

Serveis avançats per a turistes en funció de la seva localització

Dra. Aïda Valls

Dra. Yolanda Pérez

Dr. Jordi Castellà

Carles Gil

Raül Miró

Natán García

Índex de continguts

Introducció	1
1. Estudi dels projectes de recerca i de les tecnologies que s'utilitzen.....	5
1.1. Estudi dels projectes de localització i seguiment	5
1.1.1 . Pedestrian and public transport navigator (PEPTRAN).....	6
1.1.2. Personalised Tourist Services Using Geographic Information Systems via Internet (TOURSERV).....	6
1.1.3. Creation of user-friendly mobile services personalised for tourism (CRUMPET).....	7
1.1.4. GEographically relevant information for moBile users in Protected Areas (WEBPARK).....	8
1.1.5. Intelligent Mobility Agent for Complex Geographic Environments (IMAGE).....	9
1.1.6. LOcation Aware Visually Enhanced Ubiquitous Services (LOVEUS).....	9
1.1.7. EURopEAn WatErways networked information system (EUREAUWEB).....	10
1.1.8. Ambient, personalised, and context-sensitive information systems for mobile users (AMBIESENSE).....	11
1.1.9. Multimedia Geoinformation in Rural Areas with Eco-Tourism (REGEO).....	11
1.1.10. Mobile Tourism Guide (m-ToGuide).....	12
1.1.11. 3G European Location Based Advertising (ELBA).....	12
1.1.12. Cultural location based information services (M-GUIDE).....	13
1.1.13. Gaming and guiding system for museum and exhibition environments (EXPLORE). 13	
1.1.14. Intelligent mobility agents, advanced positioning and mapping technologies integration interoperable multimodal, location based services (IM@GINE IT).....	13
1.1.15. Intelligent tourism and cultural information through ubiquitous services (ITACITUS).....	14
1.2. Estudi de les tecnologies que s'estan utilitzant per localitzar i fer el seguiment de persones 15	
1.2.1. Mètodes bàsics de posicionament	15
1.2.2 Classificació de les tecnologies de posicionament.....	19
1.2.3. Tecnologies de posicionament per satèl·lit.....	19
1.2.4. Tecnologies de posicionament basades en xarxes cel·lulars	22
1.2.5 Tecnologies de posicionament en espais interiors.....	25

1.3 Bibliografia del capítol.....	28
2. Estudi dels diferents tipus de serveis que es poden oferir a partir de la localització i/o seguiment de turistes.....	31
2.1 Estudi de l'oferta de serveis de turistes segons la seva localització i/o seguiment.....	31
2.2 Identificació i classificació dels serveis que s'estan oferint a turistes i proposta de possibles nous serveis.....	33
2.3 Bibliografia del capítol.....	35
3. Estudi de les tècniques per garantir la privadesa dels turistes	45
3.1 Estudi de la privadesa en Sistemes Basats en la Localització.....	47
3.1.1 Classificació dels mètodes per a la privadesa en LBS	48
3.1.2 Privadesa en sistemes basats amb TTP no pertorbatius	49
3.1.3 Privadesa en sistemes basats amb TTP pertorbatius	51
3.1.4 Privadesa en sistemes sense TTP no pertorbatius	52
3.1.5 Privadesa en sistemes sense TTP pertorbatius	54
3.1.6 Anonimització segons l'origen de les dades.....	55
3.2 Bibliografia sobre privadesa en sistemes LBS	56
3.3 Privadesa en dades sobre trajectòries.....	58
3.3.1. Definició.....	59
3.3.2. Mètodes d'anonimització de trajectòries.....	60
3.4. Bibliografia sobre anonimització de trajectòries.....	71
4. Estudi de les tècniques d'anàlisi de dades per modelar el comportament de turistes	73
4.1. Generació de perfils d'usuari amb dades espacio-temporals	73
4.1.1 Tipus de Problemes	74
4.1.2 Mètodes d'agrupació (clustering).....	76
4.1.3 Patrons Locals Espaciotemporals.....	83
4.1.4 Predicció.....	89
4.1.5 El paper de la incertesa en la Minería de Dades Espai-Temporals	92
4.1.6 Software per a la Minería de Dades	92
4.1.7 Bibliografia sobre Minería de dades	94
4.2. Estudi de les variables que es tenen en compte per l'anàlisi del comportament i fluxos de turistes	96
5. Estudi de les noves tendències en la provisió de serveis avançats a turistes que ofereixen les tecnologies basades en localització i seguiment	99

5.1 Anàlisi dels projectes de recerca sobre mobilitat	99
5.1.1 Estudi de la privadesa en els projectes de turisme mòbil	101
5.1.2. Comparativa de la tecnologia segons el tipus de servei	102
5.2 Serveis avançats per turistes.....	104
5.2.1 Investigadors que fan estudis de mobilitat de turistes.....	104
5.2.2 Classificació dels serveis que s'estan oferint als turistes i proposta de possibles nous serveis.....	105
5.3. Anàlisi de la protecció de la privadesa dels turistes	106
5.3.1 Anàlisi dels mètodes d'anonimització en LBS.....	107
5.3.2 Anàlisi dels mètodes d'anonimització de trajectòries	108
5.4. Mètodes de mineria de dades	113
5.5. Estudi de les variables a tenir en compte	116
5.5.1. Variables en els estudis de mobilitat de turistes.....	116
5.5.2 Anàlisi de les variables.....	120
6. Conclusions i comentaris finals	125

Introducció

Aquest document recull els resultats del projecte 2008-Turisme02 fruit del *Conveni de cooperació entre la Universitat Rovira i Virgili, la Fundació URV i el Parc Científic i Tecnològic Turisme i Oci de la Costa Daurada* i finançat per la Universitat Rovira i Virgili durant l'any 2009. L'objectiu del projecte consistia en estudiar l'estat actual de provisió de serveis personalitzats a turistes, a partir d'informació de la seva localització o seguiment. Aquesta línia de recerca és d'especial interès pel Parc Científic i Tecnològic de Turisme i Oci (PCTTO) de Tarragona i les Terres de l'Ebre. En aquest projecte, que pretén ser un punt de partida per futures col·laboracions entre la URV i el PCTTO hi han participat 3 grups de recerca de diferents disciplines, i ha estat coordinat per la Dra. Aïda Valls.

