fold.

- We propose a novel transformation mechanism for the multi-label classification problem.
- We propose a novel, attentive Deep learning system, which we call BNet , which works on the new transformation method. Our system is a data-driven, end-to-end neural-based model, and it does not rely on external resources such as parts of speech taggers and sentiment or emotion lexicons.
- We evaluate the proposed system on the challenging multi-label emotion classification dataset of SemEval-2018 Task1: Affect in Tweets.
- The experimental results show that our system outperforms the state-of-the-art systems.

The rest of the chapter is structured as follows. In Section 5.2, we overview the related work on multi-label problem transformation methods and Twitter emotion analysis. In Section 5.3, we explain in detail the methodology. In Section 5.4, we report the experimental results. In Section 5.5, the conclusions are presented.

5.2 Related Works

In this section, we overview the most popular research studies related to this work. In Section 5.2.1, we summarize the most common multi-label problem transformation methods. Section 5.2.2 gives an overview of the state-of-the-art works on the problem of multi-label emotion classification on Twitter.

5.2.1 Problem Transformation Methods

Let $X = \{x_1, x_2, \ldots x_n\}$ be the set of all instances and $Y = \{y_1, y_2, \ldots, y_m\}$ be the set of all labels. We can define the set of data:

$$D = \{(x_i, \hat{Y}_i) | x_i \in X \text{ and } \hat{Y}_i \subseteq Y \text{ is the set of labels associated with } x_i\} \tag{5.1}$$

In this expression, D is called a supervised multi-label dataset.

The task of multi-label classification is challenging because the number of label sets grows exponentially as the number of class labels increases. One common strategy to address this issue is to transform the problem into a traditional classification problem. The idea is to simplify the learning process by exploiting label correlations. Based on the order of the correlations, we can group the existing transformation methods into three approaches [184, 186], namely first-order approaches, second-order approaches, and high-order approaches.

First-order approaches decompose the problem into some independent binary classification problems. In this case, one binary classifier is learned for each possible class, ignoring the co-existence of other labels. Thus, the number of independent binary classifiers needed is equal to the number of labels. For each multi-label training example $(x_i, \hat{Y}_i) \in D$, $y_k \in Y$, we construct a binary classification training set, D_k as the following: x_i will be regarded as one positive example if $y_k \in \hat{Y}_i$ and one negative example otherwise. In the first case, we will get a training example in the form $(x_i, 1) \in D_k$, which will be $(x_i, 0) \in D_k$ in the second case. Thus, for all labels $\{y_1, y_2, \ldots, y_m\} \in Y$, m training sets $\{D_1, D_2, \ldots, D_m\}$ are constructed. Based on that, for each training set D_k, one binary classifier can be learned with popular learning techniques such as AdaBoost [142], k-nearest neighbor [185], decision trees, random forests [35, 39], etc. The main advantage of first-order approaches is their conceptual simplicity and high efficiency. However, these approaches can be less effective due to their ignorance of label correlations.

Second-order approaches try to address the lack of modeling label correlations by exploiting pairwise relationships between the labels. One way to consider pairwise relationships is to train one binary classifier for each *pair* of labels [96]. Although second-order approaches perform well in several domains, they are more complicated than the first-order approaches in terms of the number of classifiers. Their complexity is quadratic, as the number of classifiers needed is $\binom{m}{2}$. Moreover, in real-world applications, label correlations could be more complex and go beyond second-order.

High-order approaches tackle the multi-label learning problem by exploring high-order relationships among the labels. This can be fulfilled by assuming linear combinations [31], a nonlinear mapping [55, 177], or a shared subspace over the whole label space [174]. Although high-order approaches have stronger correlation-modeling capabilities than their first-order and second-order counterparts, these approaches are computationally demanding and less scalable.

Our transformation mechanism, shown in Section 5.3.1, is as simple as the first-order approaches and can model, implicitly, high-order relationships among the labels if some

requirements, detailed in Section 5.3.1, are fulfilled. It requires only one binary classifier, and the number of training examples grows polynomially in terms of the number of instances and the number of labels. If the number of training examples in the multi-label training dataset is n and the number of the labels is m, then the number of the training examples in the transformed binary training set is $n \times m$.

5.2.2 Emotion Classification in Tweets

Various Machine Learning approaches have been proposed for traditional emotion classification and multi-label emotion classification. As shown in the previous chapters, most of the existing systems solve the problem as a text classification problem. Supervised classifiers are trained on a set of annotated corpora using different sets of hand-engineered features. The success of such models is heavily based on the intelligent design of features that can distinguish between the samples. With this approach, most studies have focused on engineering a set of efficient features to obtain a good classification performance [70, 71, 102]. However, it has been shown that features engineering is not an easy task, and it is time-consuming [14, 36, 56]. For example, it is well known that sentiment and emotion lexicons play an essential role in developing efficient sentiment and emotion analysis systems; however, it is challenging to create such lexicons. Also, finding the best combination of lexicons in addition to the best set of statistical features is not a straightforward task.

Recently, Deep learning models have been utilized to develop end-to-end systems in many tasks including speech recognition, text classification, and image classification. It has been shown that such systems automatically extract high-level features from raw data [56, 157].

Baziotis et al. [14], the winner of the multi-label emotion classification task of SemEval-2018 Task1: Affect in Tweets, developed a bidirectional Long Short-Term Memory (LSTM) with a deep attention mechanism. They trained a word2vec model with 800,000 words derived from a dataset of 550 million tweets. The second place winner of the SemEval leaderboard trained a word-level bidirectional LSTM with attention, and it also included non-Deep learning features in its ensemble [95]. Ji Ho Park et al. [116] trained two models to solve this problem: regularized linear regression and logistic regression classifier chain [126]. They tried to exploit labels' correlation to perform multi-label classification. With the first model, the authors formulated the multi-label classification problem as a linear regression with label distance as the regularization term. In their work, the logistic regression classifier chain method was used to capture the correlation of emotion labels. The idea is to treat the multi-label problem as a sequence of binary classification problems by taking the prediction of the previous classifier as an extra input to the next classifier.

In this work, we exploited the Deep learning-based approach to develop a system that can extract a high-level representation of the tweets and model an implicit high-order relationship among the labels. We used the proposed system alongside the proposed transformation method to train a function that can solve the problem of multi-label emotion classification in tweets. The next section explains the details of our proposed system.

5.3 Methodology

This section shows the methodology of this work. First, we explain in Section 5.3.1 the proposed transformation method, *xy*-pair-set. Afterwards, we describe the proposed system in Section 5.3.2.

5.3.1 *xy*-Pair-Set: Problem Transformation

The proposed transformation method *xy*-pair-set transforms a multi-label classification dataset D into a supervised binary dataset $\hat{D}$ as follows:

$$\forall x_i \in X, y \in Y \text{ and } (x_i, \hat{Y}_i) \in D, \exists!((x_i, y), \phi) \in \hat{D}$$
$$\text{where } \phi = \begin{cases} 1 & \text{if } y \in \hat{Y}_i \\ 0 & otherwise \end{cases} \tag{5.2}$$

Algorithm 1 explains the implementation of the proposed transformation method. It takes as inputs a multi-label dataset D (Equation (5.2)) and a set of labels Y, and it returns a transformed binary dataset. We show next an illustrative example.

Let $X = \{x1, x2\}$, $Y = \{a, b, c\}$ and $D = \{(x1, \{a, c\}), (x2, \{b\})\}$. The output of the binary relevance transformation method is a set of three independent binary datasets, one for each label. That is, $D_a = \{(x1, 1), (x2, 0)\}$, $D_b = \{(x1, 0), (x2, 1)\}$, and $D_c = \{(x1, 1), (x2, 0)\}$. In contrast, the output of our transformation method is a single binary dataset $\hat{D} = \{((x1, a), 1), ((x1, b), 0), ((x1, c), 1), ((x2, a), 0), ((x2, b), 1), ((x2, c), 0)\}$.

The task in this case, unlike the traditional supervised binary classification algorithms, is to develop a learning algorithm to learn a function $g : X \times Y \rightarrow \{0, 1\}$. The success of such an algorithm is based on three requirements: (1) an encoding method to represent an instance $x \in X$ as a high-dimensional vector V_x, (2) a method to encode a label $y \in Y$ as a vector V_y, and (3) a method to represent the relation between the instance x and the label y. These three conditions make g able to capture the inputs-to-labels and labels-to-labels relationships. In

Algorithm 1: *xy*-pair-set algorithm.

Input: Input: a multi-label classification dataset D and a set of labels Y
Output: Output: a binary classification dataset $\hat{D}$

```
1  D̂ = {};
2  foreach (x_i, Ŷ_i) ∈ D do
3      foreach y ∈ Y do
4          x̂_i = (x_i, y) ;                    ▷ a tuple of x_i and y.
5          if y ∈ Ŷ_i then
6              D̂ = D̂ ⋃ (x̂_i, 1) ;
7          else
8              D̂ = D̂ ⋃ (x̂_i, 0) ;
9          end
10     end
11 end
12 return D̂
```

this work, we take advantage of the successes of Deep learning models to fulfill the three requirements listed above. We empirically show the success of our system with respect to these conditions as reported in section 5.4.

5.3.2 BNet: System Description

This subsection explains the proposed system to solve the transformed binary problem mentioned above. Figure 5.2 shows the graphical depiction of the system's architecture. It is composed of three parts: the embedding module, the encoding module, and the classification module. We explain in detail each of them below.

Embedding Module

Let (W, y) be the pair of inputs to our system, where $W = \{w_1, w_2, \ldots, w_l\}$ is the set of the words in a tweet and y is the label corresponding to an emotion. The goal of the embedding module is to represent each word w_i by a vector v_{w_i} and the label by a vector v_y.

Our embedding module can be seen as a function that maps a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key. The query is the trainable label embedding, $E^Q(y)$, the keys are the pretrained words embeddings, $E^K(w_i)$ $\forall w_i \in W$, and the values are the trainable words embeddings, $E^V(w_i)$ $\forall w_i \in W$.

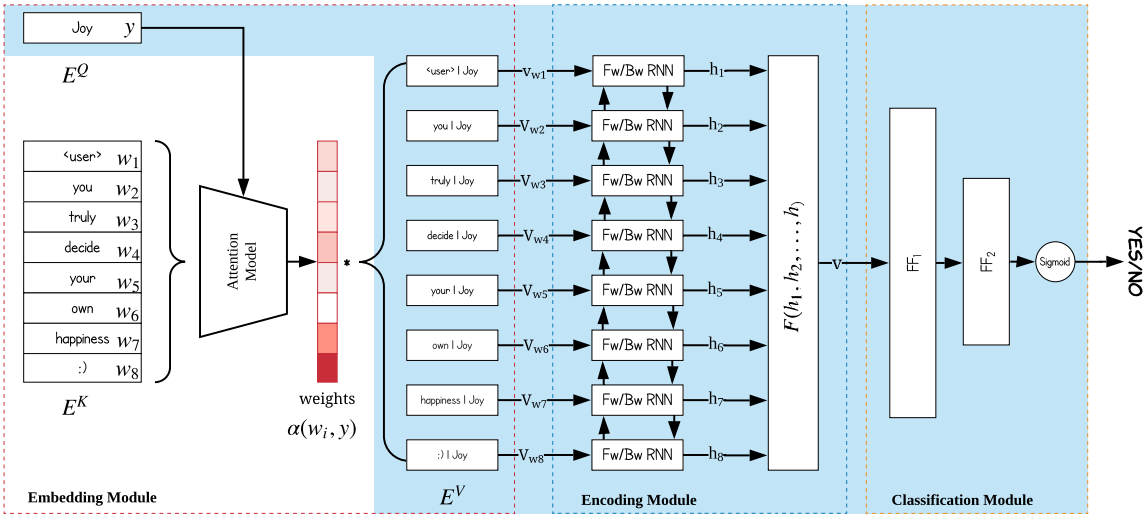


Figure 5.2: An illustration of the proposed system. Shaded parts are trainable.

As shown in Figure 5.2, we used the output of E^Q and E^K as inputs to the attention model to find the alignments, i.e., the weights α, between the label y and the words W of the input tweet. This step models the relation between the input and the label. As soon as the weights were obtained, we then multiplied each word's vector that came from the embedding E^V by its corresponding weight. Given that, the final representation of a word $w_i \in W$ is the following:

$$v_{w_i} = E^V(w_i) \cdot \alpha(w_i, y) \tag{5.3}$$

The function α is an attention-based model, which finds the strength of the relationship between the word w_i and the label y based on their semantic similarity. That is, $\alpha(w_i, y)$ is a value based on the distance $\Delta(w_i, y)$ between w_i and y as:

$$\alpha(w_i, y) = \frac{e^{\Delta(w_i, y)}}{\sum_{w_j \in W} e^{\Delta(w_j, y)}} \tag{5.4}$$

Here, $\Delta(w_i, y)$ is a scalar score that represents the similarity between the word w_i and the label y:

$$\Delta(w_i, y) = E^k(w_i) \cdot {v_y}^T \tag{5.5}$$

$$v_y = E^Q(y) \tag{5.6}$$

It is worth noting that $\alpha(w_i, y) \in [0,1]$ and:

$$\sum_{w_i \in W} \alpha(w_i, y) = 1 \tag{5.7}$$

Encoding Module

The goal of the encoding module is to map the sequence of word representations $\{v_{w_1}, v_{w_2}, \ldots, v_{w_l}\}$ that is obtained from the embedding module to a single real-valued dense vector. In As shown in Figure 5.2, the encoder used in this work is a BiRNN with two GRUs one as forward $\overrightarrow{\phi}$ and the other as backward $\overleftarrow{\phi}$ RNNs.

As we explained in chapter 2, the forward RNN , i.e., $\overrightarrow{\phi}$, reads the input sequence in a forward direction and produces a sequence of forward hidden states $(\overrightarrow{h_1}, \ldots, \overrightarrow{h_l})$, whereas the backward one , i.e., $\overleftarrow{\phi}$, reads the sequence in the reverse order $(v_{w_l}, \ldots, v_{w_1})$, resulting in a sequence of backward hidden states $(\overleftarrow{h_l}, \ldots, \overleftarrow{h_1})$.

We obtained a representation for each word v_{w_t} by concatenating the corresponding forward hidden state $\overrightarrow{h_t}$ and the backward one $\overleftarrow{h_t}$. The following equations illustrate the main ideas:

$$\overrightarrow{h_t} = \overrightarrow{\phi}(v_{w_t}, \overrightarrow{h_{t-1}}) \tag{5.8}$$

$$\overleftarrow{h_t} = \overleftarrow{\phi}(v_{w_t}, \overleftarrow{h_{t+1}}) \tag{5.9}$$

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{5.10}$$

The final input representation of the sequence is:

$$c = F(\{h_1, h_2, \ldots, h_l\}) \tag{5.11}$$

We simply chose F to be the last hidden state (i.e., $F(\{h_1, h_2, \ldots, h_n\}) = h_n$).

Classification Module

Our classifier was composed of two feed-forward layers with the *ReLU* activation function followed by a Sigmoid unit.

5.4 Experiments and Results

In this section, we first describe the dataset and the pre-processing we used, and then, we describe the experimental details. Afterwards, we introduce the state-of-the-art systems

we compared our system with, and finally, we report the empirical validation proving the effectiveness of our system.

5.4.1 Dataset

In our experiments, we used the multi-label emotion classification dataset of SemEval-2018 Task1: Affect in Tweets [102]. It contains 10,983 samples divided into three splits: training set (6838 samples), validation set (886 samples), and testing set (3259 samples). For more details about the dataset, we refer the reader to [103]. We trained our system on the training set and used the validation set to fine-tune the parameters of the proposed system. We pre-processed each tweet in the dataset as follows:

- Tokenization: We used an extensive list of regular expressions to recognize the following meta information included in tweets: Twitter markup, emoticons, emojis, dates, times, currencies, acronyms, hashtags, user mentions, URLs, and words with emphasis.

- As soon as the tokenization was done, we lowercased words and normalized the recognized tokens. For example, URLs were replaced by the token "<URL>", and user mentions were replaced by the token "<USER>". This step helped to reduce the size of the vocabulary without losing information.

5.4.2 Experimental Details

Table **??** shows the hyperparameters of our system, which was trained using Adam [84], with a learning rate of 0.0001, $\beta_1 = 0.5$, and a mini-batch size of 32 to minimize the binary cross-entropy loss function:

$$\mathscr{L}(\theta, \hat{D}) = -\mathbb{E}_{((x_i, y_i), \phi_i) \sim \hat{D}} [\phi_i \cdot g(x_i, y_i) + (1 - \phi_i) \cdot (1 - g(x_i, y_i))] \tag{5.12}$$

where, $g(x_i, y_i)$ is the predicted value, ϕ_i is the real value, and θ is the model's parameters.

The hyperparameters of our system were obtained by applying Bayesian optimization [18]. We used the development set as a validation set to fine-tune those parameters.

5.4.3 Comparison with Other Systems

We compared the proposed system with the state-of-the-art systems used in the task of multi-label emotion classification, including:

Parameter	Value
Embedding Module	E^Q: Dimensions: 11 $\times$ 310 Initialization: Uniform (-0.02, 0.02) Trainable: Yes
	E^K: Dimensions: 13,249 $\times$ 310 Initialization: Pretrained model [1]. Trainable: No
	E^V: Dimensions: 13,249 $\times$ 310 Initialization: Uniform (-0.02, 0.02) Trainable: Yes
Encoding Module	RNN Cell: GRU Hidden size: 200 Layers: 2 Encoding: last hidden state RNN dropout: 0.3
Classification Module	FF1 : 1024 units FF2: 512 units Sigmoid: 1 unit Activation: ReLU Dropout: 0.3

Table 5.1: Hyperparameters of our system.

- SVM-unigrams: a baseline Support Vector Machine system trained using just word unigrams as features [102].

- NTUA-SLP : the system submitted by the winning team of the SemEval-2018 Task1:E-c challenge [14].

- TCS : the system submitted by the second place winner [95].

- PlusEmo2Vec : the system submitted by the third place winner [116].

- Transformer: a Deep learning system based on large pre-trained language models developed by the NVIDIA AI lab [80].

5.4.4 Evaluation Metrics

We used multi-label accuracy (or Jaccard index), the official competition metric used by the organizers of SemEval-2018 Task 1: Affect in Tweets, for the E-c sub task, which can be defined as the size of the intersection of the predicted and gold label sets divided by the size of their union.

$$Jaccard = \frac{1}{|T|} \sum_{t \in T} \frac{G_t \cap P_t}{G_t \cup P_t} \tag{5.13}$$

In this expression, G_t is the set of the gold labels for tweet t, P_t is the set of the predicted labels for tweet t, and T is the set of tweets. Additionally, we also used the micro-averaged F-score and the macro-averaged F-score.

Let $\#_c(l)$ denote the number of samples correctly assigned to the label l, $\#_p(l)$ the number of samples assigned to l, and $\#(l)$ the number of actual samples in l. The micro-averaged F1-score is calculated as follows:

$$P_{micro} = \frac{\sum_{l \in L} \#_c(l)}{\sum_{l \in L} \#_p(l)} \tag{5.14}$$

$$R_{micro} = \frac{\sum_{l \in L} \#_c(l)}{\sum_{l \in L} \#(l)} \tag{5.15}$$

$$F1_{micro} = \frac{2 \times P_{micro} \times R_{micro}}{P_{micro} + R_{micro}} \tag{5.16}$$

Thus, P_{micro} is the micro-averaged precision score, and R_{micro} is the micro-averaged recall score.

Let P_l, R_l, and F_l denote the precision score, recall score, and the F1-score of the label l. The macro-averaged F1-score is calculated as follows:

$$P_l = \frac{\#_c(l)}{\#_p(l)} \tag{5.17}$$

$$R_l = \frac{\#_c(l)}{\#(l)} \tag{5.18}$$

$$F_l = \frac{2 \times P_l \times R_l}{P_l + R_l} \tag{5.19}$$

$$F1_{macro} = \frac{1}{|L|} \sum_{l \in L} F_l \tag{5.20}$$

5.4.5 Results

We submitted our system's predictions to the SemEval Task1:E-C challenge. The results were computed by the organizers on a golden test set, for which we did not have access to the golden labels.

Table 5.2 shows the results of our system and the results of the compared models (obtained from their associated papers). As can be observed from the reported results, our system achieved the top Jaccard index accuracy and macro-averaged F1 scores among all the state-of-the-art systems, with a competitive, but slightly lower score for the micro-average F1.

Table 5.2: Results of our system and state-of-the-art systems.

Model	Accuracy (Jaccard)	Micro F1	Macro F1
BNet (Our System)	**0.590**	0.692	**0.564**
SVM-Unigrams	0.442	0.57	0.443
Transformer	0.577	0.690	0.561
NTUA-SLP	0.588	**0.701**	0.528
TCS	0.582	0.693	0.530
PlusEmo2Vec	0.576	0.692	0.497

To get more insight on the performance of our system, we calculated the precision score, the recall score, and the F1 score of each label. The results of this analysis are shown in Figure 5.3. We found that our system gave the best performance on the "joy" label followed by the "anger", "fear", "disgust", and "optimism" labels. The obtained F1-score of these labels was above 70%. The worst performance was obtained on the "trust", "surprise", "anticipation", and "pessimism" labels. In most cases, our system gave a recall score higher than the precision score. It seems that the system was aggressive against the emotions "trust", "surprise", "anticipation", and "pessimism" (i.e., the system associated a low number of samples to these labels). This can be attributed to the low number of training examples for these emotions and to the Out-Of-Vocabulary (OOV) problem.

5.4.6 Attention Visualizations

We visualized the attention weights to get a better understanding of the performance of our system. The results are described in Figures 5.4–5.7, which show heat-maps of the attention weights on four example tweets from the validation set. The color intensity refers to the weight given to each word by the attention model. It represents the strength of the relationship between the word and the emotion, which reflects the importance of this word in the final

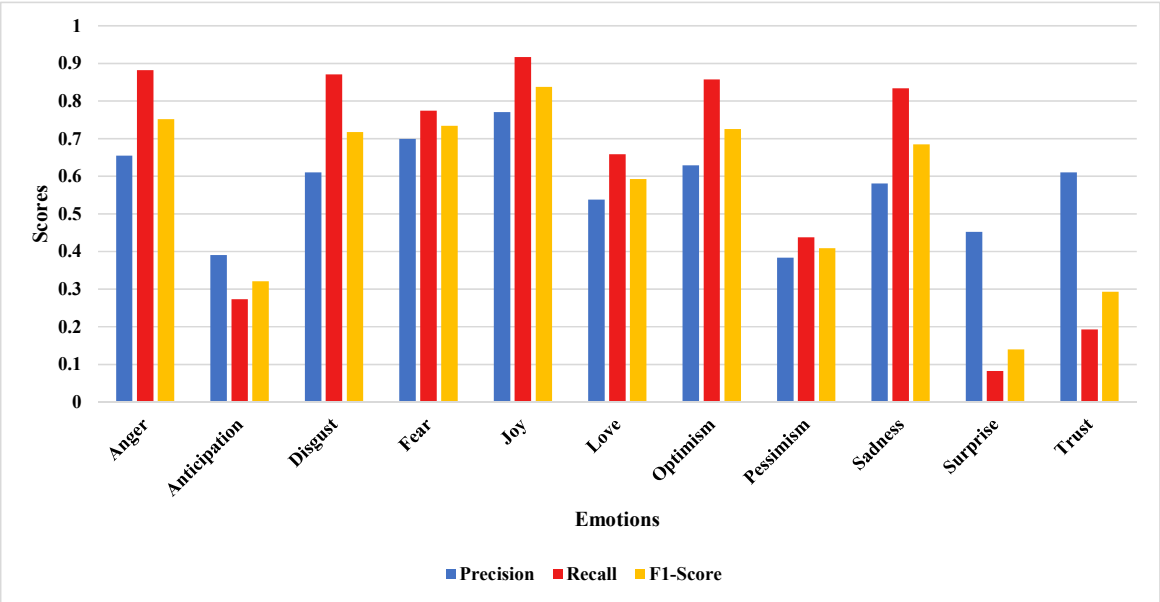


Figure 5.3: Performance analysis.

prediction. We can see that the attention model gave the important weights to the common words, such as the stop words, in case the tweet was not assigned to the emotion; for example, the word "for" in Figure 5.4, the word "this" in Figure 5.6 and the token "<user>" in Figure 5.7. Moreover, it also gave a high weight top the words and the emojis related to emotions (e.g., "cheering" and "awesome" for joy, "birthday" for love, etc.). An interesting observation is that when emojis were present, they were almost always selected as important if they were related to the emotion. For instance, we can see in Figure 5.6 that the sadness emotion relied heavily on the emoji. We also found that considering only one word to model the relation between the tweet and the emotions was not enough. In some cases, the emotion of a word may be flipped based on the context. For instance, consider the following tweet as an example: "When being #productive (doing the things that NEED to be done), #anxiety level decreases and #love level increases. #personalsexuality", the word "anxiety" is highly related to the emotion fear, but in this context, it shows optimism and trust emotions. However, our system misassociated this example with the fear emotion.

5.4.7 Correlation Analysis

Figure 5.8 shows the correlation analysis of emotion labels in the validation set. Each cell in the figure represents the correlation score of each pair of emotion labels. The reported values show exciting findings. Our system captured the relations among the emotion labels. The correlation scores of the predicted labels were almost identical to the ground-truth. There was

	im	clapping	and	cheering	for	both	teams	.	..1	..2
Anger	0.43	0.10	0.18	0.01	0.00	0.01	0.02	0.08	0.08	0.08
Anticipation	0.03	0.02	0.05	0.04	0.28	0.07	0.19	0.11	0.11	0.11
Disgust	0.15	0.08	0.26	0.01	0.01	0.02	0.04	0.14	0.14	0.14
Fear	0.01	0.01	0.02	0.00	0.23	0.01	0.00	0.24	0.24	0.24
Joy	0.01	0.07	0.00	0.74	0.13	0.04	0.01	0.00	0.00	0.00
Love	0.00	0.00	0.00	0.01	0.97	0.00	0.00	0.01	0.01	0.01
Optimism	0.01	0.01	0.02	0.13	0.50	0.18	0.09	0.02	0.02	0.02
Pessimism	0.01	0.01	0.02	0.01	0.30	0.01	0.03	0.20	0.20	0.20
Sadness	0.03	0.02	0.03	0.02	0.21	0.02	0.02	0.21	0.21	0.21
Surprise	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Trust	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00

Figure 5.4: Attention visualization example. Golden labels are {*joy*, *optimism*} and predicted labels are {*joy (0.91)*, *optimism (0.51)*}.

	i	got	a	free	dr	.	pepper	from	the	vending	machine	awesome
Anger	0.02	0.03	0.28	0.04	0.01	0.12	0.07	0.04	0.33	0.02	0.04	0.01
Anticipation	0.28	0.03	0.04	0.06	0.27	0.07	0.04	0.01	0.09	0.04	0.05	0.02
Disgust	0.02	0.03	0.22	0.04	0.02	0.14	0.04	0.09	0.31	0.04	0.04	0.01
Fear	0.11	0.00	0.16	0.00	0.00	0.12	0.00	0.00	0.58	0.00	0.01	0.00
Joy	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95
Love	0.76	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.04	0.00	0.00	0.16
Optimism	0.33	0.10	0.01	0.06	0.13	0.02	0.03	0.01	0.02	0.03	0.02	0.23
Pessimism	0.08	0.01	0.20	0.01	0.01	0.18	0.01	0.01	0.46	0.01	0.03	0.00
Sadness	0.07	0.02	0.19	0.01	0.04	0.19	0.04	0.04	0.34	0.04	0.02	0.01
Surprise	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trust	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 5.5: Attention visualization example. Golden labels are {*joy*, *surprise*} and predicted labels are {*joy (0.97)*, *optimism (0.87)*}.

an exception in the surprise and trust emotions. Our system was unsuccessful in capturing the relationships between these two emotions and the inputs or the other emotions. We attribute this apparent lack of correlation to the low number of training examples of these two emotions.

Moreover, there was always a positive correlation between related emotions such as “joy” and “optimism” (the score from the ground truth labels and from the predicted labels was 0.74). On the other side, we can see that there was a negative correlation between unrelated emotions like “anger” and “love”. The scores were -0.27 and -0.3, respectively.

	can	not	believe	zain	starting	secondary	this	year	😟
Anger	0.25	0.13	0.02	0.02	0.02	0.06	0.49	0.00	0.02
Anticipation	0.09	0.03	0.06	0.17	0.34	0.17	0.04	0.08	0.02
Disgust	0.14	0.17	0.04	0.03	0.04	0.10	0.45	0.01	0.03
Fear	0.19	0.02	0.02	0.01	0.02	0.03	0.69	0.01	0.01
Joy	0.04	0.00	0.10	0.12	0.12	0.02	0.01	0.57	0.03
Love	0.13	0.01	0.08	0.14	0.01	0.00	0.21	0.08	0.34
Optimism	0.03	0.01	0.19	0.26	0.12	0.07	0.02	0.27	0.02
Pessimism	0.34	0.06	0.03	0.03	0.02	0.04	0.35	0.01	0.13
Sadness	0.01	0.02	0.01	0.01	0.00	0.01	0.03	0.00	0.91
Surprise	0.14	0.00	0.00	0.00	0.00	0.00	0.85	0.00	0.00
Trust	0.34	0.00	0.00	0.00	0.00	0.00	0.65	0.00	0.00

Figure 5.6: Attention visualization example. Golden labels are {*sadness*, *surprise*} and predicted labels are {*love (0.74)*, *sadness (0.98)*}.

	<user>	happy	birthday	gorg	,	have	a	good	one	😘	x
Anger	0.05	0.00	0.00	0.01	0.27	0.33	0.21	0.00	0.09	0.00	0.03
Anticipation	0.49	0.01	0.04	0.01	0.05	0.14	0.06	0.03	0.07	0.04	0.07
Disgust	0.04	0.00	0.00	0.01	0.34	0.32	0.16	0.00	0.10	0.00	0.03
Fear	0.60	0.00	0.00	0.00	0.07	0.03	0.28	0.00	0.01	0.00	0.01
Joy	0.01	0.67	0.08	0.09	0.00	0.00	0.00	0.09	0.00	0.06	0.00
Love	0.02	0.01	0.73	0.04	0.00	0.00	0.00	0.00	0.00	0.20	0.00
Optimism	0.10	0.29	0.09	0.04	0.00	0.00	0.01	0.33	0.01	0.11	0.01
Pessimism	0.40	0.00	0.00	0.00	0.34	0.05	0.17	0.01	0.01	0.01	0.02
Sadness	0.16	0.01	0.03	0.02	0.42	0.16	0.13	0.01	0.02	0.01	0.02
Surprise	0.98	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
Trust	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 5.7: Attention visualization example. Golden labels are {*joy*, *love*, *optimism*} and predicted labels are {*joy (0.98)*, *love (0.91) optimism (0.95)*}.

This result further strengthened our hypothesis that the proposed system was able to, implicitly, model the relationships between the emotion labels.

	Anger	Anticipation	Disgust	Fear	Joy	Love	Optimism	Pessimism	Sadness	Surprise	Trust
Anger	1	-0.15	0.7	-0.021	-0.521	-0.27	-0.45	0.055	0.15	-0.054	-0.16
Anticipation	-0.15	1	-0.15	-0.066	0.092	-0.05	0.14	-0.051	-0.12	0.15	0.11
Disgust	0.7	-0.15	1	0.024	-0.53	-0.29	-0.49	0.015	0.2	-0.068	-0.17
Fear	-0.021	-0.066	0.024	1	-0.24	-0.15	-0.19	0.076	0.0058	-0.064	-0.029
Joy	-0.521	0.092	-0.53	-0.24	1	0.4	0.58	-0.22	-0.35	0.049	0.13
Love	0.27	-0.05	-0.29	-0.15	0.4	1	0.31	-0.14	-0.2	-0.069	0.11
Optimism	-0.45	0.14	-0.49	-0.19	0.58	0.31	1	-0.2	-0.28	0.011	0.24
Pessimism	0.055	-0.051	0.015	0.076	-0.22	-0.14	-0.2	1	0.34	-0.036	-0.081
Sadness	0.15	-0.12	0.2	0.0058	-0.35	-0.2	-0.28	0.34	1	-0.069	-0.12
Surprise	-0.054	0.15	-0.068	-0.064	0.049	-0.069	0.011	-0.036	-0.069	1	0.062
Trust	-0.16	0.11	-0.17	-0.029	0.13	0.11	0.24	-0.081	-0.12	0.062	1

(a) The ground-truth labels.

	Anger	Anticipation	Disgust	Fear	Joy	Love	Optimism	Pessimism	Sadness	Surprise	Trust
Anger	1	-0.14	0.78	0.054	-0.53	-0.3	-0.57	-0.038	0.24	0.04	-0.075
Anticipation	-0.14	1	-0.15	-0.064	0.21	-0.03	0.23	-0.044	-0.12	0.15	0.088
Disgust	0.78	-0.15	1	0.086	-0.54	-0.32	-0.59	0.015	0.31	0.04	-0.081
Fear	0.054	-0.064	0.086	1	-0.22	-0.12	-0.16	0.044	0.049	0.082	-0.039
Joy	-0.53	0.21	-0.54	-0.22	1	0.41	0.74	-0.18	-0.38	0.035	0.051
Love	-0.3	-0.03	-0.32	-0.12	0.41	1	0.46	-0.088	-0.21	0.077	0.023
Optimism	-0.57	0.23	-0.59	-0.16	0.74	0.46	1	-0.17	-0.4	0.04	0.065
Pessimism	-0.038	-0.044	0.015	0.044	-0.18	-0.088	-0.17	1	0.35	0.12	-0.027
Sadness	0.24	-0.12	0.31	0.049	-0.38	-0.21	-0.4	0.35	1	0.044	-0.048
Surprise	0.04	0.15	0.04	0.082	0.035	0.077	0.04	0.12	0.044	1	-0.0032
Trust	-0.075	0.088	-0.081	-0.039	0.051	0.023	0.065	-0.027	-0.048	-0.0032	1

(b) The predicted labels.

Figure 5.8: Correlation matrices of emotion labels of the development set.

5.5 Conclusions

In this work, we presented a new approach to the multi-label emotion classification task. First, we proposed a transformation method to transform the problem into a single binary classification problem. Afterwards, we developed a Deep learning-based system to solve the transformed problem. The key component of our system was the embedding module, which used three embedding models and an attention function. Our system outperformed the state-of-the-art systems, achieving a Jaccard (i.e., multi-label accuracy) score of 0.59 on the challenging SemEval2018 Task 1:E-c multi-label emotion classification problem.

We found that the attention function can model the relationships between the input words and the labels, which helps to improve the system's performance. Moreover, we showed that our system is interpretable by visualizing the attention weights and analyzing them. However, some limitations have been identified. Our system does not model the relationships between the phrases and the labels. Phrases play a key role in determining the most appropriate set of emotions that must be assigned to a tweet. For instance, an emotion word that reflects "sadness" can be flipped in a negated phrase or context. We consider these shortcomings in the outlines of the future works described in chapter 10.

The systems described in this chapter and the previous chapters work at the document/tweet level. An overall result is assigned to a given tweet. However, in real-world cases,

it is necessary to identify in a single document diverse targets and the attitude towards them. The next chapter proposes a novel methodology to addresses these tasks.

Chapter 6

Target-Dependent Sentiment Analysis of Tweets using Bi-directional Gated Recurrent Neural Networks

6.1 Introduction

Sentiment analysis can be done at different levels. A coarse-grained analysis attempts to extract the overall polarity on a document or sentence level, whereas, in a fine-grained level of analysis, the problem is to identify the sentiment polarity towards a certain target in a given text. This problem is known as *Target-dependent sentiment analysis* [40, 76, 166].

It is quite usual to give several opinions on different aspects of an object in a single sentence. For example, the text *"I have got a new mobile. Its camera is wonderful but the battery life is too short."* gives both positive and negative remarks about a mobile phone. It may be seen that the example contains three targets (*"mobile"*, *"camera"* and *"battery life"*) and the sentiment polarities towards them can be seen as "neutral", "positive" and "negative", respectively. Such fine-grained opinions are important for both producers and customers [94].

The importance of target information has been proven by previous studies. It has been shown [76] that about 40% of the errors of sentiment analysis systems are caused by the lack of information about the target. Thus, the target-dependent SA problem can be addressed by designing a system with two steps, shown in Figure 6.1. The first step aims to extract or identify the target in a given text, while the objective of the second step is to identify the opinion expressed in the text towards the extracted target. Previous works on those two steps are commented in the following subsections.

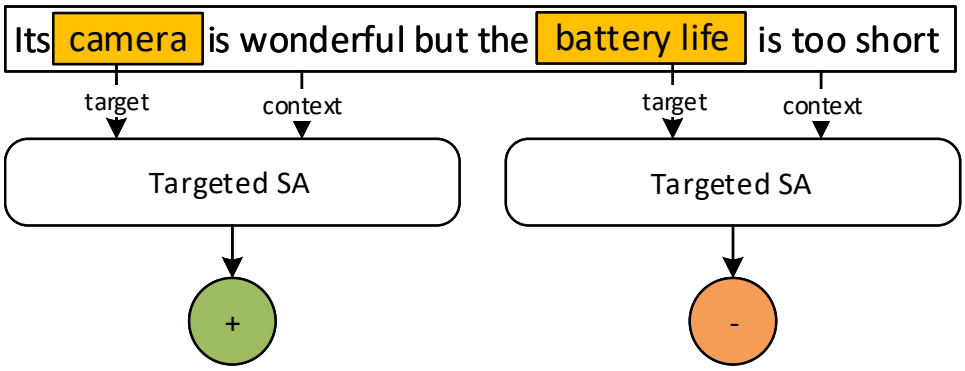


Figure 6.1: Example of the two steps. The filled rectangle represents the target extraction step, and the rounded rectangle represents the sentiment analysis step which receives as inputs the extracted target and its context.

6.1.1 Target Identification

Target-dependent sentiment analysis on Twitter is the problem of identifying the sentiment polarity towards a certain target in a given tweet. Extracting the targets from the tweets is the key task in this problem. However, all the existing studies in the field assume that the target is known. As will be seen in Section 6.2, we have developed a novel system to identify automatically the explicit targets of the tweets.

Recently, a similar problem to target identification, known as *aspect term extraction*, has been studied extensively. There are two main kinds of approaches: *supervised* and *unsupervised*. In the supervised approaches machine-learning systems are trained on manually annotated data to extract targets in the reviews. The most common techniques employed in supervised approaches are Decision Trees, Support Vector Machines, K-nearest Neighbour, Naive Bayesian classifiers and Neural Networks [81, 161]. On the other hand, unsupervised approaches aim to automatically extract product features using syntactic and contextual patterns without the need of annotated data [93, 94].

There is one particularly interesting supervised approach, which conceptualizes the aspect extraction problem as a sequence labeling problem [75]. The most successful sequence labeling systems are probabilistic graphical models such as Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) [87, 125]. However, their main drawback is that they rely heavily on a set of hand-crafted features, whose definition is very time consuming. Recently, deep neural networks have been utilized to extract automatically high-level features in many tasks such as speech recognition [59], text classification [82], image classification

[63], etc. Recurrent Neural Networks (RNNs) have been proved to be a very useful technique to represent sequential data such as text. These models have also shown great success in solving sequence labeling tasks, e.g. Named Entity Recognition (NER) and Part-Of-Speech (POS) tagging [88, 90]. Following these approaches we propose to use a bidirectional gated recurrent neural network to solve the problem of target extraction, as described in subsection 6.2.1.

6.1.2 Target-Dependent Sentiment Analysis

Target-dependent sentiment analysis (TDSA) is also regarded as a text classification problem in the literature. Standard text classification approaches such as feature-based Support Vector Machines [76, 115] can be used to build a sentiment classifier. For instance, the work presented in [76] combined manually designed target-independent features and target-dependent features with expert knowledge, a syntactic parser and external resources.

Recent studies, such as the works proposed by [40] , [166], [156] and [182], use neural network methods and encode each sentence in a continuous and low-dimensional vector space without feature engineering. Dong et al. [40] transformed a sentence dependency tree into a target-specific recursive structure, and used an *Adaptive Recursive Neural Network* to learn a higher level representation. Vo and Zhang [166] used rich features including sentiment-specific word embedding and sentiment lexicons. The work presented in [182] modeled the interaction between the target and the surrounding context using a gated neural network. Tang et al. [156] developed long short-term memory models to capture the relatedness of a target word with its context words when composing the continuous representation of a sentence. Most of these studies rely on the idea of splitting the sentence/text into target, left context and right context.

Unlike previous studies, we propose a *target-dependent bi-directional gated recurrent unit* (TD-biGRU), which is capable of modeling the relatedness between target words and their contexts by concatenating an embedded vector that represents the target word(s) with two vectors that capture both the preceding and following contextual information.

The rest of this chapter is structured as follows. In Section 6.2 the proposed models are described. The experiments and results are presented and discussed in Section 6.3. Finally, in the last section the conclusions are outlined.

6.2 System Description

We describe in this section the proposed system to tackle the problem of target-dependent SA. Figure 6.2 shows the overall architecture of the TDSA system. It is composed of two main steps. First, the targets of the tweet to be analysed are identified as described in the next subsection. After that, each extracted target is passed together with the tweet as input to the model described in subsection 6.2.2 to determine the sentiment polarity.

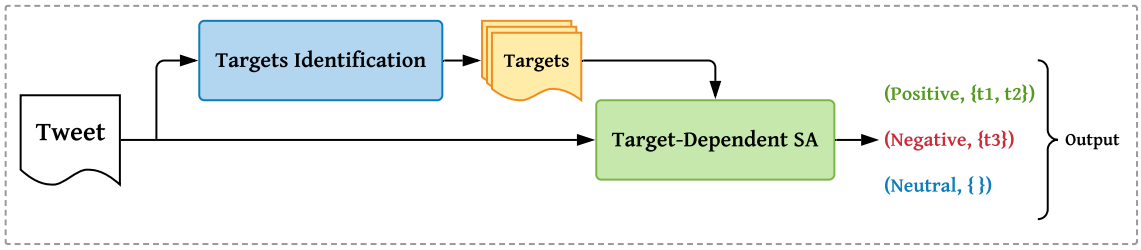


Figure 6.2: The overall architecture of the TDSA system.

6.2.1 Target Identification

In this step we aim to extract the targets on which customers expressed their opinions. Target identification can be typically regarded as a kind of sequence labeling problem in which the text (i.e. a sequence of words) can be represented using the IOB2 tagging scheme [141]. The idea is that each word in a given text is labeled by one of the tags I, O, or B, which indicates if the word is inside, outside, or at the beginning of a target respectively.

Following [75] we have used a bi-directional gated recurrent neural network to extract the targets from a given text. This model, called TI-biGRU, reads a sequence of words and predicts a sequence of corresponding IOB2 tags. Once we have the predicted sequence of IOB2 tags for a text, we can interpret it and extract the targets.

Figure 6.3 shows an example of the application of the proposed model to the problem of opinion target identification. Its main steps are the following. First, the words of the input sentence are mapped to vectors of real numbers using a pre-trained word embedding model, as explained in chapter 2, resulting in a sequence of vectors $x_1, x_2, ...x_n$. Afterwards, the resulting sequence is passed to a biGRU to produce a sequence of recurrent states $h_1, h_2, ..., h_n$. Finally, each produced sequence element h_i is passed through a softmax layer or a CRF to predict the probability distribution over the three possible output tags (I, O or B).

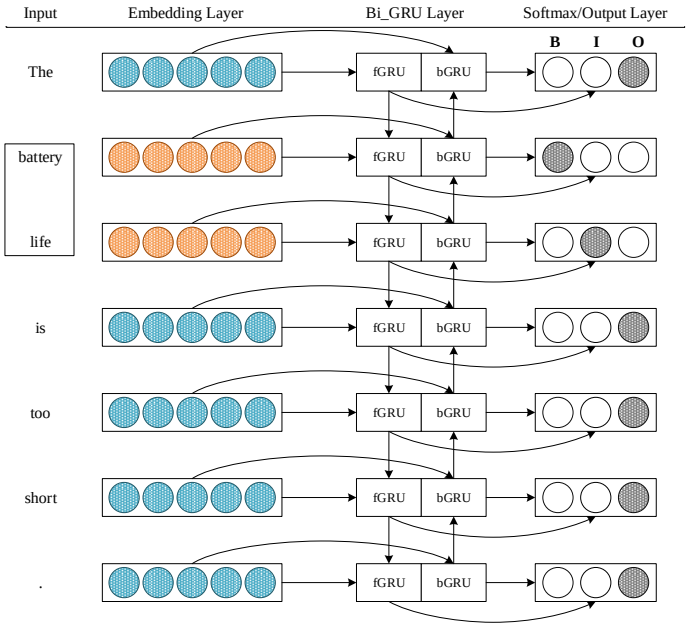


Figure 6.3: TI-biGRU model for target identification.

In the case of using softmax as the output layer, we refer to this model as TI-biGRU, the model is trained to minimize the following objective function, which is the cross-entropy between the expected tag and the predicted tag distribution of each word i:

$$J = -\sum_{s \in S} \sum_{i=1}^{n} \sum_{t=1}^{3} p_i^s(t) log(P(y = t|h_i^s)) \tag{6.1}$$

In this expression $p_i^s(t) \in \{0,1\}$ is the ground-truth function which indicates whether tag t is the correct tag for the word i in the sentence s and S is the set of the sentences in the training set. The derivative of the objective function J is taken through back-propagation with respect to the whole set of parameters of the model. These parameters are optimized using the stochastic optimization method *RMSProp*.

Below, we describe the usage of the CRF as the output layer instead of the softmax.

Sequence Tagging using CRF

In the case of using softmax as an output layer, we simply model the sequence labeling (tagging) by using the h_t's as features to make an independent tagging decision at each time t [90]. However, in such tasks it is beneficial to consider the correlation between labels in neighbourhoods, specifically when there are strong dependencies across the output tags. For

example, the tag **B** is more likely to be followed by the tag **I**. Thus, instead of modeling tagging decisions independently, we model them jointly using a CRF [87]. We refer to this version of the model as TI-RNC.

Formally, let $H = \{h_1, h_2, ..., h_n\}$ be the sequence of vectors to be labeled, which is produced by the BiGRU sub-model, and $y = \{y_1, y_2, ..., y_n\}$ is the corresponding tag sequence. Each element y_i of y is one of the B, I or O tags. Both H and y are assumed to be random variables and they are jointly modeled. The entire model can be represented as an undirected graph $G = (V, E)$ with cliques C.

In this work we employed a linear-chain CRF, where G is a simple chain or line: $G = (V = \{1, 2, ..., n\}, E = \{(i, i+1)\})$. It has two different cliques (i.e. $C = \{P, M\}$): a unary clique (P) representing the input-output connection, and a pairwise clique (M) representing the adjacent output connection. We consider P to be the matrix of output scores, where $P_{i,j}$ corresponds to the score of the jth tag of the ith word in a sentence and it is computed as follows:

$$P_{i,j} = W_{i,j} h_i + b_j \; ; \; j = 1, 2, 3 \tag{6.2}$$

In this equation, the parameters are $W_{i,j} \in \mathbb{R}^{2*d_h}$ and $b_j \in \mathbb{R}^1$, where d_h is the dimensionality size of the hidden state.

The clique M is considered to be the matrix of transition scores such that $M_{i,j}$ represents the score of a transition from the tag i to the tag j. Given that, we define the score function of the sequence of predictions as follows:

$$s(H, y) = \sum_{i=1}^{n} P_{i,y_i} + \sum_{i=0}^{n} M_{y_i, y_{i+1}} \tag{6.3}$$

In this expression y_0 and y_{n+1} denote the start and the end tags of the sentence, that we add to the set of possible tags.

A softmax over all possible tag sequences (Y^*) on a sequence H yields a probability for the sequence y as follows:

$$p(y|H) = \frac{e^{s(H,y)}}{\sum_{\tilde{y} \in Y} e^{s(H,\tilde{y})}} \tag{6.4}$$

During training, we minimize the negative log-probability of the correct tag sequence:

$$J = -log\left(p(y|H)\right) = -s(H,y) + log\left(\sum_{\tilde{y}\in Y^*} e^{s(H,\tilde{y})}\right) \tag{6.5}$$

During inference, we search for the output sequence y^* that obtains the highest probability given by:

$$y^* = \arg\max_{\tilde{y}\in Y^*} p(\tilde{y}|H) \tag{6.6}$$

In this model, Eq. 6.5 and Eq. 6.6 can be solved efficiently using dynamic programming.

The derivative of the objective function J, Eq. **??**, is taken through back-propagation with respect to the whole set of parameters of the model, which are the transition matrix M, the parameters of the BiGRU model and the parameters of the matrix P defined in Eq. 6.2. The parameters are optimized using the stochastic gradient descent (SGD) with a learning rate of 0.005. To reduce the effects of gradient exploding, we set the clipping threshold of the gradient to 5. We apply a dropout [64] between the embedding layer and the recurrent layer with probability of 0.5 to prevent over-fitting.

6.2.2 Target-Dependent Sentiment Analysis

Figure 6.4 shows the proposed model for the problem of target-dependent sentiment classification. Its main steps are the following. First, the words of the input sentence are mapped to vectors of real numbers. Then, the input sentence is represented by a real-valued vector using the TD-biGRU encoder by concatenating the vectors $\overrightarrow{h_n}, \overleftarrow{h_n}$ and x_v, formally:

$$X = [\overrightarrow{h_n}; x_v; \overleftarrow{h_n}] \tag{6.7}$$

Here x_v is the vector representation of the target word(s). If the target is a single word, its representation is the embedding vector of that word. If the target is composed of multiple words, such as "battery life", its representation is the average of the embedding vectors of the words [154].

In this way, the obtained vector summarizes the input sentence and contains semantic, syntactic and/or sentimental information based on the word vectors. Finally, this vector is passed through a softmax classifier to classify the sentence into positive, negative or neutral.

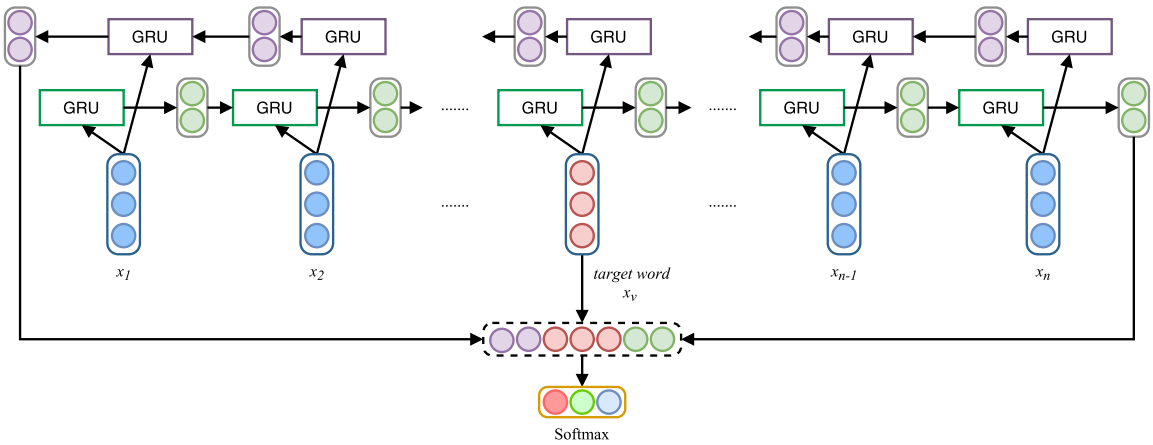


Figure 6.4: TD-biGRU model for target-dependent sentiment classification.

We trained the system to minimize the following categorical cross-entropy:

$$J = -\sum_{s \in S} \sum_{c=1}^{3} G_c(s) log(P(y = c|s)) \tag{6.8}$$

In this expression S is the training set and $G_c(s) \in \{0,1\}$ is the ground-truth function which indicates whether class c is the correct sentiment category for sentence s.

The derivative of the objective function is taken through back-propagation with respect to the whole set of parameters of the model, and these parameters are updated with the stochastic gradient descent. The learning rate is initially set to 0.1 and the parameters are initialized randomly over a uniform distribution in $[-0.03, 0.03]$. For the regularization, dropout layers [64] are used with probability 0.5 on the lookup-table output to the GRU input and on the concatenation output to the softmax input.

6.3 Experiments and Results

6.3.1 Datasets

We evaluated the effectiveness of the proposed models by using them in the supervised tasks of target identification and target-dependent sentiment classification on the benchmark dataset provided in [40]. The dataset contains 6248 training examples and 692 examples in the testing set. Each example in the dataset contains the sentence, the target and the label of sentiment polarity. In case the sentence contains more than one target with different polarities

it is repeated with each one. The numerical description of the positive, negative and neutral examples is shown in table 6.1.

	Training	Testing	Percentage
#Positives	1562	173	25%
#Neutrals	3124	346	50%
#Negatives	1562	173	25%
Total	6248	692	

Table 6.1: Numerical description of the dataset.

6.3.2 Evaluation Metrics

The evaluation metrics of the target identification problem are the precision (the number of correct targets divided by the number of all returned targets), recall (the number of correct targets divided by the number of targets that should have been returned) and F_1 (the harmonic mean of precision and recall).

The evaluation metrics of the target-dependent sentiment analysis system are the classification accuracy (the percentage of examples that are correctly classified) and the Macro-F1 measure (the averaged F1 measure over the three sentiment classes).

6.3.3 Results and Discussion

Target Identification

As stated before, all the existing models of target-dependent SA assume that the target is known. To the extent of our knowledge, this is the first target-dependent SA system that identifies and extracts the target of the tweets. We investigated the effectiveness of the proposed models, i.e., TI-biGRU and TI-RNC, that are used to automatically identify the target from a tweet, by comparing it with two baseline models.

In table 6.2 baseline-I and baseline-II denote the typical RNN and biRNN respectively. TI-GRU is the simplified version of TI-biGRU in which only the past information is considered, ignoring the bidirectionality. It is clearly shown that TI-RNC outperforms the other models. Another interesting observation from the reported result is that both baseline-II and TI-biGRU perform better than their relaxed versions (i.e. baseline-I and TI-GRU). Such conclusion confirms the usefulness of BiRNNs in this kind of tasks.

From the table, we can also notice that both TI-GRU and TI-biGRU perform better than the baselines (i.e. the standard RNN based models). In terms of recall, all the models give

Model	Precision	Recall	F_1
baseline-I	77.90	87.57	82.44
baseline-II	79.76	90.17	84.67
TI-GRU	81.18	90.89	86.10
TI-biGRU	87.39	91.18	89.25
TI-RNC	**95.30**	**94.02**	**94.66**

Table 6.2: Comparison of our model to the baselines on target identification. Best scores are shown in bold.

interesting results. We can attribute this result to the fact that achieving a recall of 100% is trivial by assuming that all words in the sentence/tweet are targets. Therefore, recall alone is not enough, and it is also necessary to measure the number of incorrectly returned targets by computing the precision.

Considering the results of TI-biGRU, the second-best model, we can see that there are remarkable improvements in terms of precision. It shows 12.8%, 9.57% and 7.65% precision improvements compared with those of baseline-I, baseline-II and TI-GRU, respectively. On the other hand, there are smaller improvements in terms of recall.

Comparing the performance of TI-RNC and TI-biGRU, we can find that there are remarkable gains in terms of all the evaluation metrics. Specifically, the percentages of the improvements are 9.05% for the precision, 3.11% for the recall and 6.06% for the F_1. Giving that, we can confirm the importance of modelling the dependencies between the labels when we solve such problems.

Target-Dependent Sentiment Analysis

We compared the proposed model with the state-of-the-art methods used in the task of target-dependent sentiment classification, including:

- **SVM-indep**: SVM classifier built with target-independent features, such as unigram, bigram, punctuations, emoticons, hashtags and the numbers of positive or negative words in the General Inquirer sentiment lexicon [76].

- **SVM-dep**: SVM-indep model extended by adding a set of features that represent the target [76].

- **Recursive RNN**: a recursive neural network is employed to learn the feature representation of the examples over a transferred target-dependent dependency tree [40].

- **AdaRNN**: extension of the recursive RNN which uses more than one composition function and adaptively selects them according to the input [40]. AdaRNN has three variations: AdaRNN-w/oE, AdaRNN-w/E and AdaRNN-comb. Unlike AdaRNN-w/oE, the AddRNN-w/E model uses the dependency type in the process of composition function selection. AddaRNN-comb combines the root vectors obtained by AdaRNN-w/E with the unigram and bigram features, and then they are fed into a SVM classifier.

- **Target-ind/Target-dep**: SVM classifiers based on a rich set of target-independent and target-dependent features [166]. This model has an extension, called **Target-dep+**, in which sentiment lexicon features have been incorporated.

- **LSTM**, **TD-LSTM**, **TC-LSTM**: these methods are based on the *long short-term memory* model (LSTM) proposed by [156]. In the LSTM model the target is ignored. The idea behind TD-LSTM is to use two LSTM neural networks, so that the left one represents the preceding context plus the target and the right one represents the target plus the following context. TC-LSTM is an extension of TD-LSTM in which a vector that represents the target is concatenated to each context word.

The values under the section "A" in Table 6.3 represent the results of the baseline model (basic bi-directional gated recurrent units - biGRU - without incorporating target information), the new TD-biGRU model in case the targets are manually given and the results when we apply the two steps of our system to analyse the tweets. Each tweet is passed to the system to first extract the targets and then identify the sentiment polarities towards these targets. Section "B" contains the results of the compared models (obtained from their associated papers). With the exception of AdaRNN, each approach presented in Table 3 has a target-independent version (which does not incorporate any information about targets) and two or three target-dependent versions. For instance, in our case biGRU is the target-independent version.

As it can be observed from the reported results, the target-independent models (SVM-indep, Target-indep, LSTM and biGRU) have a worst performance than the corresponding models that consider the target information (SVM-dep, Target-dep*, TD-LSTM, TC-LSTM and TD-biGRU). This conclusion confirms the fact that ignoring the target information causes about 40% of sentiment analysis errors [76]. It may also be noticed that neural-based models perform better than the feature-based SVM classifiers.

The TD-biGRU model outperforms the state-of-the-art models both in terms of accuracy and Macro-F1. Our end-to-end approach gives comparable results to those models, including our TD-biGRU model, that assumes the target is known.

Model	Accuracy	Macro-F1
A.Our model		
biGRU	69.94	68.40
TD-biGRU	**72.25**	**70.47**
End-To-End-TD	70.08	68.22
B. State-of-the-art systems		
SVM-indep	62.70	60.20
SVM-dep	63.40	63.30
Recursive NN	63.00	62.80
AdaRNN-w/oE	64.90	64.44
AdaRNN-w/E	65.80	65.50
AdaRNN-comb	66.30	65.90
Target-ind	67.30	66.40
Target-dep	69.70	68.00
Target-dep$^+$	71.10	69.90
LSTM	66.50	64.70
TD-LSTM	70.80	69.00
TC-LSTM	71.50	69.50

Table 6.3: Comparison of different methods on target-dependent sentiment classification. Evaluation metrics are accuracy and macro-F1. Best scores are shown in bold.

To get more insight on this result, we analyzed the confusion matrix given by the TD-biGRU model to figure out which are the most common incorrect cases. Figure 6.5 shows the confusion matrix obtained by applying TD-biGRU. As observed, the matching between the true and the predicted labels is quite high (diagonal matrix). Out of the 192 misclassified samples, 76 (39.6%) of them were misclassified between negative and neutral (i.e., either negative samples were misclassified as neutral or viceversa) and 31 (16.1%) samples were misclassified between negative and positive. The number of samples misclassified between positive and neutral is 85 (44.3%).

This analysis shows that most of the misclassified examples are related to the neutral category. We believe that this problem can be handled by adding more information (e.g. lexicon information). We leave the study of this hypothesis for the future work.

6.4 Conclusion

We have developed an end-to-end target-dependent Twitter sentiment analysis system. The proposed model has the ability of identifying and extracting the target of the tweets, repre-

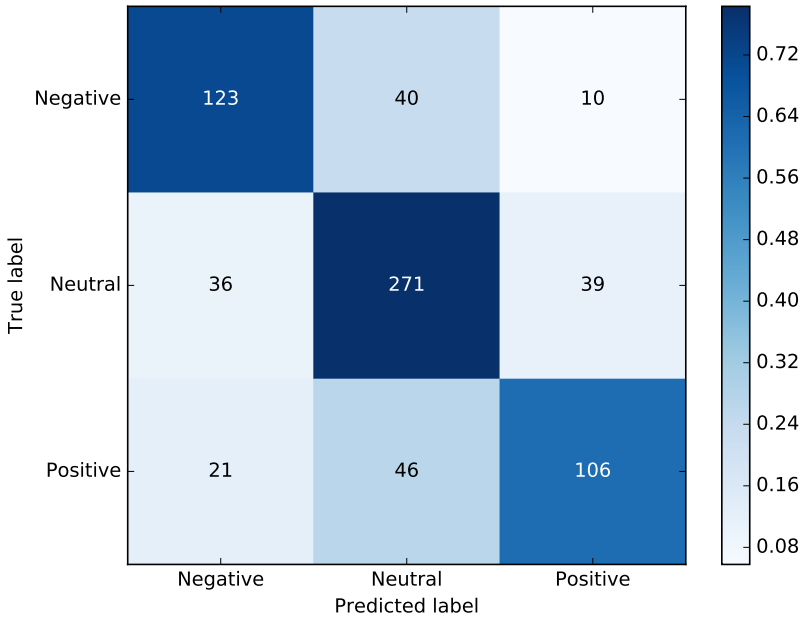


Figure 6.5: TD-biGRU Confusion matrix.

senting the relatedness between the targets and its contexts and identifying the polarities of the tweets towards the targets. The effectiveness of the proposed system has been evaluated on a benchmark of tweets, obtaining results that outperform the state-of-the-art models. The confusion matrix of the results obtained by TD-biGRU shows that most of the misclassified examples are related to the neutral category. It is worth mentioning that our system extracts only the targets that are mentioned explicitly in the tweets. We address these weaknesses in Chapter 10 and outline some future works related to the problem of target-dependent sentiment analysis.

The works presented in this chapter and the previous ones are mainly directed to the English language (Although in some cases we developed independent versions for the Arabic language). The next chapter explains our approach to develop a multi-lingual sentiment analysis system that can work with low (or even zero) resource (e.g., training data and sentiment lexicons) languages.

Chapter 7

UniSent: Universal Sentiment Analysis System for Low-Resource Languages

7.1 Introduction

It has been explained in the previous chapters that most existing sentiment analysis systems rely heavily on a large amount of human-created resources such as labeled data and sentiment lexicons. However, the distribution of sentiment resources is very unbalanced among languages. Thus, building a sentiment analysis system in low-resource languages requires a tremendous human effort to construct such resources, which is a time-consuming and expensive task. Cross-lingual Sentiment Analysis (CLSA) aims to transfer knowledge extracted from annotated sentiment resources in a rich-resource language (e.g., English) to low-resource languages. We refer to the rich-resource language as the source language, while the low-resource one is called the target language. Recently, CLSA has become a hot research topic in the fields of NLP and sentiment analysis.

The previous works on CLSA are based on two approaches: the use of an automatic machine translation system [8, 138] or the application of transfer learning methods [13, 124, 190]. Machine translation systems can transfer sentiment knowledge from a rich-resource language to low-resource languages [97, 124]. However, large parallel corpora are required to train such translation systems, which are often not available for low-resource languages [60].

In this chapter, we address the problem of CLSA. We propose a new approach, a Universal Sentiment Analysis system, called UniSent, which involves

1. pre-training of a bidirectional long short-term memory (BiLSTM) on a high-resource language, i.e., English language,

2. aligning non-English language embeddings to the English ones, and

3. fine-tuning on low-resource languages validation sets.

During the fine-tuning stage, we propose a Universal Embedding Layer which represents a word in a low-resource language with a weighted average of the most similar words to it in the English word embedding. **UniSent** is conceptually simple and empirically powerful. It can work even with zero-shot resources languages. Moreover, it is highly flexible. Adding a new language to the system does not necessarily require to re-train it.

The experimental results show the superiority of our system compared with the state-of-the-art systems in nearly all experimental settings.

The rest of this chapter is organized as follows. Section 7.2 gives a brief overview of related works on CLSA. Section 7.3 explains in detail our proposed system. In Section 7.4 we present the experimental settings and discuss the results. Finally, in section 7.5 the conclusion is presented.

7.2 Related Works

Previous works in cross-lingual sentiment analysis can be grouped into two categories: (i) machine translation-based systems and (ii) transfer learning-based systems.

Machine translation-based systems either translate the target language text into a rich-resource language, e.g., English, and then apply an English sentiment analysis system [8], or translate resources (i.e., annotated corpora, sentiment lexicons, etc.) from a rich-resource language to the target low-resource language and create a specific language sentiment analysis system [12, 138].

It has been argued that even perfect machine translation systems affect the performance quality of sentiment analysis systems [43]. As a consequence, several studies have been conducted to directly transfer and project sentiment information instead of using machine translation systems. For instance, the work presented in [189] used stacked autoencoders to create a common latent space for Chinese and English sentences in an unsupervised manner. The authors of [28] proposed an adversarial deep network to transfer the knowledge learned from a rich-resource language to low-resource languages. Chen et al., 2018b [29] proposed to use a hierarchical attention-based, long short-term network to capture the sentiment information involved in the ubiquitous emoji usage. They exploited a semi-supervised representation learning approach by solving the task of emoji prediction to learn cross-lingual representations of text that can capture both semantic and sentiment information. The authors of [42] presented a cross-lingual propagation algorithm that yields sentiment embedding

vectors for different languages. In [190], an attention-based bilingual representation learning system was proposed to learn the distributed semantics of the documents in both rich-resource languages and low-resource languages. Barnes et al. [13] developed a model to jointly represent sentiment information in a source and target language. Their model relays on a small bilingual lexicon, a source-language corpus annotated for sentiment, and monolingual word embeddings for each language.

Following this approach, we have designed a system to directly transfer sentiment information from a rich-resource language to a low-resource one as explained in the next section.

7.3 UniSent

In this section, we introduce UniSent and its implementation steps in detail. We first describe the BiLSTM architecture, the top part of Figure 7.1, in subsection 7.3.1. We then explain the word embedding alignment from a source language s to a target language t in subsection 7.3.2. Finally, we introduce in subsection 7.3.3 the proposed universal embedding layer, the sub-figure (a) of the bottom part of Figure 7.1.

7.3.1 BiLSTM

Embedding Layer

Let $W = \{w_1, w_2, ..., w_n\}$ be the set of the input words, where n is the number of input words. The goal of the embedding layer is to represent each word w_i by a low-dimensional vector $v_i \in \mathbb{R}^d$. Here, d is the size of the embedding layer.

Encoder Layer

As has been stated in preceding chapters, the goal of the encoder layer is to map the sequence of words representations, $\{v_{w_1}, v_{w_2}, ..., v_{w_n}\}$, that is obtained from the embedding layer to a single real-valued dense vector. In this work we use a BiRNN with two LSTMs (one as forward $\overrightarrow{\phi}$ RNN and the other as backward $\overleftarrow{\phi}$ RNN) to design our encoder. The representation of each word is the concatenation of the hidden states produced by $\overrightarrow{\phi}$ and $\overleftarrow{\phi}$. Once we have the sequence of the hidden states, we simply chose the last hidden state (i.e., h_n) to be the vector that represents the input.