ITAKA: Tecnologies Intel·ligents per a la Gestió Avançada del Coneixement

Web: <http://deim.urv.cat/~itaka>

Director del grup ITAKA: Dr. Antonio Moreno Ribas (antonio.moreno@urv.cat)

Responsable del grup en aquest treball: Dra. Aïda Valls Mateu (aida.valls@urv.cat)

Participants: Dr. Antonio Moreno, Dr. David Sánchez, Montserrat Batet, Carles Gil, Efraín Ricardo Sosa

CRISES: Criptografia i Secret Estadístic

Web: <http://crises-deim.urv.cat>

Director del grup CRISES: Dr. Josep Domingo Ferrer (josep.domingo@urv.cat)

Responsable del grup en aquest treball: Dr. Jordi Castellà (jordi.castella@urv.cat)

Participants: Dr. Josep Domingo-Ferrer, Dr. Agustí Solanas, Dr. Antoni Martínez-Ballesté, Dr. Josep Ma. Mateo, Raül Miró

GRATET: Anàlisi Territorial i Estudis Turístics

Web: http://www.sre.urv.cat/web/geografia/fitxa_gratet.htm

Director del grup GRATET: Dr. Salvador Antón (salvador.anton@urv.cat)

Responsable del grup en aquest treball: Dra. Yolanda Pérez (myolanda.perez@urv.cat)

Participants: Anna Julián, Natán García

El document s'estructura en capítols d'acord amb els objectius del projecte. Els capítols 1, 2, 3 i 4 expliquen l'estat actual de la recerca en tecnologies, serveis, privadesa i anàlisi intel·ligent de dades

de localització i/o seguiment de persones, amb especial interès pels estudis aplicats a turisme. El capítol 5 fa un anàlisi de l'estat actual de cadascuna d'aquests temes, i presenta noves línies de treball en els respectius àmbits. Finalment, en el capítol 6 es presenten algunes conclusions i comentaris generals. A continuació es fa una breu introducció per motivar l'interès del tema estudiat.

L'estudi de la mobilitat de les persones és una àrea de recerca emergent que planteja reptes importants a la comunitat científica.

En els darrers anys, les noves tecnologies permeten obtenir informació acurada de les seqüències d'accions realitzades per una persona. L'estudi d'aquest tipus de dades es preveu que serà molt important per que permet que es puguin dissenyar i desenvolupar nous tipus d'aplicacions avançades. Tres exemples en diferents àmbits són:

- l'estudi de les seqüències de proteïnes que descriuen la composició dels amino-àcids obre les portes a nous estudis genètics;
- les dades d'un historial mèdic permeten conèixer la història clínica d'un pacient i es podrien usar per fer medicina preventiva;
- el seguiment d'un usuari d'Internet (a partir de conèixer la seqüència de pàgines Web que visita), permet conèixer els interessos i comportament dels usuaris per tal dissenyar millors pàgines Web;

Les TIC ofereixen un nou espai d'innovació i d'oportunitats per a les empreses, fent-les més competitives en sectors com la indústria, els serveis, el comerç, l'educació, però també el turisme. La tecnologia i les seves aplicacions permeten oferir serveis basats en localització (*Location based services*, LBS) dels usuaris. Aquesta localització es pot conèixer a través d'una varietat de dispositius mòbils que són capaços d'obtenir la ubicació de l'usuari, per exemple, mitjançant l'ús d'un GPS. D'aquesta manera es poden oferir serveis personalitzats als usuaris, proporcionant informació que ha estat creada, recopilada, seleccionada o filtrada, tenint en compte la localització actual de l'usuari. Un dels serveis típics de LBS és la cerca de llocs d'interès per l'usuari, com per exemple, trobar el lloc on es fa una exposició més proper a la ubicació actual de l'usuari.

Quan aquest tipus de consultes sobre localitzacions es fan amb freqüència pel mateix individu, es pot fer el seguiment (*tracking*) de la persona. El fet de conèixer la ubicació de l'usuari i les seves rutes, juntament amb altra informació personal (informació demogràfica o els seus interessos), permet estudiar el seu comportament (model o perfil) per tal d'oferir-li nous serveis a mida o adaptar els serveis existents a les seves necessitats.

En el sector turístic el potencial d'aquest tipus de tecnologies i serveis és molt rellevant. Resulta interessant disposar de dades de seguiment i localització de turistes, per estudiar quins serveis nous i avançats es poden oferir. Per exemple, el seguiment d'un grup de persones (mitjançant dades espai-temporals) comporta poder obtenir seqüències de dades que contenen informació detallada sobre el comportament personal o vehicular de l'individu. A partir d'aquest, es poden trobar patrons que es

poden utilitzar per controlar el tràfic (p.e. congestió en carreteres o carrers d'una ciutat), millorar la sostenibilitat del transport, millorar l'accessibilitat a certs serveis o donar informació personalitzada a partir de la seva localització.

Degut a l'enorme quantitat de dades espacio-temporals dels individus, es fa impossible poder analitzar-les de forma manual. Per solucionar aquest problema, han sorgit una sèrie de tècniques que faciliten el processament i anàlisi automàtic de dades s'anomena mineria de dades (*Data Mining*). La mineria de dades estudia tècniques d'anàlisi intel·ligent de dades per tal d'extreure coneixement útil de grans volums de dades (*Knowledge Discovery*).

No obstant, el seguiment i localització dels individus atempta directament contra la seva privadesa. La privadesa dels usuaris és un dret fonamental reconegut a la declaració dels drets humans de les Nacions Unides. L'emmagatzemament de dades i l'estudi de la ubicació de l'usuari per fer-ne un seguiment atempten directament contra aquest dret fonamental. Per tant, es fa necessària la utilització de mecanismes de recollida i anonimització de les dades que tinguin en compte la privadesa, com per exemple recollir les dades només quan siguin necessàries i anonimitzar-les per tal que no es pugui relacionar la informació amb la ubicació de l'usuari. No obstant això la privadesa no ha de tenir un impacte negatiu en el sentit que la utilitat de les dades publicades ha de ser preservada.

Enfront d'aquest nou paradigma emergent, es necessita conèixer quin és l'estat actual dels avenços tecnològics en l'àrea de provisió de serveis avançats a partir d'informació de seguiment i/o localització, així com, identificar quina seria la seva possible aplicació en l'àmbit de la indústria del turisme. Això permetrà identificar quins agents poden estar interessats en oferir aquests tipus de serveis avançats i, quines noves línies de recerca es poden obrir a nivell internacional que suposin un avenç en aquest àmbit.

L'objectiu principal d'aquest projecte és estudiar l'estat actual de provisió de serveis personalitzats a turistes, a partir d'informació de la seva localització o seguiment, així com d'altres variables que es puguin conèixer.

Aquest objectiu es pot dividir en els següents sub-objectius:

O1: Estudiar els projectes i tecnologies existents per realitzar una localització dels turistes.

O2: Estudiar els diferents tipus de serveis que es poden oferir a partir de la localització i/o seguiment dels turistes

O3: Estudiar quines tècniques per garantir la privadesa dels turistes s'estan desenvolupant i en quins casos s'apliquen: en funció dels serveis a oferir o en funció de l'origen de les dades. Es poden considerar orígens en entorns tancats (dins d'un parc temàtic), en una zona concreta del territori o directament de tots els turistes que visiten el nostre país.

O4: Estudiar quines tècniques d'anàlisi intel·ligent de dades s'estan desenvolupant i aplicant per modelar el comportament dels turistes i generar perfils (*clustering*).

O5: Estudiar quines són les noves tendències en els camps de la privadesa, de l'obtenció del comportament dels turistes i de la provisió de nous serveis avançats que ofereixen les noves tecnologies. Això ha de servir per identificar les línies de recerca que es poden obrir en aquest àmbit.

El coneixement generat en aquest projecte ha de servir de punt de partida per elaborar una sol·licitud d'un projecte europeu que impliqui a grups internacionals però també a agents locals que puguin estar interessats en oferir aquests serveis personalitzats.

1. Estudi dels projectes de recerca i de les tecnologies que s'utilitzen

1.1. Estudi dels projectes de localització i seguiment

Els turistes quan viatgen solen tenir dificultats en trobar els llocs d'interés en situacions i entorns que no coneixen, com per exemple, localitzar un destí dins d'una ciutat, i també en organitzar-se el temps depenent de la ubicació dels llocs a visitar. La manera tradicional de proporcionar la informació al turista ha estat la utilització de documentació impresa i l'atenció personalitzada. L'inconvenient del mètode imprès és que no permet consultes interactives i l'inconvenient de l'atenció personalitzada és que exigeix comptar amb recursos humans qualificats suficients, amb els costos que això comporta.

Gràcies als serveis basats en la localització es pot proporcionar la informació de manera interactiva i personalitzada abaratint els costos i ampliant les funcionalitats dels serveis, aconseguint així un índex elevat de satisfacció dels usuaris.

L'estudi desenvolupat s'ha centrat en la cerca de projectes de recerca en localització i seguiment finançats per la Unió Europea (Programa Marc) que es poden trobar a la pàgina web del CORDIS [1]. El període d'estudi ha estat els 2 darrers Programes que abarquen projectes iniciats al 2000 fins als iniciats el 2006 (i que acaben al 2009). A partir de la cerca s'han agrupat segons la seva temàtica i s'han classificat en els grups següents:

- Turisme mòbil.
- Tecnologies de localització i/o seguiment.
- Serveis d'emergència.
- Coneixement geogràfic.

En aquest projecte ens centrarem amb el primer grup, el turisme mòbil. En aquest àmbit Els projectes de recerca trobats els podem veure a la Figura 1.1. A les següents seccions es descriuen breument aquests projectes. A més, per cada projecte s'identifica la tecnologia de posicionament i transmissió de dades utilitzada.