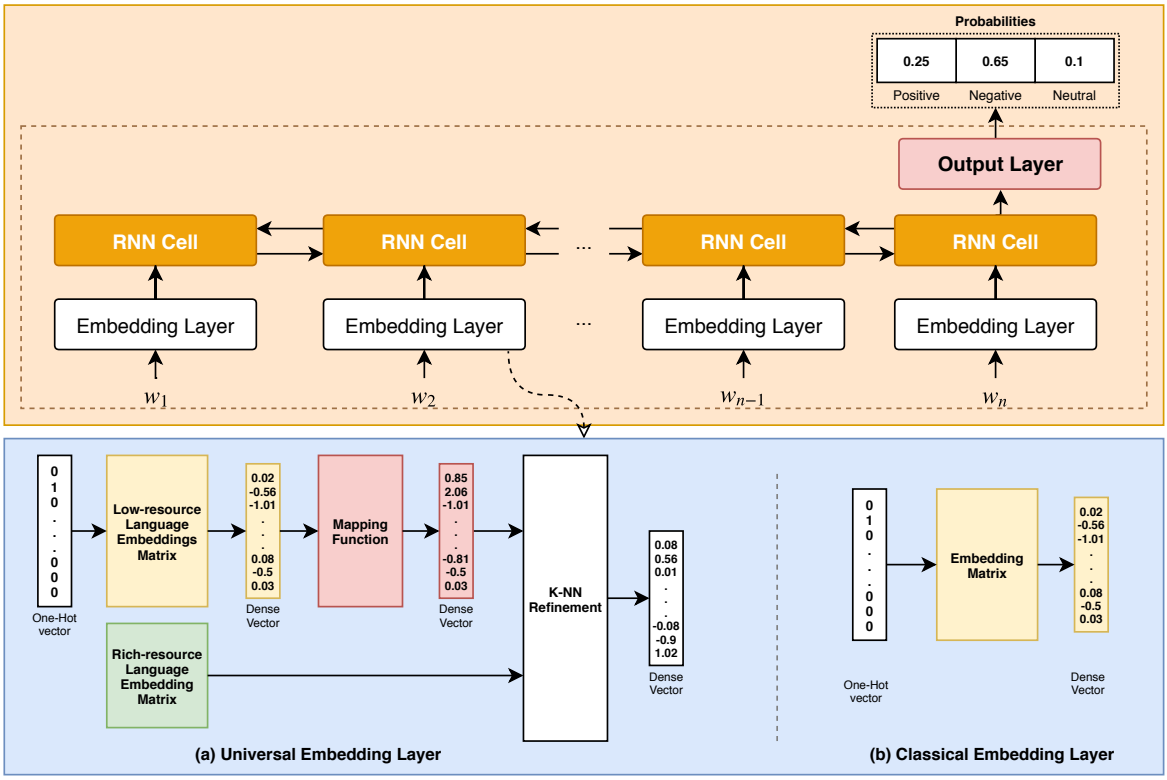


Figure 7.1: The architecture of UniSent. (a) The proposed universal embedding layer. (b) The classical one.

Output Layer

We feed the output of the BiRNN as an input vector to the output layer, i.e., the classifier. Our classifier is composed of one fully-connected layer with softmax activation function and a number of neurons equal to the number of classes, which outputs a probability distribution over all classes.

7.3.2 Embedding Alignment

Given an embedding e_w^t for a word w in a target language t, the embedding alignment aims to generate an embedding $e_w^{s \to s}$ in a source language s. In this work, our alignment method is a linear mapping matrix $W^{t \to s}$ which is obtained by applying the MUSE [37] method. MUSE is an unsupervised and domain-adversarial approach for learning $W^{t \to s}$ without parallel data. Given two sets of embeddings, i.e., t and s, trained independently on monolingual data, MUSE learns a mapping between the two sets such that translations are close in the shared space. It is based on a two-player game mechanism. The first player is a discriminator that is trained to distinguish between an element $W^{s \to t} \cdot e^s$ and an element e^t from t. In the

other hand, the second player $W^{s\to t}$ is trained to prevent the discriminator from achieving its goal. The discriminator aims at maximizing its ability to identify the origin of an embedding whilst $W^{s\to t}$ tries to prevent the discriminator from doing so by making $W^{s\to t} \cdot e^s$ as similar as possible to e^t.

7.3.3 Universal Embedding Layer

The goal of the *Universal Embedding* layer is to project vectors from embeddings of a low-resource language to a shared space that is close to the embedding space of a rich-resource language. It consists of three simple steps. First, let w_i be a one-hot sparse vector that represents the word i in the input sequence. We multiply this vector by the embedding matrix E_t of the low-resource, i.e., target, language to obtain the vector $x_i \in \mathbb{R}^d$ that represents the word in that embedding space.

$$x_i = w_i \cdot E_t \tag{7.1}$$

As soon as we get the vector x_i we use the transformation matrix $W^{t\to s}$ (subsection 7.3.2) to project it to the shared embedding space as follows:

$$z_i = x_i \cdot W^{t\to s} \tag{7.2}$$

Finally, we use Nearest Neighbor Ranking to refine the obtained vector z_i with the aim of projecting it into a subspace that captures the relevant information from the source language, i.e., the rich-resource one. For each vector z_i to be refined, the top-k semantically most similar vectors from the source embedding language, E_s are selected. The final representation vector of the input word is given by computing v_i:

$$v_i = \sum_{j=1}^{k} z_i \cdot \alpha(z_i, \hat{v}_j) \tag{7.3}$$

In this expression $\hat{v}_j$ is the jth selected vector. The function α is a normalized weighting function which finds the strength of the relationship between two vectors based on their semantic similarity. That is, $\alpha(x,y)$ is a value based on the distance $\Delta(x,y)$ between x and y as:

$$\alpha(x,y) = \frac{e^{\Delta(x,y)}}{\sum_{\hat{y}} e^{\Delta(x,\hat{y})}} \tag{7.4}$$

It is worth noting that $\alpha(x,y) \in [0,1]$ and $\sum_{\hat{y}} \alpha(x,\hat{y}) = 1$.

Figure 7.2 shows an illustration of the universal embedding layer steps. In box (a) there are two distributions of word embeddings, English words in red and Spanish words in black.

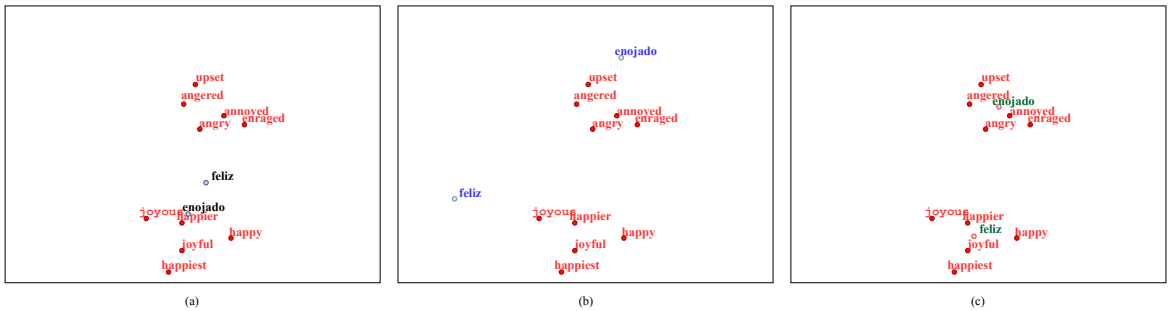


Figure 7.2: An example of the universal embedding layer steps.

Each dot represents a word in that space. Using MUSE, we project the Spanish points to a shared embedding space that is close to the English one, as shown in box (b). The final projection, obtained after refining using k-NN ranking, is shown in box (c).

7.4 Experiments and Results

7.4.1 Datasets

Labeled English Dataset. We used the dataset of SemEval from 2013 to 2017 to construct the English training and validation datasets described in Table 7.1. n this experiment, we considered only the positive and negative samples to pre-train a binary sentiment classification system.

	Training	Validation
Positive	19652	2375
Negative	7723	3972
Total	27375	6347

Table 7.1: The description of SemEval datasets.

Multi-Language Datasets. To evaluate the proposed system, we conducted experiments on the Spanish dataset of OpeNER and the MultiBooked Catalan [13]. We considered the two versions of this benchmark, the four-classes version and the binary one. The labels in the former are Strong-Negative, Negative, Positive, and Strong-Positive. The binary version is constructed by joining the strong and weak classes. Table 7.2 shows the statistical description of this benchmark. We denote this benchmark as ES-CA in this study.

	Spanish (ES)		Catalan (CA)	
	Train	Test	Train	Test
Strong-Positive	259	75	179	52
Positive	592	170	298	86
Negative	152	44	286	82
Strong-Negative	26	8	40	12
Total	1029	297	803	232

Table 7.2: The statistical description of ES-CA benchmark.

7.4.2 Experimental Settings

We used the test set of SemEval 2017 as a validation set to fine-tune the system's parameters. The macro-F1 score of this experiment was 88%. The embedding layer size is 300. We used the pre-trained word vectors for the English, Spanish and Catalan languages that were trained on Common Crawl and Wikipedia using fastText [19] with character n-grams of length five, a window of size of five and ten negative subsampling. We initialized the weights of the embedding layer with the English pre-trained word embeddings and applied a dropout of 0.2 to regularize it. All the pre-trained embedding models were used to train the embedding alignment. The encoder layer was composed of two bidirectional LSTM cells, each of size 256. Finally, we applied a dropout of 0.2 to the encoded representation. We trained the proposed system using the Adam method [84], with a learning rate of 0.0001, β=(0.9, 0.999) and a mini-batch size of 32 to minimize the cross-entropy loss function:

$$\mathscr{L} = -\mathbb{E}_{(x_i,y_i)\sim D}\left[y_i \cdot \log \hat{y}_i + (1-y_i)\cdot \log(1-\hat{y}_i)\right] \tag{7.5}$$

Here, D is the set of training samples, $\hat{y}_i$ is the predicted value obtained by passing the input sample x_i to the model, and y_i is the real value.

7.4.3 Comparison with other Systems

We compared the proposed system with the baselines and state-of-the-art systems used in the task of CLSA, including:

- **UniSent-MT**: Our English version system applied to the translated test sets.
- **SVM-MONO**: A support vector machine based system trained on the target languages using the training sets. The features used in this system are the average of the embeddings [13].

- **SVM-MT**: It is similar to the SVM-MONO, but the system was trained on the English dataset. The test sets were translated using the Google Translator API and passed to the SVM-MT system [13].
- **ARTEXTE**: An SVM-based system trained on features extracted from projected embedding spaces obtained by applying the method proposed by Artetxe et al., (2016) [13].
- **ARTEXTE-Ensemble**: An ensemble version of ARTEXTE created by training a random forest classifier on top of its predictions [13].
- **BARISTA**: An SVM-based system trained on features extracted from projected embedding spaces obtained by applying the method proposed by Gouws and Søgaard, (2015) [13].
- **BARISTA-Ensemble**: An ensemble version of BARISTA created by training a random forest classifier on top of its prediction [13].
- **BLSE**: A deep learning-based system trained to jointly represent sentiment information in a source and target language [13].
- **BLSE-Ensemble**: An ensemble version of BLSE created by training a random forest classifier on top of its predictions [13].

7.4.4 Results

In this section, we present the results of four different experiments: *binary classification* and *4-classes classifications* with the Spanish and Catalan languages. In addition, we study the effect of the size of training data and the value of k on the performance of the proposed system.

Tables 7.3 and 7.4 show the results of our system alongside with the baselines and the-state-of-the-art systems on the ES-CA benchmark. To assess the performance of our system, we use the Precision (P), Recall (R) and macro F1 evaluation metrics. The best results are in bold. The values marked by '*' indicate that our system is better than the non-ensemble versions.

In the case of the binary classification experiments, our system obtains the best F1 and P scores with the Spanish test set. The R score of BLSE is a bit better than the one of our system. Although BLSE-ensemble performs slightly better with the Catalan test set, our system offers the best P, R and F1 values with the Spanish and Catalan test sets in comparison with the non-ensemble versions of the other systems.

	Spanish			Catalan		
	P	R	F1	P	R	F1
UniSent (OurModel)	**83.7***	79.4*	**81.4***	84.3*	84.3*	84.3*
UniSent-MT	73.8	74.8	74.3	76.8	73.2	74.0
SVM-MONO	75.0	72.3	73.5	79.0	79.6	79.6
SVM-MT	82.3	76.6	79.0	78.0	76.8	77.2
BLSE	72.1	**80.1**	74.6	72.8	73.0	72.9
BLSE-Ensemble	79.5	78.7	80.3	**84.7**	**85.5**	**85.0**
Artetxe	75.0	64.3	67.1	60.1	61.2	60.7
Artetxe-Ensemble	65.3	61.3	62.6	63.1	63.3	63.2
Barista	64.7	59.8	61.2	65.3	61.2	60.1
Barista-Ensemble	60.1	55.5	56.0	63.4	62.3	62.5

Table 7.3: The results of the proposed system, baselines and the-state-of-the-art systems on Spanish (ES) and Catalan (CA) binary classification experiments.

In the case of the 4-classes experiments, our system outperforms all systems in terms of P, R and F1 metrics with the two languages (the improvement is about 4%). Generally speaking, MT-based systems perform better than the other projection-based systems (BLSE, Artetxe, and Barista). However, our system shows a clear advantage over all the non-ensemble versions including MT based systems. This finding confirms the usefulness of the proposed universal embedding layer as a projection method. It is crucial to note that our system shows a similar performance with the Spanish and Catalan test sets. The compared projection based systems, however, show slightly worse performance on the Catalan test set.

	Spanish			Catalan		
	P	R	F1	P	R	F1
UniSent (OurModel)	**74.7***	**51.2***	**54.2***	**63.3***	**53.5***	**54.3***
UniSent-MT	50.6	39.9	37.8	48.3	42.5	40.6
SVM-MONO	55.2	42.8	45.5	50.0	50.9	49.9
SVM-MT	51.8	48.5	48.8	51.8	48.5	48.8
BLSE	60.0	43.4	41.2	38.1	38.1	35.9
BLSE-Ensemble	49.5	51.2	50.3	54.1	53.9	53.9
Artetxe	40.1	36.9	34.9	21.6	29.8	23.0
Artetxe-Ensemble	43.5	44.1	43.8	46.5	48.7	47.6
Barista	44.1	37.9	39.5	36.4	38.5	36.2
Barista-Ensemble	48.3	46.6	47.1	52.8	53.7	53.0

Table 7.4: The results of the proposed system, baselines and the-state-of-the-art systems on Spanish (ES) and Catalan (CA) 4-class test version.

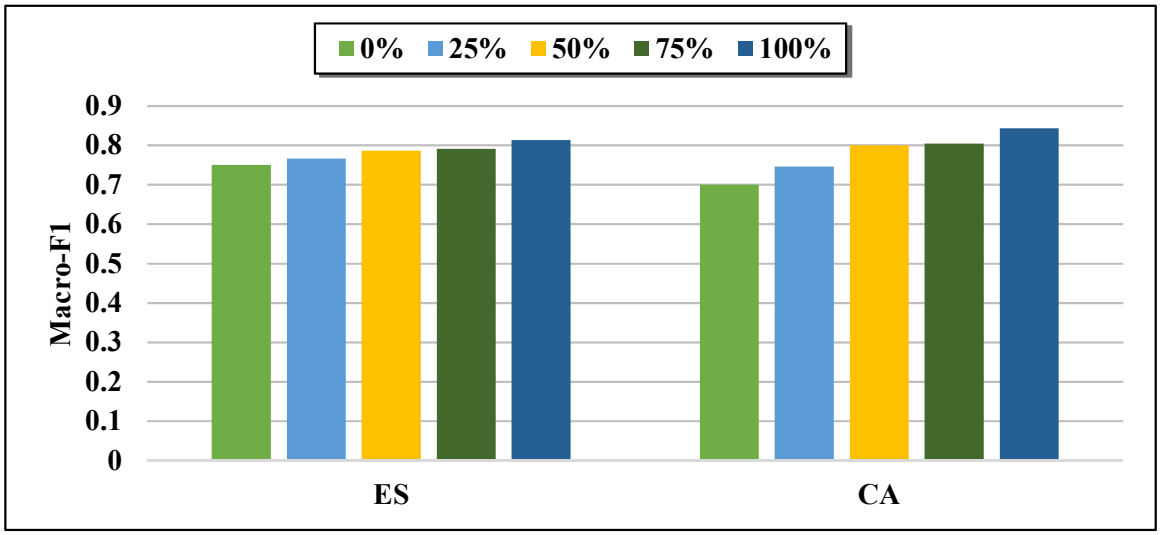


Figure 7.3: Macro F1 score vs training size.

The Effect of Training Size. In this experiment, we show the effectiveness of our system with varied training sizes. To assess that, we have conducted experiments on the binary version of the ES-CA benchmark. We have fine-tuned the system on varied training sizes arranged from 0% to 100% and reported the results on the test sets obtained in each case. These results are shown in Figure 7.3. These tests revealed that our system works with multiple languages even when there are no resources, e.g., annotated data. Moreover, increasing the size of the training data and fine-tuning the system helps to improve the systems' performance.

The Effect of k. In this analysis, we show the effect of the size of the selected neighbours in the inference phase. We tested the system on the test set of the ES-CA benchmark with different sizes of k varying from 1 to 8 and in two cases: without training (refers to the case when we use the English system and change the embedding layer) and with full training (refers to the case when we fine-tune the system by retraining it on the training sets shown in Table 7.2). The value 1 of k means that we ignore the refinement step of the universal embedding layer. Figure 7.4 shows the results of this experiment. The reported results indicate that choosing a value of k from 3 to 5 is ideal. It also points to the usefulness of the refinement step. Setting the value of k to 1 gives the worst performance in almost all cases.

7.5 Conclusion

We have presented a new system, UniSent, which is able to transfer sentiment knowledge from a rich-resource language to perform sentiment analysis on low-resource languages. The

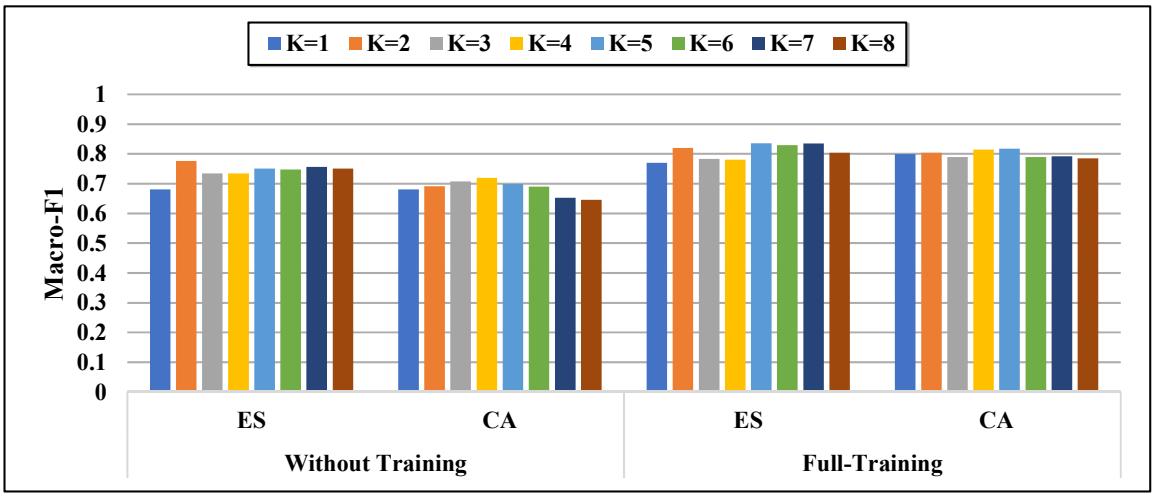


Figure 7.4: The effect of the value of k in the inference phase.

proposed system works even with languages without resources. The experimental results on the Spanish dataset of OpeNER and the MultiBooked Catalan show the effectiveness of our system. It outperforms the state-of-the-art systems in three out of four experimental setups. We have demonstrated, empirically, that our system is highly flexible, and adding a new language to it does not require to re-train it.

Chapter 8

SentiRank: Combining Aspect-Based Semantic Analysis with Multi-Criteria Decision Support Systems

8.1 Introduction

The massive spread of online platforms, such as social networks and online shopping sites, which allow sharing reviews and opinions, has changed different decision-making processes, especially the purchase process. "What other people think" has become an essential piece of information, as much as what we prefer, for most of our decision-making process. It is well known that price and other criteria have a significant influence on people' purchases; however, nowadays, online consumers' reviews have also become a common source of influence and an important input in people decision making [50, 150, 169]. Some studies have shown that 90% of consumers read online reviews before making a purchase decision [139, 148].

Commonly, service and product providers try to attract customers by publishing the best description of the features of their services. In such a case, it is easy to find most of the alternatives that fit the preferences of decision-makers. However, it is also important to consider the quality of the advertised features and the different aspects of alternatives. For instance, people usually prefer to visit restaurants with free WiFi, and we can find this information in the description of most restaurants. However, without the reviews of customers it is not possible to figure out the quality of this service.

Thus, to obtain better personalized recommendations we will consider two sources of information: the decision-maker preferences and online reviews. In this case, an important question is how to integrate these two sources to support and improve the decision-making

process. We can find in the literature numerous systems proposed for modeling decision-makers preferences and ranking alternatives to support the decision process [21, 68]. On the other hand, there are different systems for analyzing people' reviews on products or services [92]. However, there are not any works on exploiting user reviews alongside decision-maker preferences to support the decision process.

8.1.1 Objectives

The main aim of the work described in this chapter is to improve current multi-criteria decision support methods by including the information contained in online reviews about a set of alternatives. Hence, our intention in this work is to understand how we can exploit sentiment analysis to integrate the users' reviews in a multi-criteria ranking system to improve the quality of the decision-making process. Towards this objective we identify the following research questions:

- **Q1**: Which properties of user reviews can be used to improve the quality of the decision-making process?
- **Q2**: How can we extract those properties?
- **Q3**: How can we integrate the extracted properties in the decision-making problem formulation?
- **Q4**: How can we, automatically, solve the resulting decision-making problem?

8.1.2 Contributions

In response to the objectives described above, this chapter proposes the SentiRank system. It is based on a multi-criteria decision aid system to give users the ability to exploit users' reviews alongside their preferences in the process of alternatives ranking. To achieve that, we primarily examine the task of ABSA, introduced in Chapter 2. Given a collection of review documents about a set of alternatives, the goal is to automatically detect and analyze all expressions of sentiment towards a set of aspects in these documents. This problem setting involves mainly two subtasks.

- **Aspect Category Extraction**: Given a document of review about an alternative, we want to automatically extract all the aspects mentioned in the given document.
- **Aspect Category Polarity Identification**: Reviewers refer to product aspects in different contexts. They may use factual language and simply describe some aspects

(e.g., "the camera has a 3x optical zoom") or they may express their opinion towards an aspect (e.g., "the 3x optical zoom works perfectly"). Our goal is to automatically detect expressions of sentiment in customer reviews towards the extracted aspects. We further aim at analyzing the polarity of these expressions. We want to know whether an utterance is predominantly positive (e.g., "works perfectly") or negative ("is totally crap").

After that, we exploit an MCDA method, named ELECTRE-III, to develop a ranking system based on the decision-maker's preferences and the users' reviews. This is a well-known method in the area of MCDA, and it is described in Appendix A.

To summarize, the main contributions of this chapter are as follows:

1. We have developed an ABSA system to extract the set of aspect categories mentioned in reviews and identify the expressed sentiments towards the extracted aspects and used it as unit in the developed outranking system.

2. The experimental results reveal that our ABSA system achieves the state-of-the-art systems' performance.

3. We have proposed a new multi-criteria ranking model by integrating decision-makers' preferences and users' reviews, and used the ELECTRE-III method to solve the proposed problem.

4. We have evaluated the developed system in the Restaurant domain by collecting information on a set of restaurants in Tarragona and the reviews about them. After that we have used them to test the system.

The layout of this chapter is as follow. In Section 8.2, the proposed system, SentiRank, is explained step by step. We present the experiments and the results in Section 8.3. Section 8.4 presents a case study that employs the proposed system and discussed the results. Section **??** presents some conclusions.

8.2 System Description

The proposed system, SentiRank, receives as input a set of alternatives, e.g., restaurants, hotels, or laptops, to be ranked and their corresponding consumers' reviews. Figure 8.1 shows the overall architecture of SentiRank. It is composed by three primary units. The first one is the *customers' reviews* unit, which transforms the reviews about alternatives into a matrix of real numbers which we call the social performance table. The second one is the

domain analysis unit; it takes as inputs the description of the alternatives and the goals of the decision-makers and converts them to another matrix of real numbers called domain performance table. As soon as we get the output of the two units, we merge them to get the final performance table. Finally, we feed the performance table to the last unit, i.e., the *ranking unit*, to get the alternatives ranked using ELECTRE-III. The following subsections describe how these usints work.

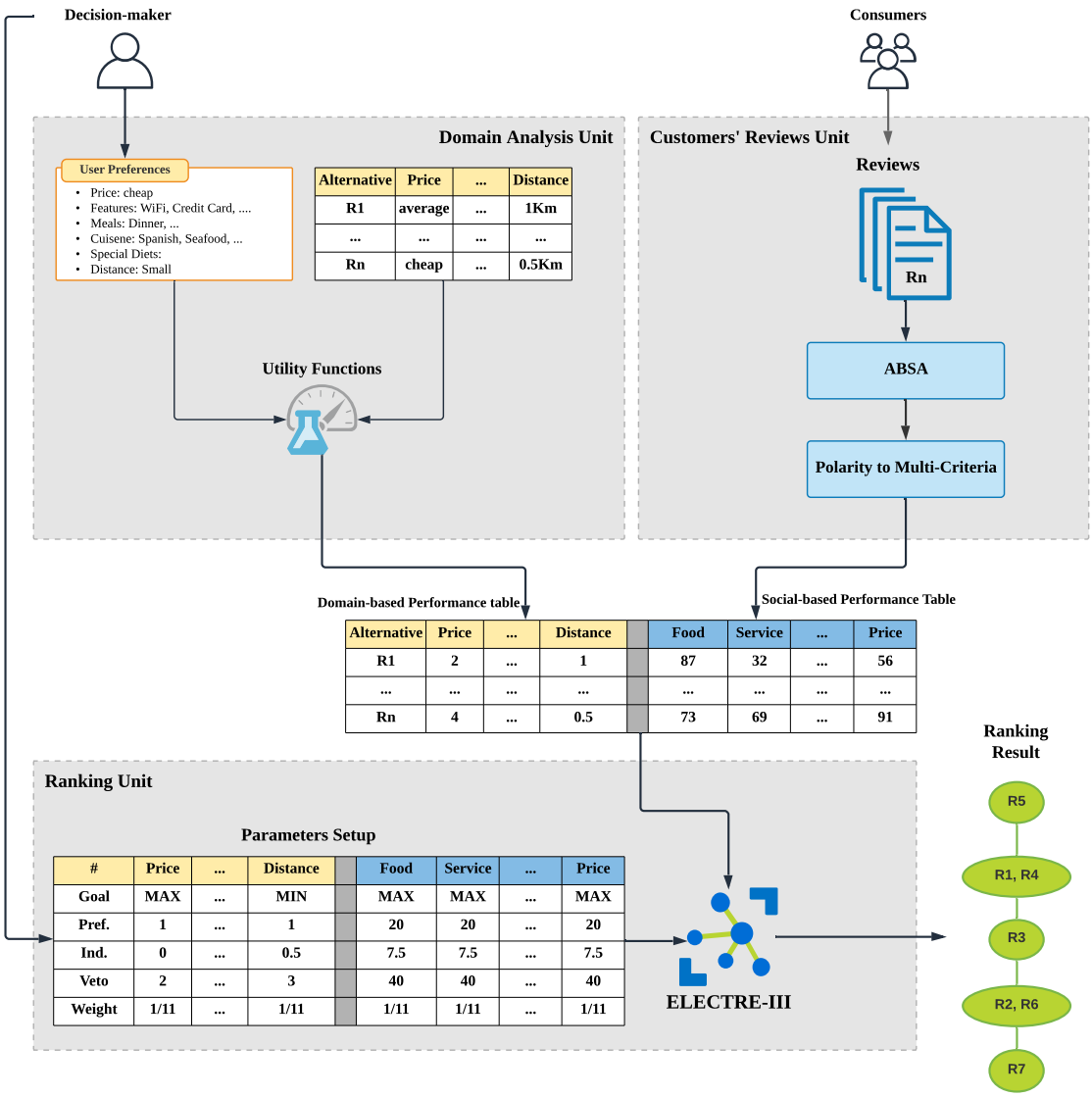


Figure 8.1: The architecture of SentiRank.

8.2.1 Customers' Reviews Unit

The input to this unit is a collection of customers' reviews about a set of alternatives, and the output is a numerical matrix called social-based performance table. Each column in the matrix represents a specific criterion. The main component of the customers' review unit is the aspect-based sentiment analysis (ABSA) model. Thus, we start by explaining how we built it below.

The ABSA model contains two sub-models. The first one aims to extract the set of aspects commented in a given review. We use the second one to identify the sentiment expressed in the review towards each extracted aspect. As shown in Figure 8.2, the first step is to preprocess the text in the review. After that, we split it into a list of sentences and tokenize each one. The next step is to use the aspect extraction sub-model to extract the aspects from each sentence. Finally, we feed the sentence and the extracted aspects to the sentiment classifier. We explain in the following sections these steps and how we built the models.

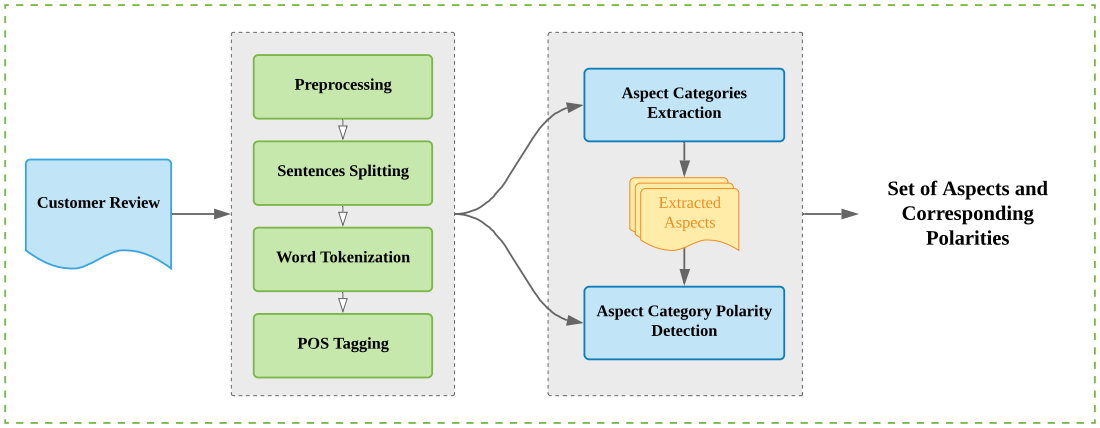


Figure 8.2: The architecture of the ABSA model.

Text Preprocessing

Text preprocessing is an essential step for most text analytics problems in general and especially for the sentiment analysis problems. The text in reviews and social posts (like tweets) is usually noisier than the one in articles and blogs. People tend to write their posts, messages, and reviews in an informal way. Thus, in our system, text cleaning and processing is essential. We preprocessed the text as explained in Chapter 3.

Aspect Category Extraction

Unlike the traditional single-label classification problem (i.e., multi-class or binary), where an instance is associated with only one label from a finite set of labels, in the multi-label classification problem [183], an instance is associated with a subset of labels. If we look to the problem of the aspect category extraction, we can notice that it is a typical multi-label classification problem, where a sentence may belong to none, one or more aspects. Thus, in this work, we regard the problem of *aspect category extraction* as a multi-label classification problem. As a result, we first present the definition of the multi-label classification problem and next explain our model.

Let $X = \{x_1, x_2, \ldots, x_n\}$ be the set of all instances and $A = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ be the set of all aspects, where n is the number of samples and m is the number of aspects. We can define the set of data:

$$D = \{(x_i, \hat{A}_i) | x_i \in X \text{ and } \hat{A}_i \subseteq A \text{ is the set of aspects associated with } x_i\} \tag{8.1}$$

In this case, D is called a supervised multi-label dataset. As we stated in Chapter 5, one common strategy to solve this problem is to transform the problem into a traditional classification problem. To this end, we use the most common method, binary relevance [126, 163, 183]. The idea of the binary relevance method is simple and intuitive. The aspect extraction problem is decomposed into multiple binary problems, one problem for each aspect. Then, an independent binary classifier is trained to predict the relevance of each one of the aspects. In order to train the binary classifier for each corresponding aspect α_j, $1 \leqslant j \leqslant m$, we first construct the following binary dataset:

$$D_j = \{(x_i, \phi(\hat{A}_i, \alpha_j)) | 1 \leqslant i \leqslant n\} \tag{8.2}$$

In this expression, $\phi(\hat{A}_i, \alpha_j)$ is a function that returns 1 if $\alpha_j \in \hat{A}_i$ and 0 otherwise.

Figure 8.3 illustrates the decomposition process.

After that, we build a SVM binary classifier $h_j : X \rightarrow \{0, 1\}$ from the training set D_j.

During the inference, and when we want to extract the aspects of a new and unseen sentence x, the system predicts as relevant labels the ones that are predicted to be 1 by the binary classifier h_j for the instance x. Hence, the function $h : X \rightarrow 2^A$ is defined as follows:

$$h(x) = \{\alpha_j | h_j(x) = 1, 1 \leqslant j \leqslant m\} \tag{8.3}$$

$A = \{a_1, a_2, a_3, a_4\}$

Ex#	Label Set
1	$\{a_1, a_2\}$
2	$\{a_2, a_4\}$
3	$\{a_1, a_3, a_4\}$
4	$\{a_4\}$

Ex#	a_1
1	true
2	false
3	true
4	false

Ex#	a_2
1	true
2	true
3	false
4	false

Ex#	a_3
1	false
2	false
3	true
4	false

Ex#	a_4
1	false
2	true
3	true
4	true

Figure 8.3: Illustration of the decomposition of a multi-label dataset into multiple binary datasets.

Aspect Polarity Detection

The goal of this sub-task is to detect the sentiment expressed towards a given aspect in a given sentence. For each input pair (sentence, aspect), the output is a single sentiment label: positive, negative, neutral, or conflict. We train one multi-class SVM classifier for each aspect category. The feature set is extended to incorporate the information about a given aspect category as explained in the next chapter. We define the aspect polarity detection function $f : X \times A \rightarrow \{positive, negative, neutral, conflict\}$ as the following:

$$f(x,\alpha) = \begin{cases} \text{positive} & \text{If, } \arg\max \sigma(x,\alpha) = 1 \\ \text{negative} & \text{If, } \arg\max \sigma(x,\alpha) = 2 \\ \text{neutral} & \text{If, } \arg\max \sigma(x,\alpha) = 3 \\ \text{conflict} & \text{If, } \arg\max \sigma(x,\alpha) = 4 \end{cases} \tag{8.4}$$

Where, $\sigma : X \times A \rightarrow \mathbb{R}^4$ is the classification support function. We summarize the aspect extraction and polarity detection steps in the algorithm 2.

Algorithm 2: Aspect extraction and polarity detection algorithm.

input : A review *rev*.
output : A list of pairs $L = \{(\alpha_1, p_1), ..., (\alpha_l, p_l)\}$, where α_k is an aspect and p_k its corresponding polarity.

```
L ⟵ Empty;
Split rev to a set of normalized sentences and store them in Sents;
for x ∈ Sents do
    Let Â be the set of aspects extracted from x using 8.3;
    for α ∈ Â do
        Let p be the polarity obtained by 8.4;
        Add the pair (α, p) to L;
    end
end
```

From Polarity to Multi-Criteria

Recall that our main goal is to integrate consumers' reviews about a set of alternatives with the decision-maker's preferences in the decision making process. Hence, we consider each aspect in the social domain as a criterion. We call these set of aspects as the social-criteria. Algorithm 3 shows the transformation steps. The inputs to the algorithm are a set of alternatives O, a set of aspects A (from a specific domain, e.g., restaurants) and a collection of reviews, RV, about each alternative. The output is a matrix with r rows and m columns, where r denotes the number of alternatives and m is the number of aspects. Each value in the cell i, j is the polarity index which represents the degree of consumers' satisfaction about the aspect $\alpha_j \in A$ of the alternative $o_i \in O$. We use the polarity index defined in [70] as the following:

$$P_{index}(P,N) = \begin{cases} 1 - \frac{N}{P} & \text{if, } P > N \\ \frac{P}{N} - 1 & \text{if, } P < N \\ 0 & otherwise \end{cases} \tag{8.5}$$

Here, P is the number of positive opinions and N is the number of negative opinions. The range of P_{index} is $[-1, 1]$, so w transform it to $[0, 100]$ by adding 1 and multipling the result by 50. In this interval 0 means the lowest satisfaction and 100 refers to the highest degree of satisfaction.

Algorithm 3: From polarity to performance table algorithm.

```
input  : A set of alternatives O, a set of aspects A and reviews RV.
output : The social performance table, M.
initialization;
r ⟵ |O|;
m ⟵ |A|;
Pos ⟵ Zeros[r,m];
Neg ⟵ Zeros[r,m];
M ⟵ Zeros[r,m];
i ⟵ 1;
while i ⩽ r do
    RV_i ⟵ RV[i];
    for rev ∈ RV_i do
        let L be the list of extracted aspects from rev using Algorithm 2;
        for (α, p) ∈ L do
            let j be the index of aspect α;
            if p == "positive" then
                Pos[i, j] = Pos[i, j] + 1;
            else
                if p == "negative" then
                    Neg[i, j] = Neg[i, j] + 1;
                end
            end
        end
    end
    i = i + 1;
end
i ⟵ 1;
while i ⩽ r do
    j ⟵ 1;
    while j ⩽ m do
        P ⟵ Pos[i, j];
        N ⟵ Neg[i, j];
        M[i, j] = P_index(P,N);
        j = j + 1;
    end
    i = i + 1;
end
```

8.2.2 Domain Analysis Unit

As shown in Figure 8.1, this unit receives as an input the decision-maker preferences and the description of the alternatives and feeds them to the utility functions to transform the preferences into numerical values. Both the description of the alternatives and the user preferences values are given for a specific criteria which we call the *domain criteria* in this work.

Decision-makers can express their preferences in different ways depending on the type of each criterion. For example, users can specify their preferences on the price criterion as a range of amounts or as a linguistic value (e.g., very cheap, cheap, average or expensive). Recall that the goal of this unit is to convert the expression of user's preferences on the domain criteria into a numerical value that reflects the satisfaction degree of the user. Hence, we define in this work three different utility functions that measure the performance of a feature of the alternative. These functions can be used in different domains, such as the price criteria in restaurants, hotels or laptops. However, it is possible to extend the functions by defining new ones.

Identity

This utility function does not need any processing, the preferred value of this criterion is given as a real value x in its range. We simply express this criterion as follows.

$$g(x) = x \tag{8.6}$$

Categorical

The range of this kind of utility function is always $[0,1]$ and the goal is to maximize it. The value 0 means the lowest degree of satisfaction and 1 means the highest. The utility function is defined as the following:

$$g(U;F) = \frac{|U \cap F|}{|U|} \tag{8.7}$$

In this expression, U refers to the items the user prefers and F is the set of available items in this criteria; it is also called the domain of the criterion.

Example : let us have a categorical criterion and the user preference is $U = \{a,b,c,d,e\}$ and what an alternative provides is $F = \{a,d,f,g,h\}$, then the utility value of this alternative given the preferred value is $g(U;F) = \frac{|\{a,d\}|}{|\{a,b,c,d,e\}|} = \frac{2}{5} = 0.4$.

Linguistic

The range of this kind of function is $[1, |L|]$, where L is the set of linguistic terms in the domain of the function and the goal is to have the maximum. The value 1 means the lowest degree of satisfaction and $|L|$ means the highest. The utility function is defined as the following:

$$g(x;p) = |L| - abs\,(pos(x) - pos(p)) \tag{8.8}$$

Where $x, p \in L$ are the alternative value and the preferred value respectively, *pos* is the position of the linguistic value in L and *abs* is the absolute value.

Example : let us have the price feature and it is represented as a linguistic variable. Let $L = \{very\ cheap, cheap, average, expensive\}$ be the linguistic values. Here, $pos(cheap)$ is 2. Assuming that the user prefers $p = cheap$ and the alternative value is $x = expensive$, then the utility value of this alternative given the user preference is $g(expensive; cheap) = 4 - abs(4-2) = 4-2 = 2$.

8.2.3 Ranking Unit

The inputs to this unit are the performance table that contains both the domain and the social-based matrices and the criteria properties setup. Each criterion, either domain-based or social-based, has the following properties: minimum value, maximum value, the goal (MAX or MIN), the preference threshold, the indifference threshold, the veto threshold and the weight. The decision-maker is responsible of setting the values of these properties. The meaning of the thresholds and weights in ELECTRE is explained in Appendix A.

The outranking procedure used in SentiRank is also explained in Appendix A. Given the performance table and the criteria' properties, we first apply the construction procedure to get the outranking relation. Afterwards, we apply the Net Flow Score procedure to perform the exploitation step to get the final ranking of the alternatives.

8.3 Experiments and Results

We have evaluated and tested the proposed system in the restaurant domain. To this end, we built an aspect extraction model and aspect polarity detection for this domain. After that, we collected the details and reviews of a set of restaurants from Tarragona as alternatives.

Finally, we tested the system with two different users' profiles (i.e., user preferences). Next subsections explain that in details.

8.3.1 Data

Table 8.1 shows the description of the domain criteria and the social-based criteria. We chose the domain criteria from the details of the restaurants on TripAdvisor. The social criteria are the aspects defined in SemEval-2014 Task 4: Aspect Based Sentiment Analysis [122].

Criterion	Utility Function	Domain	Goal
A- Domain			
Price	Linguistic	{Very Cheap, Cheap, Average, Expensive}	MAX
Cuisines	Categorical	Spanish, Seafood, Bar, ...	MAX
Special Diets	Categorical	Vegetarian Friendly, ...	MAX
Meals	Categorical	Lunch, Drinks, ...	MAX
Features	Categorical	Private Dining, Seating, WiFi, ...	MAX
Distance	Identity	$[0, \infty]$	MIN
B-Social			
Food	Identity	[0, 100]	MAX
Price	Identity	[0, 100]	MAX
Service	Identity	[0, 100]	MAX
Ambience	Identity	[0, 100]	MAX
Anecdotes	Identity	[0, 100]	MAX

Table 8.1: The description of the criteria.

Sentence	Aspect/Polarity
But the staff was so horrible to us.	(service, negative)
All the money went into the interior decoration, none of it went to the chefs.	(ambience, positive), (food, negative)
The food was delicious but do not come here on a empty stomach.	(food, conflict)
I grew up eating Dosa and have yet to find a place in NY to satisfy my taste buds.	(anecdotes, neutral)

Table 8.2: Examples of sentences from the training dataset.

We trained our aspect extraction and aspect polarity detection models on the publicly available dataset provided by the organizer of SemEval-2014 Task 4: Aspect Based Sentiment Analysis [122]. The training data was composed by 3041 English sentences, whereas the test data size was 800 English sentences. Each sentence is labeled with none, one or more aspects with their corresponding polarities as shown in Table 8.2. We present in Table 8.3 the

distribution of the polarity classes per category. We can notice that "positive" is the majority polarity class while the dominant aspect category is "food" in both the training and test datasets. Additionally, we collected the details of 23 restaurants from Tarragona as explained in section 8.4.

Aspect	Positive		Negative		Conflict		Neutral		Total	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
FOOD	867	302	209	69	66	16	90	31	1232	418
PRICE	179	51	115	28	17	3	10	1	321	83
SERVICE	324	101	218	63	35	5	20	3	597	172
AMBIENCE	263	76	98	21	47	13	23	8	431	118
ANECD	546	127	199	41	30	15	357	51	1132	234
Total	2179	657	839	159	163	52	500	94	3713	1025

Table 8.3: Aspect categories distribution per sentiment class.

8.3.2 Sentiment Analysis Model: Training Setup

We describe in this section the training setup of the aspect extraction and the aspect polarity detection models. We used LIBLINEAR[1] to build all the linear classifiers. First, we define all the features used in our system and then we show the features configuration of each model. After that, we present the evaluation metrics, the performance of the models and discuss the results.

Features

The set of features used in our system are the following:

- **Word ngrams**: presence or absence of contiguous sequences of $1,2,...,n$ tokens.
- **Character ngrams**: presence or absence of contiguous sequences of $3,4,...,n$ characters.
- **Part-Of-Speach tags**: the number of occurrences of each part-of-speech tag.
- **Lexicon-Based features**: Sentiment lexicons are lists of words with associations to positive and negative sentiments. For each token w and emotion or polarity score $score(w;l)$ in a given lexicon l, we use the sentiment/emotion score to obtain the following features: (1) the total score $\sum_w score(w;l)$; (2) the maximal score $\max_w score(w;l)$;

[1] https://www.csie.ntu.edu.tw/ cjlin/liblinear/

(3) the number of tokens in the input sentence with score(w; l) > 0; (4) the minimal score $\min_w score(w;l)$ (5) the number of tokens in the input sentence with score(w; l) < 0; (6) the score of the last token in the sentence. The set of lexicons used in this work are the following:

- Yelp Restaurant Sentiment Lexicon (Yelp-Res): this lexicon was automatically generated from the customer reviews for food-related businesses from the Yelp Dataset [2]. It has 39,274 unigram entries [86].
- NRC Hashtag Sentiment Lexicon (NRC-Hashtag): this lexicon has entries for 39,413 unigrams and 178,851 bigrams. This lexicon was constructed from a pseudo-labeled corpus of tweets [86].
- NRC Yelp Word-Aspect Association (WA-Lexicon): 183,935 reviews from the Yelp Dataset were used to generate lexicons of terms associated with the aspect categories of food, price, service, ambiance, and anecdotes. For each term w and each category c an association score was calculated using pointwise mutual information [85]. The difference between this lexicon and the others is that these lexicon is aspect associated whereas the others are general.

- **Word Clusters**: Word clusters can provide an alternative representation of text, significantly reducing the sparsity of the token space. The CMU pos-tagging[3] tool provides 1000 clusters produced with the Brown clustering algorithm on 56 million English language tweets. We use the number of occurrences of tokens from each of the 1000 clusters as clusters-based features.

Configurations

Table 8.4 shows the set of features used to build the aspect category detection models. All these configurations were obtained using K-Fold cross-validation, where $k = 5$.

	Word-Ngram	Char-Ngram	POS	Clusters	WA-Lexicon
Food	✓, N=3		✓	✓	✓
Price	✓, N=4		✓	✓	
Service	✓, N=1	✓, N=6		✓	
Ambience	✓, N=3		✓	✓	✓
Anecdotes	✓, N=3		✓	✓	✓

Table 8.4: The set of features used with each aspect category.

[2] http://www.yelp.com/dataset_challenge
[3] http://www.cs.cmu.edu/ ark/TweetNLP/clusters/50mpaths2

Following the approach used in [85], the feature set used to train the aspect-category polarity detection model was extended to incorporate the information about a given aspect category c using a domain adaptation technique as the following: each word w has two copies, *w#general* (for general domain) and *w#c* (for the specific category of the instance). For example, for the input pair ("The bread is top notch as well.", 'food') two copies of the word "top" would be used: "top#general" and "top#food". This technique allows the classifier to learn common sentiment features (e.g., the word "good" is associated with positive sentiment for all aspect categories). At the same time, aspect-specific sentiment features can be learned from the training instances pertaining to a specific aspect category (e.g., the word "delicious" is associated with a positive sentiment for the category 'food').

After sentences were tokenized and tagged, the following features were extracted: General Word-Ngrams with $N = 3$, Category-Based Word-Ngrams with $N = 3$, POS tags, Clusters and lexicon-based features. Again, we used two types of lexicons (global lexicons and target-specific lexicon). The global lexicons used are Yelp-Res and NRC-Hashtag whereas W–A Lexicon is used as a target-specific lexicon.

Evaluation Metrics

To evaluate the aspect category detection model, we use the precision (P), recall (R) and F1 measures. To evaluate the aspect-category polarity detection, we use the accuracy metric.

Results

Table 8.5 shows the performance of the proposed aspect detection model for each category. In terms of $F1$, in general, the model shows a very good performance in all the categories. The model gives the best performance with the food category followed by the price category. The worst performance is registered with the anecdotes category.

Aspect	P	R	F1	Accuracy
Food	94.43	89.23	91.76	91.63
Service	93.75	87.21	90.36	96
Price	98.61	85.54	91.61	98.25
Ambience	86.73	83.05	84.85	95.63
Anecdotes	80	75.21	77.53	87.25
Average	90.704	84.048	87.222	93.752

Table 8.5: The performance of the proposed aspect detection model.

We compare the performance of the proposed system with the top three systems in SemEval-2014 Task 4: Aspect Based Sentiment Analysis [122] as shown in tables 8.6 and

8.7. The best values are shown in bold, where the underlined values indicate that our system gives the second best performance. Our system almost achieves the performance of the best system, NRC-Can-2014 [85] in both tasks.

System	P	R	F1
NRC-Can-2014 [85]	**91.04**	**86.24**	**88.58**
UNITOR [25]	84.98	85.56	85.26
XRCE [22]	83.23	81.36	82.28
Our System	90.41	84.68	87.45

Table 8.6: The comparison of the proposed aspect detection model with the state-of-the-art.

The next section presents a case study that confirms the robustness of the proposed models.

8.4 Case Study

As a case study we collected the information and the users' feedback of 23 restaurants in Tarragona from TripAdvisor. As shown in Figure 8.4, the details of each alternative (i.e., restaurant) are the domain aspects (table 8.1) used in this study. In addition to the users' reviews, TripAdvisor provides the overall rating and the rating of each aspect, that can be seen the left side of Figure 8.4.

Table 8.8 shows the ratings of the collected alternatives. The rest of thedetails are suppressed because of the layout issue. However, readers can find this information in TripAdvisor.

We analyse the correlation between the output of our sentiment analysis system and the ratings provided by TripAdvisor as follows: first, we calculate the polarity index of each pair of alternative and aspect given a set of reviews as described in algorithm 3; then, the overall polarity index of an alternative is obtained by taking the average of all the aspects; finally, we convert the polarity index into a rating value range from 0 to 5 by dividing the polarity index

System	Accuracy
NRC-Can-2014 [181]	**82.92**
UNITOR [25]	76.29
XRCE [22]	78.14
Our System	82.44

Table 8.7: The comparison of the proposed aspect-category polarity detection model with the state-of-the-art.

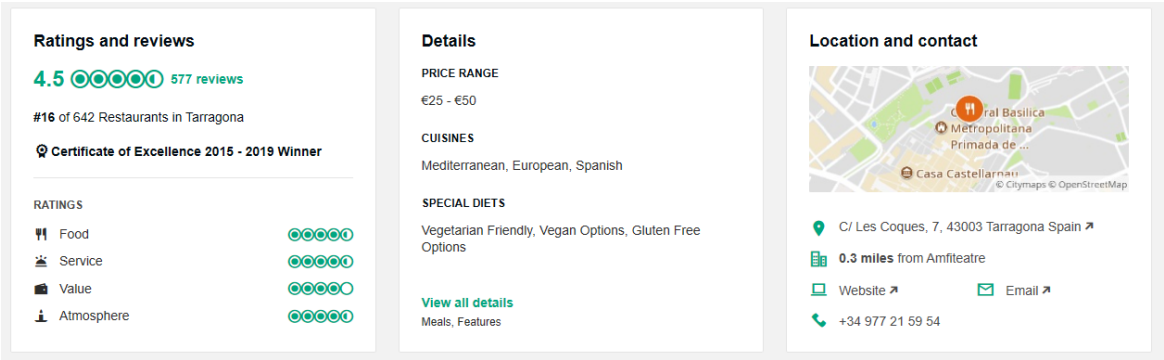


Figure 8.4: Restaurant details in TripAdvisor.

Id	Restaurant Name	Overall	Food	Service	Price	Amb.
1	AQ	4.5	4.5	4.5	4	4.5
2	ELIAN Cafe Restaurant	4.5	4.5	4.5	4.5	4.5
3	La Caleta	4.5	4.5	4.5	4	4.5
4	Tarakon	4.5	4.5	4.5	4.5	4
5	Sadoll Restaurant	4.5	4.5	4.5	4.5	4.5
6	El Taller	4.5	4.5	4.5	4	4
7	El Trull	4.5	4.5	4.5	4.5	4.5
8	Les Coques	4	4	4	4	4
9	Mas Rosello	4	4.5	4.5	4	4.5
10	Les Fonts de Can Sala	4.5	4.5	4.5	4	4.5
11	La Xarxa	4.5	4.5	4.5	4.5	4
12	El Encuentro	4.5	4	4	4.5	3.5
13	La Capital	4	4	4.5	4	3.5
14	Barquet	4.5	4.5	4	4	4
15	Arcs Restaurant	4.5	4.5	4.5	4	4.5
16	Restaurante Ca L Eulalia	4.5	4.5	4	3.5	3.5
17	Restaurante Club Nautico Salou	4.5	4.5	4.5	4	4.5
18	Octopussy	3.5	3.5	3.5	3.5	3.5
19	Palermo 1962 S.c.p.	4	4	4	4	3
20	The Cotton Club Restaurant & Cocktails	4.5	4.5	4.5	4.5	3.5
21	Lizarran Parc Central	3	3.5	3.5	3	3
22	Buffalo Grill	3	3.5	3.5	3	4
23	Indian Restaurant Mirchi Tarragona	3	2	2	3	3

Table 8.8: The alternatives and the rating based on TripAdvisor.

value by 20. Once we get the rating extracted from the reviews we perform the correlation analysis.

Figure 8.5 shows the overall correlation analysis of all aspects. We assumed that the relation between the variables is linear; the slope of the line indicates the strength of the relation. Given that, we can notice that the correlation is strong (80%) which confirms the robustness of our sentiment analysis system and indicates that it is applicable to the restaurant domain.

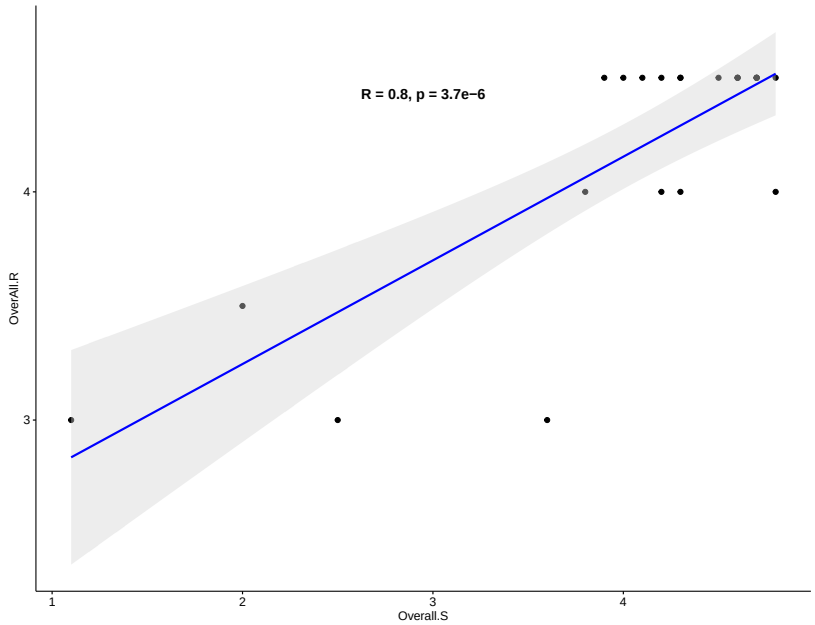


Figure 8.5: The overall correlation analysis of all aspects. Overall.S refers to the output of our system whereas Overall.R refers to the overall rating provided by TripAdvisor.

Thus, we used the proposed system, i.e., SentiRank, to combine the users' preferences with the consumers' feedback to rank the alternatives. The next subsection shows the output of SentiRank on two cases with two user profiles. In the first case, we only use the user's preferences (i.e., the domain criteria) as inputs to the ranking procedure. The second case combines the user's preferences with the consumers' feedback.

Profile 1

In this profile, we assume the decision-maker has the following preferences:

- **Price**: Cheap.
- **Cuisines**: Spanish, Seafood, Bar.
- **Special Diets**: Vegetarian Friendly, Gluten Free Options,
- **Meals**: Lunch, Drinks, Branch.

- **Features**: Private Dining, Seating, Wheelchair Accessible, Free Wifi, Accepts Credit Cards, Table Service, Parking Available, Highchairs Available, Full Bar.
- **Distance**: the distance from the town's centre must be as small as possible.

We also assume that the decision-maker is strict with the Special Diets and Meals criteria and flexible or tolerant with the Cuisine and Features criteria. Based on that the decision-maker provides the values of the indifference, preference and veto thresholds as shown in Table 8.9. All the social-based criteria have the same values of the thresholds.

Criterion	Indifference	Preference	Veto
A - Domain			
Price	0	1	2
Cuisine	0	0.4	No Veto
Special Diets	0	0	0.5
Meals	0	0	0.5
Features	0	0.4	No Veto
Distance	0.5	1	3
B - Social			
Food	7.5	20	40
Price	7.5	20	40
Service	7.5	20	40
Ambience	7.5	20	40
Anecdotes	7.5	20	40

Table 8.9: The values of the criteria thresholds.

Figures 8.6 and 8.7 show the domain and social performance tables respectively. Figure 8.8 shows the ranking results. The left side graph is the ranking when we only consider the domain criteria. The graph in the right side shows the ranking of the same alternative by adding the social-based criteria to the process. The alternatives are ranked based on the NFS value. The first indication from the results is the effect of the social-based criteria. There are clear differences in the ranking results.

The disagreement between the two cases can be attributed to the fact that most of the alternatives try to attract customers as much as possible by advertising, in the online platform, features that match the customers' preferences. On the contrary, consumers give their feedback and experiences with each alternative. This can be confirmed by checking the utility values of the domain criteria in Figure 8.6. We found most of the alternatives have similar values in most of the criteria especially with the special diets and meals criteria which we know that the user considers strictly. Moreover, we found that the most discriminative criteria in the domain criteria is the distance followed by the features.

Seq	Restaurant	Price	Cuisines	Special Diets	Meals	Features	Distance
1	AQ	2	0.33	1	0.33	0.67	2.6
2	ELIAN Cafe	3.5	0.33	1	0.33	0.33	0.5
3	La Caleta	2	0.33	1	0.33	0.89	3.5
4	Tarakon	3	0.33	0.5	0.33	0.33	1.9
5	Sadoll	3.5	0.33	1	0.33	0.67	2.1
6	**El Taller**	**3.5**	**0.33**	**1**	**0.33**	**0.78**	**1.4**
7	El Trull	3.5	0.33	1	0.33	0.67	7.6
8	Les Coques	2	0.33	1	0.33	0.56	2.3
9	Mas Rosello	2	0.33	1	0.33	0.67	6.1
10	Les Fonts de Can Sala	2	0.00	1	0.33	1.00	3
11	La Xarxa	3.5	0.67	0.5	0.33	0.78	1.6
12	El Encuentro	3	0.33	0	0.33	0.67	0.9
13	La Capital	3	0.67	0.5	0.00	0.78	1
14	**Barquet**	**3.5**	**0.67**	**1**	**0.33**	**0.56**	**1.1**
15	Arcs	2	0.33	1	0.33	0.67	2.5
16	Ca L Eulalia	2	0.67	0.5	0.33	0.56	2.3
17	Club Nautico Salou	3.5	0.67	0	0.33	0.89	14.3
18	Octopussy	3.5	0.67	0.5	0.67	0.89	2
19	Palermo 1962 S.c.p.	3.5	0.67	0.5	0.33	0.44	11.2
20	The Cotton Club & Cocktails	3.5	0.33	1	0.67	0.78	11.3
21	Lizarran Parc Central	3.5	0.33	0	0.33	0.56	1.3
22	Buffalo Grill	3.5	0.00	0	0.33	0.33	5.7
23	Indian Mirchi	3.5	0.00	0	0.00	0.11	0.75

Figure 8.6: Profile 1: the domain performance-table. The colors indicate the degree of satisfaction and range from "green", the worst, to the "red", the best.

Hence, by considering the social-based criteria we can comment the following observations:

- 'El Taller' is in position 2 in the first case (a) and in position 1 in the second one (b). Inspecting the utility values of the best two alternatives in the social performance table reveals that the alternative 'El Taller' shows better performance than 'Barquet'.

- The alternatives 'ELIAN Cafe' and 'Sadoll' are promoted to position 2.

Seq	Restaurant	Food	Service	Price	Ambience	Anecdotes
1	AQ	95.51	91.67	38.89	91.30	93.75
2	ELIAN Cafe	98.75	93.33	100.00	83.33	96.88
3	La Caleta	97.22	76.92	100.00	100.00	96.15
4	Tarakon	96.43	100.00	100.00	75.00	90.38
5	Sadoll	95.28	95.00	95.00	96.15	96.30
6	El Taller	97.83	91.67	83.33	100.00	90.91
7	El Trull	97.22	76.92	100.00	100.00	96.15
8	Les Coques	92.47	68.42	33.33	96.15	91.67
9	Mas Rosello	96.15	100.00	87.50	97.22	100.00
10	Les Fonts de Can Sala	97.22	100.00	100.00	100.00	75.00
11	La Xarxa	100.00	80.00	75.00	75.00	97.37
12	El Encuentro	92.00	75.00	87.50	66.67	83.33
13	La Capital	81.25	87.50	100.00	83.33	73.08
14	Barquet	92.50	82.50	81.25	80.77	82.00
15	Arcs	95.08	91.59	75.00	94.26	95.08
16	Ca L Eulalia	96.88	84.62	35.71	78.57	92.31
17	Club Nautico Salou	91.67	85.71	83.33	96.15	100.00
18	Octopussy	50.00	14.71	50.00	16.67	70.00
19	Palermo 1962 S.c.p.	93.65	92.31	70.00	77.27	90.48
20	The Cotton Club & Cocktails	87.01	86.50	78.89	92.22	87.59
21	Lizarran Parc Central	50.00	50.00	50.00	50.00	50.00
22	Buffalo Grill	78.95	81.25	50.00	50.00	100.00
23	Indian Mirchi	20.00	10.00	50.00	0.00	25.00

Figure 8.7: Profile 1: the social performance-table. The colors indicate the degree of satisfaction and range from "green", the worst, to the "red", the best.

- The position of the alternatives 'La Caleta' and 'Les Fonts de Can Sala' are changed from 7 and 13 to 3 respectively. If we look at the utility values of the domain criteria (shown in Figure 8.6) of these two alternatives, we can see that they show very similar performance to the best two (i.e., Barquet and El Taller) except with the distance and the price criteria. Looking at the values of the social-based criteria (i.e., the polarity index) of these two alternatives, we can notice that they have very high scores, which cause the large jump. This can indicate that price and distance are no longer critical criteria in people's decisions. This can also be observed with the alternative 'Trakon'.

- The alternatives in the worst positions have small differences.

Thus, in general, the social-based criteria impact the ranking process.

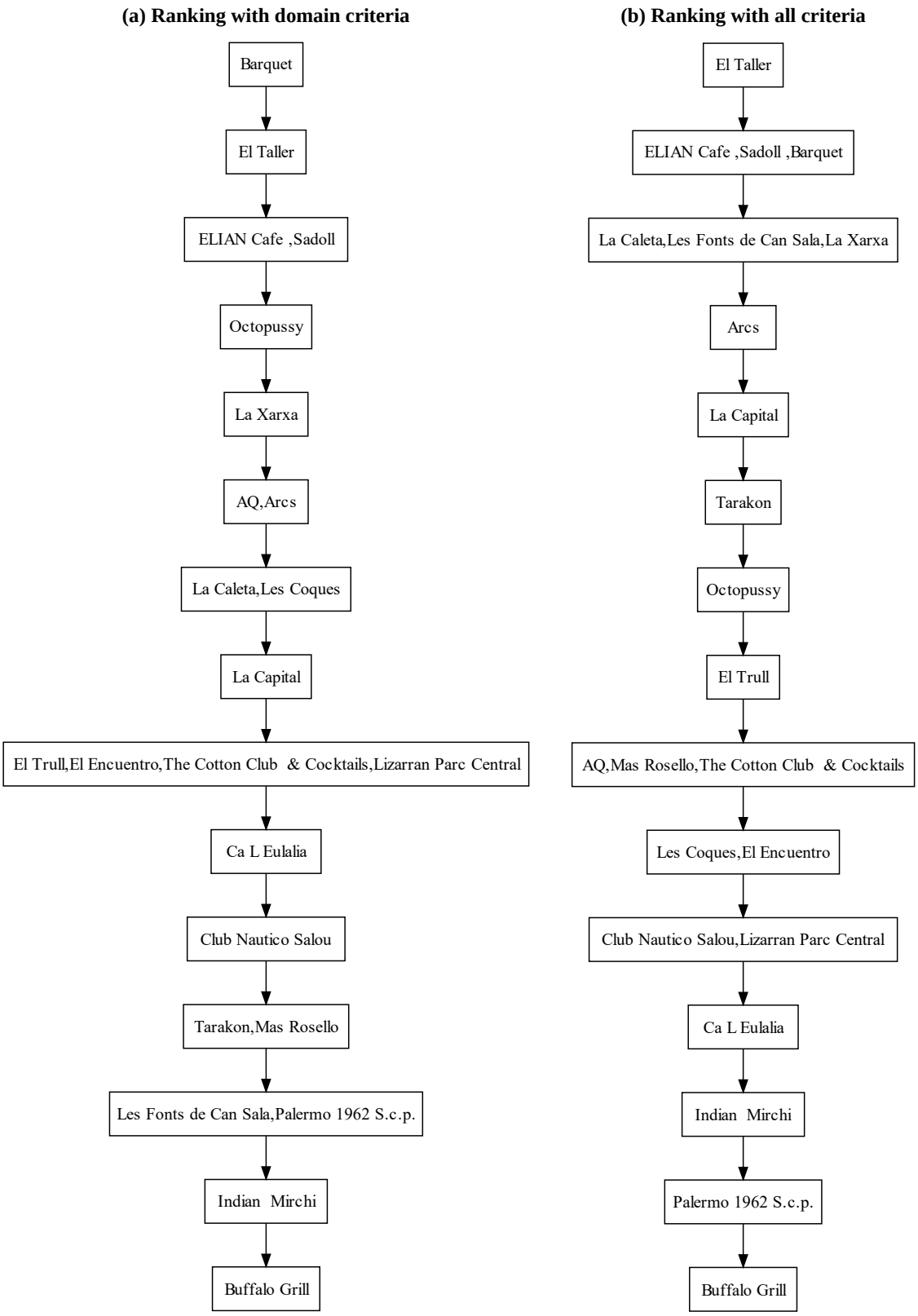


Figure 8.8: Profile 1: ranking result.