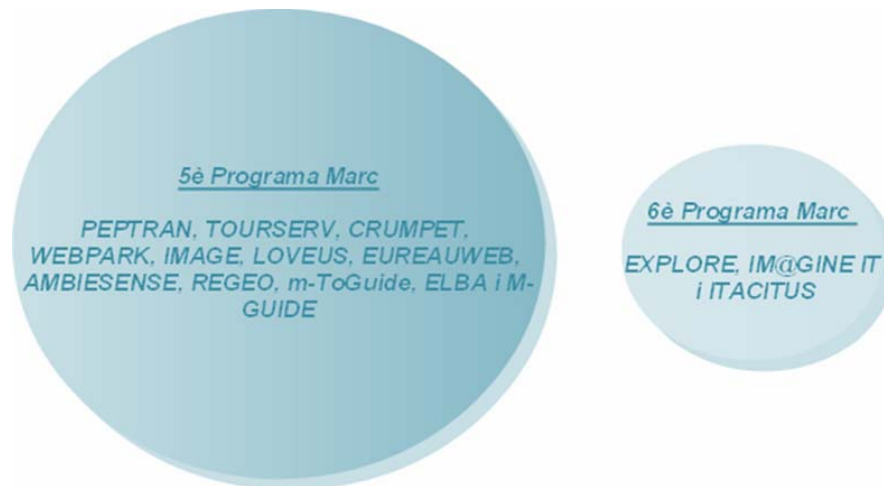


Figura 1.1. Projectes de turisme mòbil segons el Programa Marc

1.1.1 . Pedestrian and public transport navigator (PEPTRAN)

En una àrea urbana o suburbana, la insatisfacció dels habitants i dels usuaris del transport públic augmenta degut als problemes que suporten diàriament: la congestió del trànsit, la contaminació acústica i de l'aire, les esperes llargues, la manca de connexions, les infraestructures inadequades, la informació insuficient, etc.

L'objectiu de PEPTRAN [2] era desenvolupar una aplicació per guiar als usuaris d'un punt a un altre punt dins d'una ciutat (caminant o fent servir el transport públic) de la manera més eficient.

L'aplicació es va executar en dues plataformes: amb un dispositiu mòbil i amb un sistema de navegació per a automòbil amb GPS. El sistema utilitza mapes detallats dels carrers i la informació de progrés dels vehicles del transport públic (juntament amb els horaris) que poden ser descarregats del servidor central amb el sistema global per a les comunicacions mòbils (*Groupe Special Mobile*, GSM).

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GSM.

1.1.2. Personalised Tourist Services Using Geographic Information Systems via Internet (TOURSERV)

En aquest projecte es va desenvolupar un sistema que combina el turisme alpi amb informació geogràfica en 3D. Els turistes disposen d'un dispositiu amb accés a Internet i GPS (PDA o telèfons mòbils) on es poden descarregar els mapes necessaris per al funcionament del sistema. Un cop descarregats, els turistes poden saber la seva posició dins dels mapes mitjançant el GPS. Això

permet als turistes rebre contínuament la informació actualitzada (via GPRS i UMTS) sobre el seu entorn actual: rutes i pistes d'esquí de la regió, l'oferta de restaurants i hotels, les actuals previsions meteorològiques regionals, etc. GPRS és el servei general de paquets via ràdio (*General Packet Radio Service*) i UMTS és el sistema universal de telecomunicacions mòbils (*Universal Mobile Telecommunications System*, UMTS).

En aquest projecte es van adonar que les regles d'interpretació del sistema són limitades dins d'un domini d'aplicació concret ja que, per exemple, en una estació d'esquí, si una persona està a la zona de la cua d'un telecadira probablement pujarà per aquell telecadira. Per tant, el sistema podria decidir enviar la temperatura i la velocitat del vent dalt de la muntanya on acaba el recorregut el telecadira. En canvi, si canviem de domini i ens situem en un museu, no és tan trivial saber quina ha de ser la informació a enviar al turista. A [26] es descriu una arquitectura per fer front a aquest problema.

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS i UMTS.

1.1.3. Creation of user-friendly mobile services personalised for tourism (CRUMPET)

Quan es va realitzar aquest projecte els anys 2000-2002 ja es podia accedir des de qualsevol lloc a tota una gamma de serveis per a turistes nòmades. Gran part de la tecnologia ja estava disponible, per exemple, xarxes públiques, posicionament GPS, tecnologia d'agents, dispositius de mà (PDA o telèfons mòbils). Però encara hi havia problemes amb les tecnologies i la facilitat d'ús d'aquests serveis, que requerien de solucions intel·ligents que havien de sotmetre's a proves per comprovar la viabilitat i la qualitat, incloent la percepció de qualitat de l'usuari.

L'objectiu del projecte CRUMPET va ser la 'Creació de serveis personalitzats per a mòbils de fàcil ús per als turistes'. CRUMPET va tenir dos objectius principals: 1) Implementar i provar serveis de valor afegit relacionats amb el turisme per usuaris nòmades a través de xarxes mòbils i fixes. 2) Avaluar la tecnologia d'agents en termes d'acceptació de l'usuari i rendiment amb un enfocament adequat per a una ràpida creació de serveis per a usuaris nòmades.

CRUMPET va desenvolupar serveis per a turistes amb localització coneguda a partir d'un sistema multi-agent amb serveis de mediació i facilitació d'interacció. El dispositiu client és un ordinador de mà (PDA), la ubicació de l'usuari es realitza amb GPS o, alternativament, per triangulació GSM. Els PDA ofereixen una mida de pantalla i resolució suficients per mostrar mapes i pàgines HTML simples.

El sistema ofereix una interfície d'usuari senzilla per la manipulació dels serveis. Les principals característiques són les següents:

- Recomanació de serveis, per exemple, llocs d'interès turístic (sobre la base dels interessos personals de l'usuari i la ubicació actual).
- Mapes interactius (mapes de visió general de la zona, destacant la posició actual de

l'usuari, els llocs d'interès i les visites ja realitzades).

- Informació sobre llocs d'interès turístic (text curt, informació més detallada, fotografies, mapes i adreces).
- Consells proactius quan l'usuari s'apropa a un lloc que li pugui interessar.
- Itinerància en una xarxa de comunicació, és a dir, el sistema s'adapta d'una manera fàcil depenent de la qualitat de servei de les xarxes i el tipus de connexió sense fil que està disponible (WLAN o GSM).