Profile 2

In this profile, we assume the decision maker has the following preferences:

- **Price**: Expensive.
- **Cuisines**: Barbecue, Bar, Grill.
- **Special Diets**: Vegetarian Friendly, Gluten Free Options.
- **Meals**: Dinner, Drinks, Late Night.
- **Features**: Private Dining, Serves Alcohol, Table Service, Free Wifi, Accepts Credit Cards, Outdoor Seating, Parking Available, Full Bar.
- **Distance**: the distance from the town's centre must be as small as possible.

We used the same thresholds' values that were configured for Profile 1. Similarly, all the criteria are given the same importance (i.e., the same weight). The utility values of the domain criteria are shown in Figure 8.9 and the ranking results of the two cases (without and with the social-based criteria) are shown in Figure 8.10. As we can see, like the results of the profile 1, there are clear differences in the two ranking graphs. For example, the alternative 'La Caleta' is changed from position 3 to the first position. On the other hand, the alternative 'Les Coques' is downgraded from position 1 to position 8 and 'AQ' is downgraded from position 2 to position 7. We analysed the domain performance table, Figure 8.9, of this profile and found the price criterion plays a key role in the ranking process followed by the distance and features. The alternative 'Les Coques' shows good performance in terms of the social criteria (the price criterion is an exception). However, all the alternatives that outperformed the alternative 'Les Coques' by adding the social-criteria show better performance than it in most of the social criteria.

Given all that, we can conclude that adding the social-based criteria adds more constraints on the ranking process. However, we believe that the integration of the users' reviews with the decision-maker preference needs more investigation to improve the results.

8.5 Conclusion

Decision making is a very hard task, as it often requires the analysis of hundreds of potential alternatives defined on multiple and conflicting criteria. Methods based on decision rules, utility functions or outranking methods have been proposed in the multi-criteria decision aid field. All of these methods use only the utility values of, what we call in this dissertation,

Seq	Restaurant	Price	Cuisine	Special Diets	Meals	Features	Distance
1	AQ	4.00	0.00	1.00	0.33	0.63	2.60
2	ELIAN Cafe	2.50	0.00	1.00	0.33	0.50	0.50
3	La Caleta	4.00	0.00	1.00	0.33	1.00	3.50
4	Tarakon	1.00	0.00	0.50	0.33	0.25	1.90
5	Sadoll	2.50	0.00	1.00	0.33	0.50	2.10
6	El Taller	2.50	0.00	1.00	0.33	0.63	1.40
7	El Trull	2.50	0.67	1.00	0.33	0.75	7.60
8	Les Coques	4.00	0.00	1.00	0.67	0.50	2.30
9	Mas Rosello	4.00	0.00	1.00	0.67	0.63	6.10
10	Les Fonts de Can Sala	4.00	0.00	1.00	0.00	1.00	3.00
11	La Xarxa	2.50	0.00	0.50	0.33	0.63	1.60
12	El Encuentro	1.00	0.00	0.00	0.33	0.63	0.90
13	La Capital	1.00	0.33	0.50	0.00	0.75	1.00
14	Barquet	2.50	0.00	1.00	0.67	0.38	1.10
15	Arcs	4.00	0.00	1.00	0.33	0.50	2.50
16	Ca L Eulalia	4.00	0.00	0.50	0.33	0.50	2.30
17	Club Nautico Salou	2.50	0.00	0.00	0.33	0.88	14.30
18	Octopussy	2.50	0.00	0.50	0.67	0.88	2.00
19	Palermo 1962 S.c.p.	2.50	0.00	0.50	0.33	0.38	11.20
20	The Cotton Club & Cocktails	2.50	0.00	1.00	1.00	0.75	11.30
21	Lizarran Parc Central	2.50	0.00	0.00	0.33	0.38	1.30
22	Buffalo Grill	2.50	0.33	0.00	0.33	0.13	5.70
23	Indian Mirchi	2.50	0.00	0.00	0.00	0.13	0.75

Figure 8.9: Profile 2: the domain performance-table. The colors indicate the degree of satisfaction and range from "green", the worst, to the "red", the best.

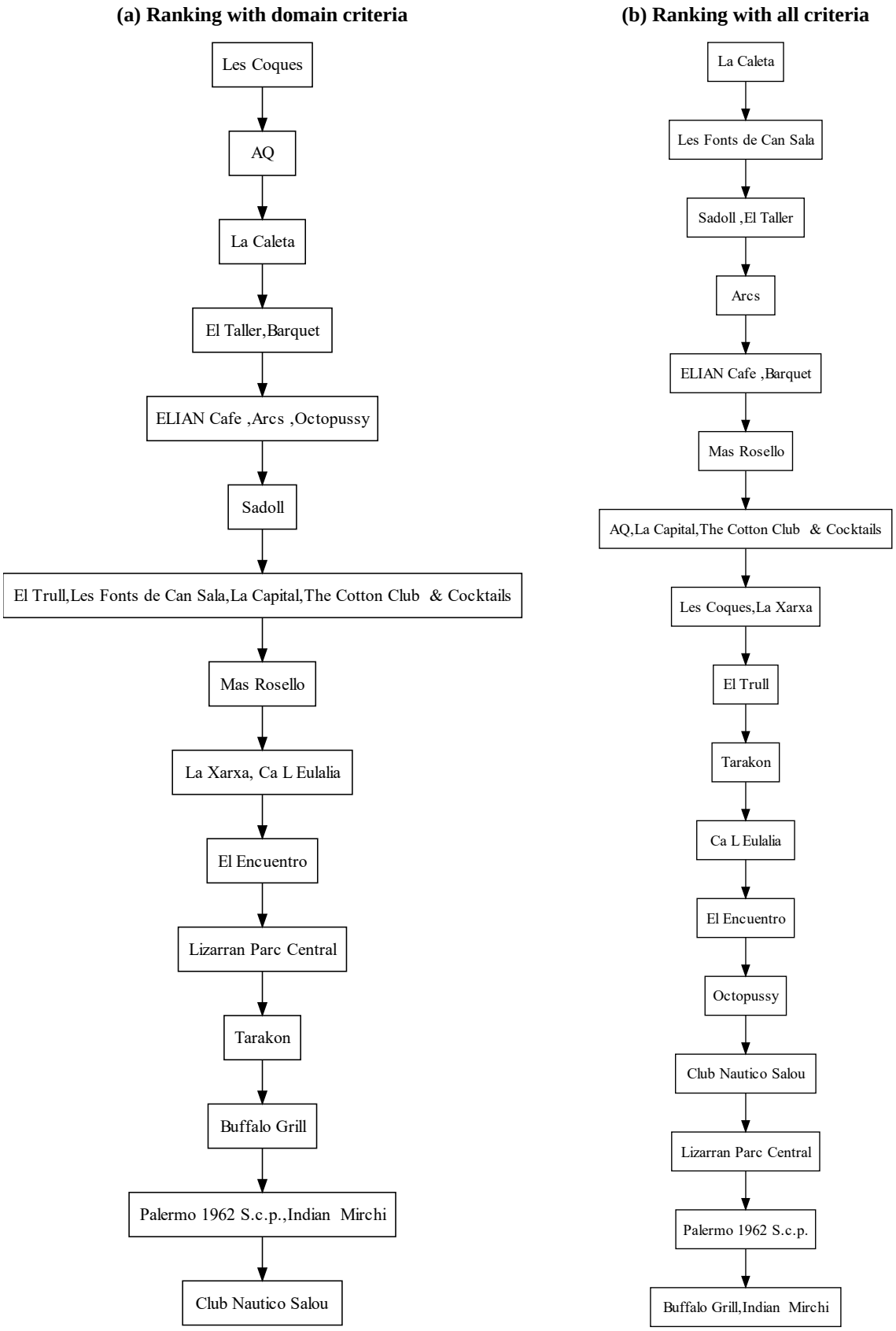


Figure 8.10: Profile 2: ranking result.

the domain criteria based on the decision maker's preferences. With the rapid growth of the web, online platforms such as online shopping websites and social networks give people a great space to share their feedback and experiences about products, services, etc. Thus, in this chapter, we have proposed an outranking system, called SentiRank, that combines the decision-maker preferences with the users' reviews about the alternatives that the decision-maker needs to rank. The contribution of this work has two sides: theoretical and practical. From the theoretical part, we have given answers to the main research questions about integrating the users' feedback with the decision-maker preferences in the ranking process. We have defined a methodology to use the aspect categories, which can be extracted from the users' reviews, as additional attributes in the decision process. From the practical side, we have implemented the system using Java and applied it it to a dataset of restaurants in Tarragona.

Chapter 9

Case Study: Analysis of the Sentiments on Destinations Expressed by Locals and Visitors through Social Media

9.1 Introduction

The main transformation that social media have caused in the communication of tourists' destinations has been their ability to allow the diffusion of stakeholders' knowledge, information and opinions [23]. Moreover, through their comments they also share their experiences and emotions [109].

Several studies [51, 172] corroborate that contents generated by users (UGC) through social media tend to be perceived as more reliable information that the one offered by more institutional sources such as official websites of tourist destinations or travel agencies. This is why all Destination Marketing Organizations (DMOs) have built and manage their social media.

DMOs are aware of the importance of UGC and they want to know what users say about their destination. In the academic field there have been studies about the impact of the content generated by tourists on the tourist market and on the demand for destinations [172]. Knowing what information users comment on or want to know about a destination can help tourist destinations to create a brand and identity. It also provides knowledge about the image of the destination, as this is shaped by the comments of all stakeholders.

The public co-creates the brand image of the destinations [151]. Their comments and experiences influence the image formed by other users. Therefore, knowing the information that stakeholders consider relevant and publish is most useful for DMOs' brand communication

strategies. However, content analysis does not suffice. It must also be analysed whether these stakeholders speak well or badly of destinations through sentiment analysis. It is important to find out whether their comments are positive or negative, because that will have a positive or negative influence on the image that other users form of the destination. Negative comments from other users' experiences spread very quickly and have a great potential to damage the destinations' image and reputation [149].

Images of destinations are created in the minds of users through the confluence of many sources of information [65]. In fact, they are created on the basis of the actions and communications of a whole set of stakeholders. Therefore, DMOs must analyse the comments and evaluations of all of them. The external publics for analysis par excellence are tourists, since their experiences and assessments are fundamental for other users who have never visited the place. Moreover, the most influential internal public are the local residents. Their comments are also highly influential because they know the region better than anyone and their views generate great credibility.

Hence, the study presented in this chapter aims to analyse whether the tweets posted by local residents and tourists of some European destinations are positive or negative and whether different results are observed according to the type of stakeholder. To do so, a new automated sentiment analysis methodology (based on the system presented in Chapter 3, i.e., SentiRich) that measures the sentiment of users' tweets has been designed and implemented. Thus, this is the first application of a fully automated state-of-the-art sentiment analysis system to compare the polarity of the comments of local people about their city with those of their visitors.

This research was done in close collaboration with Dr. Assumpció Huertas, expert in destination brand communication from URV.

The rest of the chapter contains a brief review of related works (section 9.2), a description of the new semantic analysis methodology (section 9.3), its application to a case study of 10 major European destinations (section 9.4) and some final conclusions (section **??**).

9.2 Related Works

Influence of UGC on Social Media and Credibility

Previous studies have examined the use of social media and their impact on tourism-related decisions [173]. Others have focused on knowing what users look for and are interested in [10] in these new media, or on which information they share through them [79].

The credibility users give to UGC determines the influence it will have on their tourism-related decisions as well as the use they will make of social media [180]. Credibility is given by the knowledge of the person who publishes on social media. Stakeholders' opinions have greater credibility than the information supplied by tourist destinations because they are the opinions of third parties who explain their experiences and have no commercial interest. It has been shown that the credibility of social media is basically generated by information sources and social relationships [108].

According to Neidhardt et al., 2016, users are influenced by the feelings of other users [111]. The social influence on tourism-related decisions is based on the fact that individuals adopt their thinking and behaviour according to other individuals of the social network of friends and acquaintances. Given that the social networks involve actors and connections, users' feelings, opinions and behaviour are influenced by others.

As has been seen in previous studies, the credibility that users give to social media varies, but if users' comments are negative, the impact is far higher. It is difficult to measure precisely the impact of negative comments, but it is certain that the problems, disappointments or bad experiences that a stakeholder can have at a destination or when receiving a tourist service will easily lead users to choose another destination or service. In view of this reality, it is essential to analyse whether users' comments are positive or negative.

Stakeholders in Brand Co-creation and Sentiment Analysis

Images about tourist destinations and their brands depend on the information of different sources and contents generated by diverse stakeholders: friends and acquaintances, tourist companies, suppliers, but also tourists and residents of the destination. With the emergence of social media and user-generated content, stakeholders, especially tourists and residents, have become important sources of information. They are also co-creators of brand image, because they can publish information about the destination that will influence other users' perceptions. Several authors have increasingly adopted the paradigm of co-creation [2]. They understand the brand image as a continuous, collective process of interaction between suppliers, stakeholders and consumers.

The main external stakeholders are the tourists. They share their experiences through social media [73], which has a far bigger impact on users' emotions and achieves greater credibility.

On the other hand, local residents are also, for several reasons, very important internal stakeholders in the co-creation of destination brands: first, because they know the destinations better than anyone as they live there; secondly, because their behaviour builds the identity

of the destination; and finally, because with the content they generate in social media, they co-create the destination brand with more credibility than any other stakeholder.

Given that both tourists and local residents are important brand co-creators, for a destination brand to be consistent, images of these two major stakeholders should not be contradictory, but must be complementary and consistent. A study presented in [16] revealed significant discrepancies between the tourists' and the residents' mental representation of a concrete place brand.

Considering these elements, DMOs have to integrate a more participatory vision of destination branding from the outset [2, 77] and involve these stakeholders in the process of creating and communicating target brands. In the reality of professional practice, however, very few destinations involve these stakeholders in the branding process [77].

DMOs must analyse what content the stakeholders post in social media about the destination to find out if the image they convey is the image of the destination they wish to convey. Moreover, they have to find out if what is said is positive or negative. The main problem for the analysis of users' comments and reviews in social media is their large number, which hinders manual analysis. Hence, computational and mathematical methods are being developed that allow the automated analysis of large amounts of data [111]. Studies based on sentiment analysis extract the subjectivity of a text to see if what it said is objective or subjective, and also to see if it is positive or negative [111].

The studies presented in [175] analysed the different existing approaches of sentiment analysis, its accuracy in the measurement of online reviews about tourist destinations, and its effect. The authors in [89] used text mining techniques to extract the keywords from descriptions that users made of hotels to find out what they spoke about and whether they were positive or negative.

Gräbner et al. used a lexicon-based approach to analyse tourism reviews and classify them as either positive or negative [58]. In the same vein, García et al. introduced an approach that used lexical data to calculate the sentiment of tourism reviews [52]. Currently, research in this area is evolving towards sentiment analysis focused on the texts of micro-blogs [53] and on using text mining techniques to measure users' free comments or those of travel forums [111], among others.

In this chapter we use SentiRich, explained in Cahpter 3, to analyse the polarity of the opinions of locals and visitors for 10 top european destinations.

9.3 Methodology

The methodological framework focuses on the retrieval and sentiment analysis of English tweets sent by the visitors and the local citizens of a set of destinations. The basic steps of the analysis are the following:

1. Selection of the destinations to be analysed.
2. Retrieval and pre-processing of two sets of English tweets from each destination. The first one contains the messages sent by local citizens, whereas the second one is composed of tweets sent by non-local people, which are assumed to be tourists or visitors.
3. Extraction of a set of features from each tweet and classification of the polarity of each tweet (positive, negative or neutral).
4. Comparative analysis of the sentiment of the opinions of locals and visitors in the chosen destinations (case study in section 9.4).

9.3.1 Retrieval and Pre-processing of the Tweets

A crawler tool developed by myself, twiQuery, allows retrieving tweets using different filters, like language, time and geo-location. It was used to collect the English tweets sent within a radius of 15 kilometers of the city centre of each of the selected destinations during a specified period of time. From the collected tweets, two sets of tweets were defined. The set L ("the tweets from the local citizens") is the set of tweets sent by users that declare explicitly the destination as their home location in their Twitter profile. The set T ("the tweets from the visitors") is composed of those tweets that were sent by users which have specified a home location different from the destination (a stronger condition could be used in future works, e.g. the location of the home city in a foreign country). For each destination 6000 tweets were randomly selected, 3000 for each of the sets.

In order to study if the tweets are actually refering to the destination, and not to other places, a manual analysis of 400 random tweets sent by local people from London (one of the destinations of the case study described in the next section) was made. The results were the following: 292/400 (73%) were related to local aspects of the destination (local news, weather, restaurants/food, local transport, tourist attractions and cultural activities, local stores and businesses, local sports, etc.), 11/400 (2.75%) were explicitly referring to other locations or global news, and 97/400 (24.25%) were conversational or personal life comments, etc. Thus, the references to other places seem to be minimal (in future versions of

the system, this kind of tweets could even be removed by detecting Named Entities referring to other places).

In the pre-processing stage URLs and user mentions are removed, and all the text of the tweet is converted to lower case. After that, it is tokenized and POS-tagged using the Ark Tweet NLP tool. The suffix "_NEG" is added to all the words that appear in a negated context, which is a segment of a tweet which starts with a negation (e.g. no, don't) and ends with a punctuation mark.

9.3.2 Feature Extraction and Polarity Classification

The polarity of a tweet (positive, negative or neutral) is determined by SentiRich, our sentiment analysis system described in Chapter 3, which is fed with the following features:

- Basic text features: n-grams (contiguous sequences of n tokens, with n from to 1 to 4) and negated n-grams (the same information, but only with the tokens that appear in negated contexts).

- Syntactic features: number of occurrences of each POS and bi-tagged features (combination of bi-grams with their POS tags).

- Lexicon features: it includes the estimation of the polarity of the tweet according to seven popular opinion lexicons. The information about the positive/negative polarity of each word is combined to obtain a global polarity of the tweet for each lexicon. Other lexicon-dependent features in this category include the average polarity of the positive/negative terms, the score of the last positive/negative term, and the maximum/minimum positive/negative score.

- Semantic features: each word of the tweet is mapped to a predefined cluster that groups together words that have similar meanings. Two sets of semantic clusters were used: the 1000 ones defined in the Ark Tweet NLP tool and the 4960 n-gram clusters obtained with the Word2vec tool [41].

As explained in Chapter 3, SentiRich determines the polarity at the tweet level. It is a classifier based on a Support Vector Machine (SVM). This classifier was trained using the Twitter2013 training and development sets from SemEval2013, a well-known worldwide competition of natural language processing systems based on semantic analysis.

9.4 Case Study: Top Destinations in Europe

This section describes which were the destinations selected for the case study. After that, it provides the results of the sentiment analysis system and discusses them.

9.4.1 Selection of Destinations

In our work [72] we studied how the tweets from a destination communicated the emotional values associated to the destination brand. In that study the Twitter accounts of the top 25 European destinations in 2014, according to Tripadvisor, were analysed. It was finally decided to select the 10 destinations that had sent more than 3,000 English tweets: Amsterdam, Athens, Barcelona, Berlin, Budapest, Dublin, Edinburgh, London, Rimini and Vienna. The study was further extended to the communication of emotional values by locals and visitors; thus, two sets of 3000 tweets for each of these 10 destinations were already available, and they are the ones that have been used in this case study (except Rimini, with only 1000 local tweets). These tweets were sent in the high tourism seasons of 2014 and 2015 (June 15th-Sep. 1st) and the low tourism season of 2014 (Oct. 15th- Dec. 30th).

9.4.2 Results and Discussion

The 6000 tweets of each destination (3000 local + 3000 visitors) were classified as positive, negative or neutral by SentiRich.

The results are shown in the form of radial plots in Figure 9.1 (tweets from local people) and Figure 9.2 (tweets from visitors). Each of the concentric lines represents 10% of the tweets. For example, in the case of the 3000 local tweets from Amsterdam, 1687 (56.2%) were qualified as positive, 801 (26.7%) were seen as neutral and the rest (512, 17.1%) were classified as negative.

Local Residents Tweets

The analysis can be started by looking at the sentiments expressed in the tweets of local residents shown in Figure 9.1. It can be seen that the percentage of positive views is quite homogeneous in all the destinations analysed, with values between 51 and 57%, with the single exception of Rimini, which has a much lower value (38%). This means that, in general, local residents or public mostly make positive comments about their town/city in social media. These results have very positive effects for the co-creation of the image of destinations and their brands, since the influence of the comments of the local public is very high among external users or potential tourists.

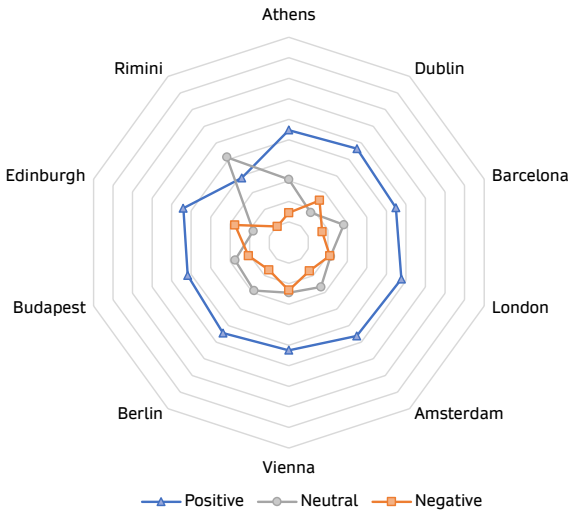


Figure 9.1: Analysis of the tweets from local citizens.

The range of neutral opinions is slightly wider, from 18% (Edinburgh, Dublin) to 30% (Athens); again, Rimini is the exception with a very high number of neutral opinions, 51%. Thus, in general, neutral comments achieve medium percentages.

Finally, the range of negative views is even wider, from only 9% in Rimini to more than 27% in Edinburgh. These big differences are interesting, since their percentages of positive comments are very uniform. It should be noted that the negative comments about tourist destinations or services greatly influence tourism-related decisions by other users and potential tourists, especially if they are made by local residents.

It is surprising that the local residents of destinations such as Dublin or Edinburgh publish more tweets with a negative tone than a neutral one. The residents of these two destinations make mostly positive tweets but, at the same time, they also express negativity with higher percentages than neutral ones. This shows the critical spirit of these local residents. However, the lower are the percentages of negative tweets, the better for the image of the destination.

On the other hand, Rimini shows quite the opposite of the two previous destinations, since it has the lowest percentage of positive and negative tweets and the highest number of neutral ones with a big difference compared to the rest of destinations.

The cities with a larger difference between positive and negative opinions are Athens (54% vs 14%, a 40% difference) and Amsterdam (39%), whereas the smaller differences are Edinburgh (26%) and Rimini and Vienna (29%). It should be noted that, the greater the difference between positive and negative tweets and the lower the percentages of the latter, the better it will be for the global image and reputation of the destinations.

The analysis now will be concentrated on the analysis of different destinations, noting the ones that convey a more positive, a more neutral and a more negative global vision through the tweets of their local residents. Athens and Amsterdam are the destinations with the most positive global tweets, because they have the highest percentages of positive and the lowest percentages of negative tweets. This can be seen graphically in Figure 9.1. The tweets by the local residents at these two destinations co-create a more positive image for these places if they are compared with other destinations.

Let us analyse the results represented graphically in Figure 9.1. It would be most desirable for the triangle to be in the outer ring, which would indicate very high positive percentages, the neutral dot between the triangle and the square, towards the centre of the circle, and the square practically in the centre of the circle, which would indicate a minimum percentage of negativity. This is what is displayed graphically for the tweets for Athens and Amsterdam.

However, the destination showing the most neutral tweets is Rimini, with a big difference from other destinations. This is clearly seen in Figure 9.1, where the dot of the neutral comments is superior to both the negative comments (square) and even the positive ones (triangle). The DMO of this destination should analyse carefully the content of the tweets of its residents and try to find out why their contents are less positive and more neutral.

Finally, the most negative tweets by their local residents are for Edinburgh and Dublin. This is clearly seen in Figure 9.1, where the squares of the negative comments exceed even the dots of the neutral ones. The DMOs at these destinations should also analyse what aspects their residents tweet about in a negative tone, improve the situation of the region and integrate local residents into the branding of the region addressing the communication of the destination brand also towards them and getting them involved in the entire process.

Tourist Tweets

The analysis of the tweets sent by visitors (Figure 9.2) shows a much more homogeneous result for all the cities in the three categories. All the destinations show between 50 and 56% of positive opinions. This means that for all destinations the majority of tweets by tourists published in the place are positive, which has a truly positive influence on the creation of the image and the reputation of the tourist destinations through social media. As mentioned previously, tourists are important co-creators of the brand image of destinations and their comments are highly influential for the tourism-related decisions of other users because they hold great credibility and reduce the risk in their tourism-related decision-making.

On the other hand, 23-29% of the comments of the tweets by tourists are neutral, showing highly homogeneous percentages. The only exceptions are Edinburgh - 21% with the lowest percentage- and Rimini - 32% with the highest one.

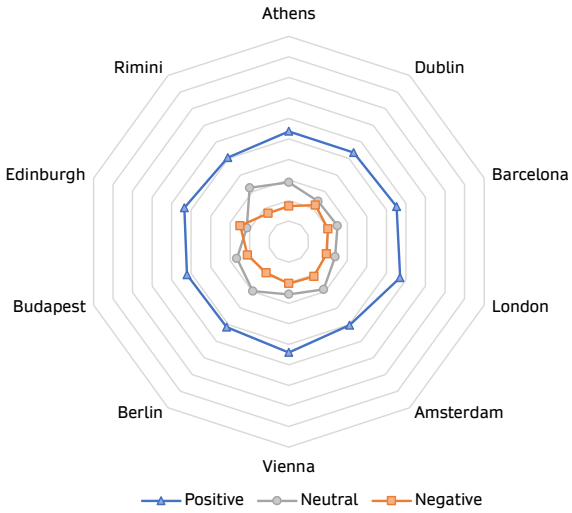


Figure 9.2: Analysis of the tweets from visitors.

Finally, Figure 9.2 shows percentages between 17 and 24% of negative tweets. Rimini and Athens are the destinations with the least negative tweets (17%) and Edinburgh scores the highest percentage (24%), followed by Dublin (with 22%). The difference between positive and negative opinions ranges from 29% (Edinburgh, Amsterdam) to 37% (London). As already mentioned, the greater the distance between the two, the better, since it will mean that the positive tweets are far more numerous than negative ones, which is what is desired.

The destinations that have the most positive comments by tourists will now be analysed. It may be seen that London and Athens show a more positive view overall, since they have the highest positive comments and the lowest negative ones. Therefore, they are the destinations for which tourists reported a more positive image with their tweets. It can be said that the tourists of these two destinations very positively co-create the brand image and reputation of these places through their comments in the tweets published in the place while they are visiting it, and this is very influential for the tourism-related decisions of other users.

Rimini is the destination that also displays the most neutral comments by its tourists, with a great difference compared to other destinations. Finally, the most negative tweets by visitors are for Edinburgh and Dublin. This is clearly seen in Figure 9.2, which shows that the negative square of Dublin almost coincides with the neutral dot, and the squares of the negative comments on Edinburgh exceed the dot of the neutral comments. The DMO of this last destination should also ascertain what aspects its tourists' negative tweets deal with and improve the situation of the region or of the tourist services which these negative comments talk about.

Comparative analysis among the tweets of local residents and tourists

It can be appreciated that, in general, the percentages of positivity, negativity and neutrality of tourists' tweets are more homogeneous among destinations than those of the local residents. In other words, the tweets of the local residents, especially neutral and negative ones, show a higher variability. This shows more diverse opinions among the local residents, which might be caused by the insatisfaction of local residents with their towns/cities or by the different involvement of local residents in the branding of destinations with the DMO.

It seems natural that residents, who know the destination much better than tourists, are more neutral or negative in their tweets than tourists, who spend only a few days in the place and usually only visit the most attractive spaces and tourist facilities. Interestingly, the percentages of tourists' neutral and negative tweets are more homogeneous than those of the local residents, but they have similar percentages.

In general, all percentages of tweets are very similar between residents and tourists. If the results are compared by destination, they coincide to a certain extent. The only difference is seen for positive tweets. The destinations with the most positive tweets by local residents are Athens and Amsterdam and for tourists Athens and London. Thus, it can be stated that Athens co-creates a very positive image with the tweets of its local residents and the tourists who visit it. This coincidence shows the satisfaction of various publics with the destination and at the same time conveys a positive view of the destination that is consistent among these different publics.

On the other hand, it is interesting that Rimini has such a high percentage of neutral tweets and that it coincides both in local residents' and tourists' tweets. However, the percentage of neutral tweets is higher among local residents. Finally, it should be pointed out that Edinburgh and Dublin coincide in having the most negative tweets both by local residents and tourists, although the residents' tweets are more negative.

9.5 Conclusions

Taking into account the enormous credibility that users give to comments and opinions in social media of local residents and tourists about a destination, whose influence is greater than the contents generated by the actual destinations themselves, it is important for these comments, made both by local residents and by tourists, to be positive. The results of the study have shown that tweets are mostly positive for both types of audiences. This means that they co-create a positive image of the destination, but with some differences and exceptions.

In addition, the fact that the percentages of local residents' tweets and those by tourists coincide to some extent is also positive, as it shows that there is no inconsistency between

them regarding the image of the place insofar as positivity or negativity are concerned. As mentioned previously, consistency in a destination image communicated by the different publics is important for the co-creation of a successful and enduring image that is accepted, shared and becomes established over time.

Social media generate relationships between users and the comments of others have a word-of-mouth influence, which is superior to the official information, especially if it comes from friends or acquaintances [73] who visit the place or live there. Therefore, negative visions and comments, in turn, have a very negative influence. For this reason, exceptions like Rimini (which is the destination with most neutral tweets) and destinations that are reflected in a more negative light, both among local residents and tourists, should analyse why this happens and seek to fix it by managing more inclusive branding and taking into account the views of all of these stakeholders.

In addition, since users' feelings and opinions are greatly influenced by those of others [172], sentiment analysis should be key for the management of DMOs and the branding of the destinations. The DMOs need to be aware of the opinions, positive or negative, of their publics in social media and in their tweets, because they will considerably influence, either positively or negatively, the formation of the image of the destinations among rest of users around the world. For this reason, one of the main contributions of this study is the development of the methodology of sentiment analysis, which can be used by DMOs to gauge the feeling and the indices of positivity-negativity of the tweets of their stakeholders.

DMOs must not only analyse the tweets of tourists and local residents, but they must involve them in the branding of the destination, especially local stakeholders, who are the main ambassadors of the place [77]. In this way a more positive sentiment of local residents will be achieved and conveyed through their social media communications which, in turn, will influence the co-creation of the destination brand image.

As a final conclusion, the study shows the importance of sentiment analysis for DMOs and their management of branding. The results show that there is consistency in communication in the positivity and negativity of the tweets published by tourists and local residents, which influences the co-creation of the generally positive destination brand image. However, DMOs must continue to analyse the sentiment of their tweets and manage their territory and its branding to achieve higher percentages of positivity and less negativity in tweets about their destination and in the comments in all social media in general.

Chapter 10

Conclusion and Future Work

10.1 Summary of Contributions

Sentiment Analysis has a wide range of applications in different fields such as commerce, public health, and social welfare. Hence, the goal of this thesis is to develop methods to automatically analyze textual content shared on social networks and identify people' opinions, emotions and feelings at different levels of analysis and in different languages. For this purpose, we used supervised hand-crafted, i.e., traditional, Machine Learning methods and also Deep Learning techniques. We tackled a wide array of problems related to sentiment analysis including the typical sentiment analysis tasks, automatically determining the intensity of emotions and the intensity of sentiment (aka valence) of the users from their tweets, target-dependent sentiment analysis and cross-lingual sentiment analysis. Below, we summarize the main contributions of this thesis.

In chapter 3, we proposed a new set of rich sentimental features for the sentiment analysis of the messages posted on Twitter. A Support Vector Machine classifier was trained using a set of basic features, information extracted from seven useful and publicly available opinion lexicons, syntactic features and clusters. The proposed system, i.e., SentiRich, was evaluated on five sets of tweets used in SemEval 2015. SentiRich outperformed the state-of-the-art systems in four out of those five sets. Extensive analysis was conducted to evaluate the effectiveness of each kind of feature set used and also to evaluate the new proposed features (i.e. bag of negated words "BonW", the polarity measure and the Bi-tagged). The obtained results confirmed the effectiveness of the sentiment lexicons as they played an important role in the improvement of the performance of the classifier.

We also developed an extended version of SentiRich, called SiTAKA. The extended version included the representation of the tweets using features extracted from pre-trained embedding models. Two versions of SiTAKA were developed, one for the Arabic language

tweets and another one for the English language tweets. We used SiTAKA to participate in the SemEval2017 international competition obtaining the 2nd position in the Arabic version and the 8th position in the English version.

Chapter 4 showed a more detailed analysis. We proposed a system that works on an array of sentiment and emotion analysis tasks in Twitter including the typical sentiment analysis tasks, automatically determining the intensity of emotions and the intensity of sentiment (aka valence) of the users from their tweets. The proposed system is an ensemble of two different approaches. The first one, called N-Channels ConvNet, is a Deep Learning approach, whereas the second one is an XGBoost regressor based on a set of embedding and lexicons-based features. The ensemble technique helped to improve the performance of the final model in all subtasks. We realized that the N-Channels ConvNet had a performance very close to the ensemble model. This observation confirms the fact that Deep Learning models, and especially ConvNets, have achieved remarkable results in many fields such as computer vision, speech recognition and natural language processing.

Chapter 4 focused on developing systems to analyze the emotions in tweets independently (i.e., one separate model for each emotion). However, in many real cases, a tweet might be assigned to multiple emotions, especially when there is a strong correlation between them.

In Chapter 5, we proposed a novel approach to solve the problem of multi-label emotion classification. First, we proposed a transformation method to change the problem into a single binary classification problem. Afterwards, we developed a Deep Learning-based system to solve the transformed problem. The critical component of our system was the embedding module, which used three embedding models and an attention function. Our system outperformed the state-of-the-art systems, achieving a Jaccard (i.e., multi-label accuracy) score of 0.59 on the challenging SemEval2018 Task 1:E-c multi-label emotion classification problem. We found that the attention function can model the relationships between the input words and the labels, which helps to improve the system's performance. Moreover, we showed that our system is interpretable by visualizing the attention weights and analyzing them.

The systems described in chapters 3-5 work at the document/tweet level. An overall result is assigned to a given tweet. However, in real-world cases, it is necessary to identify in a single document diverse targets and the attitude towards them. Chapter 6 proposed a novel methodology to perform fine-grained sentiment analysis. Specifically, we developed an end-to-end target-dependent Twitter sentiment analysis system. The proposed system is able to identify and extract the target of the tweets, representing the relatedness between the targets and its contexts and identifying the polarities of the tweets towards the targets. The

effectiveness of the proposed system was evaluated on a benchmark of tweets, obtaining results that outperformed the state-of-the-art models.

The works presented in chapters 3-6 are mainly directed to the English language (although in some cases we developed independent versions for the Arabic language). In Chapter 7, we described a multi-lingual sentiment analysis system that can work with low (or even zero) resource (e.g., training data and sentiment lexicons) languages. The proposed system, i.e., UniSent, can transfer sentiment knowledge from a rich-resource language to perform sentiment analysis on low-resource languages. The experimental results on the Spanish dataset of OpeNER and the MultiBooked Catalan showed the effectiveness of our system. It outperformed the state-of-the-art systems in three out of four experimental setups. We demonstrated, empirically, that our system is highly flexible, and adding a new language does not require to re-train it.

Chapter 8 presented a novel contribution to the fields of the multi-criteria decision aid and sentiment analysis. We proposed an outranking system, called SentiRank, that combines the decision-maker preferences with the users' reviews about the alternatives that the decision-maker needs to rank. The contribution of the work in this chapter is twofold: theoretical and practical. From the theoretical side, we gave answers to the main research questions about integrating the users' feedback with the decision-maker preferences in the ranking process. We defined a methodology to use the aspect categories, which can be extracted from the users' reviews, as additional attributes in the decision process. From the practical side, we implemented the system using Java and applied it to a dataset of restaurants in Tarragona.

Finally, in Chapter 9, we presented a case study by applying the developed systems to analyze the communication of destination brands through social networks. To this end, we collected tweets of local people, visitors, and official brand destination offices from different tourist destinations and analyzed the opinions and the emotions shared in these tweets. This work was carried out in collaboration with national and international researchers specialized in destination brand communication.

10.2 Future Work

Although we have proposed ih this dissertation efficient and accurate systems for solving various problems of sentiment analysis, there are still some challenges and shortcomings to overcome. For example, in chapter 4, we developed one separate system for each emotion to identify the intensity of that emotion expressed in the text. Hence, in our future work, we plan to overcome this issue. Recently, attention-based models, called transformers [165, 171], have been developed to solve many natural language processing tasks. Transforms are

showing impressive results in many of these tasks. We aim to adopt these models to develop a unified system for all emotions.

We identified some limitations in our proposed system of multi-label emotion classification. Our system does not model the relationships between the phrases and the labels. Phrases play a key role in determining the most appropriate set of emotions that must be assigned to a tweet. For instance, an emotion word that reflects "sadness" can be flipped in a negated phrase or context. We plan to work on solving this drawback. One possible solution is to adapt the attention function to model the relationships between different n-gram tokens and labels. Structured attention networks [83] can also be adapted and used to address this issue. Moreover, we plan to work on developing systems that perform robustly and equally on all the emotion labels by experimenting with different ideas like using data augmentation to enrich the training data or using transfer learning.

The confusion matrix of the results obtained our system for target-dependent sentiment analysis showed that most of the misclassified examples are related to the neutral category. In future work, we plan to extend our system to handle this weakness by integrating more information such as lexicon information and/or the dependency tree. It should also be noted that the proposed system extracts only the targets that are mentioned explicitly in the tweets. However, sometimes targets are mentioned implicitly in tweets, and they are detected from the context. Thus, we will consider this point in our future work by designing a system that can detect both the explicit targets and the implicit targets that are not mentioned in the tweets. Although joint learning of all subsystems has proved to be useful in natural language processing and text analysis tasks, in the work presented in Chapter 6, we trained each subsystem (i.e. the target identification and the targeted sentiment analysis) independently, and after that we combined them in the inference step. Thus, we plan to extend our system and apply this learning technique.

Our work in chapter 7 focused on utilizing the technique of cross-lingual sentiment analysis to develop a multi-lingual system. Our future work towards this line will focus on moving cross-lingual sentiment models beyond sentence-level and develop multi-lingual target-dependent sentiment analysis systems. We also aim to develop multi-lingual systems for emotion analysis.

The work done in Chapter 8 opens an interesting and promising research line that enables the use of people' feedback and reviews for decision support. For the future work, and from a practical perspective, we plan to extend the implemented system by adding more domains such as hotels, laptops, and cameras. We also plan to make the system a web-based application and implement RESTfull APIs to allow developers and users to use the system remotely. On the theoretical side, we would like to work on different levels of sentiment

analysis. For example, in this work we performed the analysis on the aspect category level, hence, in the future, we plan to perform the analysis on the aspect term level [69]. Moreover, we only considered the reviews written in English language. Thus, we would like to extend the proposed system by integrating the work presented in Chapter 7 to develop a multi-lingual version of SentiRank. Recently, Deep Learning was used in sentiment analysis and recommender systems [30], so we plan to investigate how we can use Deep Learning in future work.

We also plan to extend the work on the case studies by performing the sentiment analysis at different levels and multiple languages.

Bibliography

[1] Agarwal, B., Mittal, N., and Cambria, E. (2013). Enhancing sentiment classification performance using bi-tagged phrases. In *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on*, pages 892–895. IEEE.

[2] Aitken, R. and Campelo, A. (2011). The four rs of place branding. *Journal of Marketing Management*, 27(9-10):913–933.

[3] Al-Molegi, A., Jabreel, M., and Martinez-Balleste, A. (2018). Move, attend and predict: An attention-based neural model for peopleâ€™s movement prediction. *Pattern Recognition Letters*, 112:34–40.

[4] Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 579–586. Association for Computational Linguistics.

[5] Almeida-Dias, J., Figueira, J. R., and Roy, B. (2010). Electre tri-c: A multiple criteria sorting method based on characteristic reference actions. *European Journal of Operational Research*, 204(3):565–580.

[6] Almeida-Dias, J., Figueira, J. R., and Roy, B. (2012). A multiple criteria sorting method where each category is characterized by several reference actions: The electre tri-nc method. *European Journal of Operational Research*, 217(3):567–579.

[7] Aman, S. and Szpakowicz, S. (2007). Identifying expressions of emotion in text. In *International Conference on Text, Speech and Dialogue*, pages 196–205. Springer.

[8] Araujo, M., Reis, J., Pereira, A., and Benevenuto, F. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pages 1140–1145. ACM.

[9] Artetxe, M., Labaka, G., and Agirre, E. (2016). Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2289–2294.

[10] Ayeh, J. K., Au, N., and Law, R. (2013). Predicting the intention to use consumer-generated media for travel planning. *Tourism management*, 35:132–143.

[11] Bai, S., Kolter, J. Z., and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

[12] Banea, C., Mihalcea, R., Wiebe, J., and Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on EMNLP*, pages 127–135. Association for Computational Linguistics.

[13] Barnes, J., Klinger, R., and im Walde, S. S. (2018). Bilingual sentiment embeddings: Joint projection of sentiment across languages. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2483–2493.

[14] Baziotis, C., Nikolaos, A., Chronopoulou, A., Kolovou, A., Paraskevopoulos, G., Ellinas, N., Narayanan, S., and Potamianos, A. (2018). Ntua-slp at semeval-2018 task 1: Predicting affective content in tweets with deep attentive rnns and transfer learning. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 245–255.

[15] Baziotis, C., Pelekis, N., and Doulkeridis, C. (2017). Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 747–754, Vancouver, Canada. Association for Computational Linguistics.

[16] Beckmann, S. C. and Zenker, S. (2012). Place branding: A multiple stakeholder perspective. In *41st European Marketing Academy Conference, Lisbon, Portugal*, pages 22–25.

[17] Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166.

[18] Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., and Cox, D. D. (2015). Hyperopt: a python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, 8(1):014008.

[19] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

[20] Bollen, J., Pepe, A., and Mao, H. (2009). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Computing Research Repository - CORR*.

[21] Bouyssou, D., Marchant, T., Pirlot, M., Tsoukias, A., and Vincke, P. (2006). *Evaluation and decision models with multiple criteria: Stepping stones for the analyst*, volume 86. Springer Science & Business Media.

[22] Brun, C., Popa, D. N., and Roux, C. (2014). Xrce: Hybrid classification for aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 838–842.

[23] Buhalis, D. and Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the internetâ€"the state of etourism research. *Tourism management*, 29(4):609–623.

[24] Cambria, E., Olsher, D., and Rajagopal, D. (2014). SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In *Twenty-eighth AAAI conference on artificial intelligence*.

[25] Castellucci, G., Filice, S., Croce, D., and Basili, R. (2014). Unitor: Aspect based sentiment analysis with structured learning. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pages 761–767.

[26] Chen, L., Liu, R., Liu, Z.-P., Li, M., and Aihara, K. (2012). Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Scientific reports*, 2(1):1–8.

[27] Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *CoRR*, abs/1603.02754.

[28] Chen, X., Sun, Y., Athiwaratkun, B., Cardie, C., and Weinberger, K. (2018a). Adversarial deep averaging networks for cross-lingual sentiment classification. *Transactions of the Association for Computational Linguistics*, 6:557–570.

[29] Chen, Z., Shen, S., Hu, Z., Lu, X., Mei, Q., and Liu, X. (2018b). Ermes: Emoji-powered representation learning for cross-lingual sentiment classification. *arXiv preprint arXiv:1806.02557*.

[30] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., et al. (2016). Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*, pages 7–10. ACM.

[31] Cheng, W. and Hüllermeier, E. (2009). Combining instance-based learning and logistic regression for multilabel classification. *Machine Learning*, 76(2-3):211–225.

[32] Cherry, C., Mohammad, S. M., and De Bruijn, B. (2012). Binary classifiers and latent sequence models for emotion detection in suicide notes. *Biomedical informatics insights*, 5:BII–S8933.

[33] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

[34] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*.

[35] Clare, A. and King, R. D. (2001). Knowledge discovery in multi-label phenotype data. In *european conference on principles of data mining and knowledge discovery*, pages 42–53. Springer.

[36] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(Aug):2493–2537.

[37] Conneau, A., Lample, G., Ranzato, M., Denoyer, L., and Jégou, H. (2017). Word translation without parallel data. *arXiv preprint arXiv:1710.04087*.

[38] Cortes, C. and Vapnik, V. (1995). Support vector machine. *Machine learning*, 20(3):273–297.

[39] De Comité, F., Gilleron, R., and Tommasi, M. (2003). Learning multi-label alternating decision trees from texts and data. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pages 35–49. Springer.

[40] Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., and Xu, K. (2014). Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification. In *ACL (2)*, pages 49–54.

[41] Dong, L., Wei, F., Yin, Y., Zhou, M., and Xu, K. (2015). Splusplus: a feature-rich two-stage classifier for sentiment analysis of tweets. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 515–519.

[42] Dong, X. and de Melo, G. (2018). Cross-lingual propagation for deep sentiment analysis. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[43] Duh, K., Fujino, A., and Nagata, M. (2011). Is machine translation ripe for cross-lingual sentiment classification? In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 429–433. Association for Computational Linguistics.

[44] Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

[45] Ekman, P. (2005). *Basic Emotions*, chapter 3, pages 45–60. John Wiley Sons, Ltd.

[46] Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712.

[47] Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89.

[48] Figueira, J., Greco, S., and Ehrgott, M. (2005). *Multiple criteria decision analysis: state of the art surveys*, volume 78. Springer Science & Business Media.

[49] Figueira, J. R., Greco, S., Roy, B., and Słowiński, R. (2010). Electre methods: main features and recent developments. In *Handbook of multicriteria analysis*, pages 51–89. Springer.

[50] Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., and Freling, T. (2014). How online product reviews affect retail sales: A meta-analysis. *Journal of Retailing*, 90(2):217–232.

[51] Fotis, J. N., Buhalis, D., and Rossides, N. (2012). *Social media use and impact during the holiday travel planning process*. Springer-Verlag.

[52] García, A., Gaines, S., Linaza, M. T., et al. (2012). A lexicon based sentiment analysis retrieval system for tourism domain. *Expert Syst Appl Int J*, 39(10):9166–9180.

[53] García-Pablos, A., Duca, A. L., Cuadros, M., Linaza, M. T., and Marchetti, A. (2016). Correlating languages and sentiment analysis on the basis of text-based reviews. In *Information and Communication Technologies in Tourism 2016*, pages 565–577. Springer.

[54] Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., and Smith, N. A. (2011). Part-of-speech Tagging for Twitter: Annotation, Features, and Experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, pages 42–47, Stroudsburg, PA, USA. Association for Computational Linguistics.

[55] Godbole, S. and Sarawagi, S. (2004). Discriminative methods for multi-labeled classification. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 22–30. Springer.

[56] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

[57] Gouws, S. and Søgaard, A. (2015). Simple task-specific bilingual word embeddings. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1386–1390.

[58] Gräbner, D., Zanker, M., Fliedl, G., Fuchs, M., et al. (2012). Classification of customer reviews based on sentiment analysis. In *ENTER*, pages 460–470. Citeseer.

[59] Graves, A., r. Mohamed, A., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649.

[60] Gu, J., Hassan, H., Devlin, J., and Li, V. O. (2018). Universal neural machine translation for extremely low resource languages. *arXiv preprint arXiv:1802.05368*.

[61] Gupta, N. K., Gilbert, M., and Fabbrizio, G. D. (2010). Emotion detection in email customer care. *Computational Intelligence*, 29:489–505.

[62] Hagen, M., Potthast, M., Büchner, M., and Stein, B. (2015). Webis: An Ensemble for Twitter Sentiment Detection. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 15)*, pages 582–589. Association for Computational Linguistics.

[63] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[64] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

[65] Ho, C.-I., Lin, M.-H., and Chen, H.-M. (2012). Web usersâ€™ behavioural patterns of tourism information search: From online to offline. *Tourism Management*, 33(6):1468–1482.

[66] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

[67] Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

[68] Ishizaka, A. and Nemery, P. (2013). *Multi-criteria decision analysis: methods and software*. John Wiley & Sons.

[69] Jabreel, M., Hassan, F., and Moreno, A. (2018). Target-dependent sentiment analysis of tweets using bidirectional gated recurrent neural networks. In *Advances in Hybridization of Intelligent Methods*, pages 39–55. Springer.

[70] Jabreel, M. and Moreno, A. (2016). Sentirich: Sentiment analysis of tweets based on a rich set of features. In *Artificial Intelligence Research and Development - Proceedings of the 19th International Conference of the Catalan Association for Artificial Intelligence, Barcelona, Catalonia, Spain, October 19-21, 2016*, pages 137–146.

[71] Jabreel, M. and Moreno, A. (2017). Sitaka at semeval-2017 task 4: Sentiment analysis in twitter based on a rich set of features. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 694–699.

[72] Jabreel, M., Moreno, A., and Huertas, A. (2017). *Do Local Residents and Visitors Express the Same Sentiments on Destinations Through Social Media?*, pages 655–668. Springer International Publishing, Cham.

[73] Jacobsen, J. K. S. and Munar, A. M. (2012). Tourist information search and destination choice in a digital age. *Tourism management perspectives*, 1:39–47.

[74] Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188.

[75] Jebbara, S. and Cimiano, P. (2016). Aspect-based sentiment analysis using a two-step neural network architecture. In *Semantic Web Evaluation Challenge*, pages 153–167. Springer.

[76] Jiang, L., Yu, M., Zhou, M., Liu, X., and Zhao, T. (2011). Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 151–160. Association for Computational Linguistics.

[77] Kalandides, A., Kavaratzis, M., and Boisen, M. (2012). From â€œnecessary evilâ€ to necessity: stakeholders' involvement in place branding. *Journal of Place Management and development*.

[78] Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.

[79] Kang, M. and Schuett, M. A. (2013). Determinants of sharing travel experiences in social media. *Journal of Travel & Tourism Marketing*, 30(1-2):93–107.

[80] Kant, N., Puri, R., Yakovenko, N., and Catanzaro, B. (2018). Practical text classification with large pre-trained language models. *arXiv preprint arXiv:1812.01207*.

[81] Kessler, J. S. and Nicolov, N. (2009). Targeting sentiment expressions through supervised ranking of linguistic configurations. In *ICWSM*.

[82] Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

[83] Kim, Y., Denton, C., Hoang, L., and Rush, A. M. (2017). Structured attention networks. *arXiv preprint arXiv:1702.00887*.

[84] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[85] Kiritchenko, S., Zhu, X., Cherry, C., and Mohammad, S. (2014a). Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442.

[86] Kiritchenko, S., Zhu, X., and Mohammad, S. M. (2014b). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50:723–762.

[87] Lafferty, J., McCallum, A., Pereira, F., et al. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth international conference on machine learning, ICML*, volume 1, pages 282–289.

[88] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*.

[89] Lee, M. J., Singh, N., and Chan, E. S. (2011). Service failures and recovery actions in the hotel industry: A text-mining approach. *Journal of Vacation Marketing*, 17(3):197–207.

[90] Ling, W., Dyer, C., Black, A. W., Trancoso, I., Fermandez, R., Amir, S., Marujo, L., and Luis, T. (2015). Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1520–1530, Lisbon, Portugal. Association for Computational Linguistics.

[91] Liu, B. (2011). Opinion mining and sentiment analysis. In *Web Data Mining*, pages 459–526. Springer.

[92] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

[93] Liu, Q., Gao, Z., Liu, B., and Zhang, Y. (2015). Automated rule selection for aspect extraction in opinion mining. In *IJCAI*, pages 1291–1297.

[94] Liu, Q., Gao, Z., Liu, B., and Zhang, Y. (2016). Automated rule selection for opinion target extraction. *Knowledge-Based Systems*, 104:74–88.

[95] Meisheri, H. and Dey, L. (2018). Tcs research at semeval-2018 task 1: Learning robust representations using multi-attention architecture. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 291–299.

[96] Mencia, E. L. and Fürnkranz, J. (2008). Efficient pairwise multilabel classification for large-scale problems in the legal domain. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 50–65. Springer.

[97] Mihalcea, R., Banea, C., and Wiebe, J. (2007). Learning multilingual subjective language via cross-lingual projections. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pages 976–983.

[98] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[99] Mitchell, J., Odio, S., and Garcia, D. H. (2017). Computer-vision content detection for connecting objects in media to users. US Patent 9,672,496.

[100] Miura, Y., Sakaki, S., Hattori, K., and Ohkuma, T. (2014). TeamX: A sentiment analyzer with enhanced lexicon mapping and weighting scheme for unbalanced data. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 628–632.

[101] Mohammad, S. and Bravo-Marquez, F. (2017). Wassa-2017 shared task on emotion intensity. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 34–49.

[102] Mohammad, S., Bravo-Marquez, F., Salameh, M., and Kiritchenko, S. (2018). Semeval-2018 task 1: Affect in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 1–17.

[103] Mohammad, S. and Kiritchenko, S. (2018). Understanding emotions: A dataset of tweets to study interactions between affect categories. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

[104] Mohammad, S., Kiritchenko, S., and Zhu, X. (2013). NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, Georgia, USA.

[105] Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier.

[106] Mohammad, S. M. (2018). Word affect intensities. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan.

[107] Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. 29(3):436–465.

[108] Munar, A. M. and Jacobsen, J. K. S. (2013). Trust and involvement in tourism social media and web-based travel information sources. *Scandinavian Journal of Hospitality and Tourism*, 13(1):1–19.

[109] Munar, A. M. and Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism management*, 43:46–54.

[110] Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., and Stoyanov, V. (2016). Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval 2016), San Diego, US (forthcoming)*, pages 1–18.

[111] Neidhardt, J., Rümmele, N., and Werthner, H. (2016). Can we predict your sentiments by listening to your peers? In *Information and Communication Technologies in Tourism 2016*, pages 593–603. Springer.

[112] Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In *Workshop on'Making Sense of Microposts: Big things come in small packages*, pages 93–98.

[113] Pak, A. and Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

[114] Pang, B. and Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.

[115] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

[116] Park, J. H., Xu, P., and Fung, P. (2018). Plusemo2vec at semeval-2018 task 1: Exploiting emotion knowledge from emoji and# hashtags. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 264–272.

[117] Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

[118] Perikos, I. and Hatzilygeroudis, I. (2016). Recognizing emotions in text using ensemble of classifiers. *Engineering Applications of Artificial Intelligence*, 51:191–201.

[119] Pestian, J., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., Cohen, K., Hurdle, J., and Brew, C. (2012). Sentiment analysis of suicide notes: A shared task. *Biomedical informatics insights*, 5:3–16.

[120] Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. *Approaches to emotion*, 1984:197–219.

[121] Plutchik, R. (1988). *The Nature of Emotions: Clinical Implications*, pages 1–20. Springer US, Boston, MA.

[122] Pontiki, M., Papageorgiou, H., Galanis, D., Androutsopoulos, I., Pavlopoulos, J., and Manandhar, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. *SemEval 2014*, page 27.

[123] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[124] Rasooli, M. S., Farra, N., Radeva, A., Yu, T., and McKeown, K. (2018). Cross-lingual sentiment transfer with limited resources. *Machine Translation*, 32(1-2):143–165.

[125] Ratinov, L. and Roth, D. (2009). Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 147–155. Association for Computational Linguistics.

[126] Read, J. (2010). *Scalable multi-label classification*. PhD thesis, University of Waikato.

[127] Read, J., Pfahringer, B., Holmes, G., and Frank, E. (2011). Classifier chains for multi-label classification. *Machine learning*, 85(3):333.

[128] Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S. M., Ritter, A., and Stoyanov, V. (2015). Semeval-2015 task 10: Sentiment analysis in twitter. *Proceedings of SemEval-2015*.

[129] Rosenthal, S., Ritter, A., Nakov, P., and Stoyanov, V. (2014). Semeval-2014 task 9: Sentiment analysis in twitter. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 73–80.

[130] Rouvier, M. and Favre, B. (2016). Sensei-lif at semeval-2016 task 4: Polarity embedding fusion for robust sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), San Diego, US*.

[131] Roy, B. (1974). Critères multiples et modélisation des préférences (l'apport des relations de surclassement). *Revue d'Economie Politique*, 84(1):1–44.

[132] Roy, B. (1978). Electre iii: Un algorithme de classements fondé sur une représentation floue des préférences en présence de critères multiples. *Cahiers du CERO*, 20(1):3–24.

[133] Roy, B. and Bertier, P. (1971). *La méthode ELECTRE II: une méthode de classement en prédence de critères multiples*.

[134] Roy, B. and Hugonnard, J. (1982). Classement des prolongements de lignes de métro en banlieue parisienne (présentation dâ€™une méthode multicritère originale). *Cahiers du CERO*, 24(2-4):153–171.

[135] Roy, B. and Skalka, J. M. (1987). *ELECTRE IS: Aspects méthodologiques et guide d'utilisation*. LAMSADE, Unité associée au CNRS no 825, Université de Paris Dauphine.

[136] Saif, H., Fernández, M., He, Y., and Alani, H. (2014). On stopwords, filtering and data sparsity for sentiment analysis of Twitter. In *Proceedings of the 9th language resources and evaluation conference (LREC)*, Reykjavik, Iceland.

[137] Saif, M., Mohammad, M., and Kiritchenko, S. (2016). Sentiment lexicons for arabic social media. In *Proceedings of 10th edition of the the language resources and evaluation conference (LREC), Portorož*.

[138] Salameh, M., Mohammad, S., and Kiritchenko, S. (2015). Sentiment after translation: A case-study on arabic social media posts. In *Proceedings of the 2015 NAACL HLT*, pages 767–777.

[139] Saleh, K. (2017). The importance of online customer reviews [infographic](2017).

[140] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229.

[141] Sang, E. F. and Veenstra, J. (1999). Representing text chunks. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, pages 173–179. Association for Computational Linguistics.

[142] Schapire, R. E. and Singer, Y. (2000). Boostexter: A boosting-based system for text categorization. *Machine Learning*, 2(39):135–168.

[143] Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729.

[144] Schuster, M. and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681.

[145] Severyn, A. and Moschitti, A. (2015). UNITN: Training deep convolutional neural network for Twitter sentiment classification. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado*, pages 464–469.

[146] Shen, Y., He, X., Gao, J., Deng, L., and Mesnil, G. (2014a). A latent semantic model with convolutional-pooling structure for information retrieval. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 101–110.

[147] Shen, Y., He, X., Gao, J., Deng, L., and Mesnil, G. (2014b). Learning semantic representations using convolutional neural networks for web search. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 373–374.

[148] Shengli, L. and Fan, L. (2019). The interaction effects of online reviews and free samples on consumersâ€™ downloads: An empirical analysis. *Information Processing & Management*, 56(6):102071.

[149] Siano, A., Vollero, A., and Palazzo, M. (2011). Exploring the role of online consumer empowerment in reputation building: research questions and hypotheses. *Journal of brand management*, 19(1):57–71.

[150] Simonson, I. (2016). Imperfect progress: An objective quality assessment of the role of user reviews in consumer decision making, a commentary on de langhe, fernbach, and lichtenstein. *Journal of Consumer Research*, 42(6):840–845.

[151] Singh, S. and Sonnenburg, S. (2012). Brand performances in social media. *Journal of interactive marketing*, 26(4):189–197.

[152] Stone, P., Dunphy, D. C., Smith, M. S., and Ogilvie, D. (1968). The general inquirer: A computer approach to content analysis. *Journal of Regional Science*, 8(1):113–116.

[153] Suero Montero, C. and Suhonen, J. (2014). Emotion analysis meets learning analytics: online learner profiling beyond numerical data. In *Proceedings of the 14th Koli calling international conference on computing education research*, pages 165–169.

[154] Sun, Y., Lin, L., Tang, D., Yang, N., Ji, Z., and Wang, X. (2015). Modeling mention, context and entity with neural networks for entity disambiguation. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 1333–1339. AAAI Press.

[155] Szegedy, C., Ioffe, S., and Vanhoucke, V. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *CoRR*, abs/1602.07261.

[156] Tang, D., Qin, B., Feng, X., and Liu, T. (2015a). Target-dependent sentiment classification with long short term memory. *arXiv preprint arXiv:1512.01100*.

[157] Tang, D., Qin, B., and Liu, T. (2015b). Deep learning for sentiment analysis: successful approaches and future challenges. *WIREs Data Mining and Knowledge Discovery*, 5(6):292–303.

[158] Tang, D., Wei, F., Qin, B., Liu, T., and Zhou, M. (2014a). Coooolll: A Deep Learning System for Twitter Sentiment Classification. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 208–212, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.

[159] Tang, D., Wei, F., Qin, B., Zhou, M., and Liu, T. (2014b). Building Large-Scale Twitter-Specific Sentiment Lexicon : A Representation Learning Approach. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 172–182, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

[160] Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., and Qin, B. (2014c). Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1555–1565, Baltimore, Maryland. Association for Computational Linguistics.

[161] Toh, Z. and Su, J. (2015). Nlangp: Supervised machine learning system for aspect category classification and opinion target extraction.

[162] Tokhisa, R., Inui, K., and Matsumoto, Y. (2008). Emotion classification using massive examples extracted from the web. volume 1, pages 881–888.

[163] Tsoumakas, G. and Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3):1–13.

[164] Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Fourth international AAAI conference on weblogs and social media*.

[165] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

[166] Vo, D.-T. and Zhang, Y. (2015). Target-dependent twitter sentiment classification with rich automatic features. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, pages 1347–1353.

[167] Wang, J., Korayem, M., Blanco, S., and Crandall, D. J. (2016). Tracking natural events through social media and computer vision. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 1097–1101. ACM.

[168] Wang, P., Xu, J., Xu, B., Liu, C., Zhang, H., Wang, F., and Hao, H. (2015). Semantic clustering and convolutional neural network for short text categorization. In *Proceedings of the 53rd Annual Meeting of the ACL and the 7th IJCNLP (Volume 2: Short Papers)*, pages 352–357.

[169] Webb, E. C. and Simonson, I. (2017). Using reviews to determine preferences: How variance in customer-generated reviews affects choice. *ACR North American Advances*.

[170] Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics.

[171] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., and Brew, J. (2019). Huggingface's transformers: State-of-the-art natural language processing. *ArXiv*, abs/1910.03771.

[172] Xiang, Z. and Gretzel, U. (2010). Role of social media in online travel information search. *Tourism management*, 31(2):179–188.

[173] Xiang, Z., Wöber, K., and Fesenmaier, D. R. (2008). Representation of the online tourism domain in search engines. *Journal of Travel Research*, 47(2):137–150.

[174] Yan, R., Tesic, J., and Smith, J. R. (2007). Model-shared subspace boosting for multi-label classification. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 834–843.

[175] Ye, Q., Zhang, Z., and Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert systems with applications*, 36(3):6527–6535.

[176] Yih, W.-t., He, X., and Meek, C. (2014). Semantic parsing for single-relation question answering. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 643–648.

[177] Younes, Z., Abdallah, F., Denoeux, T., and Snoussi, H. (2011). A dependent multilabel classification method derived from the k-nearest neighbor rule. *EURASIP Journal on Advances in Signal Processing*, 2011:1–14.

[178] Yu, W. (1992). Electre tri(aspects méthodologiques et manuel d'utilisation). *Document-Université de Paris-Dauphine, LAMSADE*.

[179] Zahran, M. A., Magooda, A., Mahgoub, A. Y., Raafat, H., Rashwan, M., and Atyia, A. (2015). Word representations in vector space and their applications for arabic. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 430–443. Springer.

[180] Zeng, B. and Gerritsen, R. (2014). What do we know about social media in tourism? a review. *Tourism management perspectives*, 10:27–36.

[181] Zhang, F., Zhang, Z., and Lan, M. (2014). Ecnu: A combination method and multiple features for aspect extraction and sentiment polarity classification. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 252–258.

[182] Zhang, M., Zhang, Y., and Vo, D.-T. (2016). Gated neural networks for targeted sentiment analysis. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA. Association for the Advancement of Artificial Intelligence*, pages 3087–3093.

[183] Zhang, M.-L., Li, Y.-K., Liu, X.-Y., and Geng, X. (2018). Binary relevance for multi-label learning: an overview. *Frontiers of Computer Science*, 12(2):191–202.

[184] Zhang, M.-L. and Zhang, K. (2010). Multi-label learning by exploiting label dependency. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 999–1008.

[185] Zhang, M.-L. and Zhou, Z.-H. (2007). Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048.

[186] Zhang, M.-L. and Zhou, Z.-H. (2013). A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8):1819–1837.

[187] Zhang, X., Zhao, J., and LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Advances in neural information processing systems*, pages 649–657.

[188] Zhao, L. and Li, C. (2009). Ontology based opinion mining for movie reviews. In Karagiannis, D. and Jin, Z., editors, *Knowledge Science, Engineering and Management*, pages 204–214, Berlin, Heidelberg. Springer Berlin Heidelberg.

[189] Zhou, G., Zhu, Z., He, T., and Hu, X. T. (2016a). Cross-lingual sentiment classification with stacked autoencoders. *Knowledge and Information Systems*, 47(1):27–44.

[190] Zhou, X., Wan, X., and Xiao, J. (2016b). Attention-based LSTM network for cross-lingual sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 247–256.

[191] Zhou, Z., Zhang, X., and Sanderson, M. (2014). Sentiment analysis on twitter through topic-based lexicon expansion. In *Databases Theory and Applications*, pages 98–109. Springer.

Appendix A

Review of Multi-Criteria Decision Aid

1.1 Introduction

Decision problems [68], e.g., ranking, choice, and sorting problems, are usually complicated as they involve several criteria. People no longer consider only one criterion to make a decision. Moreover, in most of the cases, there is not any perfect alternative that matches all the criteria. An ideal option does not usually exist. Therefore, we must find a compromise. To this end, the decision-maker can make use of naive approaches such as a simple weighted sum. However, such a simple method can only be applied with the right constraints (e.g., correct normalization phase, independent criteria) to enable reasonable outputs. Moreover, it assumes the linearity of preferences which may not reflect the decision maker's preferences in many cases.

MCDA methods have been developed to support the decision-makers in their unique and personal decision process [48]. They are widely used in decision problems to find the "best possible" alternative solution. MCDA makes the process more explicit, rational, and efficient. It is a multidisciplinary field, deriving from Operations Research, and it uses mathematical approaches to deal with complex problems encountered in human activities. Nowadays, it also integrates Artificial Intelligence and economic welfare techniques.

A large number of methods have been developed to solve multi-criteria problems. In this appendix, we introduce the basic concepts required for modeling decision problems in MCDA and focus on outranking methods. We also present the ELECTRE-III method, one of the most well-known outranking methods.

1.2 Multi-Criteria Decision Problem

The formalization of a multiple criteria decision aiding procedure is made from the point of view of the so-called decision-maker, who is the person that must make a decision. The data required in a decision problem is represented using two key elements:

1. Alternatives. They are the potential actions for the decision problem from which the decision-maker has to decide. Alternatives are represented as follows: $A = \{a_1, a_2, ..., a_n\}$ is the finite set of alternatives and n is the number of alternatives in A.

2. Criteria. A criterion is a tool constructed for evaluating and comparing potential actions according to the decision-maker subjective point of view about some reference indicators. The set of criteria is denoted $G = \{g_1, g_2, ..., g_m\}$.

The type of MCDA problem depends on the goal of the decision-maker when approaching a certain situation [48]. Three main types are distinguished:

- Choice problem: it is a selection problem where the decision-maker has to choose the best option. MCDA helps the decision-maker to find among all the possibilities the best alternative(s).

- Ranking problem: in this type of problem, the decision-maker wants to get a ranking of the alternatives from the best to the worst. Two (or more) alternatives may be equivalent. The outcome of an MCDA method can be either a partial or a complete ranking of the set of alternatives.

- Sorting problem: the problem of sorting consists of assigning the alternatives to predefined categories. These categories are ordered from the worst to the best. The outcome of an MCDA sorting method is an assignment of the alternatives to the different classes.

1.3 Outranking

The outranking methods are based on pairwise comparisons of the options. This means that every option is compared to all the other options. Based on these pairwise comparisons, final recommendations can be drawn. The next subsection explains one of the most common outranking methods in MCDA and especially in ranking problems.

1.3.1 ELECTRE Methods

ELECTRE is a family of multi-criteria decision ais methods. The first ELECTRE method was presented by Benayoun, Roy, and Sussman (1966) who reported on the works of the European consultancy company SEMA for a specific real-world problem. However, the first journal article did not appear until 1968, when Roy described the method in detail [131]. Later, it was renamed to ELECTRE I. The name ELECTRE Iv (v for veto) is sometimes used when veto thresholds are considered, but it did not become an official name [48]. Several other ELECTRE methods were developed during the following two decades: ELECTRE II [133], ELECTRE III [132], ELECTRE IV [134], ELECTRE TRI [178] and ELECTRE IS [135]. ELECTRE TRI was later renamed to ELECTRE TRI-B [49], when a new version, ELECTRE TRI-C [5] was developed, but most of the literature still uses the name ELECTRE TRI for the original version. Recently, ELECTRE TRI-nC, was presented by Almeida-Dias, Figueira, and Roy (2012) [6] as an extension of ELECTRE TRI-C. All methods, except ELECTRE I, Iv and II, consider the concept of pseudo-criteria. Thanks to indifference and preference thresholds, this concept allows to model imperfect knowledge, which may be a result of uncertainty, imprecision, and ill-determination of specific data. The main advantage of the ELECTRE methods is that they avoid compensation between criteria and any normalization process, which distorts the original data. On the other hand, their main drawback is that they require various technical parameters, which means that it is not always easy to fully understand them. A more detailed history and overview of the ELECTRE methods can be found in [48]. In the following, we discuss in detail ELECTRE III as we employ it in our system.

1.3.2 ELECTRE-III

ELECTRE-III is composed by two phases. First, we construct the outranking relationships among the alternatives, and then we exploit them. The decision-maker is required to provide most of the information, including the weight of the criteria, the indifference, the preference and the veto thresholds in the first phase.

In the construction phase, ELECTRE-III makes use of outranking relations. An outranking relation, where an alternative a outranks another alternative b, denoted by aSb, expresses the fact that there are sufficient arguments to decide whether a is at least as good as b. Also, there are no essential reasons to refute this fact (Roy 1974). The exploitation phase uses one of two methods: Net Flow Score (NFS) or Distillation. In this work, we use NFS as explained below.

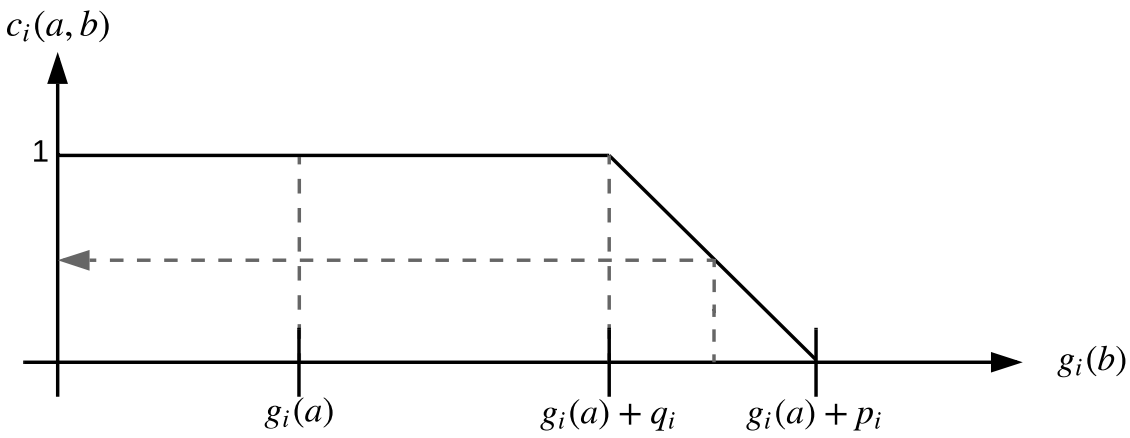


Figure 1.1: The partial concordance index $c_i(a,b)$.

We first explain the construction phase steps. We compute the outranking degree between each pair a and b of alternatives to measure or evaluate the outranking relation assertion as the following.

The credibility or outranking degree, $S(a,b)$, is a score between 0 and 1, where the closer $S(a,b)$ is to 1, the stronger the assertion. It considers two aspects: the *concordance* and the *discordance* of the statement that a outranks b. The concordance and discordance are measured respectively while incorporating the decision maker's preference on multiple, usually conflicting, criteria. The user is required to provide the indifference and preference thresholds to compute the concordance degree. The veto threshold is needed to calculate the discordance degree.

Concordance A partial concordance degree $c_i(a,b)$ measures the assertion "a outranks b" or "a is at least as good as b" on the specific criterion g_i. The uncertainty of the decision-maker preference can be represented, as shown in Figure 1.1, by the indifference, q_i, and preference, p_i, thresholds.

The *indifference* threshold indicates the largest difference between the performances of the alternatives on the considered criterion such that they remain indifferent for the decision maker. The *preference* threshold indicates the largest difference between the performances of the alternatives such that one is preferred over the other on the considered criterion. Between these two thresholds, the partial concordance degree is computed based on a linear

interpolation as shown in Figure 1.1. Given that, we can define the partial concordance index as follows:

$$c_i(a,b) = \begin{cases} 1 & \text{if, } g_i(a) \geqslant g_i(b) - q_i \\ 0 & \text{if, } g_i(a) \leqslant g_i(b) - p_i \\ \frac{p_i - g_i(a) - g_i(b)}{p_i - q_i} & otherwise \end{cases} \tag{A.1}$$

Once we get the values of all the partial concordances, the global concordance index can be computed for each pair of alternatives (a,b) as follows:

$$C(a,b) = \frac{1}{W}\sum_{i=1}^{m} w_i c_i(a,b) \tag{A.2}$$

In this expression w_i is the weight of the criterion g_i, W is the sum of the weights and m is the number of criteria.

Discordance On the other hand, the partial discordance degree $d_i(a,b)$ measures the decision-maker discordance with the assertion "a is at least as good as b" on criterion g_i. If the decision-maker, when considering criterion g_i, strongly disagrees with the assertion, the discordance degree reaches its maximum value, i.e., 1, and reflects the fact that g_i sets its veto. This is the case if the difference in performances (i.e., $g_i(b) - g_i(a)$) is higher than a so-called veto threshold, denoted by v_i. The discordance degree has minimum value 0 when there is no reason to refute the assertion. As in the case of the partial concordance degree, between these two extremes, $d_i(a,b)$ changes linearly between the preference and veto thresholds as a function of the difference $g_i(b) - g_i(a)$, as shown in Figure 1.2. We define the partial discordance index as follows:

$$d_i(a,b) = \begin{cases} 1 & \text{if, } g_i(a) \leqslant g_i(b) - v_i \\ 0 & \text{if, } g_i(a) \geqslant g_i(b) - p_i \\ \frac{g_i(b) - g_i(a) - p_i}{v_i - p_i} & otherwise \end{cases} \tag{A.3}$$

Outranking Degree Finally, a global outranking degree $S(a,b)$ summarizes the concordance and discordance degrees into one measure of the assertion "a outranks b" as shown below.

$$p(a,b) = \begin{cases} C(a,b) & \text{if, } d_i(a,b) \leqslant C(a,b), \forall i \\ \prod_{i \in I(a,b)} \frac{1 - d_i(a,b)}{1 - C(a,b)} & otherwise \end{cases} \tag{A.4}$$

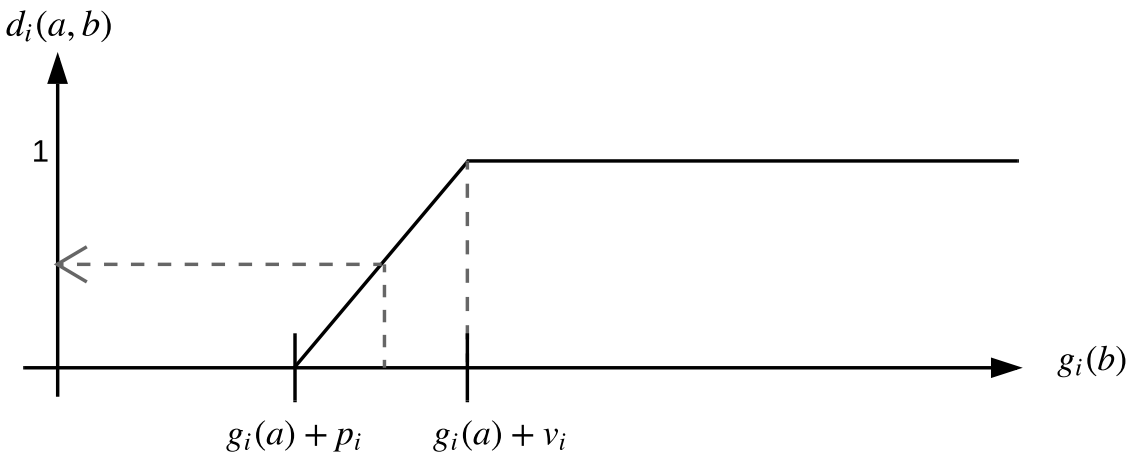


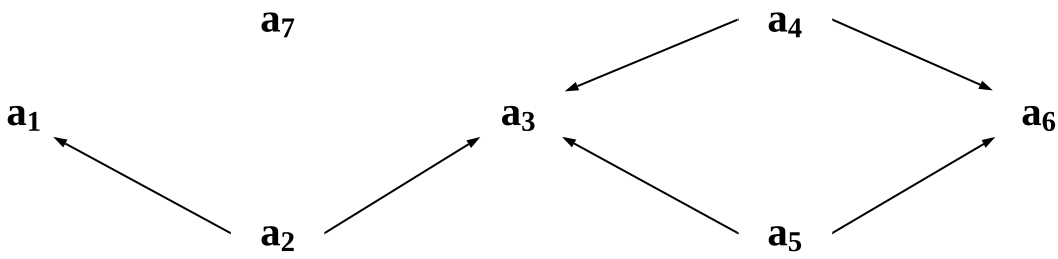
Figure 1.2: The partial discordance index $c_i(a,b)$.

Where $I(a,b)$ is the set of criteria in which $d_i(a,b) > C(a,b)$.

1.3.3 Net Flow Score

The matrix that contains the credibility values for each pair of alternatives may be interpreted as the adjacency matrix of a labelled directed graph, in which the nodes are the alternatives and the label of the edge between two alternatives a and b is the value $p(a,b)$. The NFS procedure analyzes this graph to calculate two evidences: strength and weakness. The strength of an alternative a is defined as the sum of the credibility values of the edges that leave from the node a. The weakness of an alternative a is defined as the sum of the credibility values of the edges that reach the node a. The NFS of an alternative a is the difference between its strength and its weakness. This value allows ranking the alternatives in a descending partial order. Figure 1.3 shows an example of NFS.

$$NFS(a) = \sum_{b \in A} [S(a,b) - S(b,a)] \tag{A.5}$$



S	a_1	a_2	a_3	a_4	a_5	a_6	a_7	Streangth
a_1	0	0	0	0	0	0	0	0
a_2	1	0	1	0	0	0	0	2
a_3	0	0	0	0	0	0	0	0
a_4	0	0	1	0	0	1	0	2
a_5	0	0	1	0	0	1	0	2
a_6	0	0	0	0	0	0	0	0
a_7	0	0	0	0	0	0	0	0
Weakness	1	0	3	0	0	2	0	

$NFS(a_1) = 0 - 1 = -1$ Rank = 3

$NFS(a_2) = 2 - 0 = 2$ Rank = 1

$NFS(a_3) = 0 - 3 = -3$ Rank = 5

$NFS(a_4) = 2 - 0 = 2$ Rank = 1

$NFS(a_5) = 2 - 0 = 2$ Rank = 1

$NFS(a_6) = 0 - 2 = -2$ Rank = 4

$NFS(a_7) = 0 - 0 = 0$ Rank = 2

Figure 1.3: An example of Net Flow Score.

Appendix B

Case Studies: Publications

Journal of Destination Marketing & Management 6 (2017) 170–183

Contents lists available at ScienceDirect

Journal of Destination Marketing & Management

journal homepage: www.elsevier.com/locate/jdmm

Research paper

Semantic comparison of the emotional values communicated by destinations and tourists on social media

Mohammed Jabreel [a,c,*], Antonio Moreno [a], Assumpció Huertas [b]

[a] Intelligent Technologies for Advanced Knowledge Acquisition (ITAKA), Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Av. Països Catalans, 26, 43007 Tarragona, Spain
[b] Departament d'Estudis de Comunicació, Universitat Rovira i Virgili, Av.Catalunya, 35, 43002 Tarragona, Spain
[c] Computer Science Department, Hodiedah University, Alduraihimi, 1820 Hodeidah, Yemen

ARTICLE INFO

Article history:
Received 16 December 2015
Received in revised form
21 March 2016
Accepted 22 March 2016
Available online 14 April 2016

Keywords:
Destination branding
Brand communication
Social media
Emotional values
Semantic similarity

ABSTRACT

Emotional values play a key role in the creation of destination brands. Nowadays destination management organizations (DMOs) want to make sure that they transmit a set of attractive, distinguishing values and that they are correctly perceived by their visitors. This paper presents a new methodology for the automated, unsupervised semantic analysis of large quantities of tweets sent by the DMOs and the visitors of a destination. As a case study, the results of an analysis of 60,000 tweets related to 10 major European destinations, are presented and the emotional values transmitted from the official Twitter accounts of the destinations compared with those communicated by the tourists in their personal messages. The experiment leads to two important results: the cities examined do not communicate a personalized identity and there are strong discrepancies between the emotional values transmitted by DMOs and those reflected by the comments of visitors. The framework presented in this work constitutes the first semantic methodology for a large-scale automatic analysis of the communication of emotional values by destinations through social media.

1. Introduction

In the current, globalized scenario, tourist destinations need to differentiate themselves from their competitors to stand out from the crowd and attract more tourists, investors or residents (Morgan & Pritchard, 2004). To this end, destination marketing organziations (DMOs) manage their identity and brand. Destination brands associate emotional values and tangible attractions to territories, with the intention of identifying and distinguishing them (Blain, Levy, & Ritchie, 2005; Morgan, Pritchard, & Piggott, 2003). Many attractive tourist destinations have similar strengths (e.g. five-star accommodations, cultural assets, or sun-and-beach activities), so it is the emotional side of the brand, its personality or identity, which may help them to capture the attention of potential visitors and beat their competitors (Morgan & Pritchard, 2004).

The great contribution of brands has been the establishment of relationships with tourists and the generation of connections and emotional ties with them (Laroche, Habibi, & Richard, 2013; Zhang, Pan, Smith, & Li, 2009). It has been argued (Morgan & Pritchard, 2004) that tourists mostly base their consumption decisions on these relationships and on the emotional bonds created with the territories, rather than on rational decisions or on the physical attractions featured by destinations. Therefore, the first challenge for DMOs is to communicate these emotional values, alongside its identity and brand, with the aim of generating this emotional differentiation.

In the last ten years, social media have revolutionized the communication of tourist destinations (Xiang & Gretzel, 2010). It is now commonly accepted that the comments and experiences of other users (who are supposed to lack any personal interest on a particular location) have much greater credibility to the eyes of potential tourists than the official information provided by DMOs (Fotis, Buhalis, & Rossides, 2012; Litvin, Goldsmith, & Pan, 2008; Mack, Blose, & Pan, 2008; Xiang & Gretzel, 2010) and they heavily influence their choice of travel destinations (Buhalis & Costa, 2006; Schmallegger & Carson, 2009; Yoo & Gretzel, 2010; Zhang et al., 2009). In addition, studies in the field of communication (Huertas, 2014; Macnamara & Zerfass, 2012; Valentini, 2015; Wigley & Lewis, 2012) have shown that social media allow the creation of a continuing dialogue with users and the building of a

* Corresponding author at: Intelligent Technologies for Advanced Knowledge Acquisition (ITAKA), Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Av. Països Catalans, 26, 43007 Tarragona, Spain.
E-mail addresses: antonio.moreno@urv.cat.(A. Moreno)mohammed.hamoodadullah@estudiants.urv.cat, sunsi.huertas@urv.cat.(A. Huertas)

http://dx.doi.org/10.1016/j.jdmm.2016.03.004
2212-571X/

relationship with them. This process increases the identification of users with the destination and its brand and permits them the creation of a better picture of the emotional identity of the place (Govers, Go, & Kumar, 2007; Mariné-Roig, 2013; Stepchenkova & Zhan, 2013). Social media also allow tourists to share their travel experiences and emotions through reviews, comments, photographs or videos (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004; Senecal & Nantel, 2004). All these feedback directly impact the emotional part of potential tourists and provide a better image of the destinations (Inversini & Buhalis, 2009; Marchiori & Cantoni, 2011; Xiang & Gretzel, 2010). Therefore, sharing experiences by other users via social media generates an emotional attachment to the destination and to its brand (Algesheimer, Dholakia, & Herrmann, 2005).

Therefore, social media are key tools in the emotional communication of destination brands. Despite this several recent studies (Huertas & Mariné-Roig, 2015; Huertas, Setó-Pàmies, & Míguez-González, 2015; Míguez-González & Huertas, 2015) have shown that destinations focus more on the communication of tangible tourist attractions than on the transmission of emotional values. This paper takes a deeper look to the latter: thus, the first objective is to analyze the communication of the emotional values of brands through Twitter by some of Europe's leading tourist destinations. In order to perform a comprehensive examination of a large number of tweets, the paper develops and implements a novel methodological framework in which an automated semantic analysis of the content of the tweets is made. In this analysis the adjectives used in the tweets are linked with the core emotional values of a travel destination, which have been characterized with a revised and adapted version of Aaker's brand personality scale (Aaker, 1997).

Furthermore, destination brands are key in building the image that public and potential tourists create from a destination and these images have a positive influence on the selection of the destinations to be visited (Kim, Kim, & Bolls, 2014). Thus, individuals who have created an image of a destination are more likely to visit it (Lee & Gretzel, 2012) and the chances increase even more if the image is positive. The relationship between the brand and image of a destination is so close that many academics use the two concepts interchangeably. For some authors (Anholt, 2008), the brand is merely the public image of a territory created in the users' minds from their direct perceptions and external inputs of various kinds. Although the brand may be understood as the identity or image, which can be viewed as two sides of the same coin, from the communications point of view it is of paramount importance that the identity that DMOs want to transmit matches with the image perceived by tourists. A positive matching will be a clear sign that the destination brand is well established and communicated, and that the place has achieved the desired differentiated positioning (Huertas, 2014). Therefore, a second grand challenge for DMOs is to make sure that the brand identifies the territory and that it also matches with the image perceived by the public: that is, that there is a correlation between image and identity. Thus, the second objective of this research is to analyze and compare the emotional brand values communicated through Twitter by DMOs with those described by their visitors through their personal tweets.

In summary, the main contributions of this paper are threefold. First, the paper presents a new methodology for the automated, unsupervised semantic analysis of large quantities of tweets sent by the DMOs or the visitors of a city. This framework may be useful for the researchers in the field and for the DMOs themselves, insofar as it may be used to make a self-assessment of the communication of their brand. After that, the paper presents the exhaustive results of the analysis of 30,000 tweets from 10 major European destinations, which studies the number of *emotional adjectives* they use and how many times they employ them. Finally, the same methodology is used to analyze 30,000 tweets from visitors of the same cities and the emotional values transmitted from the official Twitter accounts of the destinations is compared with those communicated by the tourists in their personal messages.

The rest of the paper is organized as follows. Section 2 includes a brief state of the art on the methods employed in destination branding studies to analyze messages in social media. The next section explains the new methodological framework defined in this work, which permits the study to associate adjectives to emotional values in a semantic and fully unsupervised fashion. Sections 4 and 5 present and discuss in detail the results of the automated analysis of 60,000 English-language tweets, which reflect the transmission and the perception of emotional values by 10 major European destinations. The final section presents the conclusions and some lines of future work.

2. State of the art

In recent years, there have been two main kinds of studies related to the evaluation of the communication of the identities and brands of territories. The first is based on numerical analysis whereas the second one focuses on content analysis. These two types of works are commented in the following subsections.

2.1. Numerical methods of analysis

This kind of evaluation methods performs a numerical analysis of the communication between DMOs and tourists, measuring aspects such as the number of followers/friends a destination has in social media, the number of comments they post, and the number of reactions generated among users (e.g. replies, retweets, and number of 'likes'). There are already many automated tools that measure these purely quantitative aspects of communication.

An example of this type of analysis is the work reported in Guerrero-Solé and Fernández-Cavia (2015), which considered various quantitative communication aspects to analyze the brand of Spain at three levels of diffusion of information. At the first level (the diffusion of messages by the state-level main touristic account, @Spain) they measured the number of messages published by this account and the number of times it mentioned other users. At the second level, two analtical measures were used: the number of retweets of the messages posted by @Spain and the number of mentions of @Spain. The number of retweets of the messages in which @Spain was mentioned was analysed at the third level. Another example of this type of works is Huertas and Mariné-Roig (2015), which measured the average reaction to top posts per destination type by calculating the total posts and the average number of likes, comments and shares for 37 Spanish destinations located in five autonomous communities (Andalusia, Canary Islands, Catalonia, Galicia, and Madrid).

This kind of numerical study is quite easy to undertake and they intend to provide a measure of the influence of a destination in the social media. However, there are works such as Guerrero-Solé and Fernández-Cavia (2014) that show that there is not a direct relationship between the dimension of the communication of a destination and its importance, showing clearly the need to make more complex analyse of the communicated content.

2.2. Content analysis methods

Content analysis refers to a family of procedures for the systematic, replicable analysis of text. These methods enable researchers to identify relevant properties of large amounts of

textual information, such as the frequencies of the most-used keywords. Within this category, it is possible to distinguish between qualitative and quantitative techniques.

2.2.1. Qualitative methods

Unstructured qualitative approaches, such as in-depth interviews, open questions and focus groups, provide rich data of a high quality but they are necessarily limited to a small sample, constraining the scope of the analysis and the generality of its results (Lai, Li, & Harrill, 2013).

For example, Zhou (2014) employed a qualitative analysis method to explore the representation of rurality in tourism by analysing 40 articles containing some description of Wuyuan and five online travel guidebooks using the qualitative analysis software, NVIVO. The work introduced in Marchiori and Önder (2015) used a purely qualitative approach to try to understand how the exposure to a set of stimuli changed the most common topics associated to a small collection of American destinations (Detroit, Kansas, Las Vegas, New Orleans, Orlando, Phoenix, San Francisco and Seattle). In that work, the same qualitative research software was used to analyse the data in three stages. In the first stage, the gathered data (before and after the exposure to the stimuli) were arranged into two files for each city. A word-frequency analysis was conducted in the 'before' files to identify the top five words (themes) in them. In the second, concurrence analysis of those words in the 'before' and 'after' files were conducted separately to see the evolution of those terms after the exposure to the stimuli. In the last stage, the conative component in the concurrence analysis was identified. In the work reported in Hays, Page, and Buhalis (2013), semi-structured interviews were used to examine the usage and impact of social media marketing strategies. Just to mention another recent example, in the work presented in Oliveira (2014), data-analysis software (ATLAS.ti 7.0) was employed to perform a qualitative analysis on 20 online articles in which Portugal had been mentioned, with the aim to identify and understand the way tourists and travelers perceive the country as a tourist destination and explore how they could support the creation of a destination brand.

2.2.2. Quantitative methods

Quantitative methods measure numerically a series of predefined attributes or properties of the destinations. There are two main types of quantitative content analysis: thematic and semantic (Roberts, 2000).

Thematic analysis methods: These methods just count the themes/categories/words that appear in the text, without interpreting them. Thus, they provide limited information on the content of the analysed messages. Analyses of word counts permit researchers to identify the predominant themes in texts (Roberts, 2000). For example, the work reported in Dickinger and Lalicic (2015) aimed to evaluate the personality of the destination brand from the comments of its visitors in the social media and to provide useful recommendations to destination managers about how to communicate a given brand personality. In that study, the WordStat program was used to count the frequency of appearance of a list of words selected by the researchers. The work presented in Huertas and Mariné-Roig (2015) used the attraction factors, activities, services and brand values of the destinations as themes of study. The most frequently mentioned ones were identified and then compared with those that generated more visitor reactions per destination type. In another recent example, the authors in Chaykina, Guerreiro, and Mendes (2014) tried to identify the brand personality attributes ascribed to Portugal by Russian-speaking visitors. They divided the most frequent words used by tourists answering a questionnaire into categories, and then computed the frequency with which they were mentioned.

Semantic analysis methods: These techniques try to discover the relationships among themes. In a semantic analysis of text, the researcher begins by constructing a template (known as a semantic grammar). Syntactic components of the text are then mapped onto the themes in the template. Semantic methods use the actual meaning of the words (provided by some external knowledge repository) to make an analysis at the conceptual level, rather than at the syntactic one. Some recent examples include Moya and Jain (2013), which analyzed how two popular tourist destinations (Mexico and Brazil) communicate the personality of their brands through Facebook, and Serna, Marchiori, Gerrikagoitia, Alzua-Sorzabal, and Cantoni (2015), in which the authors developed a system to make a semantic analysis of the image of the Basque Country, taking into account cognitive, affective and conative aspects. Just to mention another example, one of the components of the Define–Measure–Visualize framework proposed in Sevin (2014) is a 'semantic network', which represents relationships between themes. These relationships are found by clustering the most frequently used keywords using the similarity between them and their frequency of co-occurrence. Another example of a 'semantic network' analysis is given in Çakmak and Isaac (2012).

Several authors combine thematic and semantic analyses in an attempt to overcome the limitations of plain syntactic studies (Andéhn, Kazeminia, Lucarelli, & Sevin, 2014; Stepchenkova, Kirilenko, & Morrison, 2009). The first can identify the topics that are more related to the brand (by finding the most frequently used words) and, after that, the second can make a more profound analysis of their degree of association with the brand and their inter-relationships. One of the most recent works that uses this kind of analysis methods is introduced in Mariné-Roig and Clavé (2015). In that work the authors use a website content parser, 'Site Content Analyzer' to extract the keywords and count their frequencies. They then group the most frequently used keywords for each attraction value. This combination of methods has also been used to analyse opinions about destinations posted on travel blogs. For instance, the study reported in Pan, MacLaurin, and Crotts (2007) started by finding the most frequently used words and phrases in 40 travel blogs. These were then used to perform a semantic analysis to uncover the image of Charleston as a tourist destination. In recent years, some authors have gone a step further by trying to understand not only the basic themes associated to a destination but also the positive/negative views of the visitors towards them. For example, in Költringer and Dickinger (2015) the authors proposed a method to analyse and evaluate how DMOs project a destination brand through online web sources. They started by calculating the co-occurrence of the keywords of the data collected about Vienna from online Web sources (social travel guides, Anglo-American news media websites and DMOs websites) and then tried to detect the positive/negative sentiment of the sentences by measuring the number of occurrences of a predefined list of sentiment words. It is important to note that the work reported in this paper does not address this issue. The methodology used here analyzes the references to emotional values in the tweets sent by DMOs or by tourists, but it does not classify them as positive or negative.

3. Methodology

Twitter and Facebook are the social media more commonly used by tourism destination managers for their promotion. This study focuses on the analysis of English-language tweets sent by official tourist destinations and by their visitors. Official tourist destination accounts provide a real-time view of how they try to communicate their brand to potential customers (Andéhn et al., 2014) and how they conduct dialogue with them (Hvass & Munar,

2012), whereas tourists communicate their personal views about the destinations they are visiting. The steps followed in the methods of analysis were as follows:

- *Definition of the emotional values associated to a destination brand*: This step involved the adaptation of a previous brand personality scale to the specific study of destination brands.
- *Selection of the destinations to be analysed*: In this step, a set of well-known European destinations, was chosen taking into consideration some constraints regarding the language they use and a minimum quantity of messages they transmit through social media.
- *Retrieval of the set of tweets*: This phase involved the use of a new tool that permits sets of tweets to be retrieved that satisfy certain user-determined requirements.
- *Pre-processing of the set of tweets*: This step applied some simple treatments to the content of the tweets to make them easier to analyse.
- *Semantic analysis of the content of the tweets*: This step is the core of the methodology. It used a well-known ontology-based semantic similarity measure to compare the adjectives used in the tweets with the emotional values defined in the first step.
- *Interpretation of the results*: In the final step, the study analysed the results thus obtained comparing the performance of the DMOs of the different destinations and also contrasting the values transmitted by the official DMOs with the ones reflected in the opinions of the visitors.

Fig. 1 shows a graphical depiction of the main steps of this methodology.

All these method, along with the computational system that implements it, were fully designed and developed by the authors of this paper. The remainder of this section describes the following technical aspects: the definition of the emotional values, how the tweets are pre-processed and how they are semantically analysed. The selection of the destinations to be analysed in a specific case study and the retrieval of their associated tweets are described in the following section. Following that, the specific results of the semantic analysis of the case study are described and discussed.

3.1. Emotional values

Many academic studies have stressed the importance of the emotional values and the personality of the destinations in their branding (Henderson, 2000; Morgan, Pritchard, & Piggott, 2002). Destinations can be described using human personality values (Xiang & Gretzel, 2010). This study has updated and adapted the well-known brand personality scale (BPS) of Aaker (1997) to the analysis of destination brands. Indeed Ekinci and Hosany were the first authors to apply this scale to the analysis of brand destinations (Ekinci & Hosany, 2006). They showed that personality dimensions have a positive impact on the preferences of potential tourists, as a strong and well-defined personality improves the image of the destination and the intentions to visit it. They thus concluded that a proper management of emotional values is vital for the effective positioning and differentiation of destinations. In this work, Aaker's BPS has been slightly adapted to facilitate the analysis of the emotional values associated to a travel destination. The five main dimensions of analysis of the personality and the emotional values of a destination are sincerity, excitement, competence, sophistication and ruggedness. Each has been divided into a set of categories, which in turn have been refined into several sub-categories, represented by a set of terms. The whole template of analysis is shown in Table 1.

3.2. Pre-processing of the tweets

It is well known that the language used in Twitter is very casual and noisy. Tweets contain numerous punctuation errors, spelling mistakes, abbreviations, slang terms, emoticons, etc. The tweets were therefore pre-processed to mitigate these effects as follows:

- Tweets may contain URLs, usernames, hashtags and emoticons. All URLs, usernames and strange symbols were removed from the tweets.
- To reduce the dimensionality, all tweets were converted to lowercase and stop words were removed. Table 2 shows the list of stop words.
- Words with repeated letters were automatically corrected using an algorithm that performs a 'breadth first search' and analyzes all the possible ways of eliminating repeated letters in a string, checking in WordNet if they are correct.

3.3. Semantic content analysis

In previous studies on content analysis of the communication of destination brands (Stepchenkova & Morrison, 2006; Stepchenkova et al., 2009) it was pointed out that *nouns* usually provide information on the particular tourist attractions, *verbs* describe actions and types of tourism, and *adjectives* communicate the emotional responses. Thus, as the study aims to measure the association between tweets and emotional values, the analysis

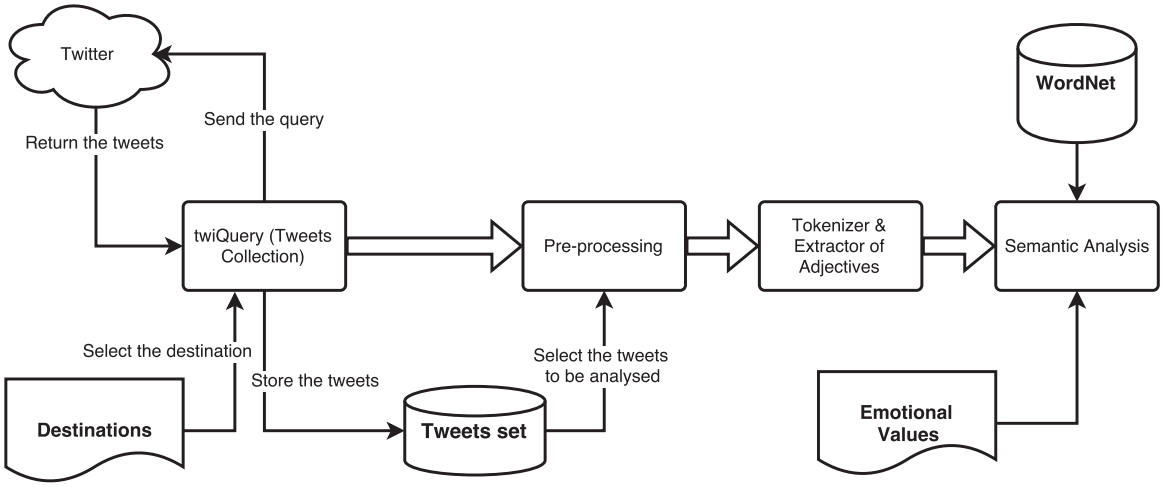


Fig. 1. Architecture of the methodological framework of analysis.

Table 1
Emotional values.

Emotional value	Category	Subcategory
Sincerity	Down-to-earth	Family-oriented
		Down-to-earth
		Sustainable
	Honest	Calm
		Real
		Traditional
		Honest
	Wholesome	Original
		Wholesome
		Quality of life
	Cheerful	Happiness
		Sentimental
		Friendly
Excitement	Daring	Trendy
		Daring
		Exciting
		Exotic
		Fashionable
	Spirited	Cool
		Spirited
		Dynamic
		Vital
		Fresh
		Young
		Sensorial
	Imaginative	Unique
		Imaginative
		Creative
	Up-to-date	Up-to-date
		Independent
		Contemporary
	Cosmopolitan	Cosmopolitan
		Tolerant
		Hospitable
Competence	Reliable	Reliable
		Hard-working
		Safe
		Rigorous
	Intelligent	Intelligent
		Technical
		Corporate
		Innovative
	Successful	Successful
		Leader
		Ambitious
		Powerful
Sophistication	Luxurious	Glamorous
		Luxurious
	Charming	Seductive
		Smooth
		Romantic
		Magical
Ruggedness	Outdoorsy	Outdoorsy
		Get-away
		Recreational
	Tough	Tough
		Rugged
		Non-conformist

focuses on the *adjectives* used by destinations and visitors.

The objective of the semantic analysis is to associate the adjectives of the set of tweets to the categories of emotional values shown in Table 1. A standard natural language parser was applied to retrieve the adjectives and count their frequency of use. A direct syntactic mapping was not possible, as most of the adjectives did not appear directly as categories/subcategories of emotional values. The idea is to use a semantic similarity measure (Sánchez, Batet, Isern, & Valls, 2012) between the adjectives and the categories/subcategories. This type of measures requires the use of some kind of external, structured knowledge base (in this case, WordNet). Ontology-based semantic similarity measures (Sánchez et al., 2012) rely on the topological structure of an ontology to calculate the degree of resemblance between two terms. The length of the path between the terms (considering hyponymy and hypernymy relationships) and their position in the hierarchy of concepts (i.e. their degree of generality) are the basic points taken into account by similarity functions. This study has used a well-known similarity measure, defined by Wu and Palmer (1994) as follows:

Table 2
Stop words list.

Stop words list
i, me, my, myself, we, our, ours, ourselves, yo, your, yours, yourself, yourselves, he, him, his, himself, she, her, hers, herself, it, its, itself, they, them, their, theirs, themselves, what, which, who, whom, this, that, these, those, am, is, are, was, were, be, been, being, have, has, had, having, do, does, did, doing, a, an, the, and, but, if, or, because, as, until, while, of, at, by, for, with, about, against, between, into, through, during, before, after, above, below, to, from, up, down, in, out, on, off, over, under, again, further, then, once, here, there, when, where, why, how, all, any, both, each, few, more, most, other, some, such, no, nor, not, only, own, same, so, than, too, very, s, t, can, will, just, don, should, now

$$Sim_{W\&P}(c1, c2) = \frac{2*N3}{N1 + N2 + 2*N3} \quad (1)$$

In this expression *N1* and *N2* are the number of hypernym links from the terms *c1* and *c2* to their least common subsumer (LCS) in the reference ontology, respectively, and *N3* is the number of hypernym links from the LCS to the root of the ontology. This measure ranges from 1 (for identical concepts) to 0 (when the LCS is the root of the ontology, so the concepts do not have any common ancestor). The main difference with this function with respect to other edge-counting measures is that it takes into account the depth of the compared concepts in the hierarchy (given the same distance, two concepts are more similar if they are more specific).

Algorithm 1 (semantic similarity) is used to calculate the semantic similarity between two words, where the GETSYNSETS function returns the list of synsets of a word in WordNet, and the function POS returns the part-of-speech tag of a synset.

Algorithm 1. Semantic similarity.

```
Function GetSemanticSimilarity (word1, word2)
  synsets1 ← getSynsets (word 1)
  synsets2 ← getSynsets (word2)
  if synsets1 is empty or synsets2 is empty then
    return 0
  else
    max _sim ← 0
    for each syns1 ∈ synsets1 do
      for each syns2 ∈ synsets2 do
        if POS (syns1) = POS (syns2) then
          sim ← Sim_W&P(syns1, syns2)
          max _sim ← MAX (max _sim, sim)
        end if
      end for
      end for
      return max _sim
    end if
end function
```

In the case of this study, the main problem of this approach is that WordNet uses hypernymy relationships between nouns, and the study's aim was to compare the adjectives found on the tweets

Table 3
Examples of automatic translations from adjectives into nouns.

Adjective	Noun
New	Newness
Best	Goodness
Wonderful	Wonder
Famous	Fame
Creative	Creativity
Traditional	Tradition
Magical	Wizard

with the subcategories associated to the emotional values (which are also mostly adjectives). Thus, before applying the Wu–Palmer semantic similarity measure, it was necessary to transform its inputs into nouns. In the case of the subcategories, these were manually translated to the equivalent nouns (e.g. 'friendly' and 'ambitious' were transformed into 'friend' and 'ambition'). Concerning the adjectives appearing in the tweets, they were automatically transformed into nouns using their derivationally related form or their attribute property in WordNet.

Table 3 shows some examples of the automatic translation of the adjectives into nouns. It may be noted that there are some good translations (e.g. 'wonderful' is translated into 'wonder') but also some bad results (e.g. 'magical' is translated into 'wizard'). It may be assumed that this is a bad translation because the adjective 'magical' in the context of tourism usually means 'beautiful or delightful in such a way as to seem removed from everyday life', whereas 'wizard' is a noun which may mean 'a man who has magical powers, especially in legends and fairy tales' or, in the context of computer science, 'a help feature of a software package that automates complex tasks by asking the user a series of easy-to-answer questions', and both of those meanings are not related to the standard use of the adjective 'magical' in a tweet of a tourist.

After this pre-processing stage it was possible to apply the Wu–Palmer similarity measure to compare the emotional subcategories and the tweet adjectives. In this way, the study could assess if the destination communicates certain emotional values and if there is a strategy behind the communication of the brand (for example, a repeated use of terms like 'love', 'honeymoon', 'passion' and 'romance' could be interpreted as a strategy towards the promotion of a 'romantic' destination).

The accuracy of the automatic semantic procedure that links adjectives to emotional values is hard to measure, since even a manual assessment is highly subjective. A common procedure to evaluate the performance of a semantic similarity measure is to calculate the correlation between its results and a human assessment of the similarity between some pairs of terms belonging to a golden standard. In the case of the Wu–Palmer similarity, Slimani (2013) obtained a 74% correlation and Budanitsky and Hirst (2006) report 82.9% and 81.9% in two different benchmarks (using the Lin measure, of which Wu–Palmer is a particular case, see Lin, 1998). Moreover, a manual study was made of the 49 adjectives appearing in the case study described in the next section that have a higher similarity with an emotional value (over or equal to 0.91). It was considered that 44 of them had been correctly classified (89.79%), whereas the other five could be debatable. These cases include words that are not actually adjectives (errors of the initial parser, like 'taste' or 'drive') and adjectives difficult to associate to the available emotional values ('lucky', 'chic' and 'able'). It also has to be taken into account that this case study only considered those adjectives that had a minimum similarity of 0.7 with an emotional adjective, so the study discarded those adjectives that were not highly related to any of the values looked for.

The case study and the detailed results of the semantic analysis of the selected destinations are shown and discussed in the following sections.

4. Case study

This section describes how the destinations to be analysed were selected and how the tweets associated to these destinations and to their visitors were obtained.

4.1. Destination selection

To select the destinations to be analysed, the researchers searched manually for the official Twitter accounts of the top 25 European destinations in 2014 according to Tripadvisor. Table 4 shows the number of tweets and followers of these accounts. In six cases it was not possible to identify such an account, and in two cases it was necessary to resort to the account of the associated region (Tuscany for Florence, and Emilia Romagna for Rimini). In five cases, marked with asterisk, most of the tweets were written in the main language of the destination and not in English. It was finally decided to select the 10 destinations that had more than 3000 English-language tweets, which are highlighted in bold in Table 4.

4.2. Retrieval of the sets of tweets

As it was the intention to compare the communication of the destination brand with the analysis of its perception, this study considered two datasets: an official set of tweets *O*, which contained the tweets that had been posted from the official accounts of the 10 selected destinations, and a tourist set of tweets *T*, which contained tweets posted by visitors of the destinations.

The researchers developed a novel tool, *twiQuery*, which was used to collect the tweets to be analysed. twiQuery is a crawler that enables users and developers to make advanced search actions on tweets, such as retrieving sets of tweets posted by a specific user, written in a specific language, sent from a certain area determined by the name of a city or geolocation and a given radius, and posted during a specific period of time. The set O is the set of all the tweets written in English and sent by the 10 official destination accounts. The set T contains the tweets in English that were sent within a radius of 15 km of the city centre of each of the 10 chosen destinations during the following periods of time:

- High tourism season of 2015: from 15-June-2015 to 01-September-2015.
- High tourism season of 2014: from 15-June-2014 to 01-September-2014.
- Low tourism season of 2014: from 15-October-2014 to 30-December-2014.

In order to make sure that only tweets sent by tourists were analyzed, set T retained only those tweets in which the location specified by the user in his/her Twitter profile is not same as the destination that was analyzed. The tweets sent by local users or by users that did not specify their home location were discarded. Also removed were the tweets sent by unusually prolific users (those that had sent over 1000 tweets in the time intervals considered). Using this process, the study obtained between 18,710 (Budapest) and 38,116 (London) tourist tweets for each of the 10 destination cities. As it was the intention to compare references to emotional values among visitors and destinations, 3000 tweets were ultimately analysed from the official destination account and 3000

Table 4
Selection of destinations.

City	Twitter account	Tweets	Followers
Istanbul	–		
Rome (*)	@Turismoromaweb	4310	16694
London	**@visitlondon**	**31,358**	**394,617**
Prague	@PragueEU	1560	2408
Paris (*)	@ParisJeTaime	13,900	78,200
Berlin	**@berlintourism**	**5466**	**22,533**
Florence	@DiscoverTuscany	1545	7203
Barcelona	**@VisitBCN_EN**	**4843**	**8937**
St Petersburg (*)	@VisitPetersburg	2142	363
Budapest	**@VisitBudapest**	**3398**	**15,689**
Lisbon	@TurismoDeLisboa	1227	5829
Venice	–		
Madrid (*)	@Visita_Madrid	8895	30,203
Amsterdam	**@Iamsterdam**	**6174**	**147,998**
Krakow	–		
Vienna	**@ViennaInfoB2B**	**4220**	**6631**
Milan (*)	@turismomilano	6169	31,197
Athens	**@CityofAthens**	**4600**	**12,387**
Zermatt	@zermatt_tourism	2953	9645
Urgup	–		
Dublin	**@VisitDublin**	**12,402**	**63,278**
Moscow	–		
Dalyan	–		
Edinburgh	**@edinburgh**	**26,565**	**57,984**
Rimini	**@ERTourism**	**18,980**	**9942**

tweets from visitors for each of the 10 cities. These tweets were randomly chosen.

5. Results

The results of the semantic analysis are shown in Table 5. This analysis only took into account 'emotional adjectives' which are those adjectives used in the tweets that have a similarity of over 0.7 with at least one of the emotional subcategories (the adjectives that do not satisfy this condition were discarded). The rows of the matrix are the subcategories of emotional values. The columns of the matrix are the 10 cities analyzed, considering the tweets of the official Twitter accounts of the destinations (on the left) and the tweets of their visitors (on the right). At the top of each column there are two numbers, which represent the total number of adjectives (on the right) and the total number of uses of those adjectives in the set of 3000 tweets (on the left). For instance, the cell at the top on the left indicates that, in the 3000 tweets sent by the official Twitter account of London, 289 emotional adjectives have been used (that could be highly related to emotional values), and those adjectives have been used 3565 times in those tweets (1.18 emotional adjectives/tweet). Each cell in the matrix, associated to an emotional value and to a city, contains two numbers with the same meaning: on the right is the number of emotional adjectives associated to that subcategory used in the tweets of the city (tweets sent by the destination or tweets sent by its visitors),

Table 5
Use of emotional adjectives by DMOs and tourists.

Subcategory	Official Account										Tourist									
	Lond	Berl	Dubl	Athe	Edin	Vien	Amst	Buda	Barc	Rimi	Lond	Berl	Dubl	Athe	Edin	Vien	Amst	Buda	Barc	Rimi
#	3565/289	1086/186	923/153	1495/197	1140/162	1251/202	1193/192	699/143	1054/171	1035/208	952/248	960/223	970/251	864/218	990/265	864/231	880/223	923/212	834/207	681/163
Family_oriented		1/1						3/1	1/1	2/2	5/4		2/2	1/1	2/2	5/2	3/2		1/1	
Down_to_earth	118/22	53/15	15/8	90/16	66/13	84/19	50/18	41/14	39/15	44/16	51/26	63/24	63/23	45/20	64/24	46/23	45/21	35/20	44/24	35/14
Sustainable	15/9	12/4	1/1	12/4	2/1	8/1	9/4	7/4	4/2	4/4	2/2	9/5	7/7	2/2	1/1	5/4	8/6	3/3	1/1	
Calm	127/9	73/4	29/6	52/4	68/5	59/4	67/4	40/2	51/4	18/4	30/6	33/8	32/5	28/8	26/10	25/9	23/7	27/9	25/4	19/7
Real	28/6	18/6	18/5	27/5	10/4	17/5	17/5	9/4	13/4	27/7	35/6	28/7	35/7	32/7	22/7	38/6	29/7	31/8	33/8	15/5
Traditional	106/21	35/9	7/3	72/12	21/6	50/11	49/13	36/5	21/10	57/16	21/10	16/11	14/10	22/12	15/10	24/8	20/6	19/8	14/7	7/5
Honest	399/10	163/5	178/7	328/7	146/5	124/8	135/8	104/6	285/5	157/9	150/10	144/5	168/8	155/8	158/6	142/6	132/6	194/8	148/4	132/4
Original	114/23	42/15	13/10	50/15	16/9	59/19	32/13	19/14	41/14	38/18	48/22	79/20	36/16	43/20	34/17	47/18	50/20	48/13	38/13	28/10
Wholesome	4/1	1/1	1/1	2/1	5/2		2/1			1/1	3/1	5/1	4/1	2/1	1/1	1/1	2/1		3/1	3/2
Quality of life																				
Happiness	57/6	11/4	8/6	6/4	40/3	22/4	29/4	16/3	15/3	7/4	25/7	20/5	33/8	36/8	40/8	22/6	20/10	21/8	26/8	66/5
Sentimental	25/10	10/4	12/5	12/8	10/4	8/5	6/4	8/5	3/3	18/7	24/12	17/7	14/11	15/7	25/15	22/11	13/8	33/11	21/12	10/6
Friendly	20/7	5/3	4/3	5/2	5/3	7/4	17/5	2/1	8/4	6/3	3/2	7/5	12/6	7/4	3/3	2/2	10/6	3/3	11/5	
Trendy	23/6	13/3	12/2	31/3		18/2	17/3	55/2	39/2	6/2	3/3	7/2	7/2	2/1	3/3	1/1	1/1	1/1	2/2	2/1
Daring	40/8	13/7	9/4	18/4	19/6	14/6	22/7	13/5	33/7	7/4	14/4	10/4	23/5	22/7	25/7	17/4	19/4	12/5	19/6	6/4
Exciting	6/1	5/2	2/2	1/1	5/3		9/2	1/1	2/2	12/2	4/3	7/2	5/3	4/2	1/1	4/2	27/1	3/1	5/2	1/1
Exotic	5/2	4/2		4/1	1/1	2/2	12/2		2/2	5/2	1/1		2/1	4/2	2/2	11/2	1/1	4/1	2/1	1/1
Fashionable																				
Cool																				
Spirited	6/2	1/1	1/1	4/1	7/3	8/3	4/3		4/3	1/1	3/2	4/3	2/2	1/1	1/1	2/2	3/2	1/1	4/3	1/1
Dynamic	3/3	9/3		1/1	2/2	20/4	4/1	1/1	23/3	4/3	1/1	5/3		2/2	8/4	2/2	2/2	3/3	1/1	1/1
Vital			1/1		8/1							1/1	2/1			1/1				
Fresh	1141/6	208/6	70/7	210/6	131/4	224/6	218/6	123/3	56/5	148/6	115/5	90/3	99/7	96/5	92/6	76/6	76/5	77/6	66/5	63/3
Young	11/2	6/2		2/1	2/1	15/2	5/1	2/1	4/1	2/2	1/1	9/4	8/2	8/2	11/1	4/1	2/1	5/1	1/1	1/1
Sensorial	38/11	11/4	66/10	52/5	20/6	15/6	14/8	9/4	7/5	17/6	24/8	18/7	23/10	12/6	31/10	14/8	25/11	16/7	20/5	27/4
Unique	124/5	25/4	14/3	33/4	41/4	42/5	53/3	20/4	36/5	38/4	13/4	7/2	13/3	9/3	9/3	13/3	13/4	12/3	10/4	13/4
Imaginative						1/1														
Creative	29/7	35/6	10/5	21/7	3/3	20/7	13/2	4/3	4/4	19/8	12/7	12/5	9/6	7/3	9/7	8/5	9/3	7/3	6/3	4/4
Up_to_date	30/2	7/2	3/1	27/2	1/1	50/2	11/2	12/2	20/2	58/2	1/1	5/3	4/2	3/1	2/1	4/2	6/2	6/2	2/2	3/2
Independent	12/3	2/2	7/2	5/1	1/1	1/1	1/1		2/2	4/2	3/2	4/2	2/1	5/1	6/5	5/2	1/1	4/2	4/2	12/2
Contemporary	333/9	63/9	42/5	32/10	71/9	60/7	43/7	31/6	45/5	46/6	49/9	57/9	66/11	43/9	67/8	54/9	50/9	49/9	50/9	18/6
Cosmopolitan				1/1																
Tolerant			1/1	1/1				1/1					1/1	1/1	1/1			1/1		
Hospitable	1/1	1/1		5/2			2/1	1/1		8/2	2/2		1/1	3/2	4/2	2/2	3/2		2/2	15/4
Reliable							1/1				2/2		1/1	1/1	2/1		1/1			
Hard-working	1/1		1/1		1/1						1/1					1/1	1/1	1/1		2/1
Safe	63/14	21/9	39/8	56/11	24/10	51/12	40/11	20/8	37/9	47/8	44/14	33/11	38/16	20/13	43/15	33/16	39/11	34/12	24/10	29/13
Rigorous		1/1							3/1						3/1				2/1	
Intelligent	23/7	22/8	8/5	23/5	7/4	11/3	30/6	3/3	16/3	14/5	7/5	15/7	12/6	9/6	8/6	9/6	7/5	11/5	8/6	1/1
Technical	9/1								1/1		3/2	1/1	2/1	2/1	1/1	2/1	2/2	2/2		2/1
Corporate	1/1	1/1	1/1	1/1	2/2		4/1	1/1			8/4	10/1	6/1	11/3	4/2	3/3	2/1	5/1	7/4	10/2
Innovative	28/3	7/1	22/2	10/3	14/3	10/3	13/4	6/2	5/2	5/3	10/5	9/3	6/4	4/2	5/4	10/5	16/5	9/1	15/3	10/2
Successful	3/2	5/2		5/2	11/2	7/3	8/3	7/2	1/1	2/1	6/2	7/2	8/3	1/1	6/2	3/2	2/1	3/1	4/2	2/1
Leader	125/20	34/9	37/6	67/21	34/10	82/12	64/11	35/16	76/11	41/10	63/13	61/12	65/13	47/11	76/17	59/11	56/12	62/14	68/16	33/9
Ambitious	4/3	2/1	1/1		1/1	2/2	1/1		5/4	5/2	3/3	3/3	3/3	2/2	2/2		1/1	5/4		2/2
Powerful	17/4	12/4	10/3	14/4	8/3	7/2	9/4	1/1	3/1	11/4	3/1	14/5	12/6	3/2	12/3	4/1	10/3	5/2	7/2	
Glamorous	76/6	49/2	105/5	89/2	111/6	58/4	93/4	17/4	33/4	74/8	52/7	52/6	43/6	53/6	44/7	44/8	56/7	68/4	51/5	30/8
Luxurious	7/2	6/3	2/2	1/1	7/2	7/3	2/1	1/1	5/2	4/2	5/2	3/2	6/3	8/2	6/2	4/4	2/2	7/3	10/3	3/2
Seductive					2/1	4/1	1/1	1/1	1/1	1/1	1/1	3/1	2/2	3/2	2/1	1/1	2/1	2/2	2/1	1/1
Smooth	17/4	9/3	18/2	28/3	8/2	12/5	6/2	11/2	46/3	8/5	19/3	23/4	16/2	16/4	20/9	19/4	28/5	17/3	18/2	25/3
Romantic	3/1	2/1	5/1	8/1	3/1		2/1		2/1	2/1	2/1	4/1	4/3	4/1		5/2	6/1	9/1	4/1	
Magical	105/2	9/1	69/2	14/1	108/2	13/1	9/2	1/1	4/1	2/1	15/1	4/1	4/1	2/1	15/2	6/1	6/1	4/1	5/1	7/2
Outdoorsy																				
Get-away	64/11	8/4	9/4	15/4	22/5	10/6	11/4	5/3	19/3	14/5	3/3	9/3	10/4	5/3	12/9	5/4	6/3	10/8	4/2	5/4
Recreational	5/3		3/1	1/1		1/1		2/1	10/2	2/1	4/2	4/1	1/1	1/1	2/1	2/1	3/1	1/1	5/1	4/1
Tough	1/1	1/1		1/1							2/2	2/1	3/1	1/1	2/1				1/1	1/1
Rugged	192/7	69/9	65/6	58/7	76/7	46/8	56/6	30/4	27/7	49/8	53/10	42/7	39/10	55/6	60/12	52/9	37/9	53/9	39/9	26/11
Non_conformist	6/4	1/1	4/4			2/2	2/1		2/1		3/3	4/3	2/2	6/3	2/1	5/3	4/1		1/1	5/1

while on the left is the number of times those adjectives have been used in the tweets. For instance, it can be seen in the table that the official Twitter account of London used 21 different adjectives related to the 'traditional' value, and those adjectives were used 106 times in the 3000 tweets analyzed.

5.1. Adjectives used by the DMO and the visitors of each city

The analysis can be started by looking at the number of adjectives used by DMOs and by tourists. It can be seen that the visitors of all cities employ a mean of 224 different adjectives, were used (on average) 891 times in the 3000 tweets/city (the main exception is Rimini, where only has 163 adjectives were used 681 times). In most cases, DMOs used a smaller variety of adjectives, although they were used more times in the tweets. More concretely, on average DMOs employed 170 different adjectives 1344 times on their tweets. The only cities in which DMOs used a wider range of emotional adjectives than their visitors were London and Rimini. The city that employed the fewest adjectives was Budapest (143), whereas London (289) was clearly the DMO that used the most different adjectives. These results are quite surprising, since if DMOs had a clear strategy for communicating the emotional values of their brands they would be expected to use a richer variety of adjectives than tourists, who would be expected to use the same kind of expressions wherever they are visiting. In this sense, it seems that London is the destination that is making a bigger effort to select different emotional adjectives. The fact that English is the native language of this destination does not seem to be very relevant, because other cities with this property, such as Dublin and Edinburgh, did not exhibit this variety.

5.2. Emotional values communicated by adjectives

The references to emotional values used by the DMO and the visitors of each city will now be analyzed in more depth. In Table 5, the gray cells show the heavier uses of adjectives to refer to an emotional value, whereas the boxed cells indicate situations in which an emotional category has been referenced only in a few cities (for instance, 'cosmopolitan' has only been linked to Athens). Black cells highlight those cases in which a city's DMO is not communicating an emotional value that has been considered by most of the other cities (for example, Budapest is the only DMO that has not used any adjective related to the value 'spirited').

5.2.1. Analysis of the number of emotional adjectives

The number of adjectives employed in the tweets (the numbers on the right in the cells of the matrix) will first be considered. As shown in the table, the emotional values that were communicated with the highest number of different adjectives by DMOs were 'original', 'down-to-earth' and 'leader', followed by 'safe', 'traditional' and 'sensory'. Almost all of the analyzed destinations communicated these values through their tweets with a large number of adjectives. It is remarkable that although all destinations want to be 'original', it may be seen in the table that they communicate mostly the same values and they are not showing a distinctive personality.

It can be observed that tourists also use the largest number of adjectives to refer to the values 'down-to-earth', 'original', 'safe' and ;leader'. However, unlike DMOs, they also use many adjectives to communicate values like 'sentimental', 'rugged' and 'contemporary' (for example, Edinburgh visitors used 15 adjectives related to 'sentimental', whereas the official tweets of this destination only used four). This disparity is interesting for DMOs which, given that these values are highlighted by their visitors, would do well to analyze them and decide if they indeed belong to their identity and if they should be more heavily communicated in their tweets. DMOs should also take into account general values such as safeness, which are referenced by visitors of all cities, and stress them in the definition of their identities.

The emotional categories that are referenced with a medium number of adjectives by DMOs are 'honest', 'calm', 'real', 'daring', 'fresh', 'unique', 'creative', 'contemporary', 'intelligent' and 'glamorous'. They are also communicated to different degrees by tourists, showing on the one hand a certain coherence between the transmitted and the perceived values. On the other hand, it also shows that all destinations seem to have many points in common in the communication of emotional values, without using specific strategies for brand personalization.

Finally, the values that have been communicated with a low number of adjectives are 'wholesome', 'exciting', 'exotic', 'spirited', 'dynamic', 'up-to-date', 'independent', cosmopolitan', 'tolerant', 'hospitable', 'reliable', ;hard-working', 'rigorous', 'seductive' and 'romantic'. Many of these values are very attractive, so it is surprising that they are only lightly communicated by a few destinations. For example, nowadays most destinations want to be regarded as sustainable, but cities such as Dublin, Edinburgh, Vienna and Barcelona use a very low number of related adjectives. Just to give another example, the number of references to a family-oriented destination is also strikingly low in all destinations: for instance, London does not use any adjective for this value, although its official website has a whole section dedicated to attractions and activities for families with children. This finding might be pointing out a lack of coherence between the values transmitted through different social media by a DMO.

It is also surprising that there are several values that have not been referenced at all neither by destinations nor by tourists: 'quality-of-life', 'fashionable', 'cool', and 'outdoorsy'. This finding should be checked more carefully, as it could be due to a poor choice of nouns to represent these values in WordNet.

By way of conclusion to the analysis of the number of adjectives used to communicate emotional values by DMOs, it may be seen that there are many commonalities in the results obtained for all the cities that are analyzed. Thus, it might be concluded that, with some exceptions, there is a lack of a proper communication strategy of the distinctive points of each destination. The same result is observed in the analysis of the adjectives used by visitors, which does not show a emotional distinction of the destinations.

Figs. 2a and 2b show the number of adjectives that have been used for each main category of emotional values by destinations and by tourists respectively. In the comparison of these two charts, the first thing that can be noticed is that 'sincerity' is the emotional value most heavily communicated by both DMOs and tourists. However, the number of adjectives used to communicate this value by tourists is considerably larger than the number of adjectives used by DMOs (with the exceptions of London and Rimini). This means that tourists give great importance to this emotional value and associate it with the destinations in their messages, whereas DMOs do not seem to have such a strong intention to communicate it.

The values 'excitement' and 'competence' are communicated by a medium number of adjectives both by tourists and DMOs, with minimal differences between them. Perhaps the cities that show a higher difference for these two emotional values are Dublin and Edinburgh, where tourists used significantly more adjectives than DMOs to communicate them. In these kinds of situation it might be argued that it is necessary for the DMOs to take this finding into account in their brand communication strategy. If tourists employ many adjectives to communicate a value, it may indicate that the destination is being associated to that emotion and it should probably be incorporated into its brand communication strategy.

Notably, the number of adjectives that communicate 'sophistication' and 'ruggedness' is the lowest in both the tweets of DMOs

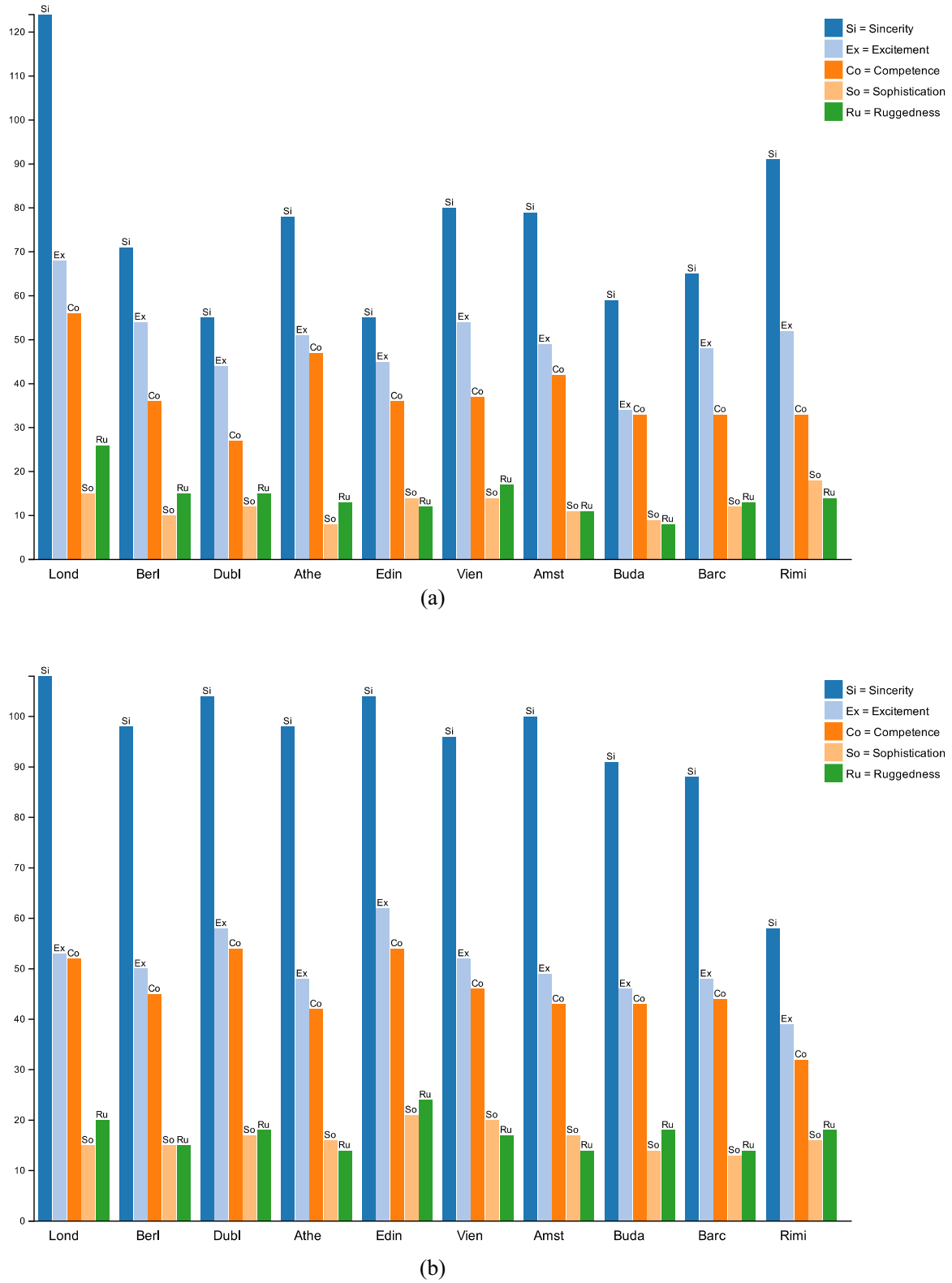


Fig. 2. Histogram of the number of adjectives used by destinations (a) and tourists (b).

and tourists, although visitors use them slightly more than destinations. For these emotional values, Edinburgh is again the city with a greatest difference between the number of adjectives used by tourists and those used by DMOs. However, there are not huge differences between them for these two kinds of emotional value.

5.2.2. *Analysis of the frequency of use of emotional adjectives*

Secondly the presentation of the findings of this study turns to the analysis of the overall frequency of use of emotional adjectives (the numbers on the left in the cells of the matrix shown in Table 5). This measure is key in the analysis of the communication of the emotional values because it shows the total number of uses of all the adjectives related to each emotional value. At the end of the

day, this is probably the most relevant finding since a destination can communicate a value more effectively by using a single adjective many times, rather than by using different adjectives each a few times.

If the absolute frequencies of use of adjectives by DMOs and tourists is observed, it can be observed that the results are the opposite to those reported in the previous section. In this case, destinations communicate the emotional values with a higher frequency than tourists (except Dublin and Budapest). This suggests that DMOs have a higher communicative intention of emotional values than visitors, although the numerical differences are not very large. It is specially remarkable in the case of London, whose DMO used 3565 emotional adjectives in the 3000 analysed tweets. This high value shows that this destination incorporated the transmission of the emotional part of its identity into its communication strategy.