Per fer front a la privadesa de l'usuari, s'utilitza un agent mediador entre el terminal mòbil de l'usuari i el proveïdor de serveis. Aquest agent mediador oculta la identitat de l'usuari al proveïdor de serveis de tal manera que el proveïdor de serveis no pot fer el seguiment de les activitats de l'usuari ni tampoc saber el perfil de l'usuari mitjançant la recopilació d'informació confidencial. L'agent mediador només revela la informació que és absolutament necessària per dur a terme la comunicació.

Es pot trobar més informació a la pàgina web del projecte [22] i al treball [27].

Tecnologia de posicionament: GPS i triangulació GSM.

Tecnologia de transmissió de les dades: WLAN i GSM.

1.1.4. GEographically relevant information for moBile users in Protected Areas (WEBPARK)

El turisme representa una part significativa de l'activitat econòmica de les comunitats rurals que gaudeixen d'atractius mediambientals, com la proximitat a parcs o entorns naturals. Els efectes d'aquest fenomen són importants econòmicament, socialment i en aspectes mediambientals.

L'ús intensiu dels parcs i de les zones naturals, planteja qüestions de sostenibilitat i crea nous problemes per a la gestió d'aquestes àrees. Un gran nombre de visitants, pot tenir efectes negatius, com el deteriorament de la qualitat mediambiental. Al mateix temps, els administradors de les àrees han de satisfer les necessitats de les grans quantitats de visitants.

En aquest projecte es va dissenyar un sistema d'informació que proporciona als visitants i als administradors d'aquestes zones, informació de l'entorn a través dels seus telèfons mòbils o PDA.

Cada usuari crea el seu perfil personalitzat. Quan l'usuari sol·licita la informació, el sistema retorna informació filtrada segons les seves preferències i la seva posició actual obtinguda via GPS. A més, els usuaris poden ser alertats quan s'acosten en algun lloc d'interès personal. La informació es pot comprar utilitzant micropagaments i/o subscripció. L'experiència de les administracions d'aquestes zones indica que els usuaris estan disposats a comprar la informació que els ajudi a prendre decisions sobre els seus plans diaris.

Aquest projecte tenia en compte la privadesa de la informació dels usuaris, permetent als usuaris declarar la seva pròpia política utilitzant el protocol P3P.

Es pot trobar més informació a la pàgina web del projecte [3].

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS i UMTS.

1.1.5. Intelligent Mobility Agent for Complex Geographic Environments (IMAGE)

IMAGE pretenia proporcionar als usuaris serveis personalitzats basats en la seva ubicació utilitzant una combinació d'agents intel·ligents amb capacitat d'auto-aprenentatge, la posició de l'usuari, la seva ruta i les seves preferències preestablertes. Un exemple seria la prestació de serveis de continguts (lístes de punts d'interès de restaurants, hotels, etc.) pròxims a la ubicació actual de l'usuari.

Aquest projecte tenia els següents objectius: 1) Dissenyar i desenvolupar una plataforma de serveis oberts i modular, per coordinar juntament les dades del proveïdor de serveis i les dades dels usuaris finals. 2) Desenvolupar serveis avançats (inclosos els de navegació, aplicacions de localització i serveis de comerç electrònic) i facilitar la integració entre ells. 3) Introduir un nou rol empresarial, l'Agent de Mobilitat, que fa d'intermediari entre els actors involucrats en la prestació de serveis. 4) Verificar la plataforma integrada i la inter-operació entre l'agent intel·ligent i dos bancs de proves (Turin, Itàlia i Tampere, Finlàndia). 5) Provar la viabilitat de la plataforma proposada a través de l'anàlisi financer i de mercat. 6) Facilitar la implementació del sistema IMAGE a través d'Europa, per elaborar un pla estratègic per a l'assoliment d'aquest, i proporcionar directrius per l'agent i els proveïdors de serveis.

Els serveis d'IMAGE estan disponibles a través d'Internet amb ordinador i en connexió sense fils a Internet (GPRS) en dispositius mòbils. En aquest últim cas, l'usuari final també pot utilitzar el posicionament automàtic a través d'un mòdul GPS. Aquests serveis són accessibles utilitzant tres dispositius, és a dir, un PDA, un mòdul GPS i un telèfon mòbil amb connexió GPRS.

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS.

1.1.6. Location Aware Visually Enhanced Ubiquitous Services (LOVEUS)

La convergència de l'evolució de les tecnologies mòbils (2.5G i 3G) i Internet ha permès l'aparició de serveis nous. L'objectiu del projecte LOVEUS era aprofitar aquestes tecnologies per proporcionar serveis personalitzats als ciutadans Europeus basats en la seva ubicació dins de llocs culturals o en entorns urbans.

Es va desenvolupar un terminal mòbil de fàcil ús que integrava la tecnologia i les característiques d'un ordinador portàtil, les d'un telèfon mòbil i les d'un assistent personal digital (PDA) en un de sol, amb un receptor GPS incorporat. Proporcionant a través d'aquesta infraestructura bàsica un nou paradigma per la promoció del turisme, el patrimoni cultural i dels serveis comercials.

A partir de la ubicació de l'usuari proporcionada pel receptor GPS, l'usuari podia buscar mitjançant el sistema, els punts d'interès pròxims a la seva ubicació que desitja visitar. D'altra banda, utilitzant la funcionalitat de visita guiada, l'usuari era guiat d'un punt d'interès a un altre acordant un ordre predefinit.