The emotional value that is mostly communicated by both destinations and tourists is 'honest'. All the DMOs and visitors use,

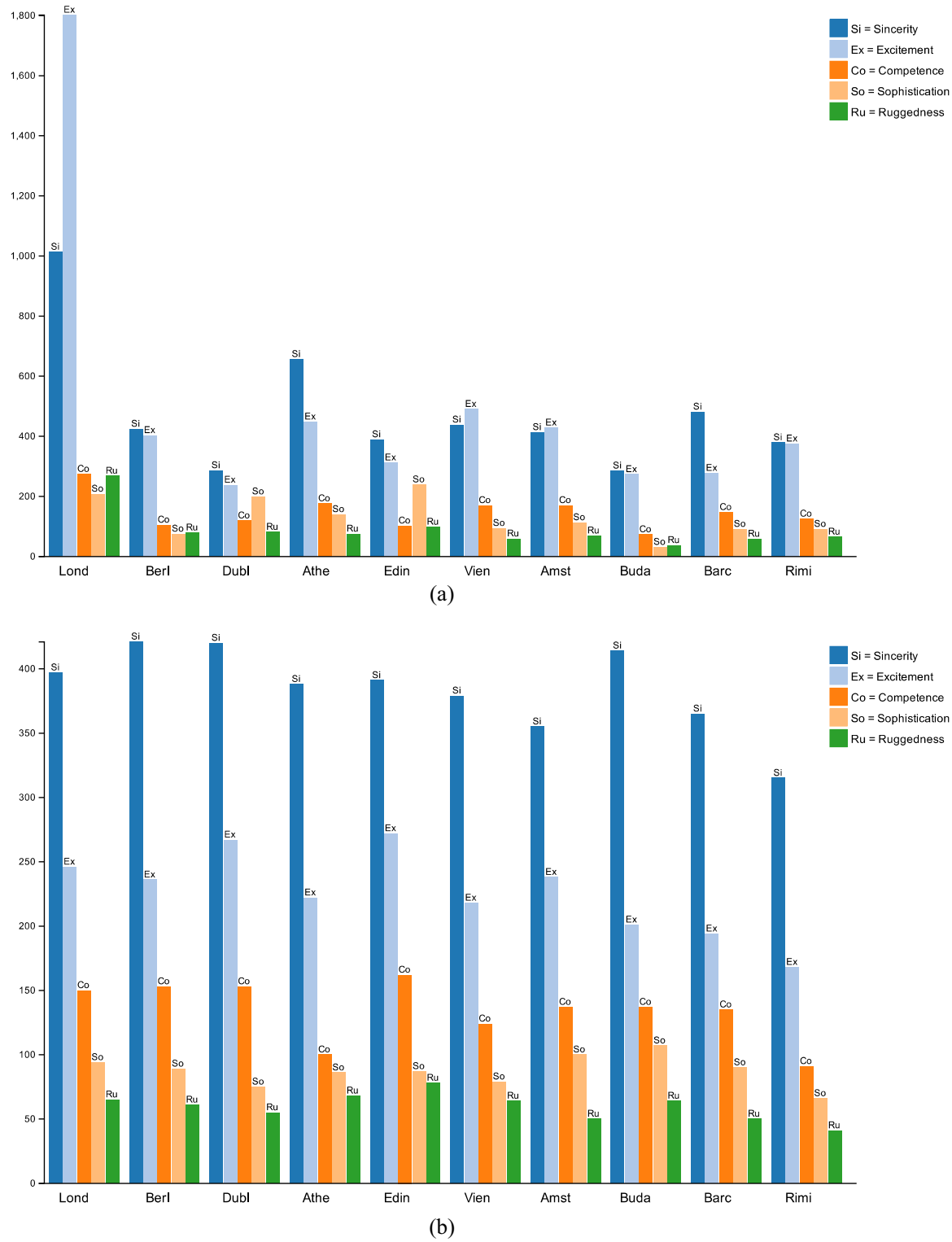


Fig. 3. Histogram of the overall use of emotional adjectives by destinations (a) and tourists (b).

in more than 100 occasions, adjectives that communicate this value, showing its importance for all of them. At the present time, in which tourism development and the evolution of the destinations have sled in many cases to the increase of mediation, virtualization and artificiality, it is observed that the value of honesty is prioritized more than ever. This may be linked to another value highly communicated by DMOs and visitors, which is 'real'. It may be concluded that tourists search for and appreciate 'real', 'honest', natural, local experiences and they turn away from artificial, common, globalized ones.

Another value well communicated by all DMOs (except Dublin and Barcelona, that do not mention it so often) is 'fresh'. This value is intuitively associated to concepts such as new, natural, green, and different. Tourists also made reference to these ideas on their tweets, but with much lower frequencies. It stands out the extremely high communication of this notion by the DMO of London (1141 times in 3000 tweets, with a short set of six adjectives), showing a clear strategy towards the intense communication of this emotional value.

The next level of communicated emotional values included 'down-to-earth', 'calm', 'traditional', 'original', 'sensory', 'contemporary', ;safe', 'leader', 'glamorous' and 'rugged'. As commented before, it is curious that interesting values such as 'family-oriented', 'quality of life', 'fashionable', 'cool', 'imaginative', 'cosmopolitan' or 'outdoorsy' were not reported at all neither by DMOs nor by tourists.

In examining whether destinations have communication strategies of specific brands and if they use social media to communicate a differentiating personality, Table 5 illustrates the emotional values reported by only one or two of them. London is the place that shows, for some values, a much higher frequency of use of adjectives than the rest of analyzed destinations. For example, it communicates with frequencies well above the rest values such as 'down-to-earth', 'calm', 'unique', 'contemporary', 'leader', 'magical' and 'rugged'. Thus, a clear communication strategy of an identifying and distinguishing mark is observed again.

Other destinations also highlight specific emotional values, albeit withless use of adjectives. For example, London and Vienna communicate 'magical' much more than the rest, or only Dublin and Edinburgh communicate 'glamorous' with higher frequencies. This shows the existence of certain differential emotional values that some DMOs try to communicate to distinguish their own identity. However, there is a very small number of differentiated associations between destinations and emotional values, and sometimes the differences among them are not very significant. This shows that few of the destinations that are analyzed show a clear strategy for communicating their brands through their official tweets.

Figs. 3a and 3b show the total number of times in which adjectives have been used by destinations and tourists respectively, to refer to each main category of emotional values. It may be observed that the use of emotional adjectives by tourists is fairly homogeneous in all destinations, whereas its use by DMOs shows more diversity. For example, tourists communicate strongly the emotional value 'sincerity' in all destinations, but this value is communicated to very different degrees by DMOs. These figures show that visitors do not perceive a clear differentiation of the emotional values transmitted by destinations.

Despite the uniformity in the opinions of the visitors, the differences between the uses of adjectives in the destinations might indicate a differentiation strategy. If they had brand communication strategies, clear differences should appear in all the cities in Fig. 3, but it is observed only in some of them. The most prominent example is London, which shows a very high level of communication of the values 'excitement' and 'sincerity'. Vienna and Athens also stand out with a high frequency of use of adjectives related to 'excitement' and 'sincerity', respectively. Dublin and Edinburgh, unlike the other destinations, give more relevance to 'sophistication' than to 'competence' and 'ruggedness'. These specific cases may indicate the existence of a strategy to communicate a distinctive, differentiated brand. However, a deeper study of the tweets of these destinations should be made in order to corroborate this assertion.

5.2.3. *Transmission and perception of emotional values*

If the emotional brand values communicated by DMOs are compared with those referenced by the adjectives employed by tourists on their tweets it is possible to detect a significant difference among them. For instance, even though the DMOs of Dublin and Edinburgh have stressed the 'glamorous' concept, it is interesting to note that the visitors of these places are the ones that have the least use of adjectives referred to this value. This fact may indicate a disparity between the values intended to be transmitted by DMOs and the ones perceived by tourists and transmitted in their tweets, which may be a very relevant information for destination managers in order to consider a reorientation of their public communication strategy.

The results reported in Table 5 show that the differences among the values communicated by the official DMOs through their tweets and those reflected by the messages of tourists are huge. Indeed, numerous examples show that the coincidence of communication of values is the exception, rather than the rule. For example, the 'original' value is strongly communicated (114 references) by London (as well as by other destinations), but its tourists only used adjectives related to this value 48 times. In contrast, Berlin communicated the same value only 42 times but its tourists used 79 adjectives related to this topic.

Other examples with lower frequencies show similar mismatches. For example the tourists of Dublin have one of the highest number of references (eight) to the value 'young', although the official DMO of this city did not use any adjective related to this characteristics in 3000 tweets. On the other side of the coin, Vienna communicated this value 15 times but it was only mentioned in four of its visitors tweets. These examples show that there is not much correlation between the communication of the identity of destinations and the images that tourists communicate through their tweets. These results provide interesting information to DMOs which might help them to improve the definition and communication of their brands and identities.

Finally, Table 6 shows the average number of emotional adjectives (and the average number of uses of those adjectives in the tweets) employed by DMOs and by tourists for each emotional value, and the difference between them. The larger differences are highlighted in italics in the rightmost column.

This table confirms the previous findings, and the different use of adjectives to communicate destination brands between DMOs and tourists is shown clearly in the last column of the table. The values on the right of that column show the differences in the number of adjectives. Although these differences are not very great, it may be seen that most of them are negative, showing that tourists use a wider set of adjectives than DMOs in their tweets. The emotional values in which these negative differences are stronger are 'safe', 'happiness', 'sentimental' and 'down-to-earth'. The values on the left of the last column correspond to the difference in the total number of times in which adjectives have been used by DMOs and by visitors on their sets of 3000 tweets. In this column, most of the values are positive, showing that DMOs employ more emotional adjectives than tourists. There are some values for which there is a huge difference of use, especially 'fresh', 'honest', 'calm' and 'unique'. The only emotional value for which there is a large negative difference (i.e. it is way more referenced by tourists than by DMOs) is 'real'.

Table 6
Average number of emotional adjectives (and their use) employed by DMOs and by tourists for each emotional value.

Subcategory	Official accounts	Tourists	Difference
Family-oriented	0.7/0.5	1.9/1.4	−1.2/−0.9
Down-to-earth	60.0/15.6	49.1/21.9	**10.9**/−**6.3**
Sustainable	7.4/3.4	3.8/3.1	3.6/0.3
Calm	58.4/4.6	26.8/7.3	**31.6**/−2.7
Real	18.4/5.1	29.8/6.8	−**11.4**/−1.7
Traditional	45.4/10.6	17.2/8.7	**28.2**/1.9
Honest	201.9/7.0	152.3/6.5	**49.6**/0.5
Original	42.4/15.0	45.1/16.9	−2.7/−1.9
Wholesome	1.6/0.8	2.4/1.0	−0.8/−0.2
Quality of life	0.0/0.0	0.0/0.0	0.0/0.0
Happiness	21.1/4.1	30.9/7.3	−9.8 /−**3.2**
Sentimental	11.2/5.5	19.4/10.0	−8.2/−**4.5**
Friendly	7.9/3.5	5.8/3.6	2.1/−0.1
Trendy	21.4/2.5	2.9/1.7	**18.5**/0.8
Daring	18.8/5.8	16.7/5.0	2.1/0.8
Exciting	4.3/1.6	6.1/1.8	−1.8/−0.2
Exotic	3.5/1.4	2.8/1.2	0.7/0.2
Fashionable	0.0/0.0	0.0/0.0	0.0/0.0
Cool	0.0/0.0	0.0/0.0	0.0/0.0
Spirited	3.6/1.8	2.2/1.8	1.4/0.0
Dynamic	6.7/2.1	2.5/1.9	4.2/0.2
Vital	0.9/0.2	0.4/0.3	0.5/−0.1
Fresh	252.9/5.5	85.0/5.1	**167.9**/0.4
Young	4.9/1.3	5.0/1.5	−0.1/−0.2
Sensorial	24.9/6.5	21.0/7.6	3.9/−1.1
Unique	42.6/4.1	11.2/3.3	**31.4**/0.8
Imaginative	0.1/0.1	0.0/0.0	0.1/0.1
Creative	15.8/5.2	8.3/4.6	7.5/0.6
Up-to-date	21.9/1.8	3.6/1.8	**18.3**/0.0
Independent	3.5/1.5	4.6/2.0	−1.1/−0.5
Contemporary	76.6/7.3	50.3/8.8	**26.3**/−1.5
Cosmopolitan	0.1/0.1	0.0/0.0	0.1/0.1
Tolerant	0.3/0.3	0.4/0.4	−0.1/−0.1
Hospitable	1.8/0.8	3.2/1.7	−1.4/−0.9
Reliable	0.1/0.1	0.7/0.6	−0.6/−0.5
Hard-working	0.3/0.3	0.6/0.5	−0.3/−0.2
Safe	39.8/10.0	33.7/13.1	6.1/−**3.1**
Rigorous	0.4/0.2	0.5/0.2	−0.1/0.0
Intelligent	15.7/4.9	8.7/5.3	7.0/−0.4
Technical	1.0/0.2	1.7/1.2	−0.7/−1.0
Corporate	1.1/0.8	6.6/2.2	−5.5/−1.4
Innovative	12.0/2.6	9.4/3.4	2.6/−0.8
Successful	4.9/1.8	4.2/1.7	0.7/0.1
Leader	59.5/12.6	59.0/12.8	0.5/−0.2
Ambitious	2.1/1.5	2.1/2.0	0.0/−0.5
Powerful	9.2/3.0	7.0/2.5	2.2/0.5
Glamorous	70.5/4.5	49.3/6.4	**21.2**/−1.9
Luxurious	4.2/1.9	5.4/2.5	−1.2/−0.6
Seductive	1.0/0.6	1.9/1.3	−0.9/−0.7
Smooth	16.3/3.1	20.1/3.9	−3.8/−0.8
Romantic	2.7/0.8	3.8/1.1	−1.1/−0.3
Magical	33.4/1.4	6.8/1.2	**26.6**/0.2
Outdoorsy	0.0/0.0	0.0/0.0	0.0/0.0
Get-away	17.7/4.9	6.9/4.3	10.8/0.6
Recreational	2.4/1.0	2.7/1.1	−0.3/−0.1
Tough	0.3/0.3	1.2/0.8	−0.9/−0.5
Rugged	66.8/6.9	45.6/9.2	**21.2**/−2.3
Non-conformist	1.7/1.3	3.2/1.8	−1.5/−0.5

In conclusion, this table shows the differences between the communication of the emotional values performed by DMOs and by tourists, which coincides with the results of other studies. Recently Mariné-Roig (2016) also showed differences between the brand communicated by DMOs through their official websites and that reflected by tourists through social networks. Nowadays it is commonly accepted that tourists co-create the destination brand through their activities in social networks (Gensler, Vlckner, Liu-Thompkins, & Wiertz, 2013; Hollebeek, Glynn, & Brodie, 2014; Labrecque, 2014; Park & Kim, 2014); thus, the communication differences between DMOs and tourists show a discrepancy that should be resolved. Therefore, it is urgent that DMOs analyze the way in which their visitors perceive their brand so that they can manage a coherent and strategic brand communication, trying to minimize such differences.

6. Conclusions

The communication through social media of the basic emotional values attached to a destination is a key aspect in the construction of a distinctive brand and personality. Up to now, most of the studies on such communication have made manual, syntactic analysis of the transmitted messages. This paper has reported the definition of a novel methodological framework, which is both automatic and fully semantic. The basic idea is to use an external corpus (WordNet) to link the meaning of the adjectives used in tweets with the emotional values attached to travel destinations. Due to its automatic nature, this analysis may be performed by DMOs as a way of self-assessing the values they transmit.

This paper has also presented a detailed case study, focused on 10 of the major touristic cities in Europe, in which 6000 tweets have been automatically analyzed for each of them (3000 tweets from the official tourism office and 3000 tweets from visitors of the city). In this way, we have been able to analyze how many emotional adjectives are used, how many times they are used and which emotional values they are associated with. The experiment has shown that most of the cities are referencing the same values, showing a lack of definition of a personalized identity. This conclusion coincides with that of a recent study carried out by Huertas and Mariné-Roig (2016) in which they manually analyzed the content of the publications of Spanish DMOs through their main official social media. That study focused on the communications of both tourist attractions and emotional values. Its main conclusion, like the one reached in this paper, was that the transmitted emotional values were very similar across all destinations and, therefore, that they were indistinctive. Although in this work the study was made on international destinations, it could be argued that this coincidence of results corroborates the findings, shows their validity and also validates the semantic methodology of analysis used in this study.

The results have also shown strong discrepancies between the values transmitted by DMOs and those reflected by the comments of visitors, proving a lack of effectiveness of the official messages in the communication of a specific distinctive brand. This difference between the values communicated by the DMOs of destinations and what tourists communicate also shows the mismatch between the identity and the brand that the destination wants to convey and the image that tourists have of the place. This shows the limitations of DMOs in the communication of their destination brands and also shows the necessity on the part of DMOs to improve the communication of their brand values through social media.

Considering these results, there does not appear to exist a strategic communication of brands in most of the official Twitter accounts of the destinations analyzed. This idea also coincides with the conclusions of Huertas and Mariné-Roig (2016), which showed that emotional values were not taken into account in the strategies of communication of content. This aspect is quite surprising because all the studies show the importance of brand and emotional communication in building relationships with the public and on influencing their travel decisions (Laroche et al., 2013; Morgan & Pritchard, 2004; Zhang et al., 2009), as well as the great potential offered by social media to communicate brands and emotional values (Govers et al., 2007; Mariné-Roig, 2013; Stepchenkova & Zhan, 2013).

Thus, the main conclusions of the study are twofold. On the technical side, it has been shown how it is possible to make a fast, automated and semantic analysis of a large number of tweets associated to a given location, avoiding the expensive cost of human manual analysis. From the communication point of view, it is possible to analyze whether DMOs make a strong effort on the definition of a personalized, distinguishing brand and if they define clear strategies on how to use social media to convey its main identity traits.

This study has important consequences for destination managers. The first is related to communication strategies. The destinations' directors of communication should design clear brand communication strategies and they must make sure that community managers follow some specific guidelines for brand communication through the social media. Given that the social media have a great potentiality to communicate destination brands, it is necessary to make a strategic use of these new tools to obtain an optimal implementation of the brand communication strategy. To achieve this goal it is necessary for directors of communication establish some brand communication guidelines to be followed by community managers, including not only broad content strategies but also concrete guidelines on how to write text as well as how to use terms and concepts.

Another implication of the study reported in this paper is that community managers of DMOs need to know that their function is not only to inform users about news or to show them the main tourist attractions of the destination but also to communicate identities, experiences and emotional values. In order to do that they have to take great care of the content they publish, to be aware of the language and words they use, as well as the images and all audiovisual material they upload. As mentioned before, other studies show the prevalence of informative content over the emotional brand content (Huertas & Mariné-Roig, 2015; Míguez-González & Huertas, 2015). A challenge for DMOs will be to address this issue. They must analyze whether there is a brand communication strategy or not, if this strategy is clearly communicated to the community managers, and if the professional profile of community managers, generally journalists, can determine an undesired prevalence of mere information over brand emotional values.

This study also shows that DMOs should continuously analyze their communication through the social media. They have to know exactly what they are communicating through these powerful new channels and to check if they are really communicating the intended identity and brand of the territory. In many occasions, what one wants to communicate is not what is actually transmitted, as the team that designs the brand does not include the social media manager. Moreover, it is necessary for DMOs to compare the destination brand they communicate with the image of the brand perceived by tourists, which they communicate through their personal social media and can have a very strong influence on the future travel intentions of potential visitors. This analysis can give two types of information to DMOs. On the one hand, they may discover which emotional values of the brand are not well communicated because they are not shared by the tourists. On the other hand, they must understand which emotional values are attached by tourists to their destinations and evaluate if it is necessary to add them to the brand or to redesign the brand communication strategy.

The proposed system has several limitations, some of which might be palliated in future versions of the system. The system only analyzes English-language tweets, so it cannot process messages in other languages. It is restricted to the analysis of textual content, so it does not analyze pictures or videos linked to the messages, which may convey a large amount of information about emotional values that the system is not able to capture. The study only considers the adjectives used in the tweet, discarding the rest of the message. As such, the study is not making any complex syntactic analysis of sentences nor any kind of sentiment analysis, so the system would not make any distinction among the tweets 'This is the most romantic city in the world!' and 'This city should definitely not be the destination of any romantic weekend'. As a future research project, it would be interesting to make a more detailed analysis of DMO communication strategies, which complements and enriches the semantic analysis of content undertaken here. In-depth interviews with should be carried out directors of communication of DMOs to find out the communicative guidelines they convey to community managers in detail. At the same time, community managers should also be interviewed to find out their profiles and the communication guidelines they have received from the director. With this qualitative information , it would be possible to better understand the results of this study and to understand the reasons of the limited brand communication through social media.

There are many other lines of future work that are currently beginning to be explored. The researchers are considering the possibility of also analyzing the tweets of the local residents, in order to try to understand their emotional perception of the city. It would also be useful to categorize the tourists on different ways (e.g. depending on their origin), so that a DMO could study the differences on image perception from different kinds of visitors. Another potentially useful line of research would be to explore how the use of nouns (especially named entities) in the tweets may be linked to different types of tourist attraction. It would also be interesting to try to compare the emotional values communicated by the same DMO through different channels (e.g. Twitter, Facebook, official web page).

Acknowledgements

This work was supported by Spain's Ministry of Economy and Competitiveness (Grant id. CSO2012-34824: *Uso e influencia de los social media y la comunicación 2.0 en la toma de decisiones turísticas y en la imagen de marca de los destinos*).

References

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research*, *34*, 347–356.

Algesheimer, R., Dholakia, U. M., & Herrmann, A. (2005). The social influence of brand community: Evidence from European car clubs. *Journal of Marketing*, *69*, 19–34.

Andéhn, M., Kazeminia, A., Lucarelli, A., & Sevin, E. (2014). User-generated place brand equity on Twitter: The dynamics of brand associations in social media. *Place Branding and Public Diplomacy*, *10*, 132–144.

Anholt, S. (2008). Place branding: Is it marketing , or isn't it? *Place Branding and Public Diplomacy*, *4*, 1–6.

Blain, C., Levy, S. E., & Ritchie, J. B. (2005). Destination branding: Insights and practices from destination management organizations. *Journal of Travel Research*, *43*, 28–338.

Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics*, *32*, 13–47.

Buhalis, D., & Costa, C. (2006). *Tourism management dynamics: Trends, management and tools*. Routledge.

Çakmak, E., & Isaac, R. K. (2012). What destination marketers can learn from their visitors' blogs: An image analysis of bethlehem, palestine. *Journal of Destination Marketing & Management*, *1*, 124–133.

Chaykina, T., Guerreiro, M., & Mendes, J. (2014). Destination brand personality of Portugal for the Russian-speaking market. *Journal of Spatial and Organizational Dynamics*, *2*, 23–40.

Dickinger, A., & Lalicic, L. (2015). 'This city is absolutely fun and trendy': A destination brand personality analysis in a Web 2.0 setting. In *Information and communication technologies in tourism 2015* (pp. 321–333). Lugano, Switzerland: Springer Science + Business Media.

Ekinci, Y., & Hosany, S. (2006). Destination personality: An application of brand personality to tourism destinations. *Journal of Travel Research*, *45*, 127–139.

Fotis, J., Buhalis, D., & Rossides, N. (2012). *Social media use and impact during the holiday travel planning process*. Springer-Verlag.
Gensler, S., Vlckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing brands in the social media environment. *Journal of Interactive Marketing*, *27*, 242–256.
Govers, R., Go, F. M., & Kumar, K. (2007). Promoting tourism destination image. *Journal of Travel Research*, *46*, 15–23.
Guerrero-Solé, F., & Fernández-Cavia, J. (2014). Activity and influence of destination brands on Twitter: A comparative study of nine Spanish destinations. In *Information and communication technologies in tourism 2014* (pp. 227–236). Dublin: Springer Science+Business Media.
Guerrero-Solé, F., & Fernández-Cavia, J. (2015). @ Spain is different: Co-branding strategies between Spanish national and regional DMOs on Twitter. In *Information and communication technologies in tourism 2015* (pp. 309–320). Lugano, Switzerland: Springer Science+Business Media.
Hays, S., Page, S. J., & Buhalis, D. (2013). Social media as a destination marketing tool: Its use by national tourism organisations. *Current Issues in Tourism*, *16*, 211–239.
Henderson, J. C. (2000). Selling places: *The new Asia-Singapore brand*. *Journal of Tourism Studies*, *1*, 36–44.
Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, *18*, 38–52.
Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, *28*, 149–165.
Huertas, A. (2014). La comunicación de los territorios, los destinos y sus marcas. Guía práctica de aplicación desde las relaciones públicas. In *Information and communication technologies in tourism 2015*. Dublin: UOC Ediciones/Colección DIRCOM.
Huertas, A., & Mariné-Roig, E. (2015). Destination brand communication through the social media: What contents trigger most reactions of users? In *Information and communication technologies in tourism 2015* (pp. 295–308). Lugano, Switzerland: Springer Science+Business Media.
Huertas, A., & Mariné-Roig, E. (2016). Differential destination content communication strategies through multiple social media. In *Information and communication technologies in tourism 2016* (pp. 239–252). Bilbao, Spain: Springer Science+Business Media.
Huertas, A., Setó-Pàmies, D., & Míguez-González, M.-I. (2015). Comunicación de destinos turísticos a través de los medios sociales. *El Profesional De La Información*, *24*, 15–21.
Hvass, K. A., & Munar, A. M. (2012). The takeoff of social media in tourism. *Journal of Vacation Marketing*, *18*, 93–103.
Inversini, A., & Buhalis, D. (2009). Information convergence in the long tail: The case of tourism destination information. In *Information and communication technologies in tourism 2009* (pp. 381–392). Amsterdam, The Netherlands: Springer Science+Business Media.
Kim, S.-B., Kim, D.-Y., & Bolls, P. (2014). Tourist mental-imagery processing: Attention and arousal. *Annals of Tourism Research*, *45*, 63–76.
Költringer, C., & Dickinger, A. (2015). Analyzing destination branding and image from online sources: A web content mining approach. *Journal of Business Research*, *68*, 1836–1843.
Labrecque, L. I. (2014). Fostering consumer relationships in social media environments: The role of parasocial interaction. *Journal of Interactive Marketing*, *28*, 134–148.
Lai, C., Li, X. R., & Harrill, R. (2013). Chinese outbound tourists' perceived constraints to visiting the United States. *Tourism Management*, *37*, 136–146.
Laroche, M., Habibi, M. R., & Richard, M.-O. (2013). To be or not to be in social media: How brand loyalty is affected by social media? *International Journal of Information Management*, *33*, 76–82.
Lee, W., & Gretzel, U. (2012). Designing persuasive destination websites: A mental imagery processing perspective. *Tourism Management*, *33*, 1270–1280.
Lin, D. (1998). An information-theoretic definition of similarity. In Proceedings of the fifteenth international conference on machine learning (ICML), Vol. 98, (pp. 296–304). Madison, Wisconsin, USA, July 24-27, 1998.
Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*, 458–468.
Mack, R. W., Blose, J. E., & Pan, B. (2008). Believe it or not: Credibility of blogs in tourism. *Journal of Vacation Marketing*, *14*, 133–144.
Macnamara, J., & Zerfass, A. (2012). Social media communication in organizations: The challenges of balancing openness, strategy, and management. *International Journal of Strategic Communication*, *6*, 287–308.
Marchiori, E., & Cantoni, L. (2011). The online reputation construct: Does it matter for the tourism domain? a literature review on destinations' online reputation. *Information Technology & Tourism*, *13*, 139–159.
Marchiori, E., & Önder, I. (2015). Reframing the image of a destination: A pre–post study on social media exposure. In *Information and communication technologies in tourism 2015* (pp. 335–347). Lugano, Switzerland: Springer Science+Business Media.
Mariné-Roig, E. (2013). From the projected to the transmitted image: The 2.0 construction of tourist destination image and identity in Catalonia (Ph.D. thesis). Universitat Rovira i Virgili.
Mariné-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, *4*, 162–172.
Mariné-Roig, E., & Clavé, S. A. (2016). Destination image gaps between official tourism Websites and user-generated content. In *Information and communication technologies in tourism 2016* (pp. 253–265). Bilbao, Spain: Springer Science+Business Media.
Míguez-González, M. I., & Huertas, A. (2015). The power of photographs in the communication and public relations of tourist destinations and their brands through Facebook and Flickr. *Catalan Journal of Communication & Cultural Studies*, *7*, 197–215.
Morgan, N., & Pritchard, A. (2004). Meeting the destination branding challenge. *Destination Branding* (pp. 59–79), 59–79.
Morgan, N., Pritchard, A., & Piggott, R. (2002). New Zealand, 100% pure. The creation of a powerful niche destination brand. *Journal of Brand Management*, *9*, 335–354.
Morgan, N. J., Pritchard, A., & Piggott, R. (2003). Destination branding and the role of the stakeholders: The case of New Zealand. *Journal of Vacation Marketing*, *9*, 285–299.
Moya, M. D., & Jain, R. (2013). When tourists are your 'friends': Exploring the brand personality of Mexico and Brazil on Facebook. *Public Relations Review*, *39*, 23–29.
Oliveira, E., & Panyik, E. (2014). Content, context and co-creation: Digital challenges in destination branding with references to Portugal as a tourist destination. *Journal of Vacation Marketing*, *21*, 53–74.
Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel blogs and the implications for destination marketing. *Journal of Travel Research*, *46*, 35–45.
Park, H., & Kim, Y.-K. (2014). The role of social network websites in the consumer-brand relationship. *Journal of Retailing and Consumer Services*, *21*, 460–467.
Roberts, C. W. (2000). A conceptual framework for quantitative text analysis. *Quality and Quantity*, *34*, 259–274.
Sánchez, D., Batet, M., Isern, D., & Valls, A. (2012). Ontology-based semantic similarity: A new feature-based approach. *Expert Systems with Applications*, *39*, 7718–7728.
Schmallegger, D., & Carson, D. (2009). Destination image projection on consumer-generated content websites: A case study of the Flinders Ranges. *Information Technology & Tourism*, *11*, 111–127.
Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, *80*, 159–169.
Serna, A., Marchiori, E., Gerrikagoitia, J. K., Alzua-Sorzabal, A., & Cantoni, L. (2015). An auto-coding process for testing the cognitive-affective and conative model of destination image. In *Information and communication technologies in tourism 2015* (pp. 111–123). Lugano, Switzerland: Springer Science + Business Media.
Sevin, H. E. (2014). Understanding cities through city brands: City branding as a social and semantic network. *Cities*, *38*, 47–56.
Slimani, T. (2013). Description and evaluation of semantic similarity measures approaches. *International Journal of Computer Applications*, *80*, 25–33.
Stepchenkova, S., Kirilenko, A. P., & Morrison, A. M. (2009). Facilitating content analysis in tourism research. *Journal of Travel Research*, *47*, 454–469.
Stepchenkova, S., & Morrison, A. M. (2006). The destination image of Russia: From the online induced perspective. *Tourism Management*, *27*, 943–956.
Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management*, *36*, 590–601.
Valentini, C. (2015). Is using social media 'good' for the public relations profession? A critical reflection. *Public Relations Review*, *41*, 170–177.
Wigley, S., & Lewis, B. K. (2012). Rules of engagement: Practice what you tweet. *Public Relations Review*, *38*, 165–167.
Wu, Z., & Palmer, M. (1994). Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on association for computational linguistics*. Association for Computational Linguistics (ACL).
Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, *31*, 179–188.
Yoo, K. H., & Gretzel, U. (2010). Antecedents and impacts of trust in travel-related consumer-generated media. *Information Technology & Tourism*, *12*, 139–152.
Zhang, L., Pan, B., Smith, W., & Li, X. R. (2009). Travelers use of online reviews and recommendations: A qualitative study. *Journal of Information Technology & Tourism* (p. 11), 11.
Zhou, L. (2014). Online rural destination images: Tourism and rurality. *Journal of Destination Marketing & Management*, *3*, 227–240.

ARTICLE IN PRESS

Journal of Destination Marketing & Management xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Journal of Destination Marketing & Management

journal homepage: www.elsevier.com/locate/jdmm

Research Paper

Emotional brand communication on Facebook and Twitter: Are DMOs successful?

Lidija Lalicic[a,*], Assumpció Huertas[b], Antonio Moreno[c], Mohammed Jabreel[c]

[a] Department of Tourism and Service Management, MODUL University Vienna, Austria
[b] Departament d'Estudis de Comunicació, Universitat Rovira i Virgili, Spain
[c] ITAKA: Intelligent Technologies for Advanced Knowledge Acquisition, Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Spain

ARTICLE INFO

Keywords:
Brand values
Semantic analysis
DMOs
Social media
Communication

ABSTRACT

The following study analyzes the emotional brand communication of the 10 most popular DMOs in Europe according to TripAdvisor (2017) by observing user responses across Facebook and Twitter. The paper presents a matrix of successful and promising values that DMOs should integrate into their social media communication strategies regarding their destination brands. Furthermore, the study visualizes a set of values that DMOs should either not include or try to avoid when aiming to successfully engage with their users. In addition to the differences between the two platforms, the type of engagement that DMOs should aim for is also examined. Overall, the paper provides DMOs with guidelines on how to effectively communicate their brands using specific emotional brand values on social media.

1. Introduction

Destination management organizations (DMOs) are facing a paradigm shift in how they have to market and communicate their destinations. The conventional function of DMOs as a dominant source of information for tourists has been challenged by the appearance of online communication tools such as Facebook and Twitter (Hays, Page, & Buhalis, 2013; Li, Robinson, & Oriade, 2017). Generally, information, communication and technologies (ICTs) have had a transformational effect on the way that DMOs market and communicate their brands, offerings and destinations (Li et al., 2017). Nowadays, wide virtual visibility is critical to attract potential first-time or inexperienced tourists (Schroeder & Pennington-Gray, 2015). This forces DMOs to expand their online presence beyond their official websites and include platforms like Facebook. In doing so, DMOs can co-produce their content with their users, be pro-active and stay competitive in online spheres (Mariani, Di Felice, & Mura, 2016). However, DMOs tend to utilize social media for traditional marketing proposes or for crisis management (i.e. service failure, natural disaster) and tend to be in the experimental phases of testing best practices for communicating with their users (Chan & Guillet, 2011; Hays et al., 2013; Leung, Schuckert, & Yeung, 2013; Sigala, 2011). Various studies show that DMOs tend to focus on functional elements in their online communication related to their destination (e.g. historic sites, beach) and limit the use of emotional adjectives (Huertas & Marine-Roig, 2016b; Michaelidou, Siamagka, Moraes, & Micevski, 2013; Xiang & Gretzel, 2010). Despite emotions being of utmost importance for experiencing a destination, they also exert the greatest influence on the overall image of a destination and on future behavioral intentions (Bigné, García, & Blas, 2009). Huertas and Marine-Roig (2015) show how the communication of specific emotional brand values on social media generates greater interactivity and, thus, enhances the overall destination brand image. Even though online platforms offer a very supportive environment to communicate the emotional values of a destination brand effectively (Munar & Jacobsen, 2014), DMOs lack strategies to communicate their brand emotional values and/or have distinct communication strategies for different social media platforms (Guerrero-Solé & Fernández-Cavia, 2013; Huertas & Marine-Roig, 2016a, b; Moreno, Jabreel, & Huertas, 2015). Uşakli, Koç, and Sönmez (2017) recently demonstrated the importance of effective online communication and engagement with users on platforms such as Facebook and Twitter, as this can increase the number of tourist arrivals. Thus, from a practical point of view, a better understanding of how to effectively be present across various social media platforms in order to enhance the destination performance is called for.

Seen from a theoretical perspective, new insights are also needed. Despite the growing research on DMOs' use of social media platforms (Uşakli et al., 2017) and DMOs' brand emotions in online spheres (Dickinger & Lalicic, 2016; Jabreel, Moreno, & Huertas, 2017; Moreno et al., 2015), there are no specific studies that analyze how successful

* Corresponding author.
E-mail address: lidija.lalicic@modul.ac.at (L. Lalicic).

https://doi.org/10.1016/j.jdmm.2019.03.004
Received 20 March 2018; Received in revised form 12 February 2019; Accepted 13 March 2019

DMOs are in communicating their respective emotional brand values across online platforms.

This study therefore aims to demonstrate which emotional brand communication strategies are used by DMOs and are most effective in terms of online popularity across Facebook and Twitter (i.e. generating user responses and re-tweets). To do so, DMOs from the top 10 European destinations according to TripAdvisor (2017) were assessed. This study presents a strategic brand communication matrix for both platforms that identifies the most receptive emotional brand values among the audiences of the two platforms as well as the values that failed to generate user engagement. In doing so, the paper provides new insight into designing effective online communication strategies. Furthermore, DMOs are recommended to critically review their online media communication strategies in order to enhance the impact and online popularity.

2. Theoretical framework

2.1. Emotional destination branding and social media communication

The emotional elements of destination brands, their personality or identity, is what differentiates tourist destinations (Ashworth & Kavaratzis, 2009; Blain, Levy, & Ritchie, 2005; Cai, 2002; Chaykina, Guerreiro, & Mendes, 2014; Ekinci & Hosany, 2006; Govers & Go, 2009; Morgan & Pritchard, 2004, 1998). A distinctive destination brand generates favorable associations and influences consumer destination preferences (Morgan, Pritchard, & Pride, 2003). As a result, many studies have focused on emotional branding in the field of tourism (Blain et al., 2005; Govers, Go, & Kumar, 2007; Huertas & Marine-Roig, 2015; Morgan et al.,2003; Morgan & Pritchard, 2004). One explicit task has been used to understand how tourists develop emotional bonds with a destination, namely, destination brand personality. This concept originates from Aaker's (1997) brand personality concept, which derives from the premise of consumers who choose brands that fit with their personal style or even complement their status. Thus, a brand personality allows businesses to create symbolic effects for consumers (Aaker, 1997). For example, a brand personality can be created where consumers describe their brand experience using human characteristics such as 'cool' and 'exciting' (Uşakli & Baloglu, 2011). According to Aaker (1997), there are generally five dimensions, known as competence, excitement, ruggedness, sincerity and sophistication, which also make up the brand personality scale (BPS). This concept can also naturally be applied to destination management. Uşakli and Baloglu (2011) state that a destination personality can be used by consumers as an avenue for self-expression or by DMOs that wish to differentiate themselves from their competitors. As a result, many studies have analyzed the concept of brand personality in the field of tourism. Some of these studies have analyzed the communication of destination brand values through the brand personality scale created by Aaker (1997) (De Moya & Jain, 2013; Ekinci & Hosany, 2006; Hosany, Ekinci, & Uysal, 2007; Pitt, Opoku, Hultman, Abratt, & Spyropoulou, 2007). These studies support the hypothesis that humanizing destination brands can lead to higher levels of brand attachment and purchase intention (Ekinci & Hosany, 2006), clearly illustrating the strategic importance of using emotional brand values in online communications.

Looking into studies that analyze the communication of destination brands in an online setting, two different approaches can be seen. The first stream of studies demonstrates the importance of social media for the communication of destination brands. For example, some studies have analyzed how social media influences the configuration of a destination brand image among users and the relationships that they create with brands (Govers et al., 2007; Laroche, Habibi, & Richard, 2013). Laroche et al. (2013) demonstrate that if users engage in a brand community, it not only has positive effects on their relationship with the brand, but also leads to higher levels of brand credibility and brand loyalty. Algesheimer, Dholakia, and Herrmann (2005) have also shown that the active participation of users on social media increases their emotional attachment to the brand.

The second stream of studies analyzes the DMOs' brand communication and user-generated content (UGC) that communicates the images that tourists create of destination brands. Some of these studies have analyzed how tourist destinations communicate their brands, especially the emotional values of the brand on social media (Huertas & Marine-Roig, 2016a; Moreno et al., 2015). De Moya and Jain (2013) explored how two tourist destinations (Mexico and Brazil) communicated their brand personality through Facebook and which personality traits their Facebook fans associated with the brands. The results showed that Mexico and Brazil communicated distinctive brand personalities on each country's official Facebook page with leading brand values of 'sincerity' and 'excitement'. This is in line with the findings of previous studies (Ekinci & Hosany, 2006; Jain & Chan-Olmsted, 2009). These results also coincide with the later research of Huertas and Marine-Roig (2016b). In their study, regardless of the type of destination, brand values related to traditional brand values such as 'honesty' and 'sincerity', as well as 'excitement', 'spirited' and 'imaginative', were found to be the most popular.

Various studies have also analyzed whether the DMOs' emotional brand communication aligns with UGC about the image of the brand. For example, De Moya and Jain (2013) demonstrate that the emotional brand values communicated by the DMO of Mexico coincided with the messages posted by its users. However, Brazil's promotional messages did not coincide with the messages posted by its users. Stepchenkova and Zhan (2013) also analyzed and compared DMOs photo content and UGC in the case of Peru and found differences among them. For example, travelers were more interested in how Peruvian people live their everyday lives, whereas the DMO focused on promoting Peruvian traditional artifacts. Dickinger and Lalicic (2016) demonstrate how the concept of destination brand personality and emotions differ across tourist services in Tripadvisor reviews. Marine-Roig (2017) also analyzes tourist destination image in user reviews on Tripadvisor by using a designative-appraisive image dichotomy similar to the dichotomy cognitive-affective. However, the designative-appraisive image dichotomy reveals the preferences and emotional opinions of users more effectively (Marine-Roig, 2017). Moreno el al. (2015) demonstrate how European DMOs communicate on Twitter and they show that DMOs often: (1) do not use specific adjectives to communicate their identity, (2) tend to use common and generic brand communications and, in so doing, they do not distinguish or identify themselves, and (3) do not have a coherent communicative strategy. Huertas and Marine-Roig (2016b) demonstrate how DMOs communicate their brands through different social media platforms. In keeping with a previous study (Moreno et al., 2015), the values were transmitted in a rather similar way and tangible elements were used more frequently in their communication, regardless of the destination matrix (Bigné et al., 2009; Michaelido et al., 2013). However, Huertas and Marine-Roig (2016b) did not find any remarkable differences in the communication of emotional brand values across the different social media platforms. Hence, no specific emotional branding strategies employed by the DMOs were identified.

As a result of these findings, the following challenges ca be seen: (1) there is a difference between what DMOs communicate about their destination brands and what tourists communicate (De Moya & Jain, 2013; Stepchenkova & Zhan, 2013), (2) DMOs do not have distinct strategies to communicate their emotional brand values (Huertas & Marine-Roig, 2016b; Moreno et al., 2015), and (3) DMOs do not have different communication brand strategies for different social media platforms in place (Huertas & Marine-Roig, 2016b). Given the dominant focus on the communication of functional and tangible elements, it would appear that emotional values are not given sufficient importance in brand communication strategies, (Bigné et al., 2009; Huertas & Marine-Roig, 2016b; Michaelidou et al., 2013). In order to achieve successful emotional brand communication, destinations should

transmit a unified identity as well as communicate those brand emotional values that also appeal to tourists' emotions (Morgan et al.,2003). Consumer engagement literature discusses in greater detail how one can measure the success of user reactions to emotional branding content. The next section will explain this in greater detail.

2.2. User engagement and successful brand communication

Successful brand communication should trigger engagement and reactions from users (Uşakli et al., 2017). On social media platforms, engagement is rendered rather transparent by the likes, shares, retweets and comments made by users (Kabadayi & Price, 2014; Oviedo-García, Muñoz-Expósito, Castellanos-Verdugo, & Sancho-Mejías, 2014). Social media therefore offers destination brands an opportunity to act as relational tools, wherein a dialogue with their users can be enacted (Lovejoy & Saxton, 2012; Saffer, Sommerfeldt, & Taylor, 2013). Thus, as Van Doorn et al. (2010) has shown, engagement with the brand can also lead to other types of relationships that go beyond merely purchasing the brand's offerings. For example, user engagement has positive effects on brand involvement and decision-making processes as well as brand satisfaction (Nusair, Bilgihan, & Okumus, 2013; Walther & Jang, 2012). Therefore, measuring online brand engagement is an important indicator of: (1) a brand's success, (2) how to generate a better destination brand image (Munar, 2011), and (3) how to maximize added value (Buhalis & Law, 2008). Indeed, a recent study from Uşakli et al. (2017) shows that positive online user engagement has a positive effect on a destination's international tourist arrivals.

Despite the importance of interaction and engagement for a destination's success, most of the previous studies suggest that DMOs generate very few interactions with users and achieve only minimal user reactions on social media (Chan & Guillet, 2011; Guillet, Kucukusta, & Liu, 2016; Wattanacharoensil & Schuckert, 2015). However, DMOs are still not using social media as a customer-service tool or engaging with their users (Hays et al., 2013; Uşakli et al., 2017). Indeed, DMOs tend to have different uses for various platforms. For example, DMOs tend to use Twitter more interactively than Instagram, Facebook and YouTube (Uşakli et al., 2017). There is limited research on whether brand communication generates more reactions (Huertas & Marine-Roig, 2016a). Of these handful of studies, Huertas and Marine-Roig (2016a) show that there is a lack of congruency between the most frequently communicated content and emotional brand values shared by DMOs and those that trigger the most reactions among users. Given the popularity of social media platforms and the ability for users to directly engage with brands on them, destination brands should use this opportunity. Based on the preceding discussion, this study aims to enrich the current studies and bolster practitioners in understanding which brand communication strategies strengthen a destination image and/or emotional brand values in order to increase tourist arrivals. Success measures will be highlighted by analyzing the most popular destination brands, which are mature and demonstrate a measurable history of brand communication. Furthermore, analysis will be performed to observe the different engagement behaviors across Facebook and Twitter, the two most popular social media platforms. Finally, DMOs will be given guidelines as to which content and emotional values trigger the greatest consumer engagement, thus, leading to the successful management of their social media outlets in order to promote their destination by engaging on an emotional level with their users.

3. Method

3.1. Sample selection

At first, the top 25 European destinations in 2017 according to TripAdvisor were considered (TripAdvisor, 2017). However, after a manual search across the official Twitter accounts of the DMOs, only those that sent at least 3000 English tweets in 2016 (a minimum average of eight to nine daily tweets) and also had an active Facebook account with English posts were included in the final sample. These conditions enabled a selection of a set of destinations that actively use these two social networks in English. This resulted in a final sample of the following 10 destinations: *Amsterdam, Barcelona, Berlin, Budapest, Dublin, Edinburgh, London, Madeira, Paris and Tenerife.* As in previous studies (Lu & Stepchenkova, 2014), the collection of the tweets and Facebook posts was supported by two tools developed by the authors. The system twiQuery (https://github.com/mhjabreel/twiQuery) was used to collect the tweets to be analyzed. It is a crawler that enables users and developers to make advanced search actions on tweets, such as retrieving sets of tweets posted by a specific user, written in a specific language, sent from a certain area (determined by the name of a city or a geolocation), posted during a specific period of time, containing a given string or hashtag, etc. The Facebook posts were retrieved by using another self-developed tool (not available to public) that accesses the DMO's Facebook official application program interface (API). It consists of the messages a company posts on its timeline, their type (video, photo, etc) and the interaction of their users (i.e. the written comments and the emotions tags Facebook makes available). All of the English tweets and posts in 2016 were retrieved and analyzed for the purpose of this paper.

3.2. Tweets and Facebook post analysis

The analysis of the tweets and posts went through various steps. First, the tweets/posts went through an initial standard pre-processing stage (Hofmann & Klinkerberg, 2013). Elongated words were also automatically corrected.

After that, the content of the tweets/posts was analyzed. For the purpose of this paper, a slight adjustment of Aaker's brand personality scale (Aaker, 1997) was made in order to apply it to destination brands. The few items that were not applicable to destination brands were removed. In this case, the following list of five dimensions (and their 54 associated emotional values) was considered:

- *Sincerity:* family-oriented, down-to-earth, calm, real, traditional, honest, original, wholesome, quality of life, happiness, sentimental, friendly.
- *Excitement:* trendy, daring, exciting, exotic, fashionable, cool, spirited, dynamic, vital, fresh, young, unique, imaginative, creative, up-to-date, independent, contemporary, tolerant, hospitable.
- *Competence:* reliable, hard-working, safe, rigorous, intelligent, technical, corporate, innovative, successful, leader, ambitious, powerful.
- *Sophistication:* glamorous, luxurious, charming, smooth, romantic, magical.
- *Ruggedness*: get-away, recreational, tough, rugged, non-conformist.

In the system that was developed, each *adjective* used in the tweets was compared semantically (i.e. at the conceptual level, not at the syntactic level) with all of the categories of emotional values. Adjectives were chosen because, as shown in previous studies on content analysis of the communication of destination brands (e.g. Lu & Stepchenkova, 2014), they are the words that tend to convey the greatest emotional responses. A standard natural language parser was applied to retrieve the adjectives. The Wu-Palmer ontology-based semantic similarity measure (Wu and Palmer, 1994) was used to check the similarity between adjectives and emotional values in WordNet. The Wu-Palmer similarity between two terms c1 and c2 is defined as

$$\mathrm{simWP}(c1, c2) = (2 \times N3) / (N1 + N2 + 2 \times N3)$$

Where N1 and N2 are the number of hypernym links from the terms c1 and c2 to their Least Common Subsumer (LCS) in WordNet and N3 is the number of hypernym links from the LCS to the root of the ontology.

Table 1
Emotional brand values ranked according to frequency of DMOs' tweet communication.

Very often used		Often used		Average usage		Hardly used	
Honest	655	Sentimental	126	Spirited	45	Ambitious	10
Fresh	500	Unique	125	Intelligent	45	Exotic	9
Glamorous	462	Independent	121	Exciting	37	Corporate	9
Rugged	272	Traditional	117	Technical	32	Non-conformist	4
Down-to-earth	262	Real	113	Recreational	32	Hard-working	3
Calm	261	Innovative	111	Dynamic	30	Vital	1
Original	211	Magical	89	Up-to-date	30	Imaginative	1
Happiness	206	Daring	71	Family-oriented	25	Tolerant	1
Creative	199	Friendly	67	Tough	23	Rigorous	1
Contemporary	194	Trendy	60	Luxurious	21	Romantic	1
Leader	187	Hospitable	56	Successful	20	Quality of life	0
Get-away	170	Powerful	55	Young	19	Fashionable	0
Smooth	157			Wholesome	14	Cool	0
Safe	156			Charming	14	Reliable	0

Table 2
Emotional brand values ranked according to frequency of re-tweets.

Highly re-tweeted		Often re-tweeted		Average re-tweeted		Hardly re-tweeted	
Happiness	72,656	Powerful	1102	Ambitious	96	Tolerant	8
Real	71,738	Daring	1054	Exotic	69	Hard-working	8
Sentimental	71,208	Friendly	912	Imaginative	64	Vital	4
Honest	12,142	Spirited	832	Wholesome	54	Quality of life	0
Glamorous	9124	Trendy	830	Romantic	45	Fashionable	0
Fresh	7919	Dynamic	760	Non-conformist	30	Cool	0
Original	7623	Magical	708			Reliable	0
Leader	6851	Recreational	652			Rigorous	0
Down-to-earth	6157	Up-to-date	526				
Creative	4729	Exciting	495				
Get-away	3749	Tough	474				
Smooth	3658	Charming	459				
Rugged	3417	Intelligent	418				
Calm	3348	Successful	370				
Contemporary	2802	Luxurious	308				
Safe	2295	Hospitable	305				
Independent	2185	Family-oriented	299				
Innovative	1906	Young	277				
Traditional	1803	Corporate	110				
Unique	1797						
Technical	1521						

Table 3
Emotional brand values ranked according to frequency of 'favorites'.

Very Popular		Popular		Rather Popular		Not Popular	
Honest	20,344	Successful	967	Corporate	132	Non-conformist	16
Glamorous	19,521	Charming	935	Wholesome	125	Hard-working	14
Fresh	11,784	Creative	858	Ambitious	111	Tolerant	10
Down-to-earth	8526	Recreational	807	Romantic	93	Vital	4
Original	5806	Exciting	775	Exotic	45	Rigorous	1
Happiness	5658	Tough	752			Quality of life	0
Calm	4858	Young	683			Fashionable	0
Get-away	4409	Technical	527			Cool	0
Safe	3955	Hospitable	518			Imaginative	0
Independent	3903	Rugged	515			Reliable	0
Innovative	3641	Up-to-date	505				
Leader	3433	Smooth	499				
Sentimental	3171	Luxurious	490				
Contemporary	3067	Intelligent	451				
Traditional	2787	Family-oriented	435				
Real	2637	Dynamic	411				
Powerful	1994	Unique	244				
Spirited	1512						
Trendy	1422						
Friendly	1369						
Daring	1284						
Magical	1205						

Table 4
Ratio re-tweet and the number of emotional brand values communicated (ranked).

Very Popular		Popular		Rather Popular		Not Popular	
Real	634.85	Imaginative	64.00	Daring	14.85	Hospitable	5.45
Sentimental	565.14	Technical	47.53	Safe	14.71	Vital	4.00
Happiness	352.70	Romantic	45.00	Luxurious	14.67	Wholesome	3.86
		Leader	36.64	Young	14.58	Hard-working	2.67
		Original	36.13	Contemporary	14.44	Quality of life	0
		Charming	32.79	Unique	14.38	Fashionable	0
		Dynamic	25.33	Trendy	13.83	Cool	0
		Creative	23.76	Friendly	13.61	Reliable	0
		Down-to-earth	23.50	Exciting	13.38	Rigorous	0
		Smooth	23.30	Calm	12.83		
		Get-away	22.05	Rugged	12.56		
		Tough	20.61	Corporate	12.22		
		Recreational	20.38	Family-oriented	11.96		
		Powerful	20.04	Ambitious	9.60		
		Glamorous	19.75	Intelligent	9.29		
		Honest	18.54	Tolerant	8.00		
		Successful	18.50	Magical	7.96		
		Spirited	18.49	Exotic	7.67		
		Independent	18.06	Non-conformist	7.50		
		Up-to-date	17.53				
		Innovative	17.17				
		Fresh	15.84				
		Traditional	15.41				

Table 5
Ratio re-tweet and the number of emotional brand value adjectives communicated (ranked).

Very Popular		Popular		Rather Popular		Not Popular	
Romantic	93.00	Get-away	25.94	Unique	19.52	Tolerant	10.00
Charming	66.79	Safe	25.35	Rugged	18.93	Hospitable	9.25
Successful	48.35	Recreational	25.22	Calm	18.61	Wholesome	8.93
Creative	43.12	Sentimental	25.17	Leader	18.36	Exotic	5.00
Glamorous	42.25	Traditional	23.82	Daring	18.08	Hard-working	4.67
Powerful	36.25	Trendy	23.70	Family-oriented	17.40	Vital	4.00
Young	35.95	Fresh	23.57	Up-to-date	16.83	Non-conformist	4.00
Spirited	33.60	Real	23.34	Technical	16.47	Rigorous	1.00
Innovative	32.80	Luxurious	23.33	Contemporary	15.81	Fashionable	–
Tough	32.70	Friendly	20.43	Corporate	14.67	Cool	–
Down-to-earth	32.54	Exciting	20.95	Dynamic	13.70	Quality of life	–
Independent	32.26			Magical	13.54	Imaginative	–
Smooth	31.78			Ambitious	11.10	Reliable	–
Honest	31.06			Intelligent	10.02		
Original	27.52						
Hapyness	27.47						

This measure ranges from 1 (for identical concepts) to 0 (when the LCS is the root of the ontology). In order to apply this measure, the terms being compared must be *nouns*. Thus, both the emotional values and the adjectives had to be transformed into nouns. In the case of the emotional values, the translation was made manually (e.g. 'ambitious' was transformed into 'ambition'). The adjectives appearing in the tweets/posts were automatically transformed into nouns by using their derivative form in WordNet. After this transformation step, it was possible to compare the emotional values and the tweet/post adjectives to assess if the destination communicates certain emotional values. Only adjectives with a similarity higher than 0.7 to an emotional value (*emotional adjectives*) were considered in the final results of the analysis.

4. Results

4.1. DMOs' emotional brand communication and user reactions - Twitter

Firstly, the emotional brand values communicated by the DMOs on their official Twitter accounts were investigated. Table 1 demonstrates the number of tweets that contain emotional brand values. Categorizing the values according to their use, it can be seen that there are 14 values that have been used very often by DMOs on Twitter. The words 'honest' (655), 'fresh' (500) and 'glamorous' (462) are the most often used words in the DMOs' tweets. There are also a set of words that DMOs do not use at all. These hardly used words include 'reliable', 'cool', 'fashionable' and 'quality of life'. This already gives a first impression of which values DMOs tend to include in their communication on Twitter and, thus, that their users should associate with their destinations.

However, of greater interest is how often DMOs' tweets including specific emotional brand values are re-tweeted by their users on their personal social networks. Table 2 provides an overview ranking the most versus the least re-tweeted emotional brand values. Here we can see a different set of top three brand emotions that users highly re-tweet; 'happiness' (72,656 re-tweets), 'real' (71,738 re-tweets) and 'sentimental' (71,208 re-tweets). Interestingly, it can be seen that the emotional brand values that DMOs hardly tweet are often also not re-tweeted.

In addition to re-tweeting, users on Twitter can also highlight a tweet from a DMO by clicking on a heart-icon, which represents another form of user engagement on Twitter. Table 3 presents a list of emotional values ranked according to the number of 'favorites' marked by the users. In comparison to Table 2, it can be seen that there is a slightly different ranking. For example, in this case, the top three emotional brand values are: 'honest' (20,344), 'glamorous' (19,521) and

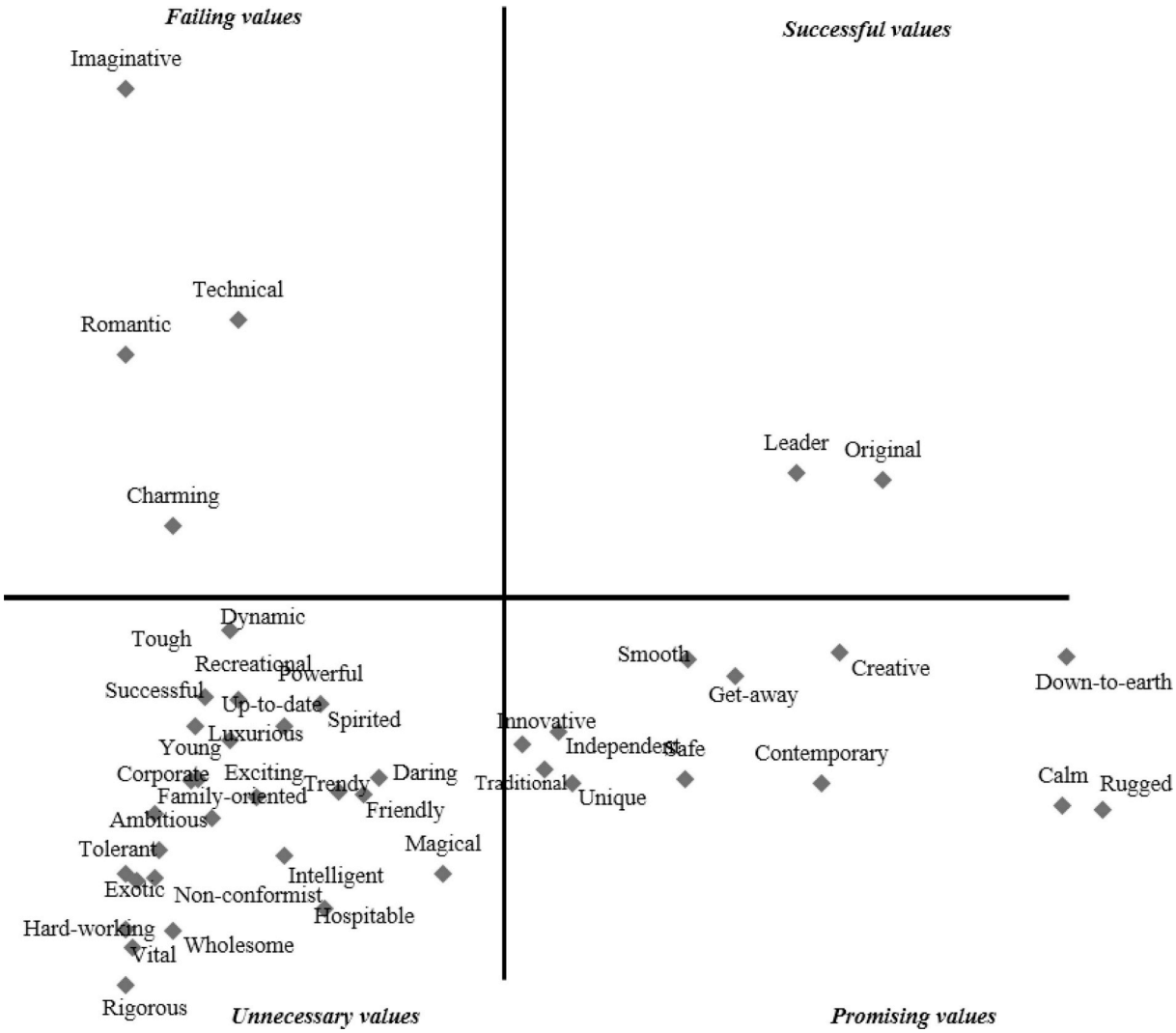


Fig. 1. Matrix of DMOs' emotional brand values in Twitter - based on re-tweets.

'fresh' (11,784). This is in line with what DMOs communicate in their tweets (see Table 1), whereas the average and hardly favorited brand emotional values are similar to the re-tweeted ones.

This study is also interested in the ratio between re-tweets and the number of emotional values as this gives us a better understanding of which values have the highest impact in terms of being the most popular in a setting like Twitter. Tables 4 and 5 show the ranking of the most popular emotional brand values based on the ratio of re-tweets and favorites. In fact, for the re-tweets ratio, three values stand out: 'real' (634.85), 'sentimental' (565.14) and 'happiness' (352.70). Thus, the popularity of these three emotional brand values for Twitter users is clearly indicated. Values such as 'hard-working', 'wholesome', 'vital' and 'hospitable' appear with a ratio lower than 10, which suggests that they are not well received by the audience. In observing Table 5, a different spread of emotional brand values can be seen in relation to the number of users marking the DMOs' message as 'favorite'. The top emotional brand values in this case are 'romantic', 'charming' and 'successful'.

In order to understand what makes a message with a specific emotional brand value successful in reaching a large audience, a matrix was developed that can support DMOs in designing successful emotional brand value communication strategies. One axis signifies the DMOs' degree of communicating their emotional brand values, which ranges from a high degree to a low degree. The other axis represents the impact in terms of engaging with users on the platforms and exhibits a similar range from low to high impact. Depending on the platform, the impact is measured differently based on the number of re-tweets, favorites and likes. The ranges of the scales depend on the categories presented in the tables. Furthermore, extreme values are taken into account and removed, allowing us to scale the remaining values This results in four dimensions: (1) *success values,* in this case the DMOs intensively communicate these emotional values and users highly engage with these messages, (2) *failing values,* the DMOs have intensively used these emotional values in their social media communication, but users do not really engage with these messages, (3) *unnecessary values*, in this case, the DMOs hardly use these values and users also tend to hardly engage with messages that contain these values, and (4) *promising values,* DMOs hardly use these emotional values, but users engage with the message. The values closer to the middle should be taken into consideration, however, no specific focus should be put on them in order to boost a DMO's online impact.

Fig. 1 is based on the impact stemming from the number of re-tweets. It can be seen that a few outstanding values that are successful. In this case, DMOs were successful when using the words 'leader' and 'original'. DMOs used a lot of words related to the words 'romantic', 'imaginative' and 'charming', however, users did not respond to these messages and effective user engagement was not achieved. There are quite a lot of words that fall into the unnecessary values category, thus, DMOs need to be aware of preventing or trying not to use these words as they do not trigger the responses needed in order to create a higher

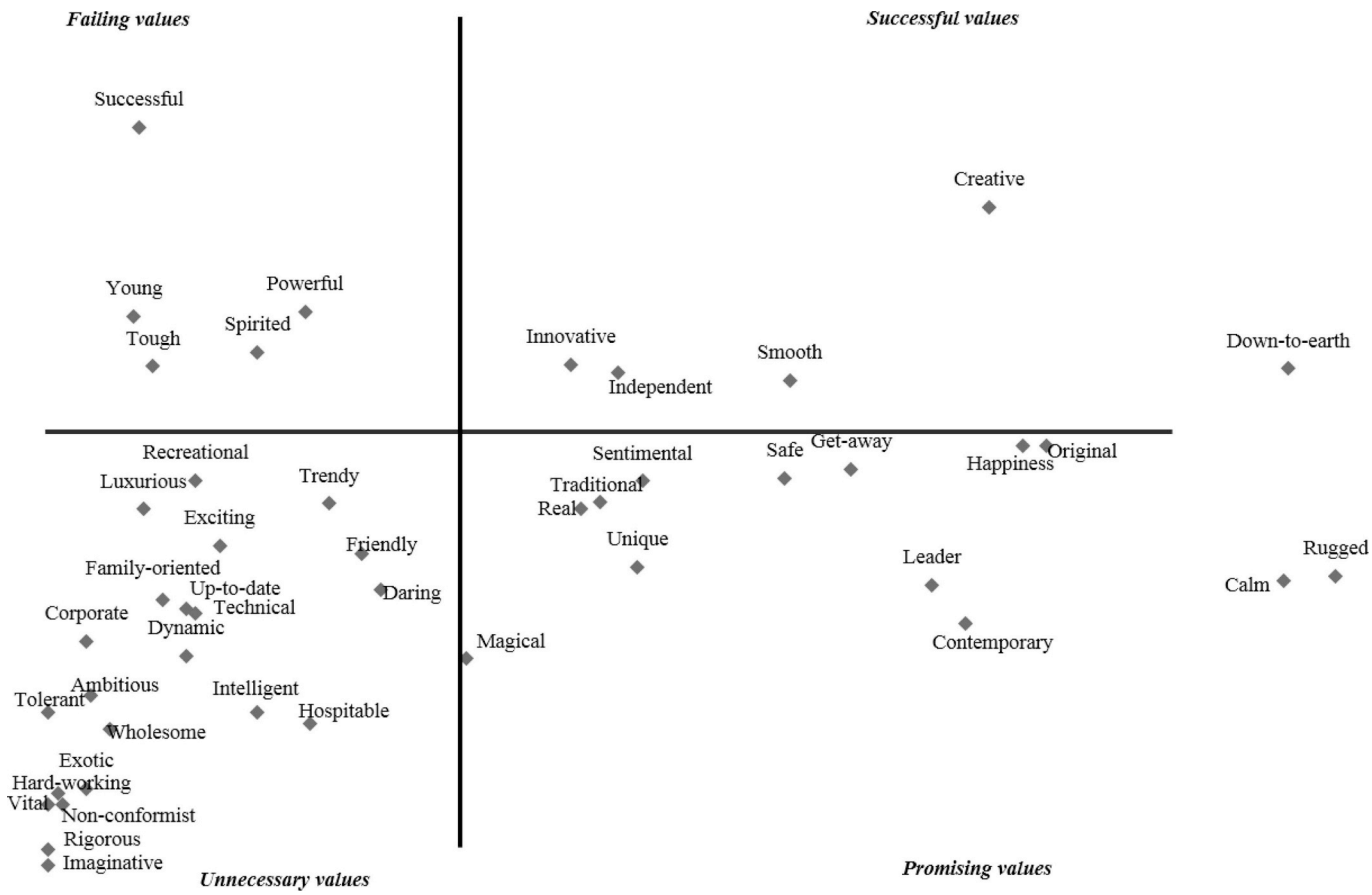


Fig. 2. Matrix of DMOs' emotional brand values Twitter - based on 'favorite'.

Table 6
Emotional brand values ranked according to frequency of DMOs (Facebook communication).

Very Often Used		Often Used		Average Usage		Hardly Used	
Honest	574	Daring	95	Powerful	44	Hospitable	9
Glamorous	258	Happiness	93	Magical	43	Dynamic	6
Down-to-earth	194	Get-away	86	Smooth	40	Wholesome	4
Fresh	170	Independent	82	Intelligent	33	Young	4
Rugged	138	Traditional	77	Tough	26	Charming	4
Leader	124	Real	66	Technical	25	Exotic	3
Safe	116	Sentimental	59	Recreational	24	Corporate	3
Original	111	Creative	58	Friendly	22	Vital	2
Unique	106			Trendy	22	Tolerant	1
Calm	105			Exciting	22	Rigorous	1
Contemporary	103			Spirited	18	Ambitious	1
Innovative	101			Luxurious	18	Romantic	1
				Up-to-date	16	Family-oriented	0
				Successful	12	Quality of life	0
						Fashionable	0
						Cool	0
						Imaginative	0
						Reliable	0
						Hard-working	0
						Non-conformist	0

level of impact online. Failing values are 'imaginative', 'technical', 'romantic' and 'charming', as they did not receive the same number of re-tweets compared to how often DMOs communicated them. However, as seen in the right lower corner of the matrix, DMOs have a set of highly promising values that users highly engage with such as 'calm', 'rugged', 'creative' and 'down-to-earth'. Consequently, they should integrate these highly promising values much more often into their tweets.

A different distribution of the emotional values (see Fig. 2) can be seen in a matrix that maps the impact stemming from users marking a DMO's tweet as a 'favorite'. The success values here relate to the word 'creative' followed by 'smooth', 'innovative' and 'independent'. One outstanding failing value is 'successful', followed by 'powerful', 'young', 'tough' and 'spirited'. Interestingly, in this case, we have 'contemporary' and 'magical' as promising values that create relatively high impact in terms of users marking the tweet as a 'favorite', whereas re-tweeting these values either are uncessary or, in the case of 'contemporary', is a promising value. In this case, there are also a number of values that fall into the unnecessary value category, in particular, 'imaginative'and 'rigorous' stand out.

4.2. DMOs' emotional brand communication and user reactions - Facebook

In this section, Facebook is analyzed on the basis of DMOs' communication strategies and user reactions. As seen in Table 6, there are two outstanding emotional brand values used in Facebook communication, namely, 'honest' and 'glamourous', followed by 'down-to-earth', 'fresh' and 'rugged'. Interestingly, eight values are not integrated at all, such as 'cool', 'non-confirmation' and 'hard-working'. The rest of the values exhibit more-or-less equal usage in DMOs' Facebook communication.

Table 7 illustrates which Facebook posts containing specific emotional values received the highest number of likes. In this case, it can be seen that 'honest' is the top runner, followed by 'glamorous' and 'original'. This aligns with Table 7 in terms of the frequency of use of these values. Other highly used values are 'original', 'down-to-earth', 'fresh', 'unique', 'contemporary', 'safe', 'innovative', 'rugged', 'leader' and 'calm'. A set of values can also be seen that are hardly communicated such as 'sentimental', 'get-away', 'wholesome' or 'luxurious'. Also

Table 7
Emotional brand values ranked according to frequency of 'likes'.

Very Popular		Popular		Rather Popular		Not Popular	
Honest	375,858	Leader	46,041	Romantic	6398	Ambitious	63
Glamorous	226,018	Traditional	39,946	Successful	5924	Rigorous	10
Original	102,792	Independent	39,086	Luxurious	5783	Family-oriented	0
Down-to-earth	102,485	Calm	37,936	Hospitable	4906	Quality of life	0
Fresh	83,542	Smooth	35,336	Real	4899	Fashionable	0
Safe	67,855	Technical	34,929	Recreational	4777	Cool	0
Innovative	63,366	Get-away	30,688	Dynamic	3659	Imaginative	0
Happiness	61,963	Powerful	30,476	Corporate	1895	Reliable	0
Rugged	58,844	Creative	27,081	Charming	1721	Hard-working	0
Contemporary	58,592	Magical	25,293	Wholesome	1276		
Daring	50,654	Exciting	22,668	Non-conformist	1065		
Unique	50,514	Intelligent	20,781	Young	709		
		Sentimental	15,201	Exotic	545		
		Trendy	14,078	Tolerant	490		
		Tough	13,589	Vital	230		
		Spirited	13,373				
		Up-to-date	13,126				
		Friendly	11,323				

Table 8
Ratio re-tweet and the number of emotional brand adjectives communicated (ranked).

Very Popular		Popular		Rather Popular		Not Popular	
Romantic	6398.00	Successful	493.67	Leader	371.30	Rigorous	10.00
Technical	1397.16	Fresh	491.42	Calm	361.30	Cool	0
Exciting	1030.36	Tolerant	490.00	Get-away	356.84	Reliable	0
Original	926.05	Independent	476.66	Luxurious	321.28	Non-conformist	0
Smooth	883.40	Unique	476.55	Wholesome	319.00	Family-oriented	0
Glamorous	876.04	Creative	466.91	Sentimental	257.64	Quality of life	0
Up-to-date	820.38	Charming	430.25	Recreational	199.04	Fashionable	0
Spirited	742.94	Rugged	426.41	Exotic	181.67	Imaginative	0
Real	742.27			Young	177.25	Hard-working	0
Powerful	692.64			Vital	115.00		
Happiness	666.27			Ambitious	63.00		
Honest	654.80						
Trendy	639.91						
Corporate	631.67						
Intelligent	629.73						
Innovative	627.39						
Dynamic	609.83						
Magical	588.21						
Safe	584.96						
Contemporary	568.85						
Hospitable	545.11						
Daring	533.20						
Down-to-earth	528.27						
Tough	522.65						
Traditional	518.78						
Friendly	514.68						

notable in Facebook communications is the observation that DMOs do not include values such as 'family-oriented', 'quality of life', 'fashionable', 'cool', 'imaginative', 'reliable', 'hard-working', or 'non-conformist'. However, to understand the ratio between the numbers of posts containing a specific value versus the number of likes for a specific value, Table 8 provides more insight. Here, a different distribution can be seen, wherein the emotional brand value 'romantic' scores highest followed by 'technical' and 'exciting'.

Fig. 3 illustrates the matrix for Facebook communication based on the impact according to the number of likes. In this case, we can see one success value, namely, 'original'. DMOs failed by using a lot of words related to 'technical', 'exciting' and 'smooth' without creating a high level of impact in the Facebook community. However, there is only one relatively promising value, namely, 'leader'. As seen from the unnecessary values category, words such as 'rigorous', 'ambitious', 'young', and 'exotic' are not often used by DMOs and are also not well-received by Facebook users. Thus, such words do not add any strength to a social media message by a DMO and do not contribute to increased engagement with users. There are various words that are in the center of the matrix which implies that DMOs can integrate them, however, they are advised to balance them with more success and promising values in order to enhance the level of impact of their communication messages on Facebook.

5. Conclusion

DMOs need to successfully plan for future online campaigns by integrating emotional brand values into their social media presence in order to generate higher levels of user engagement and enhance the positive image of their destinations (Buhalis & Law, 2008; Munar, 2011; Huertas & Marine-Roig, 2016(a,b)). This study took a critical look at the social media communication of the DMOs that represent the 10 most popular European destinations in 2017 according to TripAdvisor. The study looked specifically at emotional brand communication levels across Facebook and Twitter. Furthermore, in contrast to other studies, this study also integrated the impact of the DMOs' communication

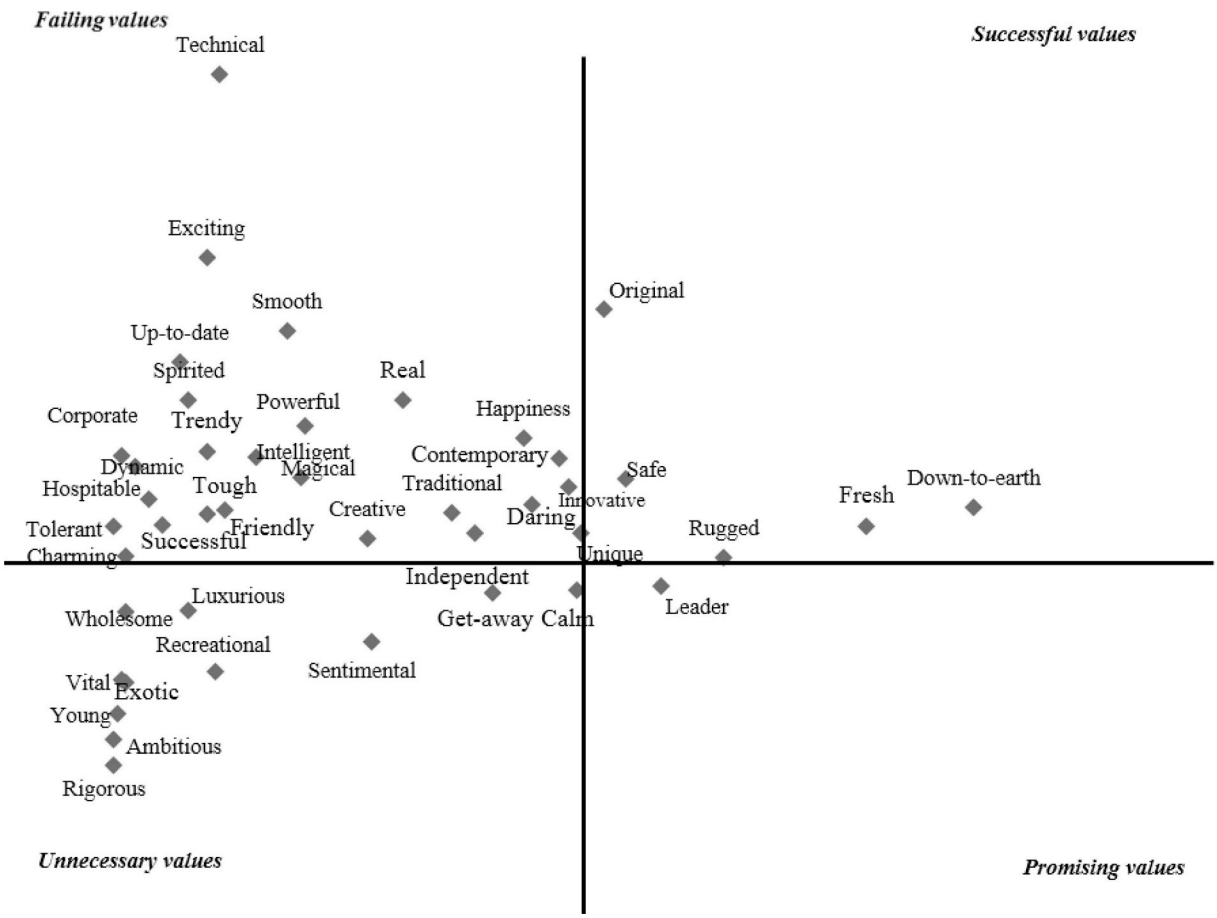


Fig. 3. Matrix of emotional brand values in Facebook (based on 'likes').

strategies by analyzing user reaction metrics across the two platforms. Consequently, a set of significant observations can be listed.

Firstly, DMOs tend to receive a higher level of user reactions (i.e. number of likes) compared to their activity on Twitter. However, the structure of the reactions is similar in both platforms. This finding is in line with studies such as Huertas and Marine-Roig's (2016(a,b)) that shows how Facebook is a very suitable platform for communicating a DMO's emotional brand values as well as generating user engagement.

Secondly, similarities can be seen in the communication of emotional brand values through Facebook and Twitter. For example, the following facts can be stated from the comparison of Table 2 (RTs) and Table 7 (likes): 11 of the 12 most liked emotional values are among the most highly re-tweeted (e.g. *honest, glamorous, fresh*). Six of the nine less popular emotional values in Facebook are hardly re-tweeted (e.g. *hard-working, reliable*). The top six popular emotional values are highly re-tweeted and nine of the remaining 12 popular emotional values in Facebook are often re-tweeted. Four of the six emotional values that are re-tweeted in an average way are rather popular in Facebook. In fact, there are only two out of 54 emotional values for which there is a difference of two columns between Tables 2 and 7: *family-oriented* and *real.* In contrast to other studies (Lalicic, Huertas, Moreno, Gindl, & Jabreel, 2018; Moreno et al., 2015), this study also demonstrates which emotional brand values trigger more user reactions and, thus, engagement. This study shows a different distribution of successful emotional brand values that trigger responses across the two platforms.

Thirdly, the study developed an emotional brand value-impact matrix. The matrix visualizes how DMOs need to design different messages based on a set of emotional brand values in order to reach their target audience and create higher levels of impact online. Furthermore, the study shows how DMOs can, through the use of specific emotional values, trigger different engagement behavior per platform. For example, in the case of Twitter, DMOs need to integrate different emotional values in order to have their messages re-tweeted compared to users who mark a post as 'favorite'.

Fourthly, the matrix demonstrates a set of values that DMOs hardly communicate but users tend to share and like in great volume. Hence, such values are promising values that DMOs are highly advised to integrate into the design of their online communication strategies. The study also identified a number of values that appear to be somewhat popular, hence, DMOs are advised to find a good mix of these values and critically evaluate which values fit their strategies and align with their distinctive destination brand. The study also discovered values that are perceived to be failing even though DMOs tend to heavily communicate their brands with these values despite the lack of user response. Examples of such values are 'exotic', 'vital', 'rigorous' and 'non-conformist'. Furthermore, unnecessary values were identified, which are values that DMOs hardly used and users hardly respond to. Thus, such values should be reconsidered or even eliminated in order to successfully and effectively communicate destination brands on social media. Overall, the suggested matrix devised by this study provides DMOs with a structured overview to: (1) re-think their online communication strategies, (2) only integrate values that matter to them, (3) successfully engage with their users, and (4) create a high online impact with their messages.

5.1. Future research and limitations

The present study demonstrates how DMOs of the 10 most popular European destinations in 2017 according to TripAdvisor communicate to a more or less similar degree certain emotional values in keeping

with previous studies on the subject (Huertas & Marine-Roig, 2016(a,b); Lalicic et al., 2018; Moreno et al., 2015). Future studies should perform in-depth analysis at a destination-level. This would allow a greater understanding of which specific emotions are communicated per destination and how each destination is successful in aligning their online social media communication with their brand identity and destination characteristics. The study attempted to integrate emerging destinations following Uskali et al.'s (2017) suggestion, however, the lack of social media activity for emerging destinations did not provide a solid benchmark for this study. In this case, a longitudinal approach to this study could reveal insight into how DMOs develop over time and whether they are able to enhance their online image (e.g. number of followers, likes) and see if there is a difference between emerging and mature destinations.

Furthermore, more research is needed on the user-engagement perspective. For example, this study did not take the user profile into account. This would also generate a better understanding of which users respond to what kind of values and in which way. This is particularly important information as DMOs tend to target their messages to specific segments. Moreover, it is also worth mentioning that the selection of a particular social network certainly imposes a bias in the analysis, so the results could have been different if other social media had been chosen (for example, it can be argued that Facebook is not a very popular platform for young people nowadays, as they prefer platforms based on pictures and videos, rather than text). Furthermore, where Facebook analysis is concerned, future studies should also integrate reactions like 'love', 'sadness', 'haha' and also comments and possible re-shares. Lastly, the study did not analyze how values co-occurred and which co-occurrences of values are successful and create higher levels of consumer engagement.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jdmm.2019.03.004.

References

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research, 34*(3), 347–356.

Algesheimer, R., Dholakia, U. M., & Herrmann, A. (2005). The social influence of brand community: Evidence from European car clubs. *Journal of Marketing, 69*(3), 19–34.

Ashworth, G., & Kavaratzis, M. (2009). Beyond the logo: Brand management for cities. *Journal of Brand Management, 16*(8), 520–531.

Bigné, E., García, I. S., & Blas, S. S. (2009). The functional-psychological continuum in the cognitive image of a destination: A confirmatory analysis. *Tourism Management, 30*(5), 715–723.

Blain, C., Levy, S. E., & Ritchie, J. B. (2005). Destination branding: Insights and practices from destination management organizations. *Journal of Travel Research, 43*(4), 328–338.

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the internet—the state of eTourism research. *Tourism Management, 29*(4), 609–623.

Cai, L. A. (2002). Cooperative branding for rural destinations. *Annals of Tourism Research, 29*(3), 720–742.

Chan, N. L., & Guillet, B. D. (2011). Investigation of social media marketing: How does the hotel industry in Hong Kong perform in marketing on social media websites? *Journal of Travel & Tourism Marketing, 28*(4), 345–368.

Chaykina, T., Guerreiro, M., & Mendes, J. (2014). Destination brand personality of Portugal for the Russian-speaking market. *Joural of Spatial and Organiztional Dynamics, 2*(1), 23–40.

De Moya, M., & Jain, R. (2013). When tourists are your "friends": Exploring the brand personality of Mexico and Brazil on Facebook. *Public Relations Review, 39*, 23–29.

Dickinger, A., & Lalicic, L. (2016). An analysis of destination brand personality and emotions: A comparison study. *Information Technology & Tourism, 15*(4), 317–340.

Ekinci, Y., & Hosany, S. (2006). Destination personality: An application of brand personality to tourism destinations. *Journal of Travel Research, 45*, 127–139.

Govers, R., & Go, F. M. (2009). Tourism destination image formation. *Handbook of Tourist Behavior* (pp. 35–49). .

Govers, R., Go, F. M., & Kumar, K. (2007). Promoting tourism destination image. *Journal of Travel Research, 46*(1), 15–23.

Guerrero-Solé, F., & Fernández-Cavia, J. (2013). Activity and influence of destination brands on Twitter: A comparative study of nine Spanish destinations. In Z. Xiang, & I. Tussyadiah (Eds.). *Information and communication technologies in tourism 2014* (pp. 227–236). Switzerland: Springer International Publishing.

Guillet, B., Kucukusta, D., & Liu, L. (2016). An examination of social media marketing in China: How do the top 133 hotel brands perform on the top four Chinese social media sites? *Journal of Travel & Tourism Marketing, 33*(6), 783–805.

Hays, S., Page, S. J., & Buhalis, D. (2013). Social media as a destination marketing tool: Its use by national tourism organisations. *Current Issues in Tourism, 16*(3), 211–239.

Hofmann, M., & Klinkerberg, R. (2013). *Rapid miner: Data mining use cases and business analytics applications.* London: CRC Press.

Hosany, S., Ekinci, Y., & Uysal, M. (2007). Destination image and destination personality: An application of branding theories to tourism places. *Journal of Business Research, 59*(5), 638–642.

Huertas, A., & Marine-Roig, E. (2015). Destination brand communication through the social media: What contents trigger most reactions of users? In I. Tussyadiah, & A. Inversini (Eds.). *Information and communication technologies in tourism 2015* (pp. 295–308). Switzerland: Springer International Publishing.

Huertas, A., & Marine-Roig, E. (2016a). User reactions to destination brand contents in social media. *Information Technology & Tourism, 15*, 291–315.

Huertas, A., & Marine-Roig (2016b). Differential destination content communication strategies through multiple social media. In A. Inversini, & R. Schegg (Eds.). *Information and communication technologies in tourism 2016* (pp. 239–252). Switzerland: Springer International Publishing.

Jabreel, M., Moreno, A., & Huertas, A. (2017). Do local residents and visitors express the same sentiments on destinations through social media? In R. Schegg, & B. Stangl (Eds.). *Information and communication technologies in tourism 2017* (pp. 655–666). Switzerland: Springer International Publishing.

Jain, R., & Chan-Olmsted, S. (2009). Is your tourist destination sincere, rugged, or sophisticated? An exploratory examination of online brand personality of nations. *In annual meeting of the Association for Education in Journalism and Mass Communication* Boston, MA.

Kabadayi, S., & Price, K. (2014). Consumer–brand engagement on Facebook: Liking and commenting behaviors. *Journal of Research in Interactive Marketing, 8*(3), 203–223.

Lalicic, L., Huertas, A., Moreno, A., Gindl, S., & Jabreel, M. (2018). Do DMOs communicate their emotional brand values? A comparison between Twitter and Facebook. In B. Stangl, & J. Pesonen (Eds.). *Information and communication technologies in tourism 2018* (pp. 159–171). Berlin, Heidelberg: Springer.

Laroche, M., Habibi, M. R., & Richard, M.-O. (2013). To be or not to be in social media: How brand loyalty is affected by social media? *International Journal of Information Management, 33*, 76–82.

Leung, R., Schuckert, M., & Yeung, E. (2013). Attracting user social media engagement: A study of three budget airlines Facebook pages. In L. Cantoni, & Z. Xiang (Eds.). *Information and communication technologies in tourism 2013* (pp. 195–206). Berlin, Heidelberg: Springer.

Li, S. C., Robinson, P., & Oriade, A. (2017). Destination marketing: The use of technology since the millennium. *Journal of Destination Marketing and Management, 6*(2), 95–102.

Lovejoy, K., & Saxton, G. D. (2012). Information, community, and action: How nonprofit organizations use social media. *Journal of Computer-Mediated Communication, 17*(3), 337–353.

Lu, W., & Stepchenkova, S. (2014). User-generated content as a research mode in tourism and hospitality applications: Topics, methods and software. *Journal of Hospitality Marketing & Management, 24*(2), 119–154.

Mariani, M. M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional destination management organizations. *Tourism Management, 54*, 321–343.

Marine-Roig, E. (2017). Measuring destination image through travel reviews in search engines. *Sustainability, 9*(8), 1425.

Michaelidou, N., Siamagka, N. T., Moraes, C., & Micevski, M. (2013). Do marketers use visual representations of destinations that tourists value? Comparing visitors' image of a destination with marketer-controlled images online. *Journal of Travel Research, 52*(6), 789–804.

Moreno, A., Jabreel, M., & Huertas, A. (2015). Automatic analysis of the communication of tourist destination brands through social networks. *Proceedings of the 10th international conference on intelligent systems and knowledge engineering (ISKE-2015)* (pp. 24–27). .

Morgan, N., & Pritchard, A. (1998). *Tourism promotion and power: Creating images, creating identities.* Hoboken, NJ: John Wiley & Sons Ltd.

Morgan, N., & Pritchard, A. (2004). Meeting the destination branding challenge. In N. Morgan, A. Pritchard, & R. Pride (Eds.). *Destination branding: Creating the unique destination proposition* (pp. 59–79). (2nd ed.). Oxford: Elsevier.

Morgan, N., Pritchard, A., & Pride, R. (2003). Marketing to the Welsh diaspora: The appeal to hiraeth and homecoming. *Journal of Vacation Marketing, 9*(1), 69–80.

Munar, A. (2011). Tourist-created content: Rethinking destination branding. *International Journal of Culture, Tourism and Hospitality Research, 5*(3), 291–305.

Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism Management, 43*, 46–54.

Nusair, K. K., Bilgihan, A., & Okumus, F. (2013). The role of online social network travel websites in creating social interaction for Gen Y travelers. *International Journal of Tourism Research, 15*(5), 458–472.

Oviedo-García, M. A., Muñoz-Expósito, M., Castellanos-Verdugo, M., & Sancho-Mejías, M. (2014). Metric proposal for customer engagement in Facebook. *Journal of Research in Interactive Marketing, 8*(4), 327–344.

Pitt, L. F., Opoku, R., Hultman, M., Abratt, R., & Spyropoulou, S. (2007). What I say about myself: Communication of brand personality by African countries. *Tourism Management, 28*(3), 835–844.

Saffer, A. J., Sommerfeldt, E. J., & Taylor, M. (2013). The effects of organizational Twitter interactivity on organization–public relationships. *Public Relations Review, 39*(3), 213–215.

Schroeder, A., & Pennington-Gray, L. (2015). The role of social media in international tourist's decision making. *Journal of Travel Research, 54*(5), 584–595.
Sigala, M. (2011). Social media and crisis management in tourism: Applications and implications for research. *Information Technology & Tourism, 13*(4), 269–283.
Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management, 36*, 590–601.
TripAdvisor (2017). *Traveller destination choice: Top 25 destinations in Europe*. Retrieved 29.06.17. from: https://www.tripadvisor.com/TravelersChoice-Destinations-cTop-g.
Uşakli, A., & Baloglu, S. (2011). Brand personality of tourist destinations: An application of self-congruity theory. *Tourism Management, 32*(1), 114–127.
Uşakli, A., Koç, B., & Sönmez, S. (2017). How 'social' are destinations? Examining european DMO social media usage. *Journal of Destination Marketing and Management, 6*(2), 136–149.
Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., et al. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research, 13*(3), 253–266.
Walther, J. B., & Jang, J. W. (2012). Communication processes in participatory websites. *Journal of Computer-Mediated Communication, 18*(1), 2–15.
Wattanacharoensil, W., & Schuckert, M. (2015). How global airports engage social media users: A study of Facebook use and its role in stakeholder communication. *Journal of Travel & Tourism Marketing, 32*(6), 656–676.

PLOS ONE |

RESEARCH ARTICLE

Semantic analysis and the evolution towards participative branding: Do locals communicate the same destination brand values as DMOs?

Mohammed Jabreel[1,3]*, **Assumpció Huertas**[2], **Antonio Moreno**[1]

1 TAKA (Intelligent Technologies for Advanced Knowledge Acquisition) research group. Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain, **2** Departament d'Estudis de Comunicació, Universitat Rovira i Virgili, Tarragona, Spain, **3** Computer Science Department, Hodeidah University, Hodeidah, Yemen

* mohammed.jabreel@urv.cat

OPEN ACCESS

Citation: Jabreel M, Huertas A, Moreno A (2018) Semantic analysis and the evolution towards participative branding: Do locals communicate the same destination brand values as DMOs? PLoS ONE 13(11): e0206572. https://doi.org/10.1371/journal.pone.0206572

Editor: Daniel Angus, The University of Queensland, AUSTRALIA

Received: October 31, 2017

Accepted: October 16, 2018

Published: November 7, 2018

Data Availability Statement: According to Twitter agreement and policy, we are restricted to redistribute or publish the data. So, we provided the dataset of IDs of analysed tweets in this GitHub repository: (https://github.com/mhjabreel/BrandSemData). The tweets can be downloaded given their IDs using the following API: (https://developer.twitter.com/en/docs/tweets/post-and-engage/apireference/get-statuses-show-id). Another option is to use the TwitterCrawler from this GitHub repo: (https://raw.githubusercontent.com/ywwbill/TwitterCrawler).

Abstract

Participative branding is a process by which DMOs, locals and visitors contribute, through their activity in social media, to the definition of the emotional values associated to a destination. From the communicative point of view it is particularly important that DMOs and local citizens work together in the transmission of a set of coherent values that compose a personalised identity. This paper presents a novel methodology, based on semantic similarity measures, that permits to make an automated analysis of the emotional values transmitted by official tourist offices and by local citizens through social media. A study of 54,000 tweets from 9 main European tourist destinations highlights the lack of a strategy towards the communication of a distinctive brand and a strong gap between the official view transmitted by the DMOs of the destinations and the one communicated by their residents.

Introduction

The concept of *destination brand*, that represents a unique combination of destination characteristics and added values, both functional and non-functional, which have taken on a relevant meaning [1] and try to convey a favorable image [2], has been heavily studied in tourism and marketing. The research works that have been conducted on this subject have followed two main directions. On the one hand there is *emotional branding*, which highlights the importance of brand emotional values in the communication of tourist destinations [3–7]. On the other hand, a more recent trend is the study of *interactive and participatory branding* [8–11].

Emotional branding is based on the idea that many tourist destinations have similar attractions, so at the end it is the emotional aspect of the brand, its personality or identity, that distinguishes it [2, 4, 12–16]. The great contribution of destination brands is the creation of emotional ties with both locals and visitors, which are generated through the relationships they establish with them [17, 18]. These relationships and emotional ties created by destination

Funding: The authors acknowledge the support of Univ. Rovira i Virgili through a Martí Franquès PhD grant, the assistant/teaching grant for the Department of Computer Engineering and Mathematics and the Research Support Funds 2016PFR-URV-B2-60.

Competing interests: The authors have declared that no competing interests exist.

brands have a key influence in the tourists' decisions about the places to visit [1]. Thus, *Destination Management Organisations* (DMOs), which are organisations that promote places in order to increase their number of visitors and develop their marketing and services, are posed a double challenge: to communicate the emotional values that make up the identity of the destination and their brand to distinguish themselves from the rest, and to establish good emotional relationships with the tourists in order to create a good image and appeal.

The technological evolution and the emergence of social media have enabled the appearance of *User Generated Content* (UGC), which has provoked an evolution towards *participatory branding* [19]. The communication of a brand, traditionally seen as a unidirectional process, is now understood as a process of interactive dialogue between local and external actors [20, 21]. In particular, their participation in the creation of destination brands is crucial [11, 19, 22, 23]. Thus, DMO managers must create the appropriate interaction channels and encourage their participation in the branding process, especially that of internal or local users [9, 19, 24].

The impact of social media in the communication of destination brands has been enormous. When a tourist visits an unknown place, the word-of-mouth of other users with no vested interest in the destination generates substantial credibility and reduces the risks in the decision making process [7]. Several studies have confirmed that the opinions and experiences shared by users through social media are perceived as being more reliable than the institutional information provided by DMOs of destinations [25–28].

On the basis of these studies it can be said that users, and especially the local ones, play a key role in the branding of destinations. Locals are directly affected by brand decisions and the touristic strategies of the region. Hence, they ought to participate in the branding process. In addition, they are important ambassadors and communicators of this brand and, therefore, their involvement in the entire branding process is crucial [24, 29]. So, the participation of locals is necessary to create sustainable, consistent, non-artificial and successful brand images [10, 29, 30]. However, in professional practice, locals do not always participate in the creation of the brand [9]. In fact, only a few brand creation processes involve them [2, 31].

Given this scenario, the main aim of this paper is to present a methodology that permits to analyse in an unsupervised, automatic and efficient way the brand communicated by a given destination and to compare it with the brand transmitted by its locals.

The analysis of the communication of emotional values through social media presents two main challenges. On the one hand they are intangible and abstract, so their analysis is complex and the researcher may introduce a subjective bias even if specific analysis templates are used to measure them. On the other hand, the amount of messages and comments about a given destination may be huge, making a manual analysis unfeasible. To overcome these limitations this paper introduces a new automated semantic analysis methodology that enables the efficient and objective analysis of a large number of tweets.

Thus, the contributions of this article are twofold. Firstly, it provides a new semantic analysis methodology that allows analysing the communication of brand emotional values posted by DMOs and by their locals, through the analysis of adjectives used in tweets posted by the official destinations and by their residents. Secondly, the proposed methodology permits to find out if the DMOs and locals communicate a coherent brand of the destinations, which will allow us to know if DMOs involve locals in the branding process and if a coordinated participatory branding strategy exists. We understand DMOs and locals have different interests and images of destinations and they will not talk about the same issues in their social media, but it is important that all of them communicate a coherent or, at least, a non-contradictory image of the place. The research is needed by DMOs because would help them to know which image of the destination is communicated by tweets of locals in comparison with theirs.

The rest of the paper is structured as follows. The next section comments the state of the art on participative branding and on the semantic analysis of social media content. After that, the emotional values model is presented. Section 4 describes the methodology used to retrieve, process and semantically analyse a set of tweets. Then, the results of an analysis of 54,000 tweets from 9 major European destinations are presented and discussed. The paper closes with some conclusions and lines of future work.

1 Literature review

This section is divided in two parts. The first one discusses two main research trends on collaborative destination branding and their impact on the brand management activities of DMOs. The second one introduces the main lines of work in semantic analysis of social media content.

1.1 Participative branding, co-creation of the brand and implications for tourist destinations

Social media have transformed not only the communication of destination brands, but also their definition, turning it into a process involving multiple *stakeholders* (all the publics that are impacted by and also impact a destination, like organisations, politicians, companies, locals, tourists, mass media, etc.) [19]. In fact, several authors claim that one may speak of the existence of a new conceptualization of destination branding, which requires the co-creation of destination brands by all these actors [9, 17, 23, 32, 33]. Two types of branding strategies, focused on shared participation and on co-creation, are commented in the following subsections.

1.1.1 Participative branding studies. At the present moment of technological transformations and social changes, academic studies are evolving towards relational, interactive and participatory branding in the fields of marketing [21, 34–36], public relations [20, 37, 38] or tourism [8–11, 17, 19, 39, 40].

These studies request the participation of locals in the whole process of regional brand creation. They argue that only if relations are struck up with locals and they are involved with the brand of the destination will they accept it and will a sense of identity and of belonging to the destination and to the brand be aroused in them [9].

Participation by locals involves the expression of different points of view on a region, often with conflicting interests; thus, it requires dialogue, disagreement, debate and also negotiation, that is, the establishment of interactions and relationships [10]. The creation of the brand, therefore, must be a democratic process involving the locals and taking their interests into account [3, 41, 42]. Only with the involvement of locals, residents and organizations, who are the ones who really know the region, can a consistent and successful destination brand be defined [43]. Such a commonly agreed-upon brand generates a sense of place among the locals, facilitates their acceptance, and transmits authenticity to tourist.

Despite the wealth of academic publications on the importance of participatory destination branding, in their professional activities DMOs are still assimilating the changes in communication and they are gradually adapting to new technologies and to the idea of participatory branding. Tourist destinations still communicate their brands through slogans and advertising campaigns and they hardly involve their locals in the branding process [2, 19, 31]. Thus, DMOs need to adapt quickly to this new reality, to rethink the role that locals should take in the whole process of destination branding and to create the channels that allow their participation.

1.1.2 Co-creation branding studies. Interestingly, in parallel with these tourism marketing researchers who stress the need of local participation in destination branding, another school of authors from the field of general marketing argues that all the stakeholders are already involved in the creation of brands via their comments in social media [21, 34–36], so their participation in the co-creation of brands is already an inevitable reality. Despite being apparently contrary visions, they are complementary and they move in the same direction. While the latter claim that stakeholders, and locals among them, are already co-creators of the brand via their comments, the former affirm that DMOs should involve locals and make them participate from the beginning of the branding process.

These latter researchers affirm that there has been a great transformation in the field of brand communication. In recent years, many studies have been conducted on communication disciplines concerning brand-consumer relationships [44, 45]. Consumers are currently very active and, furthermore, they are expected to be even more so in the future [46, 47]. They have already begun to participate in the definition of products and their brands, co-creating them on the basis of their comments and experiences. This process leads to a greater implication with the brand, a greater loyalty towards it, a better image and a positive influence on other potential consumers [21].

1.1.3 Implications of these studies for tourist destinations. Thus, a brand is no longer owned by an institution that defines it to convey certain preestablished emotional values in its own interest, but it is created on the basis of all the interactions and relations created in the minds of the stakeholders. Therefore, we may speak of a collective process of brand co-creation [9]. All stakeholders of a brand, from internal/local to external/consumers, are active co-creators of the brand.

This process has also happened in tourism. DMOs have lost the control over the destination brands, which have become the property of a community or a collective. Thus, the meaning of the brand of a destination is constantly being co-created by the participation and comments of the locals, companies, visitors, etc, and, as a result, brands are now more dynamic, authentic and collective [11, 48].

DMOs must bear in mind that the image of a destination in the minds of potential visitors is not only the result of the destination's communication and advertising campaigns. This situation is due to the emergence of social media. Their capacity to allow users to interact and to create dialogues has transformed branding. Moreover, due to the ease of using social media to share experiences, users have become co-creators of *brands* [34]. Most interestingly, it has been shown that the content generated by users through social media have far greater impact and credibility than those broadcast through traditional channels [49]. Therefore, the co-creation of content by users cannot be ignored by DMOs.

The importance and influence of user-generated content through social media in the creation of destination brands have been previously demonstrated [50, 51]. The images of tourist destinations are created from the interactions and the information of DMOs, locals, other tourists, etc. Social media have become the primary sources of information for potential tourists, and UGC generates information, experiences and emotions that create the image of the brand [52–57].

So, the reality is that whether locals are involved early in the process of brand creation or not, they are always co-creating and communicating the destination brand in any case. Then, it is better to have them involved from the outset. Moreover, DMOs should constantly analyse UGC to know what image of the destination is being communicated by locals, and consequently manage the brand coherently [57].

In the particular case of the work reported in this paper, Twitter was chosen as the social network of study. Many authors have highlighted the increasing importance of social media in

tourism information search [50, 58, 59] and, as [60] have shown, Twitter is the fourth main social medium where users search for information, following Tripadvisor, tourism blogs and Facebook. However, the aim of this work is not to analyse how users look for information in the social media when they are organizing a trip, but rather to think about how the comments and experiences that DMOs and locals of destinations share through social media influence in the process of branding co-creation. DMOs should take into account that the image of destinations is not created only by the rational, practical, tangible information that tourists search when they prepare a trip. It is actually created by any content (photographs, videos, emotional comments) made by individual people at any time, based in their own experiences and emotions or those of other users. Moreover, DMOs should also keep in mind that experiences shared by friends and acquaintances are more trustworthy and have more impact in the creation of image destinations than those that are impersonal [61–63], and Facebook and Twitter are currently the most prominent social networks to receive constant information, updates and comments from family and friends. Furthermore, they are also the two social media more commonly used by DMOs to promote destinations. The main advantage of Twitter over Facebook, which led to its final selection, is the sheer amount of messages that are created daily and the capacity to obtain easily messages with given characteristics (e.g. tweets sent from a specific location in a certain language at a particular point of time by users with given characteristics). In any case, the methodology of analysis described in this paper could also be applied to messages distributed through Facebook or any other social network.

1.2 Semantic analysis to measure emotional branding

Emotional branding is causing a substantial change in destination management and communication. Destination branding involves the association of distinctive attributes and emotional values to the regions [3, 5, 29]. Therefore, DMOs try to communicate an identity, a personality and a set of values that have an impact on the emotions of the users and generate an attraction for the destination [2]. Destination brands create relationships with stakeholders, whether local or external, which lead to emotional bonds with them [18]. In addition, tourists base their decisions to visit a destination rather on emotional aspects and ties than on rational decisions based only on the main attractions [4].

The images about a place in the minds of potential tourists have a decisive influence on the choice of destination [64, 65]. Many factors influence on the creation of these images, including the expected experiences of the potential tourists, the information received from social media and the behaviour of their locals among others [6, 66, 67]. Therefore, DMOs, with their communicative actions through the various media, seek to transmit an identity and a brand, with the objective of promoting a desired brand image that also is coherent with the vision of the locals [29].

This paper seeks to analyse whether DMOs communicate via their official tweets the same brand and the same emotional values as the locals of destinations. A new methodology of semantic analysis has been designed and developed with this purpose. To date, two types of methodologies have been used to analyse the communication of the emotional content of a destination brand: *qualitative* and *quantitative*.

Qualitative studies are based on methods such as in-depth interviews and focus groups. They provide high quality information, but they are necessarily limited to small samples, and therefore they are not representative of the entire universe. Thus, despite the wealth of information, the results cannot be easily generalized [68]. Many of these studies use qualitative analysis software such as ATLAS.ti or NVIVO. For example, NVIVO was used to analyse how a series of stimuli change the attributes associated with a sample of North American destinations

[69], and ATLAS was employed to analyse qualitatively 20 on-line articles about Portugal, in order to ascertain what image tourists have of this destination and how they can help create the destination brand [70]. Another recent example of a qualitative study is the use of semi-structured interviews to determine the use and impact of social media on destination communications [71].

Regarding quantitative studies, they can be mainly classified in two types: *thematic* and *semantic* [72]. Thematic analyses involve the recording of issues, attributes, values or categories by counting the frequency of the words that appear in the text. They permit to know which are the topics or categories that are most relevant in texts, but they are not able to interpret their meaning. Thus, we cannot know if positive or negative things are being said of these items, if there is a touch of irony in the text, what is said precisely about them, etc. Despite these limitations, the measurement of the frequency of words has been heavily used in content analysis. Some recent examples include the analysis of the attributes communicated and associated with the Portugal brand by Russian visitors [13], of the brand attraction factors and emotional values conveyed through social media by DMOs in a sample of Spanish tourist destinations [7] and of the destination brand communication through tourists' comments in social media [73].

The main contribution of semantic analysis to content analysis is the discovery of the relationships between subjects or categories, creating a more conceptual study rather than a merely syntactic one. Therefore, it can go deeper into the meanings of the words and understand more profoundly what is being said. Just to cite some recent examples, semantic analysis have been employed to assess how the brands of the destinations of Brazil and Mexico are communicated through Facebook [74], to define a "semantic network" to analyse the relations between topics [75], to measure the image of the Basque Country, taking cognitive, affective and conative aspects into account [76] and to study the emotional qualities in tourists' UGC through the analysis of their perceptions and experiences [77]. Various authors have combined thematic and semantic analysis in their studies [78–80].

This work belongs to this novel line of semantic studies. We have designed and developed a methodology that allows to analyse the communication of something as subjective as brand emotional values through semantic analysis, avoiding the subjectivity and the bias that a manual analysis might involve. Furthermore, it permits to consider a large volume of data, as thousands of tweets can be analysed at a time. This framework of analysis, which is explained in detail in section 4, could certainly be highly useful in the brand management and communication tasks of the DMOs of tourist destinations. Before describing the methodology of analysis, the following section explains which are the emotional values that we want to measure.

2 Emotional Model

The importance of the emotional values and the personality of the destinations in their branding has been stressed by many academic studies [1, 81]. Destinations can be described using human personality values [50]. We have applied a very slightly modified version of the well-known *Brand Personality Scale* (BPS) of Aaker to the analysis of destination brands [82]. It has been shown that personality dimensions have a positive impact on the preferences of potential tourists, as a strong and well-defined personality improves the image of the destination and the intentions to visit it [16].

In this work Aaker's BPS has been used to analyse the emotional values associated to a travel destination. The five main dimensions of analysis of the personality and the emotional values of a destination are *sincerity*, *excitement*, *competence*, *sophistication* and *ruggedness*. Each of them has been divided into a set of categories, which in turn have been refined into several

sub-categories, represented by a set of terms. The whole template of analysis is shown on Table 1.

3 Methodology

Twitter and Facebook are the social media more commonly used by tourism destination managers for their promotion. The short updates sent by users in Twitter (*tweets*, limited to a maximum of 140 characters) provide a rich source for finding out the opinions, feelings and emotions of the public. In this work we have focused on the analysis of English tweets sent by official tourist destinations and by their locals. The final aim is to study how the emotional values defined in the previous section are communicated. The steps followed in the methodology of analysis have been the following:

- *Selection of the destinations to be analysed*. In this step we have chosen a set of well-known European destinations, specifying some constraints on the language they use and a minimum quantity of messages transmitted through social media.
- *Retrieval of the set of tweets*. This phase involves the use of a new tool that permits to retrieve sets of tweets that satisfy certain user-given requirements.
- *Pre-processing of the set of tweets*. This step applies some simple treatments on the content of the tweets to make them easier to analyse.
- *Semantic analysis of the content of the tweets*. This step is the core of the methodology. It uses a well-known ontology-based semantic similarity measure to compare the adjectives used in the tweets with the emotional values defined in section 3.
- *Interpretation of the results*. In the final step we can analyse the obtained results, comparing the performance of the DMOs of the different cities and also contrasting the values transmitted by the official DMOs with the ones reflected in the opinions of their locals.

Fig 1 shows a graphical depiction of the main steps of this methodology.

In the remainder of this section we describe the technical aspects of the study: how destinations were selected and how their associated tweets were retrieved, pre-processed and semantically analysed.

3.1 Destination selection

To select the destinations to be analysed we searched manually for the official Twitter accounts of the top 25 European destinations in 2014, according to Tripadvisor (http://www.tripadvisor.co.uk/TravelersChoice-Destinations-cTop-g4).

In 8 cases (Istanbul, Venice, Florence, Krakow, Urgup, Moscow, Dalyan, and Rimini) we were not able to identify an account. In 5 cases (Rome, Paris, St Petersburg, Madrid and Milan) most of the tweets were written in the main language of the destination and not in English. In some cities (Prague, Lisbon, Zermatt) the number of tweets was very low. It was finally decided to select the nine destinations that had more than 3,000 English tweets, which are shown in Table 2.

3.2 Retrieval of the sets of tweets

As we want to compare the communication of the destination brand from the DMO and the local residents' points of view, in this study we consider two datasets: an *official set of tweets O*, which contains the tweets that have been posted from the official accounts of the selected

Table 1. Emotional Model.

Emotional value	Category	Subcategory
Sincerity	Down-to-earth	Family-oriented
		Down-to-earth
		Sustainable
	Honest	Calm
		Real
		Traditional
		Honest
	Wholesome	Original
		Wholesome
		Quality of Life
	Cheerful	Happiness
		Sentimental
		Friendly
Excitement	Daring	Trendy
		Daring
		Exciting
		Exotic
		Fashionable
	Spirited	Cool
		Spirited
		Dynamic
		Vital
		Fresh
		Young
		Sensorial
	Imaginative	Unique
		Imaginative
		Creative
	Up-to-date	Up-to-date
		Independent
		Contemporary
	Cosmopolitan	Cosmopolitan
		Tolerant
		Hospitable
Competence	Reliable	Reliable
		Hard-working
		Safe
		Rigorous
	Intelligent	Intelligent
		Technical
		Corporate
		Innovative
	Successful	Successful
		Leader
		Ambitious
		Powerful

(*Continued*)

Table 1. (Continued)

Emotional value	Category	Subcategory
Sophistication	Luxurious	Glamorous
		Luxurious
	Charming	Seductive
		Smooth
		Romantic
		Magical
Ruggedness	Outdoorsy	Outdoorsy
		Get-away
		Recreational
	Tough	Tough
		Rugged
		Non-Conformist

https://doi.org/10.1371/journal.pone.0206572.t001

destinations, and a *local set of tweets L*, which contains tweets posted by the locals of the destinations.

We have developed a novel tool, *twiQuery*, which has been used to collect the tweets to be analysed. twiQuery is a crawler that enables users and developers to make advanced search actions on tweets, such as retrieving sets of tweets posted by a specific user, written in a specific language, sent from a certain area determined by the name of a city or geolocation and a given radius, posted during a specific period of time, etc. The set O is the set of all the tweets written in English and sent by the nine official destination accounts. The set L contains the tweets in English that have been sent within a radius of 15 kms. of the city center of each of the chosen destinations. In both cases the following periods of time were considered:

- High tourism season of 2015: from 15-June-2015 to 01-September-2015.
- High tourism season of 2014: from 15-June-2014 to 01-September-2014.
- Low tourism season of 2014: from 15-October-2014 to 30-December-2014.

In order to make sure that the tweets in the set L have been actually sent by locals, and not by visitors, the system keeps only those tweets that have been sent by users that declare explicitly the destination as their home location in their Twitter profile.

After removing the tweets sent by strangely prolific users (those that have sent over 1,000 tweets in the considered time intervals), we obtained between 11,000 (Budapest) and 370,000 (London) local tweets per each of the nine cities. As we want to compare easily the references to emotional values from locals and from destinations, we finally decided to analyse 3,000 tweets from the official destination account and 3,000 tweets from locals for each of the nine cities (Amsterdam, Athens, Barcelona, Berlin, Budapest, Dublin, Edinburgh, London and Vienna). The tweets were randomly chosen. Table 3 shows the number and percentage of different local users for each destination.

It has to be noted that, even after filtering the locality of the users, it may not be guaranteed that they are talking about the destination. In order to study this issue, a manual analysis of 400 random tweets sent by local people from London was made. This study showed that 292 tweets (73%) were related to local aspects of the destination (local news, weather, restaurants/food, local transport, tourist attractions and cultural activities, local stores and business, local sports, etc.), 97 tweets (24.25%) were conversational or personal life comments, and only 11 (2.75%) were referring to other locations or global news. In the light of these results, we believe that the

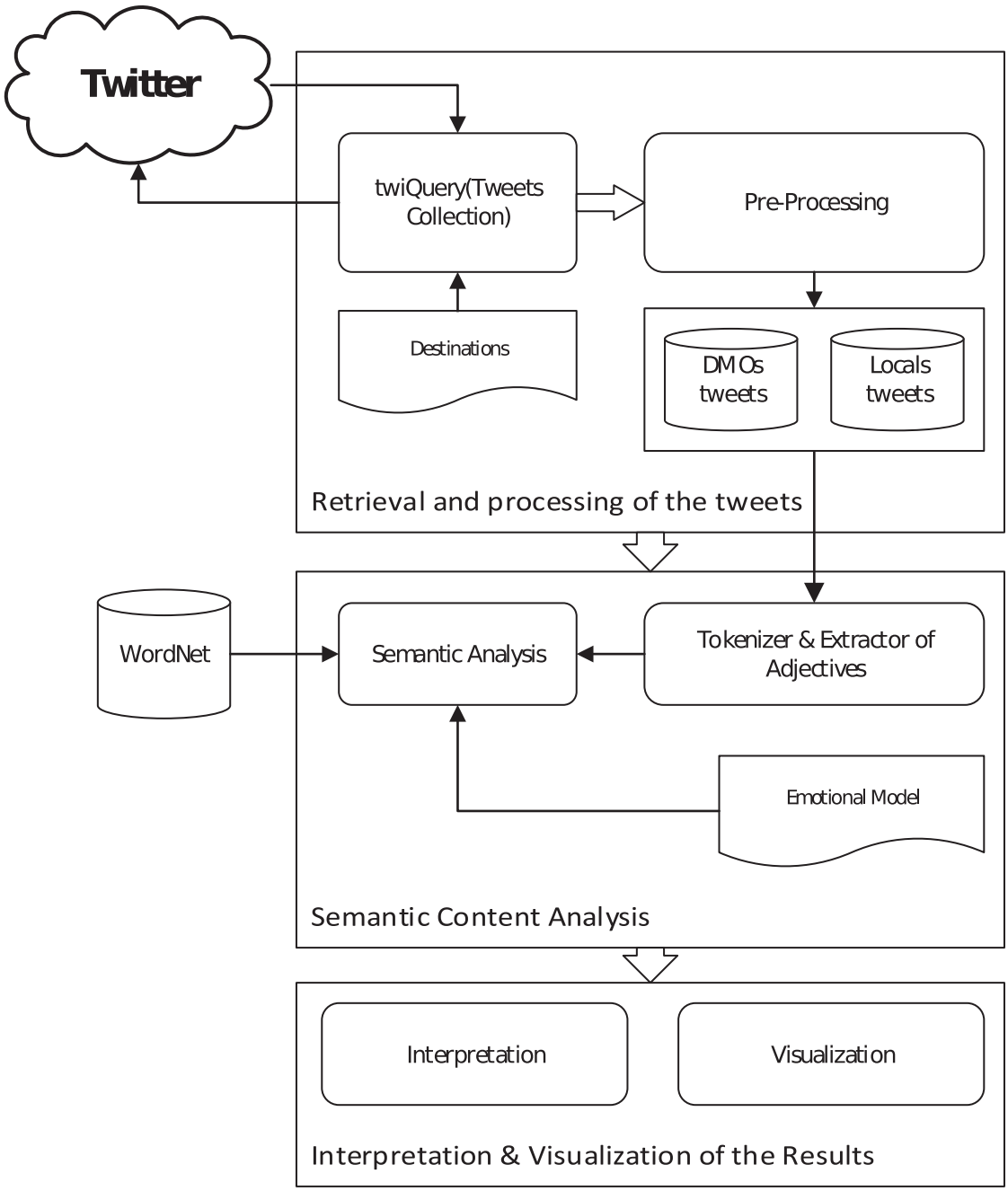


Fig 1. Architecture of the methodological framework of analysis.

https://doi.org/10.1371/journal.pone.0206572.g001

set of local tweets is representative enough of the local aspects of the destination (however, in the future work it could be possible to devise more precise ways of filtering the set of local tweets).

3.3 Pre-processing of the tweets

It is well-known that the language used in Twitter is very casual and noisy. Tweets contain numerous punctuation errors, spelling mistakes, abbreviations, slang terms, emoticons, etc. The tweets were pre-processed to mitigate some of these effects as follows:

Table 2. Selection of destinations.

City	Twitter account	Tweets
London	@visitlondon	31,300
Berlin	@berlintourism	5,400
Barcelona	@VisitBCN_EN	5,300
Budapest	@VisitBudapest	3,400
Amsterdam	@Iamsterdam	6,100
Vienna	@ViennaInfoB2B	4,200
Athens	@CityofAthens	4,600
Dublin	@VisitDublin	12,400
Edinburgh	@edinburgh	26,500

https://doi.org/10.1371/journal.pone.0206572.t002

Table 3. Unique users in each analysed destination.

Destination	Unique Twitter users	Percentage
London	2555	85%
Berlin	1256	42%
Edinburgh	1419	47%
Dublin	1918	64%
Athens	1012	34%
Vienna	958	32%
Amsterdam	1347	45%
Budapest	781	26%
Barcelona	1333	44%

https://doi.org/10.1371/journal.pone.0206572.t003

- Tweets may contain URLs, usernames, hashtags and emoticons. All URLs, usernames and strange symbols were removed from the tweets.
- To reduce the dimensionality, all tweets were converted to lowercase and stop words were removed. Table 4 shows the list of stop words.
- Words with repeated letters were automatically corrected using an algorithm devised and implemented by the authors that performs a breadth-first search and analyses all the possible ways of eliminating repeated letters in a string, checking in Wordnet (see http://wordnet.princeton.edu) if they are correct.

3.4 Semantic content analysis

In previous studies on content analysis of the communication of destination brands it was pointed out that *nouns* usually provide information on the particular tourist attractions, *verbs* describe actions and types of tourism, and *adjectives* communicate the emotional responses

Table 4. Stop words list.

Stop words list
i, me, my, myself, we, our, ours, ourselves, yo, your, yours, yourself, yourselves, he, him, his, himself, she, her, hers, herself, it, its, itself, they, them, their, theirs, themselves, what, which, who, whom, this, that, these, those, am, is, are, was, were, be, been, being, have, has, had, having, do, does, did, doing, a, an, the, and, but, if, or, because, as, until, while, of, at, by, for, with, about, against, between, into, through, during, before, after, above, below, to, from, up, down, in, out, on, off, over, under, again, further, then, once, here, there, when, where, why, how, all, any, both, each, few, more, most, other, some, such, no, nor, not, only, own, same, so, than, too, very, s, t, can, will, just, don, should, now

https://doi.org/10.1371/journal.pone.0206572.t004

[78, 83]. Thus, as we want to measure the association between tweets and emotional values, we have focused the analysis on the *adjectives* used by destinations and visitors.

The objective of the semantic analysis is to associate the adjectives found in a set of tweets with the categories of emotional values shown in Table 1. We used the Penn Treebank POS-tagger from the NLTK library [84] to retrieve the adjectives and count their frequency of use. A direct syntactic mapping is not possible, as most of the adjectives do not appear directly as categories/subcategories of emotional values. The idea is to use a *semantic* similarity measure [85] between the adjectives and the categories/subcategories. This type of measures requires the use of some kind of external structured knowledge base (in our case, WordNet). Ontology-based semantic similarity measures rely on the topological structure of an ontology to calculate the degree of resemblance between two terms. The length of the path between the terms (considering hyponymy and hypernymy relationships) and their position in the hierarchy of concepts (i.e. their degree of generality) are the basic points taken into account by similarity functions. In this work we have used a well-known similarity measure, defined by Wu and Palmer [86] as follows:

$$Sim_{W\&P}(c1, c2) = \frac{2 * depth(c3)}{length(c1, c3) + length(c2, c3) + 2 * depth(c3)} \quad (1)$$

In this expression *c3* is the *Least Common Subsumer* (LCS) of *c1* and *c2* in the reference ontology, *length* is a function that returns the number of hypernym links among two concepts and the *depth* of a concept is the number of hyperlinks that separate it from the root of the ontology. This measure ranges from 1 (for identical or synonym concepts) to 0 (when the LCS of the concepts *c1* and *c2* is the root of the ontology, so they do not have any common ancestor). The main difference of this function with respect to other edge-counting measures is that it takes into account the depth of the compared concepts in the hierarchy (given the same distance, two concepts are more similar if they are more specific). The semantic similarity between two words is calculated as the maximum similarity between the synsets associated to each of the words in WordNet (which represent their different senses). The adjectives with a similarity higher than 0.7 were considered as *emotional adjectives* and counted in the analysis, whereas the rest were dismissed.

The main problem of this approach is that WordNet uses hypernymy relationships between nouns, and we wanted to compare the adjectives found on the tweets with the subcategories associated to the emotional values (which are also mostly adjectives). Thus, before applying the Wu-Palmer semantic similarity measure we had to transform its inputs into nouns. In the case of the subcategories, they were manually translated to the equivalent nouns (e.g. *friendly* and *ambitious* were transformed into *friend* and *ambition*, respectively). Concerning the adjectives appearing in the tweets, they were automatically transformed into nouns using their *derivationally related form* or their *attribute* property in WordNet.

Table 5 shows some tweets posted by local people from London and analysed using our model. The table shows the tweet (with the analysed adjective in bold face), the emotional value that has been considered more similar to the adjective, and the degree of similarity between them. Note that there may be parsing errors (e.g. "*love*" is labelled as an adjective in the second tweet).

Let us consider a concrete example, using the last tweet of that table. The adjective "beautiful" is transformed automatically into a noun ("beauty") using the *attribute* property of WordNet. Then, it can be compared with the nouns that represent the different emotional values. Some results (the interested reader may find in http://ws4jdemo.appspot.com/ a detailed

Table 5. Example of tweets analysed using our model.

Tweet	Emotional Value	Similarity
Amazing **creative** buildings on #EastLondon #lovelondon @ Shoreditch High Street <URL>	Creative	1.0
Summer **love**. @ St James's Park <URL>	Romantic	0.93
Another **joyful** day in this amazing city!!! #London #londonpride, #LoveWins #Friends #Pride2015. . . <URL>	Exciting	0.93
Beautiful morning for a ride. #mangobikes #cycling #lovetheweekend @ Botley Hill Farmhouse <URL>	Glamorous	0.92

https://doi.org/10.1371/journal.pone.0206572.t005

explanation of how these results are obtained after analysing the WordNet noun hierarchy) are the following:

- $Sim_{W\&P}$(beauty, glamour) = 0.93
- $Sim_{W\&P}$(beauty, charm) = 0.87
- $Sim_{W\&P}$(beauty, honesty) = 0.66
- $Sim_{W\&P}$(beauty, original) = 0.63
- $Sim_{W\&P}$(beauty, innovation) = 0.55

Thus, the system considers that the term *beautiful* is highly related to emotional values such as *glamorous* or *charming*, but it is quite unrelated to the values *honest*, *original* and *innovative*.

The accuracy of the automatic semantic procedure that links adjectives to emotional values is hard to measure, since even a manual assessment is highly subjective. The common procedure to evaluate the performance of a semantic similarity measure is to calculate the correlation between its results and a human assessment of the similarity between some pairs of terms belonging to a golden standard. In the case of the Wu-Palmer similarity, Slimani [87] obtained a 74% correlation and Budanitsky and Hirst [88] report 82.9% and 81.9% in two different benchmarks (using the Lin measure, of which Wu-Palmer is a particular case [89]).

4 Results and discussions

In Tables 6 and 7, which present the results of the analysis of the communication of the brand emotional values by DMOs and by locals or residents of destinations through their tweets, two figures appear for each emotional value and destination. The first one is the total number of times that an emotional value has been communicated (possibly using a certain number of adjectives), whereas the second figure is the number of different adjectives used to communicate an emotional value in the analysed tweets. For example, the official tourist account of London used 9 adjectives 33 times to convey the *calm* value. The first row shows the accumulated values for each city (e.g. Berlin locals have used 186 emotional adjectives 1086 times in their 3000 tweets). The nine columns in the left part of the figure correspond to the analysis of the tweets sent by the official tourist accounts, whereas the ones on the right side provide the results of the analysis of the tweets sent by local people. The orange cells highlight the positions with higher values. Green cells mark situations in which a city does not communicate a value that is transmitted by most of the other destinations (for example, Dublin's DMO is the only one that did not communicate the value *wholesome*). Yellow cells indicate a value that is transmitted by a low number of destinations (e.g. the local people from Athens were the only ones that referred to the *cosmopolitan* concept, albeit only once).

Table 6. Use of emotional adjectives by DMOs (orange: high values—green: usual emotional value not communicated by the destination—yellow: emotional value communicated only by a few destinations).

Subcategory	Lond	Berl	Dubl	Athe	Edin	Vien	Amst	Buda	Barc
#	1103/261	945/222	1011/245	840/219	1082/266	893/206	943/229	885/199	760/192
Family-oriented			3/1	2/2				1/1	
Down-to-earth	57/25	55/25	60/22	52/20	54/19	44/18	48/25	64/24	43/22
Sustainable	6/6	7/6	4/2	8/7	9/8	3/3	3/3	1/1	3/1
Calm	33/9	30/7	33/7	23/6	37/8	24/9	28/7	18/6	26/9
Real	39/6	31/8	28/8	32/6	31/7	25/5	27/6	32/7	26/6
Traditional	17/10	14/9	18/11	21/10	9/9	15/10	11/7	12/7	9/8
Honest	162/10	141/4	163/9	150/7	184/11	141/7	163/8	145/6	147/6
Original	45/26	43/20	37/21	33/17	37/20	25/11	26/14	24/10	32/11
Wholesome	6/1	6/1		3/1	4/1	2/1	7/1	7/1	6/1
Quality of life									
Hapyness	31/8	28/8	36/7	39/8	48/12	35/8	24/7	57/7	32/6
Sentimental	27/13	21/9	22/14	18/13	19/11	19/12	16/11	24/14	14/11
Friendly	3/3	6/3	11/4	5/4	9/4	1/1	3/2	4/3	12/6
Trendy	2/2	1/1	5/4	4/2	6/4	1/1	2/2	1/1	2/2
Daring	22/7	17/4	17/6	9/7	11/5	14/5	14/5	10/4	12/7
Exciting	6/2	4/2	5/1	7/2	2/2	6/2	4/1	1/1	3/1
Exotic	1/1		2/1			3/2	2/2	6/2	1/1
Fashionable						1/1			
Cool									
Spirited	4/3	1/1	1/1	2/2	3/2		4/2	3/2	6/2
Dynamic	1/1	1/1		1/1	1/1	2/2	2/2	2/2	2/2
Vital		1/1	2/1						
Fresh	156/7	143/5	94/5	109/7	134/6	129/5	158/4	91/8	98/4
Young	12/3	11/2	7/2	6/2	4/1	2/2	5/1	8/2	5/1
Sensorial	20/6	25/7	34/9	16/5	23/12	25/8	25/7	20/7	8/3
Unique	23/4	9/2	12/2	19/3	19/5	9/2	13/4	11/3	15/2
Imaginative					1/1				
Creative	14/6	10/5	11/5	11/6	6/3	10/6	16/6	6/2	3/1
Up-to-date	3/2	2/2	4/1	2/1	4/2	7/2	1/1	2/2	7/2
Independent	2/2	3/2	2/2	1/1	7/2	5/1	1/1	6/2	3/2
Contemporary	57/9	60/9	67/10	40/9	64/8	78/9	45/9	82/7	25/5
Cosmopolitan					1/1	1/1			
Tolerant	2/1		1/1			1/1			2/1
Hospitable	1/1		2/2	1/1		2/1	1/1	3/1	2/1
Reliable	2/1		2/1	1/1	2/1		1/1	2/1	
Hard-working						4/1		1/1	1/1
Safe	57/15	43/11	37/15	22/8	46/15	23/10	31/14	31/12	26/11
Rigorous	1/1	1/1					1/1	1/1	
Intelligent	11/6	9/7	12/4	5/5	16/8	6/4	11/6	1/1	3/2
Technical	2/2	2/2	2/1	1/1	5/3	1/1	1/1	1/1	3/1
Corporate	3/1	3/2	4/3	4/2	1/1	8/3	7/1	3/2	3/2
Innovative	7/3	20/4	6/2	9/2	10/5	8/3	14/3	7/3	4/1
Successful	5/3	2/1	5/2	6/2	8/2	6/2	4/3	3/2	1/1
Leader	66/16	61/9	73/14	54/12	72/16	71/13	65/15	57/11	49/12
Ambitious	4/2	1/1	7/4	4/2	6/3	1/1	2/2	1/1	5/4

(*Continued*)

Table 6. (Continued)

Subcategory	Lond	Berl	Dubl	Athe	Edin	Vien	Amst	Buda	Barc
#	**1103/261**	**945/222**	**1011/245**	**840/219**	**1082/266**	**893/206**	**943/229**	**885/199**	**760/192**
Powerful	2/2	16/6	11/3	7/3	12/5	4/3	12/4	7/3	5/2
Glamorous	66/6	41/6	39/6	44/6	44/6	58/5	41/4	42/6	33/4
Luxurious	9/3	5/2	10/3	4/3	14/4	7/2	6/3	3/2	6/3
Charming	1/1		3/2	1/1	4/2	1/1	1/1		1/1
Smooth	22/2	22/3	21/4	12/5	19/5	23/4	24/6	19/3	17/2
Romantic	5/1	3/1	4/1	11/1	3/1	7/1	8/1	6/1	9/1
Magical	12/2	9/2	8/1	5/1	20/2	1/1	6/2	12/1	2/1
Outdoorsy									
Get-away	7/5	7/5	8/5	9/4	10/6	4/4	13/7	5/3	7/6
Recreational	3/2	2/1	4/1	4/1	4/2		4/2	1/1	3/2
Tough	2/1		3/1	1/1	1/1	1/1	2/1	1/1	2/1
Rugged	60/10	22/10	60/9	20/7	52/10	26/8	38/11	36/8	32/9
Non-conformist	4/2	6/4	11/4	2/1	6/3	3/2	2/1	4/1	4/1

https://doi.org/10.1371/journal.pone.0206572.t006

The figures on the right of each cell show that, in general, DMOs use a greater number of different emotional adjectives in their tweets than the locals. This greater number of adjectives could be intentional on the part of the DMOs, because they could be trying to use more adjectives to communicate a value. If we look at the exceptions for particular emotional values, it can be seen that locals use a (very slightly) higher number of adjectives than DMOs just for the values *daring*, *spirited*, *dynamic* and *creative*. If we look at the exceptions for destinations, only the locals of London use a higher number of emotional adjectives than the destination. Therefore, there are very few exceptions indeed.

The leftmost figure in each cell is a more appropriate datum for the evaluation of the communication of the emotional values, because it shows the total number of times that a value is actually communicated. For example, a value can be communicated only with three different adjectives, but they can be used many times. If we look at these figures, with the exception of Dublin and Budapest, locals communicate more emotional values than DMOs. This poses two realities.

First, it may lead one to think that most DMOs of tourist destinations do not actually have an established brand communication strategy. These results coincide with others from previous studies [37, 90, 91] that have analysed various social media and various samples of destinations. All their results show the inexistence of brand communication strategies in the management of social media.

Secondly, if we focus only on the adjectives used in the tweets, these results show that locals or residents are better communicators of emotional values than DMOs. This fact would also coincide with the studies of participatory branding that claim that all relevant stakeholders (including residents) are important brand co-creators [21, 34–36]. Therefore, one of the contributions of this study, which was quite unexpected, is to demonstrate quantitatively a reality that was known, but that was difficult to measure or had not been measured in such a clear and quantifiable way.