El funcionament intern es basa en la construcció de mapes personalitzats. Quan el component de navegació detecta que es necessita una certa part d'un mapa, el sistema sol·licita al servidor la construcció d'aquest mapa d'acord amb el perfil de l'usuari, la seva posició actual, la seva direcció i el seu context d'aplicació. La informació (incloent els mapes) s'envia mitjançant una connexió GPRS o UMTS. Es pot trobar més informació a la pàgina web del projecte [19].

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS i UMTS.

1.1.7. EUROPEAN Waterways networked information system (EUREAUWEB)

Aquest projecte està centrat en els usuaris (en llanxes velers, canoes i altres vehicles d'aigua) de vies navegables interiors europees. Tots ells necessiten informació el dia abans d'un viatge (per exemple, per a la planificació del seu itinerari), durant un viatge (per exemple, informació sobre llocs d'interès o de la previsió del temps), o després d'un viatge (quan els usuaris vulguin compartir la seva experiència de navegació). La manera habitual que tenien els usuaris d'informar-se era mitjançant material imprès, com ara mapes, guies de rius, guies turístiques, etc. Encara que alguns d'ells ja utilitzaven Internet per a obtenir informació, suposava un gran esforç seleccionar la informació de diverses fonts, ja que no hi havia un sistema d'informació integrat adaptat pels usuaris de les vies navegables.

En aquest projecte es volia proporcionar als usuaris informació personalitzada sobre oci i turisme de les vies navegables d'Europa segons la seva ubicació. D'aquesta manera es podria ajudar als turistes a planificar millor el seus viatges.

A través de xarxes de comunicació mòbil o sense fil, com GPRS i UMTS, es facilita la connexió a Internet. Aquestes tecnologies són efectives i estan disponibles sense importar la ubicació on ens trobem. Per saber la ubicació de l'usuari s'utilitza un receptor GPS que podria estar incorporat al dispositiu.

Els usuaris mitjançant un portal web o el seu telèfon mòbil poden crear i modificar el seu perfil d'usuari. Els perfils d'usuari ofereixen un mitjà eficaç per al filtratge d'informació en funció dels seus interessos personals. Els usuaris poden sol·licitar qualsevol tipus d'informació pròxima a la seva ubicació actual.

Es pot trobar més informació a la pàgina web del projecte [11] i al treball [17].

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS i UMTS.

1.1.8. Ambient, personalised, and context-sensitive information systems for mobile users (AMBIENSENSE)

Les necessitats d'informació de les persones canvien a mesura que es troben noves situacions. Satisfer aquestes necessitats d'informació en un context personalitzat i sensible és un dels reptes de la informàtica i de la intel·ligència ambiental (és a dir, centrada en l'ambient de l'usuari).

Al projecte AMBIENSENSE es va desenvolupar un sistema flexible on es podia accedir a la informació de dues maneres diferents. La primera, a través d'etiquetes RFID col·locades a diferents punts d'una ciutat, proporcionava als usuaris mòbils (dispositius mòbils) informació rellevant segons la seva ubicació (per exemple, una etiqueta col·locada a una estàtua proporcionava informació rellevant de la estàtua). També es podia accedir a la informació a través d'un proveïdor de serveis de continguts (*Content Service Provider*, CSP) on la informació era proporcionada directament a l'usuari (normalment amb un cost per l'usuari mòbil).

Es pot trobar més informació a la pàgina web del projecte [9] i al treball [15] que descriu sistemes d'informació sensibles al context de l'usuari i la seva personalització.

Tecnologies: Bluetooth, WLAN, etiquetes RFID.

1.1.9. Multimedia Geoinformation in Rural Areas with Eco-Tourism (REGEO)

Les regions rurals que participaven al projecte REGEO volien atraure nous visitants i competir amb les principals empreses de la indústria del turisme.

REGEO va desenvolupar un sistema de gestió de continguts geo-multimèdia per a les regions rurals. Enfocat especialment al turisme ecològic, inclosos els aspectes dels Parcs Naturals i Nacionals (per exemple, planificar rutes de senderisme en 3D).

El sistema proporcionava funcionalitats simples de Sistemes d'informació Geogràfica (SIG) per fer ús dels continguts de dades espacials. A més, utilitzava noves tecnologies per a la presentació i avaluació de les dades i un gestor de continguts geo-multimèdia per a diferents grups d'usuaris: administració, empreses públiques i privades o institucions, així com els clients individuals com a visitants. L'accés a la informació podia ser en diferents tecnologies (Internet, GPRS/UMTS) o bé amb sistemes fora de línia (utilitzant CD-ROM). La localització de l'usuari era proporcionada per un receptor GPS.

Els principals objectius del projecte eren: 1) desenvolupar una base de dades de continguts geo-multimèdia descentralitzada. 2) aplicar els sistemes d'informació turístics existents i millorar els

mitjans de presentació d'ambdós dispositius, en línia i fora de línia. 3) el desenvolupament d'aplicacions mòbils per a realitzar un servei basat en la localització. 4) obrir la base de dades per ús local o d'administracions de parcs naturals i 5) provar el sistema en quatre regions Europees.

Es pot trobar més informació a la pàgina web del projecte [14] i al treball [5] que descriu el disseny de l'arquitectura del sistema.

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS i UMTS.

1.1.10. Mobile Tourism Guide (m-ToGuide)

El projecte m-ToGuide estava dirigit al mercat turístic europeu i s'oferia als turistes una àmplia gamma d'informació i serveis. Els turistes demanaven informació actual i precisa, relacionada en multituds de temes, per exemple, allotjaments, restaurants, espectacles, rutes, horaris, etc. Aquest projecte abarcava clarament el gran potencial d'ús per part dels turistes de serveis relacionats amb la seva localització.