4.1 Analysis per emotional value

In this subsection we first analyse the number of adjectives used to communicate each emotional value (the right numbers on the cells of Tables 6 and 7). After that, we focus on the total

Table 7. Use of emotional adjectives by locals (orange: high values—green: usual emotional value not communicated by the destination—yellow: emotional value communicated only by a few destinations).

Subcategory	Lond	Berl	Dubl	Athe	Edin	Vien	Amst	Buda	Barc
#	3565/289	1086/186	923/153	1495/197	1140/162	1251/202	1193/192	699/143	1054/171
Family-oriented		1/1						3/1	1/1
Down-to-earth	118/22	53/15	15/8	90/16	66/13	84/19	50/18	41/14	39/15
Sustainable	15/9	12/4	1/1	12/4	2/1	8/1	9/4	7/4	4/2
Calm	127/9	73/4	29/6	52/4	68/5	59/4	67/4	40/2	51/4
Real	28/6	18/6	18/5	27/5	10/4	17/5	17/5	9/4	13/4
Traditional	106/21	35/9	7/3	72/12	21/6	50/11	49/13	36/5	21/10
Honest	399/10	163/5	178/7	328/7	146/5	124/8	135/8	104/6	285/5
Original	114/23	42/15	13/10	50/15	16/9	59/19	32/13	19/14	41/14
Wholesome	4/1	1/1	1/1	2/1	5/2		2/1		
Quality of life									
Hapyness	57/6	11/4	8/6	6/4	40/3	22/4	29/4	16/3	15/3
Sentimental	25/10	10/4	12/5	12/8	10/4	8/5	6/4	8/5	3/3
Friendly	20/7	5/3	4/3	5/2	5/3	7/4	17/5	2/1	8/4
Trendy	23/6	13/3	12/2	31/3		18/2	17/3	55/2	39/2
Daring	40/8	13/7	9/4	18/4	19/6	14/6	22/7	13/5	33/7
Exciting	6/1	5/2	2/2	1/1	5/3		9/2	1/1	2/2
Exotic	5/2	4/2		4/1	1/1	2/2	12/2		2/2
Fashionable									
Cool									
Spirited	6/2	1/1	1/1	4/1	7/3	8/3	4/3		4/3
Dynamic	3/3	9/3		1/1	2/2	20/4	4/1	1/1	23/3
Vital			1/1		8/1				
Fresh	1141/6	208/6	70/7	210/6	131/4	224/6	218/6	123/3	56/5
Young	11/2	6/2		2/1	2/1	15/2	5/1	2/1	4/1
Sensorial	38/11	11/4	66/10	52/5	20/6	15/6	14/8	9/4	7/5
Unique	124/5	25/4	14/3	33/4	41/4	42/5	53/3	20/4	36/5
Imaginative						1/1			
Creative	29/7	35/6	10/5	21/7	3/3	20/7	13/2	4/3	4/4
Up-to-date	30/2	7/2	3/1	27/2	1/1	50/2	11/2	12/2	20/2
Independent	12/3	2/2	7/2	5/1	1/1	1/1	1/1		2/2
Contemporary	333/9	63/9	42/5	32/10	71/9	60/7	43/7	31/6	45/5
Cosmopolitan				1/1					
Tolerant			1/1	1/1				1/1	
Hospitable	1/1	1/1		5/2			2/1	1/1	
Reliable							1/1		
Hard-working	1/1		1/1		1/1				
Safe	63/14	21/9	39/8	56/11	24/10	51/12	40/11	20/8	37/9
Rigorous		1/1							3/1
Intelligent	23/7	22/8	8/5	23/5	7/4	11/3	30/6	3/3	16/3
Technical	9/1								1/1
Corporate	1/1	1/1	1/1	1/1	2/2		4/1	1/1	
Innovative	28/3	7/1	22/2	10/3	14/3	10/3	13/4	6/2	5/2
Successful	3/2	5/2		5/2	11/2	7/3	8/3	7/2	1/1
Leader	125/20	34/9	37/6	67/21	34/10	82/12	64/11	35/16	76/11
Ambitious	4/3	2/1	1/1		1/1	2/2	1/1		5/4

(*Continued*)

Table 7. (Continued)

Subcategory	Lond	Berl	Dubl	Athe	Edin	Vien	Amst	Buda	Barc
#	**3565/289**	**1086/186**	**923/153**	**1495/197**	**1140/162**	**1251/202**	**1193/192**	**699/143**	**1054/171**
Powerful	17/4	12/4	10/3	14/4	8/3	7/2	9/4	1/1	3/1
Glamorous	76/6	49/2	105/5	89/2	111/6	58/4	93/4	17/4	33/4
Luxurious	7/2	6/3	2/2	1/1	7/2	7/3	2/1	1/1	5/2
Charming					2/1	4/1	1/1	1/1	1/1
Smooth	17/4	9/3	18/2	28/3	8/2	12/5	6/2	11/2	46/3
Romantic	3/1	2/1	5/1	8/1	3/1		2/1		2/1
Magical	105/2	9/1	69/2	14/1	108/2	13/1	9/2	1/1	4/1
Outdoorsy									
Get-away	64/11	8/4	9/4	15/4	22/5	10/6	11/4	5/3	19/3
Recreational	5/3		3/1	1/1		1/1		2/1	10/2
Tough	1/1	1/1		1/1					
Rugged	192/7	69/9	65/6	58/7	76/7	46/8	56/6	30/4	27/7
Non-conformist	6/4	1/1	4/4			2/2	2/1		2/1

https://doi.org/10.1371/journal.pone.0206572.t007

number of emotional values that have been transmitted or communicated by the adjectives employed by DMOs or locals in their tweets (the values on the left on the same table).

Firstly, the numbers in Tables 6 and 7 show that the values communicated by a larger number of different adjectives are virtually the same in the tweets of destinations and in those of locals. Those that use a higher number of adjectives to be communicated are *down-to-earth*, *original*, *secure/safe* and *leader*. These values are indeed interesting both to destinations and locals alike, especially with regard to security/safety, leadership or originality. In theory, all destinations would like to be original and secure/safe. On the other hand, there are emotional values that are hardly communicated by any adjective (or not communicated at all) in the tweets of destinations and locals. It is surprising that values such as *quality of life*, *fashionable*, *cool* or *cosmopolitan* are not transmitted by any adjective, since they are values with which many destinations would probably like to be associated. This fact will have to be analysed in further detail, since it might be due to a poor choice of the nouns that represent these values in WordNet.

Secondly, if we look now at the values most communicated by tourist destinations and by locals, independently of the number of adjectives used, analysing the left figure of each cell, *honest* and *fresh* stand out, and they are far more communicated by locals than by DMOs. The locals of London, for example, communicate the value *fresh* 1141 times in their 3000 tweets when the destination only communicates it 156 times. This means that the locals of London truly have an image of the city to which they associate this value and the DMO should take this fact into account in its brand communication strategy.

Other values that are also quite communicated are *down-to-earth*, *calm*, *original*, *contemporary*, *secure/safe*, *leader*, *glamorous* and *rugged*. The majority of them are more communicated by locals than by tourist destinations, although there are some exceptions. The value *traditional*, for example, which is communicated by all destinations with low scores, shows very high scores in the tweets of the locals of London, Athens or Vienna. Thus, the locals of these three destinations consider that their cities are traditional while their DMOs hardly communicate this fact at all. The *glamorous* emotional value is also communicated with medium frequency by all DMOs of the analysed destinations; however, the residents of Edinburgh, Dublin and Amsterdam associate this value with their cities with frequencies that triple those of the destinations. Just to mention another example, the residents of London communicate the

contemporary value six times more frequently than the destination itself. In these cases, in which there is a distinguishing feature attributed by locals to some tourist destinations, these emotional values acquire a distinctive role of authenticity, because the locals know the identities of the destinations better than anyone. Therefore, the DMOs should be highly receptive to these values and take them into account in the management of the identity and the brand of their destination. Although locals and DMOs can have different images of a destination, it is important to guarantee a certain level of coherence between them, which will help to build a strong brand image and a successful positioning.

The opposite situation appears in the value *happiness*, that is communicated with average scores by all of the DMOs whereas it is hardly communicated by the locals of Athens and Dublin. In the case of Athens, the severe economic crisis the country has suffered during the last few years added to the current immigration problems may have been the reasons why the locals of Athens do not associate happiness with this city, because the score is really very low compared to those of the locals of the rest of the analysed cities.

In some cases it may be seen that a destination communicates an emotional value much more frequently than the locals (e.g. the *contemporary* value in Budapest). This effect probably occurs when a DMO aims to promote an image that is desired but does not correspond fully to reality. Instead, locals offer a far more real image of the city.

4.2 Analysis per destination

If we now analyse the number of adjectives that communicate emotional values in each destination, it may be noted that the DMOs that use the highest number of adjectives are Edinburgh, London and Dublin (These are the English-speaking destinations in the dataset. The effect of the native tongue of the destination should probably be studied more carefully). Conversely, the locals who use most adjectives are those of London and Vienna. However, as said before, these figures only show a communicative intention, but they do not reflect all the communication of emotional values.

So, now we focus on the first figure of each cell, which shows the total number of times an emotional value has been communicated. In this case, the DMOs that communicate more emotional values with their tweets are again London, Edinburgh and Dublin, in this order, and the locals that communicate more values are those of London, Athens and Vienna. The locals of London triple the values communicated by the London DMO and those of Athens double the values of the destination. Thus, it seems that the DMOs of these three destinations (London, Edinburgh and Dublin) have a communication strategy through Twitter. Now we will analyse which are the values they communicate, whether they are differential values, and if they are shared by their locals.

London communicates the values *honest* and *fresh* with high frequencies, and *down-to-earth*, *original*, *contemporary*, *leader* and *rugged* with medium frequencies. All these emotional values are the most communicated by almost all major tourist destinations. Therefore, the existence of a differentiated brand communication strategy that aims to separate the destination from the rest and identify it with original values is not apparent. However, the locals do communicate emotional values with very different frequencies from those of the DMO and other destinations. For example, the locals of London associate the destination with the value *rugged* with a frequency that triples the one of the rest of cities and that of the DMO. This does seem to be a characteristic value of the destination according to its locals. The same thing happens with the values *leader* and *original*.

The Edinburgh DMO, for example, also communicates the same generic emotional values than the rest of DMOs. Therefore, it does not show either a strategic and differentiated brand

communication strategy. However, the tweets of locals do communicate differentiating values. For example, *calm* is communicated with average frequency, but locals double this frequency. They also double the frequencies of the values *glamorous* and *unique/diverse*. Thus, they communicate that their city is quiet, unique, diverse and glamorous. On the other hand, the DMO communicates the values *secure/safe* and *leader* with medium and high scores and the locals show frequencies much lower than those of the DMOs. Thus, the locals of Edinburgh do not communicate that their city is safe or a leader. The DMO should take all of this into account and analyse whether it communicates the reality of the territory, try to solve the existing problems, reposition the identity and the brand of the destination and then communicate these values powerfully.

Something very similar is seen in the communication of emotional values by the city of Dublin. The values communicated by the DMO are those shared by most destinations. However, the different values communicated by its locals distinguish the territory. The DMO communicates with medium-high frequencies the values *down-to-earth* and *contemporary* but its locals communicate these values with low frequencies. In contrast, the DMO communicates *sensorial* and *glamorous* with medium-low frequencies whereas locals communicate these values with high frequencies. Thus, the locals of Dublin communicate differential and unique brand values while the DMO merely communicates undistinguishing values.

We have also measured the Pearson correlation between the data of the DMOs for each emotional value, studying in a separate way the number of adjectives and the total number of uses of the adjectives. The results are shown in the symmetric matrices depicted in Fig 2 in a graphical and a numerical form.

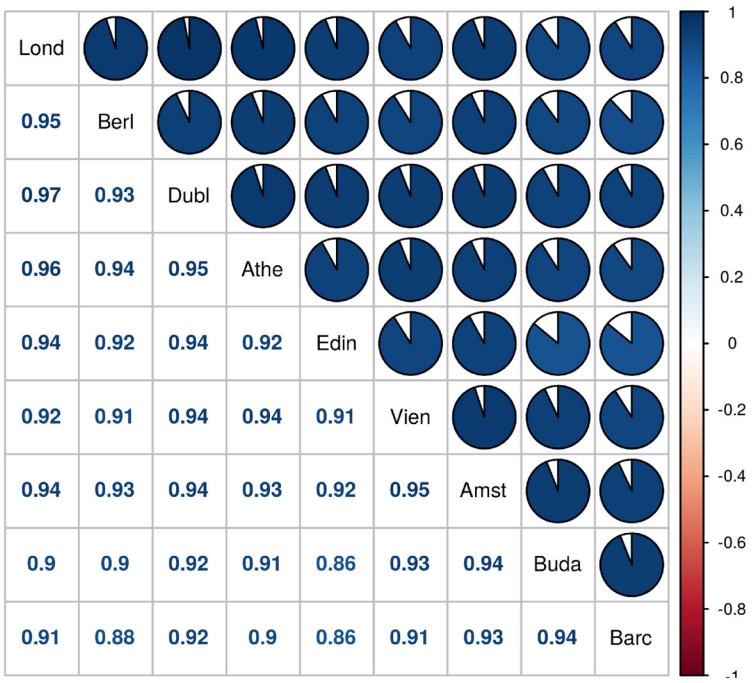
It may be seen that the correlation between the number of adjectives used for each emotional value is quite high (between 0.86 and 0.97), whereas the correlation between the total number of uses of the adjectives is even higher (between 0.92 and 0.98). Thus, it may be concluded that there are not very significant differences between DMOs in this respect.

4.3 Comparison between DMOs and locals

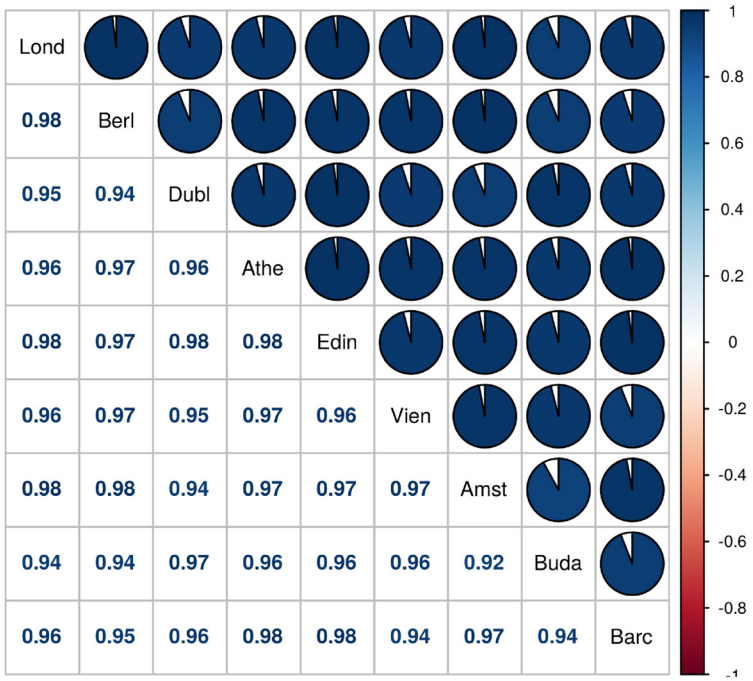
Figs 3 and 4 provide a graphical representation of the difference in the communication of emotional values among DMOs and local people. The first one shows, for each emotional value, the difference between the average numbers of adjectives used by DMOs and by locals. It is easily seen that, in the majority of cases, DMOs use more adjectives, being *down-to-earth*, *sentimental*, *happiness* and *calm* the most prominent cases. There are only a few values for which DMOs have a slightly higher number of emotional adjectives, being *traditional* and *unique* the top ones.

In turn, Fig 4 makes the same analysis but with the difference on the average number of total uses of adjectives referring to each one of the emotional values. As commented before, in this case the superiority is on the DMO's side, very specially in the *fresh* concept. Other values heavily mentioned by DMOs include *honest*, *calm*, *traditional*, *unique*, *rugged* and *magical*. On the other hand, there are a few values more mentioned by locals, in special *happiness*, *real* and *sentimental*.

As in the case of DMOs, we also measured the Pearson correlation between the different cities, taking into account the number of adjectives and the uses of those adjectives for each emotional value. The results are shown in Fig 5. The differences among cities are much more significant than in the case of DMOs commented in the previous section. Concretely, the correlation between the number of adjectives used for each emotional value ranges from 0.75 (between Dublin and Athena/Budapest) to 0.97 (between Barcelona and Vienna). The differences are even higher when the total number of uses of the adjectives is considered. In this



(a) Correlation between the number of adjectives.



(b) Correlation between the number of uses.

Fig 2. Correlation between the data of each DMO.

https://doi.org/10.1371/journal.pone.0206572.g002

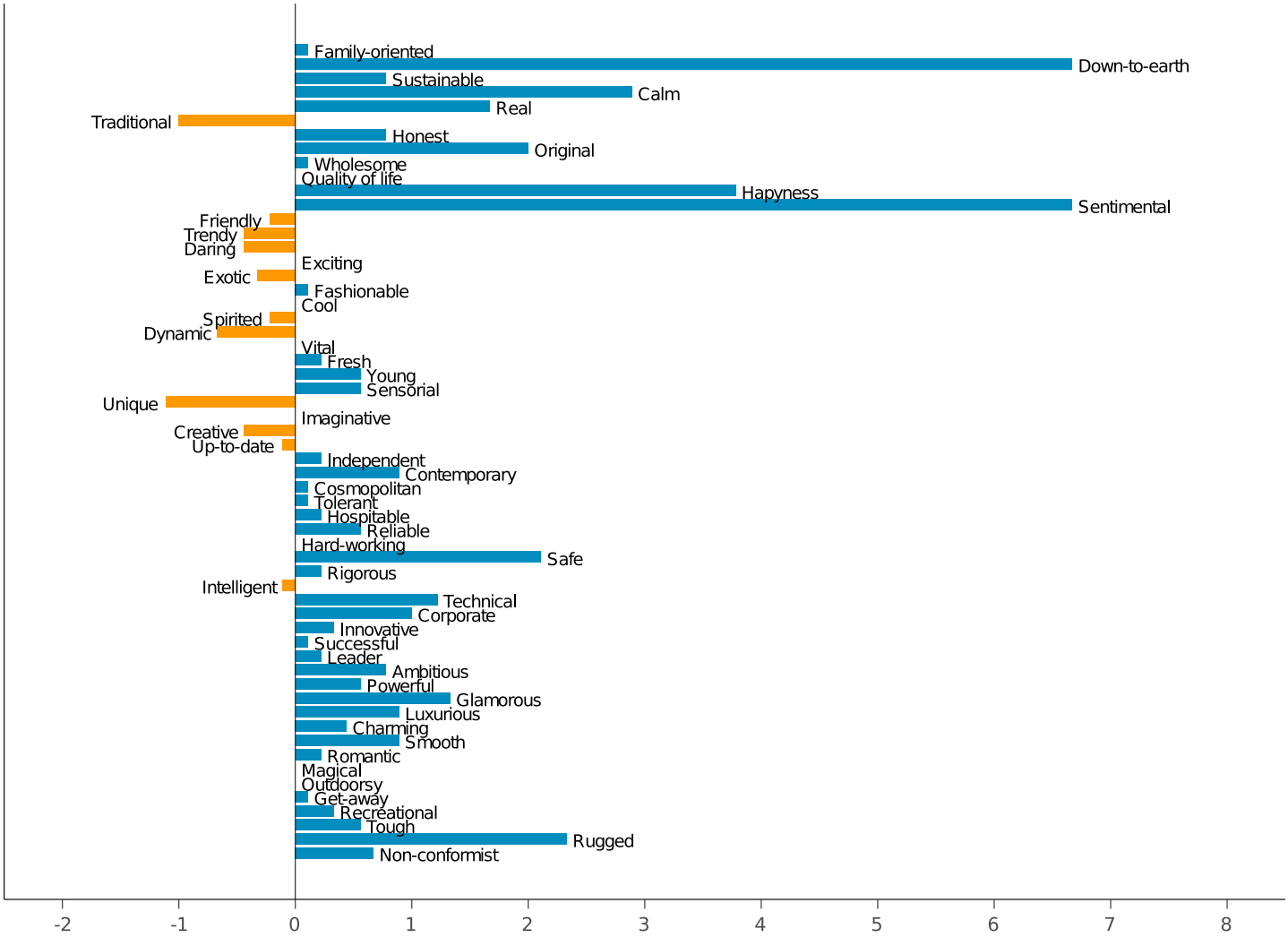


Fig 3. Difference on the number of emotional adjectives between DMOs and local citizens.

https://doi.org/10.1371/journal.pone.0206572.g003

case, the lowest correlation is 0.48 (between London and Barcelona) and the highest one is 0.95 (between Amsterdam and Berlin/Vienna).

Conclusions

The results of the study show that DMOs, despite using a greater variety of adjectives, generally communicate less emotional values than locals. In addition, no great differences are observed in the use of these values. That is, the values communicated by all DMOs have a high coincidence. Thus, few distinctive features are observed. As a result, the first conclusion of the study seems to be that DMOs do not have a clearly established brand communication and differentiation strategy. The reasons may be that they prefer to communicate activities and agenda than the brand in their online communication, that they do not focus on emotional values in brand communication, or that they do not establish guidelines for content communication specifying the terms to be used. Other studies have also shown that many DMOs prefer to communicate the tourist attractions or the city agenda than the emotional values and, therefore, values end up being less communicated [7, 90].

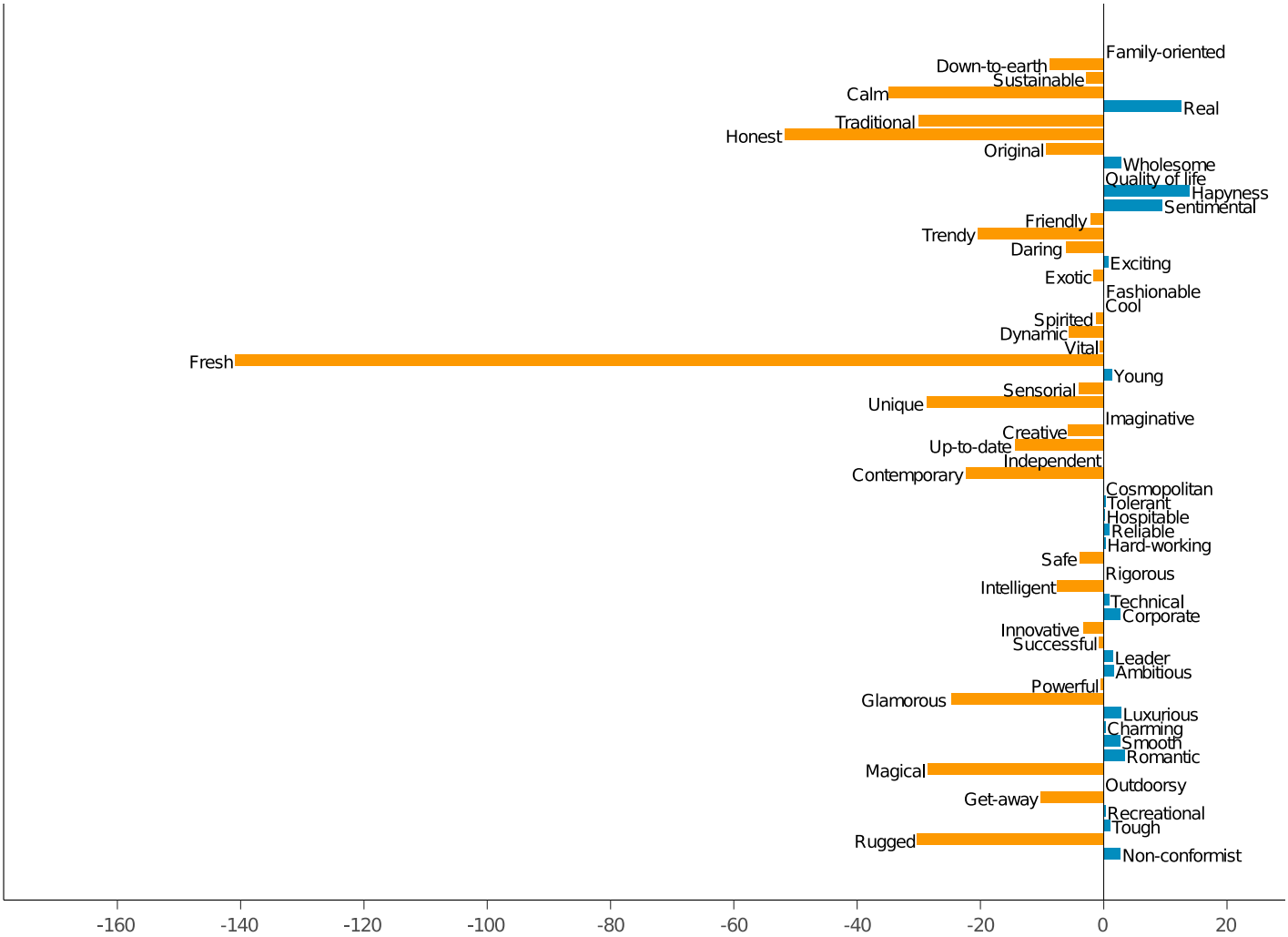
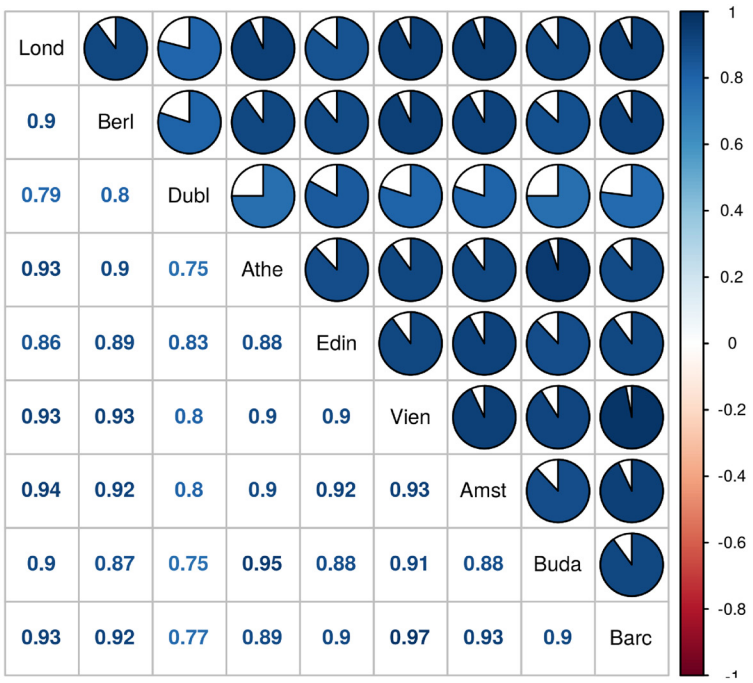


Fig 4. Difference on the uses of emotional adjectives between DMOs and local citizens.

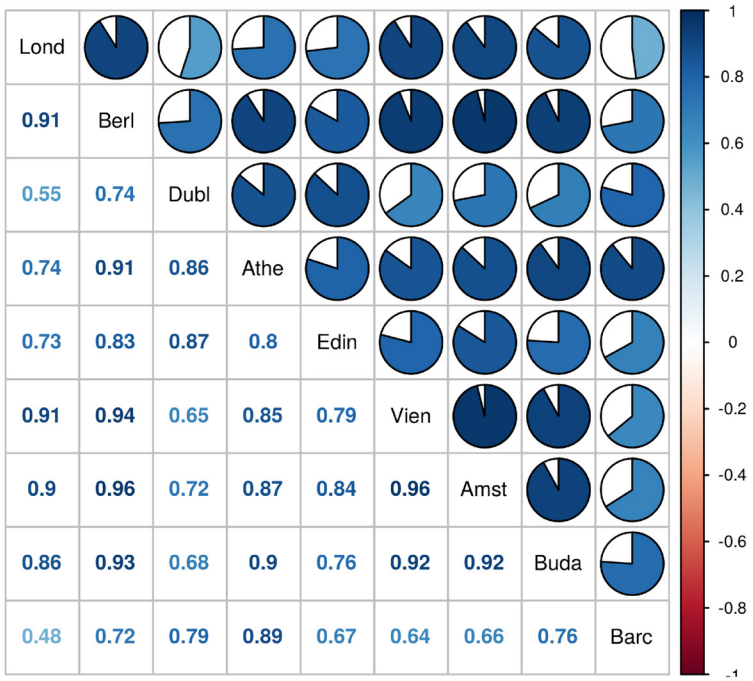
https://doi.org/10.1371/journal.pone.0206572.g004

So, the first recommendation for DMOs is to take great care of the communication of the destination brand. The brand and the emotional values it associates to the regions are critical in distinguishing the destination from others [2, 4, 12, 14–16] and in creating partnerships and relationships with the locals [17, 18]. There is a need to take care of the communication of emotional values through all communication channels and especially through social media. As mentioned above, these new communication channels have great potential for emotional communication and they should not be underused [50, 92]. Moreover, for the excellent communication of certain emotional values there must be a strategy to communicate specific content, which sets the terms to use in communications through social media. Given that Twitter only allows sharing 140 characters, the communication through this microblog must take extreme care of each of the employed words.

Furthermore, the results show that locals, contrary to what we expected to find, communicate more emotional values than DMOs. In addition, they do so using fewer adjectives, and they sometimes show special differential emotional values that set their destination apart from the rest and are not shared by all destinations. These results, as mentioned above, confirm the



(a) Correlation between the number of adjectives.



(b) Correlation between the number of uses.

Fig 5. Correlation between the data of local residents.

https://doi.org/10.1371/journal.pone.0206572.g005

results of studies on participatory branding that assert that users are important brand co-creators [21, 34–36]. In fact, another contribution of this study has been to demonstrate that this participation in the on-line creation of a distinctive brand is indeed real and, moreover, quantifiable and measurable. It should also be borne in mind that the comments and the image that locals have of their town/city provide a twofold additional credibility. First, because they know the region better than anyone, and second, because, in principle, they have no vested economic or commercial interest in the association of emotional values to the territory.

All this leads to a second implication of the study for DMOs of tourist destinations. Since locals co-create brands through their comments and experiences in social media, and because their knowledge of the area is extensive and real, the DMOs should listen to and analyse what locals communicate, especially regarding the communication of the brand and emotional values, to carry out a more coherent destination branding and brand communication.

In addition, the results of the study show the existence of major differences between the brand emotional values communicated by DMOs and those communicated by the locals themselves. Therefore, DMOs need to analyse and consider the communication of brand values by locals, assess them and take into account them for branding strategy. In fact, to avoid inconsistencies in communication, as many studies and authors say [8–11, 19, 39, 40], the branding of tourist destinations should be more participatory, and this poses another challenge for DMOs. Therefore, this work is also in line with these previous studies. Another implication that can be drawn is that DMOs should include the participation of all the locals in the branding process from the outset. Only in this way will they ensure more coherent branding and communication of brand emotional values.

Finally, it is worth pointing out some limitations of our current approach that might lead to lines of future research:

- The system only analyses tweets in English. It could be conceivable to think about the possibility of using automatic translation mechanisms in order to analyse tweets in other languages. This idea is not very far-fetched if only adjectives are analysed. It would also be interesting to make more exhaustive independent studies for English-speaking and non-English-speaking destinations, to assess the possible differences in the usage of adjectives. In this study it may be observed that the DMOs of the English-speaking destinations (London, Edinburgh and Dublin) have the three top positions in the number of adjectives and the number of uses of adjectives; however, when the tweets of locals are considered, then London keeps the first place but Edinburgh and Dublin have very low positions (e.g. 7th and 8th in the number of adjectives).
- We are only taking into account the adjectives present in the tweets, disregarding other information in the text that could be important. In particular, we are not analysing whether the reference to the emotional value is positive or negative. Moreover, we are not studying the possible irony implicit in the tweet, which could be radically changing the meaning of the message. It would also be interesting to study the adverbs present in the tweets.
- Concerning the tweets sent from locals, in the present study it is not possible to ensure that they are actually talking about the destination itself (although a manual analysis of a representative sample indicated that almost 3 out of 4 tweets did indeed refer to local aspects of the destination). Therefore, it would be very interesting to think about possible fliters that could be applied to the dataset to remove tweets that are unrelated to the city.
- We are not analysing other pieces of information which Twitter users employ to transmit feelings and emotions, like images or videos. A picture (for example, a couple kissing in the beach at sunset) can convey a much stronger emotional impact than 140 characters. It might

be the case that some of the analysed cities prefer using images to text, and the current system is not able to analyse this information.

- The initial selection of WordNet nouns to refer to each emotional value is quite relevant, and a change of words could lead to an important change in the final results. Moreover, different steps in the analytic process (tweet parsing, part-of-speech tagging, automatic association of nouns to the adjectives appearing in the tweets) certainly may introduce errors in the process.
- The study only analyses how destinations and locals transmit emotional values. One of our immediate research lines is the study of the tweets of visitors to analyze which emotional values are actually perceived in the destination.
- It would also be possible to apply the same methodology to analyze tweets referring to attractions, restaurants, hotels, etc.

Author Contributions

Conceptualization: Assumpció Huertas, Antonio Moreno.

Data curation: Mohammed Jabreel.

Formal analysis: Antonio Moreno.

Funding acquisition: Antonio Moreno.

Investigation: Mohammed Jabreel.

Methodology: Mohammed Jabreel, Antonio Moreno.

Resources: Assumpció Huertas.

Software: Mohammed Jabreel.

Supervision: Antonio Moreno.

Visualization: Mohammed Jabreel.

Writing – original draft: Mohammed Jabreel, Assumpció Huertas, Antonio Moreno.

Writing – review & editing: Antonio Moreno.

References

1. Morgan N, Pritchard A, Piggott R. New Zealand, 100% Pure. The creation of a powerful niche destination brand. Journal of Brand Management. 2002; 9(4):335–354. https://doi.org/10.1057/palgrave.bm.2540082
2. Govers R, Go F. Place branding: virtual and physical identities, glocal, imagined and experienced. Basingstoke, UK: Palgrave Macmillan; 2009.
3. Morgan NJ, Pritchard A, Piggott R. Destination branding and the role of the stakeholders: The case of New Zealand. Journal of Vacation Marketing. 2003; 9(3):285–299. https://doi.org/10.1177/135676670300900307
4. Morgan N, Pritchard A. Meeting the destination branding challenge. Destination branding. 2004; p. 59–79.
5. Blain C. Destination Branding: Insights and Practices from Destination Management Organizations. Journal of Travel Research. 2005; 43(4):328–338. https://doi.org/10.1177/0047287505274646
6. Govers R, Go FM, Kumar K. Promoting Tourism Destination Image. Journal of Travel Research. 2007; 46(1):15–23. https://doi.org/10.1177/0047287507302374

7. Huertas A, Mariné-Roig E. Destination Brand Communication Through the Social Media: What Contents Trigger Most Reactions of Users? In: Information and Communication Technologies in Tourism 2015. Springer Science + Business Media; 2015. p. 295–308. Available from: http://dx.doi.org/10.1007/978-3-319-14343-9_22.
8. Hatch MJ, Schultz M. Toward a theory of brand co-creation with implications for brand governance. Journal of Brand Management. 2010; 17(8):590–604. https://doi.org/10.1057/bm.2010.14
9. Aitken R, Campelo A. The four Rs of place branding. Journal of Marketing Management. 2011; 27(9-10):913–933. https://doi.org/10.1080/0267257X.2011.560718
10. Houghton JP, Stevens A. City branding and stakeholder engagement. In: City branding. Springer; 2011. p. 45–53.
11. Kavaratzis M, Kalandides A. Rethinking the place brand: the interactive formation of place brands and the role of participatory place branding. Environment and Planning A. 2015; 47(6):1368–1382. https://doi.org/10.1177/0308518X15594918
12. Cai LA. Cooperative branding for rural destinations. Annals of tourism research. 2002; 29(3):720–742. https://doi.org/10.1016/S0160-7383(01)00080-9
13. Chaykina T, Guerreiro M, Mendes J. Destination Brand Personality of Portugal for the Russian-Speaking Market. Journal of Spatial and Organizational Dynamics. 2014; 2(1):23–40.
14. Blain C, Levy SE, Ritchie JB. Destination branding: Insights and practices from destination management organizations. Journal of travel research. 2005; 43(4):328–338. https://doi.org/10.1177/0047287505274646
15. Kavaratzis M, Ashworth GJ. City branding: an effective assertion of identity or a transitory marketing trick? Tijdschrift voor economische en sociale geografie. 2005; 96(5):506–514. https://doi.org/10.1111/j.1467-9663.2005.00482.x
16. Ekinci Y, Hosany S. Destination Personality: An Application of Brand Personality to Tourism Destinations. Journal of Travel Research. 2006; 45(2):127–139. https://doi.org/10.1177/0047287506291603
17. Hankinson G. Relational network brands: Towards a conceptual model of place brands. Journal of Vacation Marketing. 2004; 10(2):109–121. https://doi.org/10.1177/135676670401000202
18. Laroche M, Habibi MR, Richard MO. To be or not to be in social media: How brand loyalty is affected by social media? International Journal of Information Management. 2013; 33(1):76–82. https://doi.org/10.1016/j.ijinfomgt.2012.07.003
19. Kavaratzis M. From "necessary evil" to necessity: stakeholders' involvement in place branding. Development. 2012; 5(1):7–19.
20. Aragón EP, Domingo D. Developing Public Relations 2.0: Practitioners' perceptions on the implementation of interactive communication strategies. Public Relations Review. 2014; 40(3):559–561. https://doi.org/10.1016/j.pubrev.2014.04.001
21. Hollebeek LD, Glynn MS, Brodie RJ. Consumer Brand Engagement in Social Media: Conceptualization, Scale Development and Validation. Journal of Interactive Marketing. 2014; 28(2):149–165. https://doi.org/10.1016/j.intmar.2013.12.002
22. Baker B. Destination branding for small cities: The essentials for successful place branding. Destination Branding Book; 2007.
23. Warnaby G. Towards a service-dominant place marketing logic. Marketing Theory. 2009; 9(4): 403–423. https://doi.org/10.1177/1470593109346898
24. Kavaratzis M, Braun E, Zenker S. My city—my brand: the role of residents in place branding. In: 50th European Regional Science Association Congress, Jönköping; 2010. p. 19–23.
25. Murphy L, Mascardo G, Benckendorff P. Exploring word-of-mouth influences on travel decisions: friends and relatives vs. other travellers. International Journal of Consumer Studies. 2007; 31(5): 517–527. https://doi.org/10.1111/j.1470-6431.2007.00608.x
26. Litvin SW, Goldsmith RE, Pan B. Electronic word-of-mouth in hospitality and tourism management. Tourism Management. 2008; 29(3):458–468. https://doi.org/10.1016/j.tourman.2007.05.011
27. Fotis J, Buhalis D, Rossides N. Social media use and impact during the holiday travel planning process. Springer-Verlag; 2012.
28. Leung D, Law R, van Hoof H, Buhalis D. Social Media in Tourism and Hospitality: A Literature Review. Journal of Travel & Tourism Marketing. 2013; 30(1-2):3–22. https://doi.org/10.1080/10548408.2013.750919
29. Huertas A. La comunicación de los territorios, los destinos y sus marcas. Guía práctica de aplicación desde las relaciones públicas. UOC Ediciones/Colección DIRCOM; 2014.
30. Merrilees B, Miller D, Herington C. Antecedents of residents' city brand attitudes. Journal of Business Research. 2009; 62(3):362–367. http://dx.doi.org/10.1016/j.jbusres.2008.05.011.

31. Ashworth G, Kavaratzis M. Beyond the logo: Brand management for cities. Journal of Brand Management. 2009; 16(8):520–531. https://doi.org/10.1057/palgrave.bm.2550133
32. Hanna S, Rowley J. Towards a strategic place brand-management model. Journal of Marketing Management. 2011; 27(5-6):458–476. https://doi.org/10.1080/02672571003683797
33. Kavaratzis M, Hatch MJ. The dynamics of place brands: An identity-based approach to place branding theory. Marketing Theory. 2013; 13(1):69–86. https://doi.org/10.1177/1470593112467268
34. Gensler S, Völckner F, Liu-Thompkins Y, Wiertz C. Managing brands in the social media environment. Journal of Interactive Marketing. 2013; 27(4):242–256. https://doi.org/10.1016/j.intmar.2013.09.004
35. Labrecque LI. Fostering Consumer—Brand Relationships in Social Media Environments: The Role of Parasocial Interaction. Journal of Interactive Marketing. 2014; 28(2):134–148. https://doi.org/10.1016/j.intmar.2013.12.003
36. Park H, Kim YK. The role of social network websites in the consumer—brand relationship. Journal of Retailing and Consumer Services. 2014; 21(4):460–467. https://doi.org/10.1016/j.jretconser.2014.03.011
37. Moreno A, Jabreel M, Huertas A. Automatic Analysis of the Communication of Tourist Destination Brands through Social Networks. In: 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). Institute of Electrical & Electronics Engineers (IEEE); 2015. p. 546—553. Available from: http://dx.doi.org/10.1109/iske.2015.22.
38. Valentini C. Is using social media "good" for the public relations profession? A critical reflection. Public Relations Review. 2015; 41(2):170–177. https://doi.org/10.1016/j.pubrev.2014.11.009
39. Gregory A. Involving Stakeholders in Developing Corporate Brands: the Communication Dimension. Journal of Marketing Management. 2007; 23(1-2):59–73. https://doi.org/10.1362/026725707X178558
40. Ind N, Bjerke R. Branding governance: A participatory approach to the brand building process. John Wiley & Sons; 2007.
41. Konecnik M, Lehu JM. Tourism destination brand identity: The case of Slovenia. Journal of Brand Management. 2008; 15(3):177–189. https://doi.org/10.1057/palgrave.bm.2550114
42. Ruzzier MK. Country brands and identity: Slovenia. In: Morgan N, Pride AP, editors. Destination Brands (Third Edition). third edition ed. Oxford: Butterworth-Heinemann; 2011. p. 291–302.
43. Buhalis D. Marketing the competitive destination of the future. Tourism Management. 2000; 21(1): 97–116. https://doi.org/10.1016/S0261-5177(99)00095-3
44. Aaker D, Kumar V, Day G. Marketing research. Wiley; 2004.
45. Aaker J, Fournier S, Brasel SA. When Good Brands Do Bad. Journal of Consumer Research. 2004; 31(1):1–16. https://doi.org/10.1086/383419
46. Prahalad CK. The cocreation of value. Journal of Marketing. 2004; 68(1):23.
47. Nambisan S, Baron RA. Interactions in virtual customer environments: Implications for product support and customer relationship management. Journal of Interactive Marketing. 2007; 21(2):42–62. https://doi.org/10.1002/dir.20077
48. Klijn EH, Eshuis J, Braun E. The Influence of Stakeholder Involvement on The Effectiveness of Place Branding. Public Management Review. 2012; 14(4):499–519. https://doi.org/10.1080/14719037.2011.649972
49. Hennig-Thurau T, Malthouse EC, Friege C, Gensler S, Lobschat L, Rangaswamy A, et al. The Impact of New Media on Customer Relationships. Journal of Service Research. 2010; 13(3):311–330. https://doi.org/10.1177/1094670510375460
50. Xiang Z, Gretzel U. Role of social media in online travel information search. Tourism Management. 2010; 31(2):179–188. https://doi.org/10.1016/j.tourman.2009.02.016
51. Munar AM, Jacobsen JKS. Motivations for sharing tourism experiences through social media. Tourism Management. 2014; 43:46–54. https://doi.org/10.1016/j.tourman.2014.01.012
52. Buhalis D. Strategic use of information technologies in the tourism industry. Tourism Management. 1998; 19(5):409–421. https://doi.org/10.1016/S0261-5177(98)00038-7
53. Buhalis D, Licata MC. The future eTourism intermediaries. Tourism Management. 2002; 23(3):207–220. https://doi.org/10.1016/S0261-5177(01)00085-1
54. Pan B, Fesenmaier DR. Online information search: vacation planning process. Annals of Tourism Research. 2006; 33(3):809–832. https://doi.org/10.1016/j.annals.2006.03.006
55. Buhalis D, Law R. Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tourism Management. 2008; 29(4):609–623. https://doi.org/10.1016/j.tourman.2008.01.005

56. Wu SI, Wei PL, Chen JH. Influential factors and relational structure of Internet banner advertising in the tourism industry. Tourism Management. 2008; 29(2):221–236. https://doi.org/10.1016/j.tourman.2007.03.020

57. Llodrà-Riera I, Martínez-Ruiz MP, Jiménez-Zarco AI, Izquierdo-Yusta A. A multidimensional analysis of the information sources construct and its relevance for destination image formation. Tourism Management. 2015; 48:319–328. https://doi.org/10.1016/j.tourman.2014.11.012

58. Huang CY, Chou CJ, Lin PC. Involvement theory in constructing bloggers' intention to purchase travel products. Tourism Management. 2010; 31(4):513–526. https://doi.org/10.1016/j.tourman.2009.06.003.

59. Xiang Z, Magnini VP, Fesenmaier DR. Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. Journal of Retailing and Consumer Services. 2015; 22:244–249. https://doi.org/10.1016/j.jretconser.2014.08.005.

60. Pabel A, Prideaux B. Social media use in pre-trip planning by tourists visiting a small regional leisure destination. Journal of Vacation Marketing. 2016; 22(4):335–348. https://doi.org/10.1177/1356766715618998

61. Prebensen N, Skallerud K, Chen JS. Tourist Motivation with Sun and Sand Destinations: Satisfaction and the Wom-Effect. Journal of Travel & Tourism Marketing. 2010; 27(8):858–873. https://doi.org/10.1080/10548408.2010.527253

62. Munar AM, Jacobsen JKS. Trust and Involvement in Tourism Social Media and Web-Based Travel Information Sources. Scandinavian Journal of Hospitality and Tourism. 2013; 13(1):1–19. https://doi.org/10.1080/15022250.2013.764511

63. Ladhari R, Michaud M. eWOM effects on hotel booking intentions, attitudes, trust, and website perceptions. International Journal of Hospitality Management. 2015; 46:36–45. https://doi.org/10.1016/j.ijhm.2015.01.010.

64. Lee W, Gretzel U. Designing persuasive destination websites: A mental imagery processing perspective. Tourism Management. 2012; 33(5):1270–1280. https://doi.org/10.1016/j.tourman.2011.10.012

65. Kim SB, Kim DY, Bolls P. Tourist mental-imagery processing: Attention and arousal. Annals of Tourism Research. 2014; 45:63–76. https://doi.org/10.1016/j.annals.2013.12.005

66. Gartner WC. Image formation process. Journal of travel & tourism marketing. 1994; 2(2-3):191–216. https://doi.org/10.1300/J073v02n02_12

67. Pan B, Li XR. The long tail of destination image and online marketing. Annals of Tourism Research. 2011; 38(1):132–152. https://doi.org/10.1016/j.annals.2010.06.004

68. Lai C, Li XR, Harrill R. Chinese outbound tourists' perceived constraints to visiting the United States. Tourism Management. 2013; 37:136–146. https://doi.org/10.1016/j.tourman.2013.01.014

69. Marchiori E, Önder I. Reframing the Image of a Destination: A Pre-Post Study on Social Media Exposure. In: Information and Communication Technologies in Tourism 2015. Springer Science + Business Media; 2015. p. 335–347. Available from: http://dx.doi.org/10.1007/978-3-319-14343-9_25.

70. Oliveira E, Panyik E. Content, context and co-creation: Digital challenges in destination branding with references to Portugal as a tourist destination. Journal of Vacation Marketing. 2015; 21(1):53–74. https://doi.org/10.1177/1356766714544235

71. Hays S, Page SJ, Buhalis D. Social media as a destination marketing tool: its use by national tourism organisations. Current Issues in Tourism. 2013; 16(3):211–239. https://doi.org/10.1080/13683500.2012.662215

72. Roberts CW. A conceptual framework for quantitative text analysis. Quality and Quantity. 2000; 34(3):259–274. https://doi.org/10.1023/A:1004780007748

73. Dickinger A, Lalicic L. "This City Is Absolutely Fun and Trendy" A Destination Brand Personality Analysis in a Web 2.0 Setting. In: Information and Communication Technologies in Tourism 2015. Springer Science + Business Media; 2015. p. 321–333. Available from: http://dx.doi.org/10.1007/978-3-319-14343-9_24.

74. Moya MD, Jain R. When tourists are your "friends": Exploring the brand personality of Mexico and Brazil on Facebook. Public Relations Review. 2013; 39(1):23–29. https://doi.org/10.1016/j.pubrev.2012.09.004

75. Sevin HE. Understanding cities through city brands: City branding as a social and semantic network. Cities. 2014; 38:47–56. https://doi.org/10.1016/j.cities.2014.01.003

76. Serna A, Marchiori E, Gerrikagoitia JK, Alzua-Sorzabal A, Cantoni L. An Auto-Coding Process for Testing the Cognitive-Affective and Conative Model of Destination Image. In: Information and Communication Technologies in Tourism 2015. Springer Science + Business Media; 2015. p. 111–123. Available from: http://dx.doi.org/10.1007/978-3-319-14343-9_9.

77. Serna A, Gerrikagoitia JK, Bernabé U. Discovery and Classification of the Underlying Emotions in the User Generated Content (UGC). In: Information and Communication Technologies in Tourism 2016.

Springer Science + Business Media; 2016. p. 225–237. Available from: http://dx.doi.org/10.1007/978-3-319-28231-2_17.

78. Stepchenkova S, Kirilenko AP, Morrison AM. Facilitating Content Analysis in Tourism Research. Journal of Travel Research. 2009; 47(4):454–469. https://doi.org/10.1177/0047287508326509

79. Andéhn M, Kazeminia A, Lucarelli A, Sevin E. User-generated place brand equity on Twitter: The dynamics of brand associations in social media. Place Branding and Public Diplomacy. 2014; 10(2): 132–144. https://doi.org/10.1057/pb.2014.8

80. Költringer C, Dickinger A. Analyzing destination branding and image from online sources: A web content mining approach. Journal of Business Research. 2015; 68(9):1836–1843. https://doi.org/10.1016/j.jbusres.2015.01.011

81. Henderson JC, et al. Selling places: the new Asia-Singapore brand. Journal of Tourism Studies. 2000; 1(11):36–44.

82. Aaker JL. Dimensions of brand personality. Journal of marketing research. 1997; 34(3):347–356. https://doi.org/10.2307/3151897

83. Stepchenkova S, Morrison AM. The destination image of Russia: From the online induced perspective. Tourism Management. 2006; 27(5):943–956. https://doi.org/10.1016/j.tourman.2005.10.021

84. Bird S, Klein E, Loper E. Natural language processing with Python. " O'Reilly Media, Inc."; 2009.

85. Sánchez D, Batet M, Isern D, Valls A. Ontology-based semantic similarity: A new feature-based approach. Expert Systems with Applications. 2012; 39(9):7718–7728. https://doi.org/10.1016/j.eswa.2012.01.082

86. Wu Z, Palmer M. Verbs semantics and lexical selection. In: Proceedings of the 32nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics (ACL); 1994. p. 133–138. Available from: http://dx.doi.org/10.3115/981732.981751.

87. Slimani T. Description and Evaluation of Semantic Similarity Measures Approaches. International Journal of Computer Applications. 2013; 80(10):25–33. https://doi.org/10.5120/13897-1851

88. Budanitsky A, Hirst G. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. Computational Linguistics. 2006; 32(1):13–47. https://doi.org/10.1162/coli.2006.32.1.13

89. Lin D. An information-theoretic definition of similarity. vol. 98; 1998.

90. Míguez-González MI, Huertas A. The power of photographs in the communication and public relations of tourist destinations and their brands through Facebook and Flickr. Catalan Journal of Communication & Cultural Studies. 2015; 7(2):197–215. https://doi.org/10.1386/cjcs.7.2.197_1

91. Huertas A, Mariné-Roig E. Differential Destination Content Communication Strategies Through Multiple Social Media. In: Information and Communication Technologies in Tourism 2016. Springer Science + Business Media; 2016. p. 239–252. Available from: http://dx.doi.org/10.1007/978-3-319-28231-2_18.

92. Inversini A, Buhalis D. Information Convergence in the Long Tail: The Case of Tourism Destination Information. In: Information and Communication Technologies in Tourism 2009. Springer Science + Business Media; 2009. p. 381–392. Available from: http://dx.doi.org/10.1007/978-3-211-93971-0_32.

Do DMOs Communicate Their Emotional Brand Values? A Comparison Between Twitter and Facebook

Lidija Lalicic[1](✉), Assumpció Huertas[2], Antonio Moreno[3], Stefan Gindl[4], and Mohammed Jabreel[3]

[1] Department of Tourism and Service Management, MODUL University Vienna, Vienna, Austria
lidija.lalicic@modul.ac.at

[2] Departament d'Estudis de Comunicació, Universitat Rovira i Virgili, Tarragona, Spain
sunsi.huertas@urv.cat

[3] Department d'Enginyeria Informàtica i Matemàtiques, Intelligent Technologies for Advanced Knowledge Acquisition (ITAKA), Universitat Rovira i Virgili, Tarragona, Spain
{antonio.moreno,mohammed.jabreel}@urv.cat

[4] Department of New Media and Technology, MODUL University Vienna, Vienna, Austria
stefan.gindl@modul.ac.at

Abstract. Communication through social media is an effective way to position a destination brand. In particular, the emotional values of a brand trigger a positive reaction from potential visitors. It is important for destinations to align their emotional communication strategies on different social media platforms to enhance their online image. The lack of comprehensive research in this area led this study to analyse the usage of the two most used social media platforms (Facebook and Twitter) among popular European tourist destinations. The study shows how destinations communicate their emotional values differently in Facebook and Twitter. The methodology of analysis allows destinations to compare the values they communicate with those of their competitors, so that they can improve their positioning and present a distinctive attractive personality of their destination.

Keywords: Emotional values · Communication strategies · Social media DMOs

1 Introduction

Facebook and Twitter are by far the most popular social media platforms used by firms to engage with consumers (Shively, 2014). These platforms are popular marketing tools, they are easy to maintain, they have a low cost and they present an effective large reach. Many *Destination Management Organizations* (DMOs) believe in the effectiveness of social media marketing for building a strong brand. However, the right

B. Stangl and J. Pesonen (eds.), *Information and Communication Technologies in Tourism 2018*, https://doi.org/10.1007/978-3-319-72923-7_13

communication of a destination brand is a challenge for them. Guerrero-Solé and Fernández-Cavia (2013) analyzed the performance on Twitter of some Spanish destinations and proved that there is room for improvement, in particular in the alignment of the destination image with the social media communication strategies. The literature on *destination branding* has highlighted the role of communication for the successful use of a brand. Brands are vital in the development of desirable destination images and reputation (Moreno et al., 2015). Furthermore, various users of the destination brand can associate and identify themselves with elements represented by the brand (Huertas & Mariné-Roig, 2016a), subsequently influencing the attachment to the brand and the intentions to visit the destination. Destinations can communicate the brand in two ways, through *functional* and *emotional* elements. Both of them should be taken into account when communicating destination brands in order to compete with other destinations (Hosany, Ekinci & Uysal, 2007; Huertas & Mariné-Roig, 2015, 2016b). In particular, Bigné et al. (2009) showed that the psychological and emotional components exercise the greatest influence on the overall image of the destination and on future behavior intentions. By developing a solid brand destination personality, destinations are able to stay competitive in a market with similar destinations (Ekinci & Hosany, 2006; Govers & Go, 2009). The communication of specific emotional brand values through social media generates greater interactivity and, thus, it enhances the brand image for tourism destinations (Huertas & Mariné-Roig, 2015). However, Bigné et al. (2009) and others (e.g., Michaelidou et al., 2013; Huertas & Mariné-Roig, 2016a, b) have demonstrated that DMOs tend to communicate tangible elements more often than the emotional values through social media. This leads to a gap among the communicated brand values and the image that users have about them (De Moya & Jain, 2013; Stepchenkova & Zhan, 2013; Míguez-González & Huertas, 2015). Often, DMOs do not have specific strategies to communicate their brand values nor brand communications strategies for different social media (Moreno et al., 2015; Huertas & Mariné-Roig, 2016a, b). DMOs should speak in a consistent manner in social media platforms in order to create a positive image among all stakeholders. However, despite the importance of social media platforms, there is a lack of studies performing cross-over platforms analysis, particularly in relation to brand communication strategies. In fact, there are few studies analyzing the use of brand emotions in online spheres (e.g., Dickinger & Lalicic, 2016). Hence, this study aims to close this gap by analyzing the emotional communication strategies of the DMOs of popular European destinations (TripAdvisor, 2017) in the two most important social media platforms: Facebook and Twitter. First, the study analyzes the brand emotional values communicated by each destination on each platform, by looking at the semantic relationship between the adjectives used by the destination in its posts and some basic categories of emotional values. Second, the study identifies in which way the communication significantly differs across destinations and across platforms. In doing so, DMOs are provided with practical recommendations on how to align their communication strategies and speak with one single emotional brand message, without having inconsistencies in the overall brand message transmitted to various stakeholders.

2 Literature Review

2.1 Destination Branding and Personal Values

A distinctive destination brand generates favorable associations in the mind of stakeholders, influencing their destination preferences and their final decisions (Hosany et al., 2007). As a result, a growing number of studies related to destination branding have focused on emotional branding. Hankinson (2004) was one of the first that highlighted the creation of emotional ties between the destination brand and its stakeholders. Later, authors such as Laroche et al. (2013), proved the importance of these emotional ties for creating a positive destination image, brand loyalty and tourist purchase decisions. Aaker (1996) introduced the hierarchical model of Customer-Based Brand Equity (CBBE), showing how companies can influence the effectiveness of their branding processes by building their brand but still managing the various brand dimensions, such as brand identity, meaning, responses and relationships. Aaker's (1996) hierarchy is founded by brand salience, which refers to the strength of the presence in the mind of the consumer; however, a firm's aim should be to increase familiarity with the brand through the right and repeated exposure, and creating strong associations with the brand. As a follow-up, there are also a number of studies that have analyzed the communication of destination brands. They have mainly focused on the cognitive, functional or tangible elements of the brand instead of the emotional elements or brand values (Xiang & Gretzel, 2010). The studies that focus on the analysis of the communication of brand values generally define a template to analyze the association of certain values with the content published by the destinations. Another stream of studies (Ekinci & Hosany, 2006; Hosany et al., 2007; Pitt et al., 2007; De Moya & Jain, 2013) have used the *Brand Personality Scale* (BPS) of Aaker (1997), which is based on five dimensions (sincerity, sophistication, excitement, ruggedness and competence). In doing so, the values of human personality have been used to analyze the way in which destinations communicate the emotional values of their brand. Various studies have shown how personality dimensions have a positive impact on the preferences of potential tourists, and generate behavioral intentions to re-visit (Ekinci & Hosany, 2006; Hosany, Ekinci & Uysal, 2007). A variety of studies have analyzed the communication of destination brands through social media (De Moya & Jain, 2013; Oliveira & Panyik, 2015; Huertas & Mariné-Roig, 2016b; Garay & Cànoves, 2017) and most of them also use Aaker's BPS as a reference. In general, the most common methodology of analysis relies on an automatic content analysis, either of a quantitative (De Moya & Jain, 2013) or a more qualitative nature (Oliveira & Panyic, 2015; Garay & Cànoves, 2017). One limitation of this kind of analysis is the difficulty in measuring the communication of emotional values using content analysis software. Other studies have analyzed the communication of the emotional values through a template manually (Huertas & Mariné-Roig, 2016b). However, there are also limitations in the manual approach, given the researchers' subjectivity and the limited number of cases that may be analyzed. One approach to overcome these problems is the development of automatic web analysis tools. For example, some studies have already analyzed the communication of emotional values or the image perception by making advanced automatic web content analysis with sentiment detection (Költringer

& Dickinger, 2015) or semantic analysis (Moreno et al., 2015). However, it is necessary to do more research on the integration of these innovative methods to capture the online communication of emotional brand values of DMOs in a more robust manner.

2.2 Communication of Brand Emotional Values in Social Media

The majority of studies have analyzed how social media influences the creation of a destination brand image among its users and the relationships they create with brands (Govers & Go, 2009; Laroche et al., 2013; Hudson et al., 2015). Laroche et al. (2013), using surveys, demonstrated that brand communities through social media have positive effects on the user-brand relationship, credibility and brand loyalty. Algesheimer et al. (2005) also showed that active participation of users in social media increases the emotional attachment to the brand and the brand loyalty. These authors developed a conceptual model to analyze how different aspects of the relationship of the customers with the brand community influence their intentions and behaviors. Among the second type of studies, some have analyzed how tourist destinations communicate their brand (Huertas & Mariné-Roig, 2016a, b; Moreno et al., 2015). Other types of studies have analyzed the brand destination images created in the mind of users (Jabreel et al., 2017), based upon user-generated content. In some cases, users' surveys and reviews have been used and compared for difference perceptions (Dickinger & Lalicic, 2016). Results showed that the dimensions of excitement, sophistication and competence were significantly more represented in social media than in the conventional survey (Dickinger & Lalicic, 2016). Further studies made a comparative analysis of the contents of both DMOs and its users (De Moya & Jain, 2013; Stepchenkova & Zhan, 2013). For example, De Moya & Jain, (2013) explored how México and Brazil communicate their brand personality through Facebook, and which personality traits their Facebook fans or users associate with them. The results show that these countries communicate distinctive brand personalities on their official Facebook pages. In line with previous studies, the most communicated brand values were sincerity and excitement (Ekinci & Hosany, 2006; Jain & Chan-Olmsted, 2009). These results also coincide with a later research of Huertas and Mariné-Roig (2016a, b) in which, for four types of destinations in Spain, the most mentioned values were related to honesty and sincerity. Interestingly, De Moya and Jain (2013) showed that emotional values communicated by the DMO of México coincided with the messages posted by its users, whereas the brand personality traits in the promotional messages of the DMO of Brazil did not coincide with the messages posted by its users. Stepchenkova and Zhan (2013) also analyzed and compared contents generated by DMOs and users and found differences among them. Interestingly, travelers were more interested in how Peruvian people live their everyday lives, the DMO focused on promoting the distinctive Peruvian culture, traditions and art instead. Then, Moreno et al. (2015) analyzed the destination brand values communicated by the main European DMOs through Twitter. The results showed that the destinations do not use specific adjectives to communicate their identity, but common and generic ones that many destinations want to be associated with. Moreover, the destinations do not seem to have a coherent communicative strategy to try to communicate emotional brand values using Twitter. Lastly, Huertas and Mariné-Roig (2016b) explored how the cognitive elements and emotional values of

the destination brands are communicated through different social media platforms. First, they did not find any remarkable differences in the communication of emotional values through different social media platforms. Second, they did not identify specific strategies concerning emotional content. Their results coincide with those of Moreno et al. (2015), who showed that, despite the different destination types, very similar emotional values were communicated, and there is still a dominant focus on the functional elements.

3 Method

The aim of this work is to analyse how emotional values are communicated by different destinations through Twitter (tweets) and Facebook (posts). The study focuses on the semantic analysis of English tweets and posts sent by official tourist destination accounts on these two social networks. The basic steps of the analysis, described in the following subsections, are the following: (1) definition of the emotional values associated to a destination brand, (2) selection of the destinations to be analysed, (3) retrieval and pre-processing of English tweets and posts sent by official tourist destinations on Facebook and Twitter, (4) semantic analysis of the content of the tweets (this step is the core of the methodology, and it aims to link the adjectives used in the tweets/posts with the emotional values defined in the first step) and (5) interpretation of the results.

3.1 Emotional Values

In this work we have made a slight adaptation of the emotional values defined in Aaker's *Brand Personality Scale* (Aaker, 1997) to its use in the case of destination brands. We have considered the five categories shown in the following list, which are further decomposed in specific subcategories: (1) *Sincerity*: family-oriented, down-to-earth, sustainable, calm, real, traditional, honest, original, wholesome, quality of life, happiness, sentimental, friendly, (2) *Excitement*: trendy, daring, exciting, exotic, fashionable, cool, spirited, dynamic, vital, fresh, young, sensorial, unique, imaginative, creative, up-to-date, independent, contemporary, cosmopolitan, tolerant, hospitable, (3) *Competence*: reliable, hard-working, safe, rigorous, intelligent, technical, corporate, innovative, successful, leader, ambitious, powerful, (4) *Sophistication*: glamorous, luxurious, seductive, smooth, romantic, magical and (5) *Ruggedness*: outdoorsy, get-away, recreational, tough, rugged, non-conformist.

3.2 Selection of the Destinations to Be Analysed

The top 25 European destinations in 2017, according to the user feedback and booking interest measured by Trip Advisor, were considered (TripAdvisor, 2017). We searched manually for their official Twitter destination accounts and we selected those destinations that had sent at least 3000 English tweets. All of these destinations also have corresponding Facebook accounts with posts in English. The 10 destinations that were finally analysed are Amsterdam, Barcelona, Berlin, Budapest, Dublin, Edinburgh,

London, Madeira, Paris and Tenerife. The other 15 destinations do not have official Twitter accounts or they send their tweets mainly in their local language.

3.3 Retrieval and Pre-processing of the Sets of Messages

This study focuses on the Twitter tweets and the Facebook posts that were posted from the official accounts of the 10 selected destinations during 2016. Tweets were retrieved with a tool developed by the authors that allows obtaining tweets that satisfy different constraints (language, time, geo-location, etc.). Facebook posts were obtained with another self-developed tool that accesses Facebook's Graph API. It is designed to provide insights into the activity of public pages. The tool downloads and analyses the posting history of selected pages as well as the fan interactions with the posts (i.e., comments and reactions). The analysis helps to assess the efforts and successes of the page owners, e.g. companies or entities of public interest, by measuring their activity and the engagement and loyalty of their fans. In the pre-processing stage all URLs, usernames, non-alphabetic symbols and stop words were removed. All tweets/posts were converted to lowercase. Words with repeated letters were automatically corrected using a simple algorithm developed by the authors that performs a breadth first search to analyse all the possible ways of eliminating repeated letters in a string, checking in WordNet if they are correct.

3.4 Semantic Content Analysis

The aim of this step is to link the adjectives appearing in the tweets/posts with the categories of emotional values defined previously. Adjectives were chosen because, as shown in previous studies on content analysis of the communication of destination brands (Stepchenkova et al., 2009), they are the words that tend to convey the emotional responses. A standard natural language parser was applied to retrieve the adjectives and count their frequency of use. A direct syntactic mapping was not feasible, as most of the adjectives did not appear directly as categories/subcategories of emotional values. The Wu-Palmer ontology-based semantic similarity measure (Wu and Palmer, 1994) was used to check the similarity between adjectives and emotional values in WordNet. The similarity between two terms c_1 and c_2 is defined as

$$simWP(c_1, c_2) = (2 * N_3)/(N_1 + N_2 + 2 * N_3),$$

where N_1 and N_2 are the number of hypernym links from the terms c_1 and c_2 to their *Least Common Subsumer* (LCS) in WordNet, and N_3 is the number of hypernym links from the LCS to the root of the ontology. This measure ranges from 1 (for identical concepts) to 0 (when the LCS is the root of the ontology, so the concepts do not have any common ancestor). In order to apply this measure, the compared terms must be *nouns*. Thus, both the emotional values and the adjectives had to be transformed into nouns. In the case of the emotional values, these were manually translated to the equivalent nouns (e.g. 'ambitious' was transformed into 'ambition'). The selection of nouns is certainly important. We tried to select nouns as semantically close to the adjective as possible, and we chose a particular meaning (synset) of the noun in the

case of polysemous words. Concerning the adjectives appearing in the tweets/posts, they were automatically transformed into nouns using their derivationally related form or their attribute property in WordNet. After this pre-processing stage, it was possible to apply the Wu-Palmer similarity measure to compare the emotional values and the tweet/post adjectives. In this way, we can assess if the destination communicates certain emotional values and if there is a strategy behind the communication of the brand. Only adjectives with a similarity higher than 0.7 to an emotional value (*emotional adjectives*) were considered in the final steps of the analysis.

4 Results

The number of tweets and Facebook posts and the use of different adjectives are summarized in Table 1, in which the 10 destinations are ranked according to their activity on Twitter and Facebook.

Table 1. Volume of tweets/posts and use of adjectives in the selected destinations

	Twitter			Facebook	
	#Tweets	#Adjectives		# Posts	#Adjectives
Paris	2058	851	London	1323	610
Edinburgh	1935	665	Madeira	1006	1936
Barcelona	1828	654	Edinburgh	847	524
Berlin	1797	621	Dublin	402	251
Dublin	1722	456	Tenerife	398	637
Madeira	1273	514	Amsterdam	386	372
Tenerife	1102	490	Barcelona	319	1337
Amsterdam	940	521	Paris	223	229
London	633	277	Budapest	223	197
Budapest	159	123	Berlin	45	42

In the case of Twitter, the majority of the DMOs are quite active, sending 1000–2000 tweets in 1 year (an average of 3–6 daily tweets). Paris (2058 tweets) and Edinburgh (1935 tweets) are the most active DMOs, whereas Budapest is the least active (only 159 tweets). The use of adjectives follows a similar ratio. DMOs' Facebook activity, on the other hand, shows a different ranking, in which the posting activity of London, Madeira and Edinburgh stands out. Interestingly, the use of adjectives is not aligned with the number of posts. For example, Barcelona posted only 319 times on Facebook during 2016, but it used 1337 different adjectives, whereas the top runner, London, was much more moderate in the use of adjectives (only 610 in 1323 posts). Thus, DMOs not only use the two platforms differently in terms of activity, but they also tend to have a different usage intensity of adjectives in their online communication across the platforms. In order to reveal more insights into the specific use of adjectives, their relationship with the 5 brand dimensions on Twitter and Facebook was analyzed. Taking into account the percentage of emotional adjectives

employed for each category of emotional values, there are only small differences between Facebook and Twitter: sincerity (FB: 38%, Tw: 37%), excitement (FB: 28%, Tw: 26%), competence (FB: 17%, Tw: 11%), sophistication (FB: 10%, Tw: 14%) and ruggedness (FB: 10% and Tw: 9%). Hence, the dimensions of competence and sophistication differ slightly across the platforms. Furthermore, an analysis was performed to see how often the adjectives are actually used in the Facebook posts and tweets. In this case, we can see that this happens almost identically across the two platforms: sincerity (FB: 41%, Tw: 41%), excitement (FB: 23%, Tw: 22%), competence (FB: 18%, Tw: 18%), sophistication (FB: 8%, Tw: 9%) and ruggedness (FB: 10% and Tw: 10%). Lastly, the average number of uses of each emotional adjective by every DMO among the two platform was calculated. A Chi-square analysis showed that significant differences occur across the platforms (see Table 2). For example, Paris uses each emotional adjective more in its communication through Twitter (5.30 uses of each adjective) than for Facebook posts (only 3.17 uses of each emotional adjective) ($p < 0.05$). London, in contrast, uses the emotional adjectives more on Facebook (5.30) than on Twitter (4.28) ($p < 0.05$). Destinations like Barcelona and Edinburgh tend to use similar emotional expressions on both platforms ($p < 0.001$).

Table 2. Average number of uses of each emotional adjective—chi-square analysis

	Ratio Twitter	Ratio Facebook	*p*-value
Paris	5.30	3.17	<0.05
Edinburgh	5.32	5.95	<0.001
Barcelona	4.38	4.94	<0.001
Berlin	2.08	1.21	0.130
Dublin	6.53	3.79	0.209
Madeira	2.83	4.22	0.23
Tenerife	3.30	3.44	<0.05
Amsterdam	3.57	2.78	<0.001
London	4.28	5.30	<0.05
Budapest	1.57	1.87	<0.001

Figures 1 and 2 display which dimensions are the most often represented in DMOs' online brand communication across Twitter and Facebook, respectively. As seen in Fig. 1, Barcelona is one of the outstanding DMOs in the communication of feelings of sincerity through Twitter. The majority of DMOs send tweets with feelings of excitement, particularly destinations like Paris, Edinburgh, Amsterdam and Barcelona. Expressions related to competence are especially used by the destinations Barcelona, Edinburgh and Amsterdam. The dimension of sophistication is highly used only by the DMO of Barcelona. Lastly, feelings related to the dimension of ruggedness are limited in online communication. The DMOs of Edinburgh, Barcelona and Tenerife are the ones that include relatively more feelings of ruggedness in their tweets.

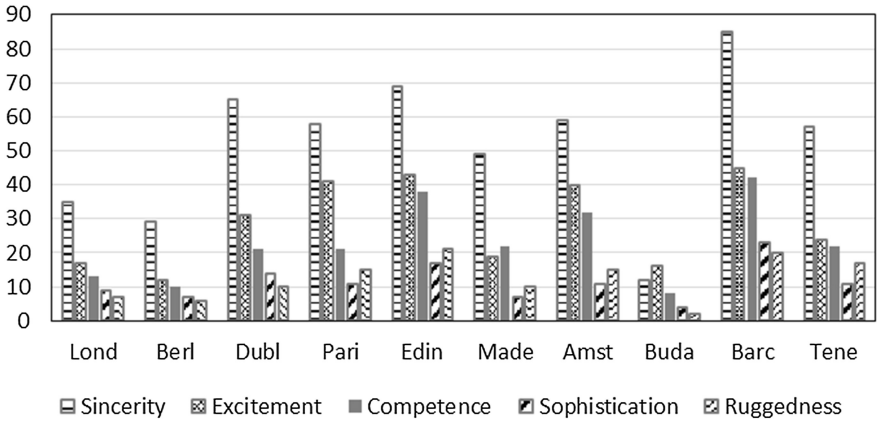


Fig. 1. Brand personality dimensions—tweets

Figure 2 illustrates the most common dimensions of emotional values mentioned in the Facebook posts. The communication of sincerity is much higher than the one of other values among the majority of the DMOs. The dimension of excitement is more homogeneous, than in the case of tweets. Words that reflect the dimension of competence tend to be more popular in Facebook posts by the DMOs of Tenerife, Madeira, London and Edinburgh. The dimensions of sophistication and ruggedness, like in Twitter, are not heavily used among the DMOs. Destinations like London, Edinburgh and Tenerife are the top communicators with regards to ruggedness.

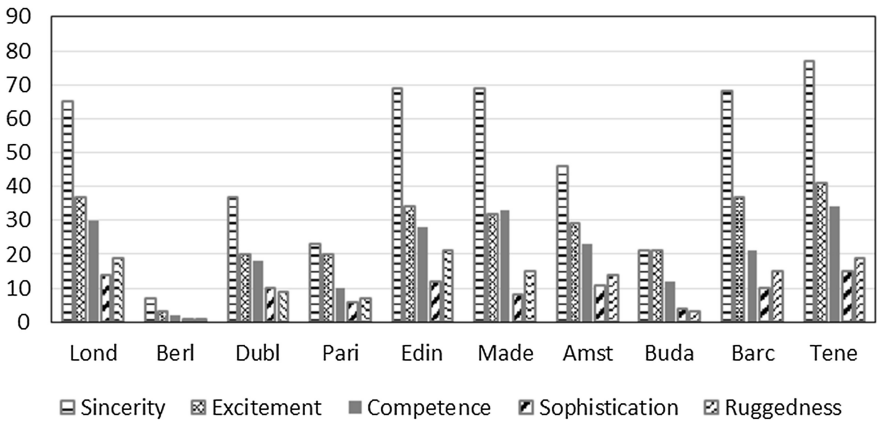


Fig. 2. Brand personality dimensions—Facebook posts

Lastly, in order to understand in which way the DMOs communicate their emotional brand values, also the usage of the specific categories and subcategories in relation to their frequency are listed. Given the space limit, the five most popular subcategories communicated through social media are listed in Table 3.

Table 3. Top 5 subcategories of emotional values mentioned in Twitter and Facebook

Competence		Sophistication		Ruggedness	
Twitter	Facebook	Twitter	Facebook	Twitter	Facebook
Leader	Reliable	Glamourous	Glamourous	Rugged	Rugged
Safe	Hard-working	Smooth	Luxurious	Get-away	Get-away
Innovative	Safe	Magical	Charming	Recreational	Recreational
Powerful	Responsible	Luxurious	Smooth	Tough	Tough
Intelligent	Intelligent	Charming	Romantic	Non-conformist	Non-conformist

Sincerity		Excitement	
Twitter	Facebook	Twitter	Facebook
Honest	Honest	Fresh	Fresh
Down-to-earth	Down-to-earth	Contemporary	Unique
Calm	Original	Creative	Daring
Original	Calm	Unique	Contemporary
Happiness	Traditional	Daring	Creative

Interestingly, there are only slight differences in the reference to emotional value categories in Facebook and Twitter (e.g., the expressions related to ruggedness are identical). However, the dimension of competence tends to be communicated in different ways across the two platforms (concepts such as 'leader' and 'safe' are the top runners in the tweets, but 'reliable' and 'hard-working' are more referenced in Facebook posts). Also, for expressions related to sophistication, the emotion 'glamourous' is highly referenced in both platforms, but the next running emotions are different (tweets communicate feelings of 'magical' and 'smooth', whereas Facebook posts focus on feelings like 'luxurious' and 'charming').

5 Conclusion and Recommendations

The emotional values of a DMO are an important differentiator to stand out in social media and remain competitive. As DMOs tend to communicate their tangible brand aspects more often, this study aimed to explore the communication of emotional brand aspects through Facebook and Twitter. First, the study demonstrates significant differences in the use (activity) and communication (intensity) of emotional brand values through the two platforms. The number of tweets/posts published by DMOs, the number/uses of adjectives and the communicated emotional brand values were analysed. In the case of Twitter, the number of tweets, adjectives and referenced emotional values are quite homogeneous among the observed DMOs. Interestingly, Facebook posts demonstrate discrepancies among the destinations. Second, the study also shows that the number of adjectives and emotional brand values communicated through Twitter is smaller than in Facebook. DMOs do not use the platforms equally, particularly in relation to the number of adjectives and the values communicated per post/tweet. This coincides with the results of previous studies, showing that social media have different characteristics that make them more or less suitable to

communicate brand emotional values as well as showing that Facebook possesses more emotional brand values in their posts than Twitter (Huertas & Mariné-Roig, 2016b). Third, the emotional brand values communicated are very similar for all destinations analyzed. They do not show distinctive strategies of brand communication, as Moreno et al. (2015) showed. The most communicated emotional values are sincerity and excitement, as detected in other studies (Ekinci & Hosany, 2006; Jain & Chan-Olmsted, 2009; Huertas and Marine-Roig, 2016a, b). However, feelings of ruggedness are communicated only in a limited way by a few destinations. This is also in line with other studies (i.e., Dickinger & Lalicic, 2016). However, the study identified differences between the types of emotional brand values communicated across the two platforms. For example, DMOs' communicate through Twitter magical and smooth values, whereas the Facebook communication focuses on luxurious and charming aspects. These results do not coincide with the previous studies, where DMOs did not have distinct strategies to communicate their brand values or communication brand strategies for different social media (Moreno et al., 2015; Huertas & Marine-Roig, 2016a, b). Overall, this study shows that there is a shift, where DMOs (at least the most popular destinations in Europe), are aware of the different potentialities of each social media platform and adapt accordingly different brand communication strategies. However, the study also shows that destinations still do not have a distinctive communication strategy regarding emotional brand values and they still do not take advantage of all the communicative potential that social media has in the communication of emotive destination brands. The study advises destinations to compare their communicated emotional values with competing destinations. In doing so, they can improve their differentiation and present a distinctive attractive personality of their destination online. DMOs are also recommended to verify how their online communicated emotional brand values coincide with their brand strategies and adapt them accordingly. Future research could also include a comparison of DMOs' brand emotional strategies and their online communication behavior, and include new methods such as storytelling. Furthermore, other engagement measurements, such as retweets or likes, can be included to test the effect of communicating brand emotions.

References

Aaker, A.: Measuring brand equity across products and markets. Calif. Manag. Rev. **38**(3), 102–120 (1996)

Aaker, J.L.: Dimensions of brand personality. J. Mark. Res. **34**, 347–356 (1997)

Algesheimer, R., Dholakia, U.M., Herrmann, A.: The social influence of brand community: evidence from European car clubs. J. Mark. **69**(3), 19–34 (2005)

Bigné, E., García, I.S., Blas, S.S.: The functional-psychological continuum in the cognitive image of a destination: a confirmatory analysis. Tourism Manag. **30**(5), 715–723 (2009)

De Moya, M., Jain, R.: When tourists are your "friends": exploring the brand personality of Mexico and Brazil on Facebook. Publ. Relat. Rev. **39**, 23–29 (2013)

Dickinger, A., Lalicic, L.: An analysis of destination brand personality and emotions: a comparison study. Inf. Technol. Tourism **15**(4), 317–340 (2016)

Ekinci, Y., Hosany, S.: Destination personality: an application of brand personality to tourism destinations. J. Travel Res. **45**, 127–139 (2006)

Garay Tamajón, L., Cànoves Valiente, G.: Barcelona seen through the eyes of TripAdvisor: actors, typologies and components of destination image in social media platforms. Curr. Issues Tourism **20**(1), 33–37 (2017)

Guerrero-Solé, F., Fernández-Cavia, J.: Activity and influence of destination brands on Twitter: a comparative study of nine Spanish destinations. In: Xiang, Z., Tussyadiah, I (eds.) Information and Communication Technologies in Tourism 2014, pp. 227–236. Springer International Publishing, Switzerland (2013)

Govers, R., Go, F.M.: Tourism destination image formation. Handb. Tourist Behav. 35–49 (2009)

Hankinson, G.: Relational network brands: towards a conceptual model of place brands. J. Vacation Mark. **10**(2), 109–121 (2004)

Hosany, S., Ekinci, Y., Uysal, M.: Destination image and destination personality: an application of branding theories to tourism places. J. Bus. Res. **59**(5), 638–642 (2007)

Hudson, S., Roth, M.S., Madden, T.J., Hudson, R.: The effects of social media on emotions, brand relationship quality, and word of mouth: an empirical study of music festival attendees. Jourism Manag. **47**, 68–76 (2015)

Huertas, A., Mariné-Roig, E.: Destination brand communication through the social media: What contents trigger most reactions of users? In: Tussyadiah, I., Inversini, A. (eds.) Information and Communication Technologies in Tourism 2015, pp. 295–308. Springer International Publishing, Switzerland (2015)

Huertas, A., Mariné-Roig, E.: User reactions to destination brand contents in social media. Inf. Technol. Tourism **15**, 291–315 (2016a)

Huertas, A., Mariné-Roig, E.: Differential destination content communication strategies through multiple social media. In: Inversini, A., Schegg, R. (eds.) Information and Communication Technologies in Tourism 2016, pp. 239–252. Springer, Switzerland (2016b)

Jabreel M., Moreno A., Huertas A.: Do local residents and visitors express the same sentiments on destinations through social media? In: Schegg, R., Stangl, B. (eds.) Information and Communication Technologies in Tourism 2017, pp. 655–668. Springer International Publishing, Switzerland (2017)

Jain, R., & Chan-Olmsted, S.: Is your tourist destination sincere, rugged, or sophisticated? An exploratory examination of online brand personality of nations. In: Annual Meeting of the Association for Education in Journalism and Mass Communication. Boston, MA (2009)

Költringer, C., Dickinger, A.: Analyzing destination branding and image from online sources: a web content mining approach. J. Bus. Res. **68**(9), 1836–1843 (2015)

Laroche, M., Habibi, M.R., Richard, M.-O.: To be or not to be in social media: how brand loyalty is affected by social media? Int. J. Inform. Manag. **33**, 76–82 (2013)

Michaelidou, N., Siamagka, N.T., Moraes, C., Micevski, M.: Do marketers use visual representations of destinations that tourists value? Comparing visitors' image of a destination with marketer-controlled images online. J. Travel Res. **52**(6), 789–804 (2013)

Míguez-González, M.I., Huertas, A.: The power of photographs in the communication and public relations of tourist destinations and their brands through Facebook and Flickr. Catalan J. Commun. Cult. Stud. **7**(2), 197–215 (2015)

Moreno, A., Jabreel, M., Huertas, A.: Automatic analysis of the communication of tourist destination brands through social networks. In Proceedings of the 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE-2015), pp. 24–27 (2015)

Oliveira, E., Panyik, E.: Content, context and co-creation: digital challenges in destination branding with references to Portugal as a tourist destination. J. Vacation Mark. **21**(1), 53–74 (2015)

Pitt, L.F., Opoku, R., Hultman, M., Abratt, R., Spyropoulou, S.: What I say about myself: communication of brand personality by African countries. Tourism Manag. **28**(3), 835–844 (2007)

Shively, K.: New study: 80% of Fortune 500 companies active on Facebook and Twitter. Retrieved from https://simplymeasured.com/blog/new-study-80-of-fortune-500-companies-active-on-facebook-and-twitter/#sm.0000ij1r4unaffq4v8n1qvxpqd83u on July, 3 2017 (2014)

Stepchenkova, S., Kirilenko, A.P., Morrison, A.M.: Facilitating content analysis in tourism research. J. Travel Res. **47**, 454–469 (2009)

Stepchenkova, S., Zhan, F.: Visual destination images of Peru: comparative content analysis of DMO and user-generated photography. Tourism Manag. **36**, 590–601 (2013)

TripAdvisor.: Traveller destination choice-top 25 destinations in Europe. Retrieved https://www.tripadvisor.com/TravelersChoice-Destinations-cTop-g, June 29, 2017 (2017)

Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics (ACL) (1994)

Xiang, Z., Gretzel, U.: Role of social media in online travel information search. Tourism Manag. **31**(2), 179–188 (2010)

Information Technology & Tourism (2019) 21:63–81
https://doi.org/10.1007/s40558-018-0134-7

ORIGINAL RESEARCH

Which emotional brand values do my followers want to hear about? An investigation of popular European tourist destinations

Lidija Lalicic[1] · Assumpció Huertas[2] · Antonio Moreno[3] · Mohammed Jabreel[3]

Received: 2 May 2018 / Revised: 31 October 2018 / Accepted: 29 November 2018 /
Published online: 5 December 2018

Abstract
In this paper, the communication of destination brands through social media (Twitter and Facebook) is investigated with a focus on the use of emotional adjectives. Based on ten of the most popular destinations in Europe, more than 15,000 tweets and 6000 Facebook posts and users' reactions to those messages were analysed. The study shows that DMOs are active and communicate using various emotional values about their brands on both platforms. The most popular emotional brand values for Twitter are 'glamorous' and 'happiness', while 'honest' and 'trendy' appear most frequently on Facebook. Interestingly, besides the fact that users actively engage on both platforms, significant differences were detected based on the values that users respond to as opposed to those that destinations use in their communication. Users respond much more strongly to messages that contain values related to the adjectives 'sentimental' and 'happiness' on Twitter, whereas on Facebook, adjectives such as 'getaway' and 'young' generate better responses. The study demonstrates that destinations can improve user reaction rates on social media and facilitate a positive destination brand image through the use of specific emotional brand values. Overall, the findings provide valuable as well as directly applicable implications for DMOs to use different values that align better with user expectations of the destination, which will ultimately lead to more effective online marketing strategies and stronger competitive profiles. This is an extended version of a conference paper entitled "Do DMOs communicate their emotional brand values? A comparison between Twitter and Facebook", previously published in the proceedings of Information and Communication Technologies in Tourism 2018 Conference (ENTER 2018) held in Jönköping, Sweden, January 24–26, 2018.

Keywords Emotional brand values · Social media communication strategies · User engagement · DMOs

✉ Lidija Lalicic
lidija.lalicic@modul.ac.at

Extended author information available on the last page of the article

1 Introduction

Social media allows users to interact with brands with the least amount of effort when compared to traditional marketing information outlets (Amaro et al. 2016; Li et al. 2017). Moreover, User Generated Content (UGC) has been shown to significantly influence the level of co-creation of destination images (Költringer and Dickinger 2015; Lu and Stepchenkova 2015) and the decisions of other users in the network (Liu and Park 2015) as opposed to formal destination communication platforms (Litvin et al. 2008; Papathanassis and Knolle 2011). Hence, Destination Management Organizations (DMOs) have to respond quickly to these changing consumer trends. At the very least, they need to optimize their online social media communication and, at the same time, build strong emotional brands that stand out in the online world (Roque and Raposo 2015).

Various authors highlight the importance of developing online communication strategies that address the emotional component of brands and work on the visibility of the brand (Laroche et al. 2013). Gobe (2010), introducing the concept of emotional branding, refers to this as an opportunity for brands to focus on sensory experiences and imaginative ways to capture the attention of their users. Also, it allows brands to create stories and metaphors about their products and services, stimulate their users' sensors and trigger responses, ultimately establishing a relationship (Gobe 2010). In the case of destinations, emotions play a significant role in enhancing the experience (Bigné et al. 2003) and facilitating higher and better expectations of a destination, which has proven to lead to higher levels of satisfaction and destination attachment (Govers et al. 2007; Schroeder and Pennington-Gray 2015). As a consequence, the increased awareness and brand engagement also leads to higher sales and competitive positioning of a given destination (Schroeder and Pennington-Gray 2015).

However, the critical question for DMOs is how to effectively communicate emotional brand values while positioning themselves online. Research shows that firms can manage online communication by focusing on the content of information and the style of writing (Mariani et al. 2016), even though the style appears to be more influential (Huertas and Mariné-Roig 2016a). Platforms, such as Facebook, can be very supportive environments, where DMOs can experiment with their various communication strategies (Munar and Jacobsen 2014; Ashley and Tuten 2015). DMOs need to approach these platforms in a way that differs from traditional marketing strategies as the transfer of the creative appeal of a brand may not be so easy and effective among users (Moran and Gossieaux 2010).

However, those DMOs that attempt to communicate their brands in online spheres tend to mainly focus on the functional parts of their brands (i.e., communicate words related to basic sights and facilities) (Michaelidou et al. 2013; Huertas and Mariné-Roig 2016b). Conversely, the focus on the functional brand elements fails to provide a complete picture of what the destination has to offer and fails to create the right expectations for the users who follow the destination brand online (De Moya and Jain 2013; Hays et al. 2013; Stepchenkova and Zhan 2013; Míguez-González and Huertas 2015). Other studies investigated if and

how DMOs communicate the emotional components of their destinations brands (Jabreel et al. 2017; Moreno et al. 2015). Interestingly, those DMOs that communicate their emotional brand values also tend to generate a successful online image (Huertas and Mariné-Roig 2015; Mariné-Roig 2017) and even see higher levels of tourist arrivals (Usakli et al. 2017). Recenlty, Lalicic et al. (2018) investigated DMOs' Facebook and Twitter communication, integrating Aakers' brand personality scale (1997) in order to show the distinct differences across the two platforms in terms of intensity of emotional brand communication approaches. However, the user metrics, which are an indication of success online (Usakli et al. 2017), have been left out of the analysis and would enhance our understanding of this subject (Lalicic et al. 2018).

As a response to the aforementioned discussion, this paper aims to create a better understanding of how DMOs communicate their brands on social media platforms, placing a specific focus on the emotional brand experience. Furthermore, the paper aims to understand which specific emotional values used in DMOs' communication can explain user responses. In doing so, the paper also investigates which values tend to elicit stronger responses among users and whether there is a discrepancy between the intensity of DMOs' communication of these values and the frequency of user responses. If there is a high match, for example, between the intensity of DMOs' emotional communication and the number of responses, these values will be grouped as effective. As opposed to values that are hardly communicated by DMOs but receive a high level of response. Such values can be grouped as promising for a DMOs' future online communication. Thus, such insights can help DMOs to align their online presence with their audience and become more effective in their online social media communication.

In this case, the responses are recorded by the number of 'likes' and other reactions that Facebook provides and the number of retweets or tweets marked as 'favourite'. Using self-developed tools, Facebook posts, tweets and user metrics are retrieved based on data collected from the most popular destinations in Europe according to TripAdvisor (2017). In doing so, this paper sheds light on the current communication practices of highly visited destinations on the two most popular social media websites. Furthermore, destinations are assessed on whether they are successful in communicating their emotional brand elements. As online popularity is crucial for a DMOs' brand appearance and recognition (De Vries et al. 2012), the outcomes of this study will also help DMOs to reconsider their communication strategies and what they want to be known for online. From a theoretical perspective, this paper enhances the discussion of DMOs' social media engagement (De Vries et al. 2012), their online successes (Usakli et al. 2017; Mariani et al. 2016) as well as what can be improved to enhance user responses (Lalicic et al. 2018), thereby also providing some insight into relationship management (Roque and Raposo 2015).

The paper is structured as follows: firstly, the most important literature related to emotional branding for DMOs and online commination as well as social media engagement will be discussed. Then, the method and sample is introduced, followed by the findings, which illustrate DMOs' online communications techniques as well as the relationships between the communication and numbers of responses that are

tested. Finally, a critical discussion and contributions to theory and practical implications are provided along with a set of limitations and future research avenues.

2 Theoretical framework

2.1 Emotional branding and online communication

According to Gobe (2010), emotional branding allows consumers to form emotional connections with brands in a hidden manner. Thus, emotional branding focuses on the various sensory experiences consumers can have with a brand (Gobe 2010). In tourism, emotional branding works well as the industry subsists largely on experience-based products. As a result, a large amount of tourism research has been devoted to the topic of emotional branding (Morgan et al. 2003; Morgan and Pritchard 2004; Blain et al. 2005; Govers et al. 2007; Huertas and Mariné-Roig 2015). One significant observation is that communicating the emotional elements of a brand appear to successfully enhance a destination's overall image among users as well as its strategic position (Bigné et al. 2009; Ekinci and Hosany 2006; Huertas and Mariné-Roig 2015, 2016b). Many researchers reference the brand personality scale from Aaker (1997) in order to understand how tourists develop emotional connections with destinations and which specific values stand out (i.e., Hosany et al. 2006, 2007; Pitt et al. 2007; De Moya and Jain 2013; Chaykina et al. 2014). For example, values related to the brand personality dimensions of *excitement* and *sophisticated* create much higher levels of satisfaction, brand attachment and purchase intention (Ekinci and Hosany 2006).

Also, those destinations that include brand personality values in their social media strategies mainly focus on words related to 'honesty', 'sincerity' and 'excitement' (De Moya and Jain 2013; Huertas and Mariné-Roig 2016b). Huertas and Mariné-Roig (2016b) demonstrate that this holds true regardless of the destination type. There still seems to be a difference between specific destinations and the use of emotional branding approaches. For example, Uşkali et al. (2017) show that mature destinations tend to focus more on emotional values, whereas immature destinations according to Usakli et al. (2017), tend to emphasize functional-focus communication strategies more frequently at the start. Also, there are significant discrepancies between what users communicate about a destination brand compared to what a DMO communicates about itself. For example, De Moya and Jain (2013) compared the DMO Facebook fan pages of Brazil and Mexico in this manner. In this case, UGC coincided with Mexico's DMOs' emotional brand values, whereas Brazil's promotional messages did not coincide with the messages posted by its users. In analysing photo materials used by DMOs and those posted by users, Stepchenkova and Zhan (2013) also found differences. The DMO heavily focused on promoting Peruvian culture and art, whereas the tourists displayed more interest in the Peruvian way of life. Also, significant differences in the emotional brand values across tourist services in Vienna (Austria) were observed on TripAdvisor (Dickinger and Lalicic 2016). The aforementioned examples all show a clear identity-image gap, where users perceive the destination differently than the DMOs wants them to experience,

which hinders them in developing a closer connection with their users (Költringer and Dickinger 2015).

Twitter has also been analysed in this fashion (Moreno et al. 2015; Lalicic et al. 2018). According to Moreno et al. (2015), European DMOs do not use specific adjectives and tend to prefer generic communication to promote their brands. Other studies demonstrate the lack of a coherent message or strategy across the different platforms (Huertas and Mariné-Roig 2016b). Instead, research shows that DMOs exhibit high similarity and tangible elements in their communication styles, and, therefore, they do not distinguish themselves. Thus, for practitioners as well as for researchers, the question of what kind of emotional brand communication is needed online to engage with users remains unanswered. Therefore, this study will provide more insights into this issue.

2.2 Social media engagement

The engagement with social media to communicate about the brand and build up a fan base can be challenging for many firms (Xiang and Gretzel 2010; Lovejoy and Saxton 2012; Saffer et al. 2013; Wattanacharoensil and Schuckert 2015). Another challenge is the amplifying power of users, given that users can share their enthusiasm for brands on brand pages in addition to their own social media networks (Kabadayi and Price 2014). The transparency of social media easily visualizes the popularity of a brand post according to the volume of 'likes', shares, retweets and comments made by users (Guillet et al. 2016; Oviedo-García et al. 2014; Míguez-González and Huertas 2015). Thus, brands are encouraged to design activities and communication techniques that allow them to engage with as many users as possible on a continuous basis. In fact, if brands are successful, they can develop deeper relationships that go beyond acquiring the brand's offerings (Chan and Guillet 2011; Van Doorn et al. 2010; Nusair et al. 2013; Walther and Jang 2012). Su et al. (2015) suggest that firms need to help users find the brand's benefits and matches between their aspirations, insights and experiences they can/will have with the brand. Furthermore, Su et al. (2015) state that firms need to understand that if a message is highly emotional, the psychological characteristics of consumers can be addressed and, in doing so, this approach may be suitable for less involved consumers. Thus, elements like experiential appeals, animations and social causes can lead to a better performing brand online (Su et al. 2015).

In the case of tourism, DMOs exhibit different online behaviour, depending on the platform with which they engage. For example, when compared to Instagram or Facebook, Twitter is used more interactively by DMOs (Usakli et al. 2017). Lalicic and Gindl (2018) demonstrate that DMOs from popular destinations are either not active on Facebook or not active in a consistent manner. These DMOs posted irregularly and did not work towards building a loyal fan base. Interestingly, those that were successful demonstrated continuous engagement, good timing of posts to engage with their users and developed a loyal user base over time (Lalicic and Gindl 2018). Also, it has been proven that DMOs that post photos and videos on Facebook lead to a higher level of user reactions (Mariani et al. 2016; Lalicic and Gindl 2018).

However, the question of what kind of communication, especially emotional brand communication, generates user engagement has not been answered fully (Huertas and Mariné-Roig 2016a). According to Huertas and Mariné-Roig (2016a), there is a lack of congruency between the most frequently communicated content and emotional brand values shared by DMOs and those that trigger the most reactions among users. Therefore, in this paper, we aim to bolster practitioners' understanding of the use of emotional brand values and their effect on user engagement and, thus, effective relationship marketing tactics.

3 Methods

3.1 Sample

At first, the authors were interested in the top 25 European destinations in 2017 according to TripAdvisor (TripAdvisor 2017).[1] However, due to a lack of online activity, we decided to include only those destinations that sent at least 3000 English tweets between January 1, 2016 and December 31, 2016 on their official Twitter accounts. In addition, we also checked those that have an active Facebook account with English posts, which resulted in a final sample of the following destinations: *Amsterdam, Barcelona, Berlin, Budapest, Dublin, Edinburgh, London, Madeira, Paris and Tenerife.*

The self-developed system twiQuery[2] was used to collect the tweets for analysis. It is a crawler that enables users and developers to make advanced search actions in order to retrieve tweets that pinpoint distinct criteria, such as specific users, languages, regions (determined by the name of a city or a geolocation), periods of time, strings of communication or hashtags, etc. The Facebook posts were retrieved by using another self-developed tool that accesses Facebook's Graph API and collects information related to posting history as well as fan interactions (i.e. comments and reactions). All of the tweets and posts sent in 2016 were retrieved and analysed.

3.2 Data analysis

First, the tweets and posts were pre-processed: all URLs, usernames, non-alphabetic symbols and stop words were removed, and the content was converted to lowercase letters. Words with repetitive letters were corrected by using a self-developed algorithm (the algorithm performs a breadth first search to analyze all possible ways of eliminating repetitive letters in a string and then cross-references those words in WordNet to verify if they are correct) (Lalicic et al. 2018). The second step was the analysis of the content of the tweets/posts. To do so, we employed a semantic analysis built

[1] Paris, London, Rome, Crete, Barcelona, Prague, Istanbul, Lisbon, Majorca, Santorini, Florence, Venice, Rhodes, Tenerife, Amsterdam, Corsica, Budapest, Madeira, Berlin, Edinburgh, St. Petersburg, Vienna, Dublin, Krakow and Sochi.

[2] TwiQuery may be downloaded on this web page: https://github.com/mhjabreel/twiQuery.

on Aaker's brand personality scale (Aaker 1997), which provided us with a categorization of *emotional values*. *Sincerity* is expressed by adjectives, such as 'family-oriented', 'down-to-earth', 'sustainable', 'calm', 'real', 'traditional', 'honest', 'original', 'wholesome', 'quality of life', 'happiness', 'sentimental' and 'friendly'. The *excitement* category is represented by adjectives, like 'trendy', 'daring', 'exciting', 'exotic', 'fashionable', 'cool', 'spirited', 'dynamic', 'vital', 'fresh', 'young', 'sensorial', 'unique', 'imaginative', 'creative', 'up-to-date', 'independent', 'contemporary', 'cosmopolitan', 'tolerant', 'hospitable'. *Competence* is expressed by adjectives, such as 'reliable', 'hard-working', 'safe', 'rigorous', 'intelligent', 'technical', 'corporate', 'innovative', 'successful', 'leader', 'ambitious' and 'powerful'. *Sophistication* is represented by adjectives, like 'glamorous', 'luxurious', 'seductive', 'smooth', 'romantic', and 'magical'. The last dimension, *ruggedness*, includes adjectives like 'outdoorsy', 'get-away', 'recreational', 'tough' and 'ruggedness'.

The *adjectives* were then compared semantically with all of the categories of emotional values. Previous work on the analysis of content communicated by destinations on social media shows that adjectives are the words that tend to convey the greatest emotional load. In this case, the standard natural language parser to retrieve the adjective was applied. The Wu-Palmer ontology-based semantic similarity measure was applied (Wu and Palmer 1994) in order to compute the similarity between adjectives and emotional values. The Wu-Palmer similarity between two terms, $c1$ and $c2$, is defined as $simWP(c1, c2) = (2 \times N3)/(N1 + N2 + 2 \times N3)$, where $N1$ and $N2$ are the number of hypernym links from the terms $c1$ and $c2$ to their Least Common Subsumer (LCS) in WordNet, and $N3$ is the number of hypernym links from the LCS to the root of the ontology. This measure ranges from 1 (for identical concepts) to 0 (when the LCS is the root of the ontology; for example, the similarity between 'love' and 'romantic' is 0.7619). However, the terms must be *nouns* in order to be compared using this measure. We manually translated emotional values into their equivalent nouns (e.g., 'ambitious' was transformed into 'ambition'), whereas the adjectives were automatically transformed into nouns based on their derivative form in WordNet. Only adjectives with a similarity higher than 0.7 to an emotional value (*emotional adjectives*) were considered in the final steps of the analysis (Lalicic et al. 2018). This analysis allowed us to assess if a destination communicates certain emotional values that distinguish it from other destinations and shows that there is a strategy behind the communication of the brand. Then, we collected all the engagement measures, which for Twitter were 'favourites' and retweets and for Facebook 'like', 'love', 'haha', 'sad' and 'anger'. Concerning, the collection of comments and re-shares, only a few DMOs in the sample received comments or a number of re-shares, thus we decided to not include this form of engagement as a part of the analysis. Through Chi-square analyses, significant relationships between the engagement metrics and brand emotional values were tested.

Table 1 Destination overview of Twitter communication

	Tweets	Retweets	RTs/tw	Favourites	Favs/tw
Paris	2058	51,900	25.22	71,425	34.71
Edinburgh	1935	12,344	6.38	9857	5.09
Barcelona	1828	16,070	8.79	4204	2.30
Berlin	1797	553	0.31	892	0.50
Dublin	1722	21,209	12.32	53,007	30.78
Madeira	1273	2252	1.77	1299	1.02
Tenerife	1102	14,2830	129.61	2199	2.00
Amsterdam	940	14,291	15.20	24,329	25.88
London	633	7800	12.32	19,960	31.53
Budapest	159	274	1.72	557	3.50

4 Results

4.1 DMOs' emotional brand communication and user reactions—Twitter

This section illustrates the DMOs' tweet behaviour and responses. Hence, Table 1 shows, for each destination, the following data:

- Number of tweets sent by the destination in 2016 that contain at least one emotional value.
- Number of retweets of those tweets.
- Average number of retweets of each tweet (rate of the first two columns).
- Number of times that the tweets were marked as 'favourite' by the users.
- Average number of favourites of each tweet (rate of columns four and one).

In total, 13,477 tweets were analysed. As Table 1 shows, Paris was the most active DMO with 2058 tweets, whereas the DMO of Budapest was the least active, exhibiting 159 tweets. However, observing the numbers of retweets (users re-sending the message sent by the DMO), which totals 269,532 retweets, we see a different order. Tenerife stands out significantly (142,830), followed by Paris with 51,900 retweets of the tweets containing emotional values ($p < 0.001$). The ratio indicates how often a tweet has on average been retweeted, which ultimately gives an indication of the DMOs' ability to reach its audience. The DMO of Tenerife was very successful with a 129.61 retweet ratio, meaning that, on average, one tweet was retweeted by Twitters users 129.61 times. The DMO of Berlin was the least successful (0.31 ratio). The 'favourite' (heart icon, 'liking' the tweet), shows that, in total, DMOs' tweets received 187,729 'favourite' distinctions and is, thus, a less popular activity to do for users than retweeting. In this case, Paris stands out significantly with a total of 71,425 'favourites' ($p < 0.001$), and the ratio as well (34.71). Again, the DMO of Berlin was the least successful in this case (ratio 0.50).

Hence, retweeting seems to be a more common response to a DMO's tweet than marking it as 'favourite' is, which also implies that users are sharing within their

network, and, thus, a larger crowd is reached. This highlights the importance of effective communication even more.

However, it is more interesting to understand which explicit adjectives DMOs communicated and which of those triggered user engagement, like retweets and 'favourites'. Firstly, a general tendency to use the following emotional adjectives in tweets is observed: 'honest', 'fresh' and 'glamorous'. Interestingly, there are also a set of adjectives that DMOs hardly used while tweeting, which included: 'reliable', 'rigorous' and 'tolerant'.

Furthermore, the paper investigates which emotional brand values were retweeted by their users in their personal social networks. In general, tweets that predominantly included the adjectives of 'happiness', real' and 'sentimental' were highly retweeted by users. The emotional brand values that the DMOs hardly tweeted were rather similar to those once tweeted by DMOs (e.g., 'rigorous', 'tolerant').

Tweets that contained the following emotional brand values were among the highest marked with 'favourite': 'honest', 'glamorous' and 'down-to-earth' ($p < 0.001$). Adjectives, such as 'rigorous','imaginative' and 'ambitious' were hardly marked as 'favourite'.

This study is also interested in which specific values triggered the most effective user reactions on Twitter. It may be of note that in order to create a successful impact, the following values in ratio to the use and number of retweets were calculated. In the case of retweet ratios, DMO messages that used the adjectives 'real', 'sentimental' and 'happiness' were most successful. Thus, the popularity of these three emotional brand values for Twitter users is clearly indicated. Interestingly, tweets containing adjectives related to 'ambitious, 'magical' and 'hospitable' appeared with a ratio lower than ten, which suggests that they are not well received by the audience and should be avoided in online communication, when aiming to engage effectively with a user base.

However, there is a different order of emotional values in ratio to the number of users marking the DMOs' message as 'favourite'. The most popular adjectives are as follows: 'charming', 'hardworking' and 'successful'. Thus, it is important that DMOs integrate these values into their communication strategies in order to engage with their users, create an impact online and create value for users to interact with them while discussing the destination. For 'favourite', tweets including adjectives such as 'rigorous', 'imaginative' and 'ambitious' are not well received.

Table 2 summarizes the values related to the frequency of the DMOs' communication, user responses and the level of impact that they generate as well as the Aakers' dimensions. As seen, DMOs tend to use a variety of adjectives that cover three dimensions (*sincerity*, *excitement* and *sophistication*). Interestingly, the dimension of *competence* is neither used nor does it often generate a response among users. Considering the ratios with a high impact, these are the values that DMOs should continue to use. Users respond most strongly to 'sincerity'-related adjectives in terms of re-tweet activity, whereas the words listed in Table 2 as 'low', in this case 'competence' and 'sophistication', receive hardly any responses. In terms of increasing the success of a tweet (i.e., getting many followers to mark it as 'favourite'), the dimension of *competence* as Table 2 shows, works much better, whereas *excitement* does not lead to an increased effect of marking the tweet as 'favourite'.

Table 2 Communication and responses on Twitter—emotional brand values

Communication and responses	Emotional brand values	Aakers' dimensions
DMO's most communicated emotional value	*Honest, fresh, glamorous*	*Sincerity, excitement, sophistication*
DMO's least communicated emotional value	*Reliable, rigorous, tolerant*	*Competence*
'retweets' (high)	*Real, happiness, sentimental*	*Sincerity*
'retweets' (low)	*Ambitious, rigorous, tolerant*	*Competence*
'favourites' (high)	*Honest, glamorous, down-to-earth*	*Sincerity, sophistication*
'favourites' (low)	*Rigorous, imaginative, ambitious*	*Ruggedness, competence*
Ratio-'retweets' (high)	*Real, sentimental, happiness*	*Sincerity*
Ratio-'retweets '(low)	*Ambitious, hospital, magical*	*Sophistication, competence*
Ratio-'favourites' (high)	*Charming, hard-working, successful*	*Competence*
Ratio-'favourites' (low)	*Rigorous, imaginative, ambitious*	*Excitement, competence*

Table 3 Facebook communication—user reactions

	Posts	'Likes'	'Haha'	'Angry'	Likes/post	Haha/post	Angry/post
Paris	229	65,502	1099	68	286.03	4.80	0.30
Edinburgh	524	442,074	316	48	843.65	0.60	0.09
Barcelona	1337	215,082	133	104	160.87	0.10	0.08
Berlin	42	5044	4	1	120.10	0.10	0.02
Dublin	251	150,632	157	37	600.13	0.63	0.15
Madeira	1936	98,957	112	18	51.11	0.06	0.01
Tenerife	637	44,261	9	7	69.48	0.01	0.01
Amsterdam	372	212,046	669	77	570.02	1.80	0.21
London	610	462,072	1312	142	757.50	2.15	0.23
Budapest	197	6305	1918	1	32.01	9.74	0.01

4.2 DMOs' emotional brand communication and user reactions – Facebook

In this section, Facebook is analysed based on DMOs' communication and user reactions. Table 3 shows the following data of each destination:

- Number of posts sent in 2016 containing at least one emotional adjective.
- Columns 2, 3 and 4 show the number of user reactions to those posts ('likes', 'haha', 'laughter' and 'angry', respectively).
- Columns 5, 6 and 7 show the average number of each type of reaction per post.

In total, DMOs generated 6135 Facebook posts. The DMO of Madeira (1936 posts) was very active on Facebook, followed by the DMO of Barcelona (1337 posts). Concerning user responses, the 'like' was the most popular one, followed by 'haha' at a considerable distance. In total 1701,975 'likes' were given. In this case, the DMOs of London (462,072 likes) and Edinburgh (442,074) were rather successful in triggering responses from their posts that contained emotional values ($p < 0.001$). The DMOs from Berlin and Budapest tended to be rather passive on Facebook, and yet they still generated more than 5000 'likes' for their posts that contained emotional adjectives. It is remarkable that the number of 'haha' and 'angry' reactions is almost negligible. In total, 5729 'haha' distinctions were made, primarily by users responding to the DMOs of Paris and London. The 'angry' responses were relatively infrequent (503), and the posts made by the DMOs of Barcelona and London DMOs received most of them.

Also where Facebook communication is concerned, adjectives used by DMOs in relation to user reactions were analysed. Overall, posts containing the following adjectives were the most successful in reaching their users: 'honest', 'traditional' and 'trendy'. Interestingly, some values were not used by DMOs at all in their posting communication, such as 'getaway', 'recreational', and 'tough'. The rest of the values exhibited rather equal use in DMOs' Facebook communication and its impact on user responses.

In analysing the posts that received the most reactions from users, we have three reactions to consider. Firstly, for the user reaction 'like', the following set of values stands out; 'honest', 'trendy' and 'vital' ($p<0.001$), whereas posts that merely contained words related to 'charming', 'recreational' and 'ambitious' did not receive many 'likes'. Secondly, the 'haha' icon was given to posts containing adjectives like 'exciting', 'daring' and 'vital' ($p<0.05$). Finally, the 'angry' user reaction was recorded most in posts that used the following adjectives: 'down-to-earth', 'calm' and 'family-oriented'. However, no significant differences ($p=0.251$) were identified.

The ratio between the values mentioned in a post and user engagement was observed. This provides rather valuable insight for DMOs, as they are able to use this information to understand which values are best received by their respective audiences. The emotional values communicated in the messages that received the most 'likes' are 'getaway', 'tough' and 'young' ($p<0.001$), whereas adjectives like 'ambitious, 'reliable' and 'innovative' did not receive many 'likes', even though the DMOs communicated these words relatively frequently. The reaction of 'haha' is mostly in keeping with adjectives such as 'exotic', 'daring' and 'exciting' ($p<0.05$). For the 'angry' reaction, there is very little impact from the values related to the adjectives 'getaway', 'family-oriented' and 'real', however, it is not significant ($p=0.438$). Depending on the type of responses required, the use of values and the frequency of the values used may help to trigger the desired impact on the brand community. Table 4 summarizes these values in relation to the frequency of communication, user responses and the level of impact they generate. Once again, the values are linked to the Aakers' dimensions, and in doing so, Facebook posts are dominantly communicating values related to the dimension of *sincerity* and the least related to the dimension of *ruggedness*. Users tend to respond to posts focusing on *excitement* much more. Thus, in terms of impact, those posts that receive a higher number of reactions tend to focus on *excitement* and *ruggedness*.

5 Discussion

In conducting this research, we aimed to provide a more detailed analysis of brand emotional values used by DMOs and their effects on user engagement across Facebook and Twitter. This paper was interested in analysing the most popular destinations according to the TripAdvisor ranking from 2017, as those destinations are likely to have successful online profiles and strategies that stand out and engage with a large fan base. Lalicic et al. (2018) demonstrated that DMOs use the platforms in significantly different ways in terms of activity, but that they tend to use similar brand values. However, insight into which values are used precisely and can trigger users to engage with a DMO's posts or tweets remained unanswered.

This study has several interesting discussion points that enhance the current literature on DMO communication and user engagement in social media spheres. Firstly, the study shows that DMOs, even if they are not very active, have a significant amount of users responding to their activities on both platforms. However, the paper shows that DMOs are more successful with regards to receiving reactions

Table 4 Communication and responses on Facebook—emotional brand values

Communication and responses	Emotional brand values	Aakers' dimensions
DMOs' most communicated emotional value	*Honest, traditional, trendy*	*Sincerity*
DMOs' least communicated emotional value	*Getaway, recreational, tough*	*Ruggedness*
'likes' (high)	*Honest, trendy, vital*	*Excitement*
'likes' (low)	*Charming, recreational, ambitious*	*Ruggedness, sophistication, competence*
'haha' (high)	*Exciting, daring, vital*	*Excitement*
'haha' (low)	*Though, get-away, leader*	*Ruggedness and competence*
'angry' (high)	*Down-to-earth, calm, and family-oriented*	*Sincerity*
'angry' (low)	*Spirited, glamorous, creative*	*Excitement*
ratio-'likes' (high)	*Get-away, tough, young*	*Ruggedness, excitement*
ratio-'likes' (low)	*Ambitious, reliable, innovative*	*Competence*
ratio-'haha' (high)	*Exotic, daring, exciting*	*Excitement*
ratio-'haha' (low)	*Reliable, leader, rugged*	*Ruggedness, competence*
ratio-'angry' (high)	*Get-away, family-oriented, real*	*Ruggedness*
ratio-'angry' (low)	*Creative, spirited, dynamic*	*Excitement*

on Facebook as opposed to Twitter. This is in line with Huertas and Marine-Roig's (2016a) argument that Facebook is a very suitable platform for communicating DMO emotional brand values as well as generating user engagement.

Secondly, this study demonstrates similar to Lalicic et al. (2018), that DMOs use similar brand adjectives across the platforms, such as 'honest' and 'glamorous'. However, the study shows that the distribution of successful emotional brand values in terms of user reactions is significantly different. In doing so, new insights are provided. For example, on Twitter, users respond much more strongly to tweets containing adjectives related to 'happiness' and 'sentimental'. While on Facebook, users 'like' posts with adjectives related to 'honest' and 'vital' much more. Furthermore, the study calculated the ratio of emotional values and specific users' responses. This demonstrates that for Twitter and Facebook, two different emotional brand communication styles are valued by the users (Twitter: 'real' and 'happiness' and Facebook: 'tough' and 'young'*)*. Thus, DMOs were successful in using the 'right' adjectives on Twitter, whereas on Facebook, new adjectives could be integrated to create more effective posts. Linking this back to the overall personality dimensions by Aaker, for Twitter, DMOs focus on the dimension of *excitement* and *sincerity*, whereas the most effective messages are focused on the *sincerity* and *competence* adjectives. We also see a variance on Facebook, where DMOs communicate about *sincerity* and users respond much more strongly to adjectives related to *excitement* and *ruggedness*. This is similar to other studies analysing DMOs' communication and usage of brand personality values (Ekinci and Hosany 2006; De Moya and Jain 2013; Huertas and Mariné-Roig 2016b; Dickinger and Lalicic 2016). In doing so, the study also implies that if DMOs communicate a wide range of brand vales, users only respond to certain values that fit with their image as well as self-congruity with the destination.

Thirdly, the study demonstrates compared to other studies (Huertas and Mariné-Roig 2016b), that some emotional values can trigger different responses from users. For example, if a tweet includes adjectives related to *honesty*, followers tend to re-share this much more as opposed to marking it as 'favourite'. For Facebook, the type of response also tells us something about user opinions about the post. Thus, words related to 'daring' and 'exciting' seem to create a laughter response among users rather than just a 'like'. In doing so, the study adds news insights to existing studies capturing user engagement of tourism brands (Su et al. 2015; Usakli et al. 2017).

Fourthly, the study's observations allow us to speculate about so called 'effective values' that DMOs should continue to use in order to effectively reach out to their users (those with high ratio values). Compared to other studies, such practical recommendations is new (Jabreel et al. 2017). In addition, the study also identified some 'ineffective values', meaning that DMOs used specific adjectives (e.g., 'tough' and 'sentimental' on Twitter), but did not seem to spark any significant responses from the user base. This would link to the dimension of *ruggedness*, which collides with the preferred (emotional) destination experiences users have in mind, following Ekinci and Hosany's (2006) line of reasoning.

Overall, and in contrast to other studies, which have analysed DMOs' communication on a general level (Huertas and Mariné-Roig 2016a) or integrated the concept of Aakers' online brand personality (Dickinger and Lalicic 2016), this study

provides new insights for theories explaining DMOs' online successes but also detailed and hands-on implications are given. In fact, the study enhances the understanding of how to tactically use platforms, like Facebook or Twitter, to successfully implement emotional branding strategies while reaching out to a large group of followers.

Thus, with regards to the practical implications, DMOs are provided with a better idea of how to communicate their destination brands online while attempting to integrate the emotional brand experience. The concept of brand personality has proven to be a successful approach to connect with travellers and enhance the overall experience. However, the formation of an image, which nowadays is merely done online, should not be underestimated (Amaro et al. 2016; Fournier and Avery 2011). The power of social media and the tools available to stand as a brand are enormous (De Vries et al. 2012). It is, however, important to observe which elements of the brand communication users respond to. The suggested 'effective and ineffective values' is a first attempt to support practitioners in understanding and designing their communication strategies carefully. As seen, some values might trigger different responses than others. For example, retweeting behaviour has a much wider impact as opposed to users who mark posts as a 'like' or 'favourite'. Such engagement metrics need to be understood in relationship to the effect of brand awareness and also the brand popularity online in a larger network (Su et al. 2015). The paper mainly focuses on general words and provides examples of specific values per destination. Hence, DMOs are also advised to verify whether the values correspond to their brand and brand experience, and if users tend to like a specific value more than a DMO communicates it, which might also help them to consider a repositioning of their brand to specific consumers to some extent. If the values do not correspond with the DMO's brand at all, DMOs can consider investigating why consumers tend to respond to those values, and whether there are new ways to offer the destination brand experience. Of course, this type of discussion opens up many new thoughts and avenues for future research.

For example, destinations are advised to take a more longitudinal approach, which would allow them to understand the possible effects of specific marketing tactics on revenue or other metrics of success. Furthermore, there is much more to explore. In order to create a holistic understanding, one could analyse how values co-occur, and which co-occurrences of values are successful in terms of consumer engagement. In addition, besides focusing solely on communication, future studies should also include other elements, such as whether a photo, video or link was included in the messages. This will also provide more insight into designing effective messages (Mariani et al. 2016). Information on the characteristics of the users could also further enhance our understanding. In particular, if DMOs wish to address specific segments, more interaction among users needs to be investigated. Lastly, the use of experiments to control for various variables and their effects on user behaviour in specific segments can be integrated in order to generate a better understanding of user segmentation techniques in an online setting.

This study also acknowledges a set of limitations. Firstly, in terms of Facebook, the metric of likes and likewise emotional responses have been recorded, but no comments are shares were analysed. This form of engagement is more active and

could be integrated by future research as well. Another issue with social media is the existence of possible bots. Thus, such issues could be considered more carefully. The present study also only collected messages exclusively in English, not in other languages. Furthermore, only the textual content of the message was considered, meaning that other important sources of information (e.g. attached pictures) were not considered, despite the fact that they would provide more insights into the effectiveness of the overall presentation of DMOs' online presence. Also, the manual translation of Aaker's emotional values to WordNet nouns is entirely subjective, and it may lead to errors and unexpected results, especially where polysemic words are considered. The automatic translation of adjectives into nouns can also suffer from the same problem. Lastly, a full semantic analysis of the sentences has not been undertaken nor were linguistic issues, such as the detection of irony in statements (quite common in Twitter in general, although probably not in DMOs' messages), considered.

References

Aaker JL (1997) Dimensions of brand personality. J Mark Res 34(3):347–356

Amaro S, Duarte P, Henriques C (2016) Travelers' use of social media: a clustering approach. Ann Tour Res 59:1–15. https://doi.org/10.1016/j.annals.2016.03.007

Ashley C, Tuten T (2015) Creative strategies in social media marketing: an exploratory study of branded social content and consumer engagement. Psychol Mark 32(1):15–27. https://doi.org/10.1002/mar.20761

Bigné E, García IS, Blas SS (2009) The functional-psychological continuum in the cognitive image of a destination: a confirmatory analysis. Tour Manag 30(5):715–723

Blain C, Levy SE, Ritchie JB (2005) Destination branding: insights and practices from destination management organizations. J Travel Res 43(4):328–338. https://doi.org/10.1177/0047287505274646

Chan NL, Guillet BD (2011) Investigation of social media marketing: how does the hotel industry in Hong Kong perform in marketing on social media websites? J Travel Tour Mark 28(4):345–368. https://doi.org/10.1080/10548408.2011.571571

Chaykina T, Guerreiro M, Mendes J (2014) Destination brand personality of Portugal for the Russian-speaking market. J Spat Organ Dyn 2(1):23–40

De Moya M, Jain R (2013) When tourists are your "friends": Exploring the brand personality of Mexico and Brazil on Facebook. Public Relat Rev 39:23–29. https://doi.org/10.1016/j.pubrev.2012.09.004

De Vries L, Gensler S, Leeflang PS (2012) Popularity of brand posts on brand fan pages: an investigation of the effects of social media marketing. J Interact Mark 26(2):83–91. https://doi.org/10.1016/j.intmar.2012.01.003

Dickinger A, Lalicic L (2016) An analysis of destination brand personality and emotions: a comparison study. Inf Technol Tour 15(4):317–340. https://doi.org/10.1007/s40558-015-0044-x

Ekinci Y, Hosany S (2006) Destination personality: an application of brand personality to tourism destinations. J Travel Res 45:127–139. https://doi.org/10.1177/0047287506291603

Fournier S, Avery J (2011) The uninvited brand. Bus Horiz 54(3):193–207. https://doi.org/10.1016/j.bushor.2011.01.001

Gobe M (2010) Emotional branding: the new paradigm for connecting brands to people. Allworth Press, New York

Govers R, Go FM, Kumar K (2007) Promoting tourism destination image. J Travel Res 46(1):15–23. https://doi.org/10.1177/0047287507302374

Guillet B, Kucukusta D, Liu L (2016) An examination of social media marketing in China: how do the top 133 hotel brands perform on the top four Chinese social media sites? J Travel Tour Mark 33(6):783–805. https://doi.org/10.1080/10548408.2015.1064337

Hays S, Page SJ, Buhalis D (2013) Social media as a destination marketing tool: its use by national tourism organisations. Curr Issues Tour 16(3):211–239. https://doi.org/10.1080/13683500.2012.662215

Hosany S, Ekinci Y, Uysal M (2006) Destination image and destination personality: an application of branding theories to tourism places. J Bus Res 59(5):638–642

Hosany S, Ekinci Y, Uysal M (2007) Destination image and destination personality: an application of branding theories to tourism places. J Bus Res 59(5):638–642

Huertas A, Mariné-Roig E (2015) Destination brand communication through the social media: what contents trigger most reactions of users. In: Tussyadiah I, Inversini A (eds) Information and communication technologies in tourism 2015. Springer, Basel, pp 295–308

Huertas A, Mariné-Roig E (2016a) User reactions to destination brand contents in social media. Inf Technol Tour 15:291–315. https://doi.org/10.1007/s40558-015-0045-9

Huertas A, Mariné-Roig E (2016b) Differential destination content communication strategies through multiple social media. In: Inversini A, Schegg R (eds) Information and communication technologies in tourism 2016. Springer, Basel, pp 239–252

Jabreel M, Moreno A, Huertas A (2017) Do local residents and visitors express the same sentiments on destinations through social media? In: Schegg R, Stangl B (eds) Information and communication technologies in tourism 2017. Springer, Basel, pp 655–668

Kabadayi S, Price K (2014) Consumer–brand engagement on Facebook: liking and commenting behaviors. J Res Interact Mark 8(3):203–223. https://doi.org/10.1108/JRIM-12-2013-0081

Koltringer C, Dickinger A (2015) Analyzing destination branding and image from online sources: a web content mining approach. J Bus Res 68(9):1836–1843. https://doi.org/10.1016/j.jbusres.2015.01.011

Lalicic L, Gindl S (2018) DMOs' Facebook success stories: a retrospective view. In: Stangl B, Pesonen J (eds) Information and communication technologies in tourism 2018. Springer, Basel, pp 172–185. https://doi.org/10.1007/978-3-319-72923-7_14

Lalicic L, Huertas A, Moreno A, Gindl S, Jabreel M (2018) Do DMOs communicate their emotional brand values? A comparison between Twitter and Facebook. In: Stangl B, Pesonen J (eds) Information and communication technologies in tourism 2018: Springer, Basel, pp 159–171. https://doi.org/10.1007/978-3-319-72923-7_13

Laroche M, Habibi MR, Richard MO (2013) To be or not to be in social media: how brand loyalty is affected by social media? Int J Inf Manag 33:76–82. https://doi.org/10.1016/j.ijinfomgt.2012.07.003

Li SC, Robinson P, Oriade A (2017) Destination marketing: the use of technology since the millennium. J Destin Mark Manag. https://doi.org/10.1016/j.jdmm.2017.04.008

Litvin SW, Goldsmith RE, Pan B (2008) Electronic word-of-mouth in hospitality and tourism management. Tour Manag 29(3):458–468. https://doi.org/10.1016/j.tourman.2007.05.011

Liu Z, Park S (2015) What makes a useful online review? Implication for travel product websites. Tour Manag 47:140–151. https://doi.org/10.1016/j.tourman.2014.09.020

Lovejoy K, Saxton GD (2012) Information community and action: how nonprofit organizations use social media. J Comput-Mediat Commun 17(3):337–353. https://doi.org/10.1111/j.1083-6101.2012.01576.x

Lu W, Stepchenkova S (2015) User-generated content as a research mode in tourism and hospitality applications: topics, methods, and software. J Hosp Mark Manag 24(2):119–154. https://doi.org/10.1080/19368623.2014.907758

Mariani MM, Di Felice M, Mura M (2016) Facebook as a destination marketing tool: evidence from Italian regional destination management organizations. Tour Manag 54:321–343. https://doi.org/10.1016/j.tourman.2015.12.008

Mariné-Roig E (2017) Measuring destination image through travel reviews in search engines. Sustainability 9(8):1425. https://doi.org/10.3390/su9081425

Michaelidou N, Siamagka NT, Moraes C, Micevski M (2013) Do marketers use visual representations of destinations that tourists value? Comparing visitors' image of a destination with marketer-controlled images online. J Travel Res 52(6):789–804. https://doi.org/10.1177/0047287513481272

Míguez-González MI, Huertas A (2015) The power of photographs in the communication and public relations of tourist destinations and their brands through Facebook and Flickr. Catalan J Commun Cult Stud 7(2):197–215

Moran E, Gossieaux F (2010) Marketing in a hyper-social world: the tribalization of business study and characteristics of successful online communities. J Advert Res 50(3):232–239. https://doi.org/10.2501/S0021849910091397

Moreno A, Jabreel M, Huertas A (2015) Automatic analysis of the communication of tourist destination brands through social networks. In: Proceedings of the 10th international conference on intelligent systems and knowledge engineering (ISKE-2015), pp 24–27

Morgan N, Pritchard A (2004) Meeting the destination branding challenge. In: Morgan N, Pritchard A, Pride R (eds) Destination branding: creating the unique destination proposition. Butterworth-Heinemann, Oxford, pp 59–79

Morgan N, Pritchard A, Pride R (2003) Marketing to the Welsh diaspora: the appeal to hiraeth and homecoming. J Vacat Mark 9(1):69–80

Munar AM, Jacobsen JKS (2014) Motivations for sharing tourism experiences through social media. Tour Manag 43:46–54. https://doi.org/10.1016/j.tourman.2014.01.012

Nusair KK, Bilgihan A, Okumus F (2013) The role of online social network travel websites in creating social interaction for Gen Y travellers. Int J Tour Res 15(5):458–472. https://doi.org/10.1002/jtr.1889

Oviedo-García MA, Muñoz-Expósito M, Castellanos-Verdugo M, Sancho-Mejías M (2014) Metric proposal for customer engagement in Facebook. J Res Interact Mark 8(4):327–344. https://doi.org/10.1108/JRIM-05-2014-0028

Papathanassis A, Knolle F (2011) Exploring the adoption and processing of online holiday reviews: a grounded theory approach. Tour Manag 32(2):215–224. https://doi.org/10.1016/j.tourman.2009.12.005

Pitt LF, Opoku R, Hultman M, Abratt R, Spyropoulou S (2007) What I say about myself: communication of brand personality by African countries. Tour Manag 28(3):835–844

Roque V, Raposo R (2015) The usage of social media by the destination management organizations. The case study of the Portuguese DMO. Dos Algarves 26(2):87–114. https://doi.org/10.18089/DAMeJ.2015.26.2.5

Saffer AJ, Sommerfeldt EJ, Taylor M (2013) The effects of organizational Twitter interactivity on organization–public relationships. Public Relat Rev 39(3):213–215

Schroeder A, Pennington-Gray L (2015) The role of social media in international tourist's decision making. J Travel Res 54(5):584–595. https://doi.org/10.1177/0047287514528284

Stepchenkova S, Zhan F (2013) Visual destination images of Peru: comparative content analysis of DMO and user-generated photography. Tour Manag 36:590–601. https://doi.org/10.1016/j.tourman.2012.08.006

Su N, Reynolds D, Sun B (2015) How to make your Facebook posts attractive: a case study of a leading budget hotel brand fan page. Int J Contemp Hosp Manag 27(8):1772–1790. https://doi.org/10.1108/IJCHM-06-2014-0302

TripAdvisor (2017) Traveller destination choice—top 25 destinations in Europe. https://www.tripadvisor.co.uk/TravelersChoice-Destinations-cTop-g4 Accessed 29 Jun 2017)

Uşaklı A, Koç B, Sönmez S (2017) How'social' are destinations? Examining European DMO social media usage. J Destin Mark Manag 6:136–149. https://doi.org/10.1016/j.tmp.2014.01.001

Van Doorn J, Lemon KN, Mittal V, Nass S, Pick D, Pirner P, Verhoef PC (2010) Customer engagement behavior: theoretical foundations and research directions. J Serv Res 13(3):253–266. https://doi.org/10.1177/1094670510375599

Walther JB, Jang JW (2012) Communication processes in participatory websites. J Comput-Mediat Commun 18(1):2–15. https://doi.org/10.1111/j.1083-6101.2012.01592.x

Wattanacharoensil W, Schuckert M (2015) How global airports engage social media users: a study of Facebook use and its role in stakeholder communication. J Travel Tour Mark 32(6):656–676. https://doi.org/10.1080/10548408.2014.955245

Wu Z, Palmer M (1994) Verbs semantics and lexical selection. In: Proceedings of the 32nd annual meeting on association for computational linguistics. Association for Computational Linguistics, pp 133–138

Xiang Z, Gretzel U (2010) Role of social media in online travel information search. Tour Manag 31(2):179–188. https://doi.org/10.1016/j.tourman.2009.02.016

Affiliations

Lidija Lalicic[1] · Assumpció Huertas[2] · Antonio Moreno[3] · Mohammed Jabreel[3]

Assumpció Huertas
sunsi.huertas@urv.cat

Antonio Moreno
antonio.moreno@urv.cat

Mohammed Jabreel
mohammed.jabreel@urv.cat

[1] Department of Tourism and Service Management, MODUL University Vienna, Vienna, Austria

[2] Departament d'Estudis de Comunicació, Universitat Rovira i Virgili, Tarragona, Spain

[3] Departament d'Enginyeria Informàtica i Matemàtiques, ITAKA: Intelligent Technologies for Advanced Knowledge Acquisition, Universitat Rovira i Virgili, Tarragona, Spain

UNIVERSITAT
ROVIRA i VIRGILI