La seva utilització permet als usuaris obtenir accés immediat a la informació o serveis que estan buscant, per exemple, atraccions, compres, llocs d'interès, adreces. El principal objectiu era promoure la utilització de les xarxes cel·lulars 2.5-3G amb serveis basats en la localització. Altres objectius eren: 1) Mantenir actualitzada la informació. 2) Serveis basats en la localització relacionats amb el turisme. 3) Format multimèdia d'alta qualitat. 4) Planificació de ruta. 5) Serveis de reserva i venda de bitllets. 6) Promocions, ofertes especials i descomptes.

El Kit del terminal de l'usuari proporciona a l'usuari informació rellevant i rep també les seves aportacions. El terminal utilitza GPS per tal de trobar la localització de l'usuari i un mòdem GPRS per poder comunicar-se amb el sistema. La comunicació entre el terminal i el sistema es realitza a través de l'assistent intel·ligent (*Smart Wizard*). L'assistent intel·ligent és el cor del sistema m-ToGuide perquè és el responsable de proporcionar les respostes a cada sol·licitud del terminal, per exemple mapes, informació, rutes, gires, etc. i la comunicació entre els diferents subsistemes.

En definitiva, el projecte m-ToGuide era una guia d'informació turística mòbil que proporcionava accés a la informació per als turistes sobre serveis mòbils.

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS.

1.1.11. 3G European Location Based Advertising (ELBA)

El projecte ELBA (Publicitat basada en la localització europea) tenia per objectiu desenvolupar i validar un sistema de publicitat basada en la localització. La visió del projecte ELBA era demostrar les possibilitats i potencialitats de les aplicacions i serveis mòbils 2.5-3G (GPRS i UMTS respectivament).

Els usuaris mitjançant un dispositiu mòbil (PDA o telèfon mòbil) poden sol·licitar manualment al proveïdor de serveis, els anuncis o la informació que ells estan interessats, per exemple, mentre un usuari està en una ciutat determinada sol·licita al proveïdor de serveis quina és la botiga de roba més propera a la seva ubicació. Seguidament, el proveïdor de serveis (via GPRS o UMTS) respon a l'usuari amb el resultat de la seva sol·licitud. També hi ha l'opció que l'usuari es registri a un tipus d'informació determinada i l'usuari rebi la informació automàticament, per exemple, els descomptes d'una marca de roba quan l'usuari passeja a prop d'una de les seves botigues. Es pot trobar més informació a la pàgina web del projecte [10].

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS i UMTS.

1.1.12. Cultural location based information services (M-GUIDE)

Al projecte M-GUIDE van desenvolupar i desplegar un sistema integrat per a la prestació de serveis d'informació multilingüe basats en la localització per a visitants (turistes) d'Atenes (Grècia i Turquia) i Finlàndia.

El sistema oferia principalment informació turística (informació cultural) sobre llocs d'interès del centre de la ciutat, com per exemple, teatres i centres d'exposicions i informació dels transports públics. Hi ha molt poca informació d'aquest projecte a la seva pàgina web [13].

1.1.13. Gaming and guiding system for museum and exhibition environments (EXPLORE)

L'objectiu d'aquest projecte era desenvolupar una sistema de guia virtual i entreteniment per museus i entorns d'exposició.

Els passos a seguir pel funcionament del sistema eren: la instal·lació d'una xarxa de punts d'accés Bluetooth al museu, la instal·lació d'una aplicació als telèfons mòbils dels visitants i la descarrega dinàmica dels continguts durant la visita al museu. La localització (posicionament) i la comunicació és a través de Bluetooth.

Tecnologia de posicionament: Bluetooth.

Tecnologia de transmissió de les dades: Bluetooth.

1.1.14. Intelligent mobility agents, advanced positioning and mapping technologies integration interoperable multimodal, location based services (IM@GINE IT)

Aquest projecte estava centrat en la falta serveis d'informació basada en la mobilitat personalitzats i intel·ligents per cobrir tot el viatge sencer. Fins al moment, els serveis existents eren solament a nivell local, eren incomplets i oferien diferents interfícies d'usuaris i plataformes de prestacions de serveis.

Aquest projecte tenia com a objectiu desenvolupar un punt d'accés únic a través del qual l'usuari pot obtenir informació basat en la localització, com per exemple, informació de transport multimodal dinàmic o estàtic (és a dir, combinant el cotxe, autobús o a peu), la cartografia i l'enrutament, la navegació i altres serveis relacionats, tenint en compte les preferències personals de l'usuari. Es pot trobar més informació a la pàgina web del projecte [12].

Tecnologia de posicionament: GPS i Cell ID.

Tecnologia de transmissió de les dades: GSM i GPRS.

1.1.15. Intelligent tourism and cultural information through ubiquitous services (ITACITUS)

Aquest projecte vol oferir serveis a mida per a cada viatger, en l'entorn del turisme cultural. El sistema es basa en una interfície d'usuari avançada en dispositius mòbils que suporta el reconeixement d'imatges i veu i té la capacitat d'utilitzar qualsevol sistema de comunicació: Wi-Fi, Bluetooth, UMTS / GPRS, i IrDA. Mitjançant aquest sistema, l'usuari és capaç de localitzar i trobar la informació que pugui ser útil en el seu actual context cultural.

Des del punt de vista turístic, el sistema ITACITUS consisteix en dues aplicacions integrades:

- Una guia de la ciutat i una plataforma basada en web que consta de:
 - El gestor d'itineraris (*Itinerary Manager*, IM), per planificar un itinerari i després utilitzar-lo dinàmicament.
 - El mòdul d'informació del patrimoni cultural (*Cultural Heritage Information Module*, CHIMO), per entregar serveis i informació juntament al IM durant la planificació del itinerari i la visita.
- Un visualitzador que proporciona una sèrie de continguts avançats d'àudio i vídeo per millorar la interfície amb el turista.

El gestor d'itinerari ajuda als usuaris a planificar la visita, suggereix llocs i esdeveniments segons el seus perfils, així com els horaris (horari d'obertura i de tancament, durada mitja de la visita, etc.) i espacial (distància entre dos llocs, els transports disponibles, etc.). Durant la visita l'objectiu del sistema es fer el seguiments dels usuaris per tal de verificar que la visita es realitza amb el temps inicialment previst, i ajustar-la en cas de retard o anticipació. Per exemple, l'usuari pot sol·licitar explícitament modificar el itinerari o el sistema podria notificar a l'usuari un retard durant la visita a causa de la meteorologia o canvis de transport.

Es pot trobar més informació a [23]. Aquest treball cobreix la recerca i el desenvolupament del sistema multiagent que ofereix la informació basada amb la localització sobre dispositius mòbils, amb perfils d'usuaris intel·ligents.

Tecnologia de posicionament: GPS.

Tecnologia de transmissió de les dades: GPRS/UMTS, WIFI, IrDA i Bluetooth.

1.2. Estudi de les tecnologies que s'estan utilitzant per localitzar i fer el seguiment de persones

Avui en dia, existeixen moltes tecnologies per localitzar i/o fer el seguiment de persones. En aquesta secció s'expliquen totes les tecnologies disponibles per a poder-ho dur a terme.

1.2.1. Mètodes bàsics de posicionament

Abans d'explicar aquestes tecnologies farem una petita introducció dels mètodes de posicionament bàsics per a obtenir la posició espacial de l'objectiu, que són:

- Sensors de proximitat
- Lateració
- Angulació
- Coincidència de patrons
- *Dead Reckoning*
- Propostes híbrides

1.2.1.1 Sensors de proximitat

Aquest mètode [20] és el més estès per obtenir la posició d'un terminal mòbil i també la manera més fàcil. Està basat en un rang de cobertura limitat de senyals de ràdio, d'infrarojos o d'ultrasons. La posició del terminal mòbil s'obté a partir de les coordenades de l'estació base, ja sigui rebent les senyals pilot (les senyals utilitzades per mesurar posicions) d'un terminal per l'enllaç de pujada (*uplink*) o que les senyals pilot siguin rebudes pel terminal per l'enllaç de baixada (*downlink*).

La posició es pot aconseguir detectant el contacte físic o per identificació de cel·la (*Cell-ID*). En el segon cas s'ha de tenir en compte la precisió que dependrà de la cobertura de la senyal a una zona determinada.

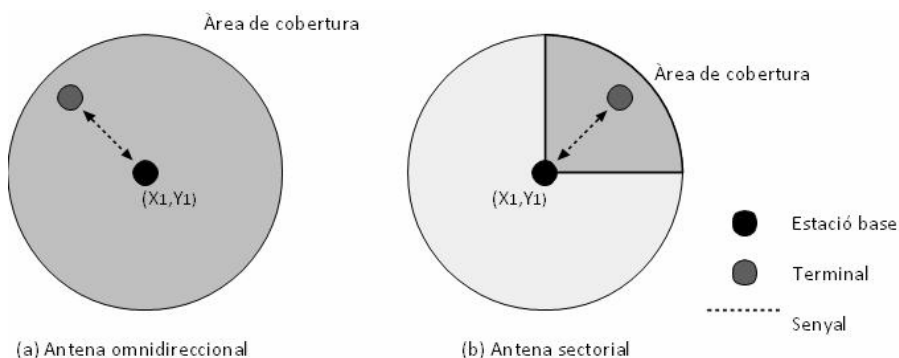


Figura 1.2. Sensors de proximitat

Les antenes que s'utilitzen a les estacions base poden ser omnidireccionals o sectorials. Les antenes omnidireccionals capten la mateixa senyal en qualsevol direcció, és a dir, als 360°, i les antenes sectorials capten la senyal en un rang de direccions limitat (entre 90° i 180°). Les antenes sectorials tenen més intensitat (abast) de senyal que les antenes omnidireccionals però per contra cobreixen una zona més restringida. Combinant varies antenes sectorials es pot donar cobertura a tot el pla horitzontal, amb més intensitat que una antena omnidireccional. L'elecció del tipus d'antena dependrà de l'escenari concret d'aplicació. Aquest mètode és aplicable a tots els sistemes de comunicacions cel·lulars. A la Figura 1.2 (a) es mostra una antena omnidireccional i a la Figura 1.2 (b) una antena sectorial.

1.2.1.2 Lateració

Amb la tècnica de lateració [20] la posició dels terminals mòbils es determina utilitzant les diferències de distàncies a tres o més llocs fixos de referència (estacions base). Aquestes diferències de distàncies es poden calcular de dues maneres diferents:

1. Mesurant el temps d'arribada de les senyals (*Time Of Arrival*, TOA).
2. Mesurant les diferències de temps d'arribada de les senyals (*Time Difference Of Arrival*, TDOA).

En el primer cas, la informació de posicionament s'obté del temps absolut que tarda una ona en arribar entre el transmissor i el receptor o viceversa. Això implica que el receptor sap l'hora exacta de transmissió. Com alternativa a aquesta proposta, es podria incloure la mesura del temps d'anada i tornada de la senyal, i així d'aquesta manera ja no ens fa falta la sincronització entre el transmissor i el receptor. Al segon cas, el problema de tenir els rellotges sincronitzats al transmissor i al receptor es resol mitjançant l'ús de diversos transmissors sincronitzats a una base de temps comuna, i mesurant la diferència de temps d'arribada al receptor.

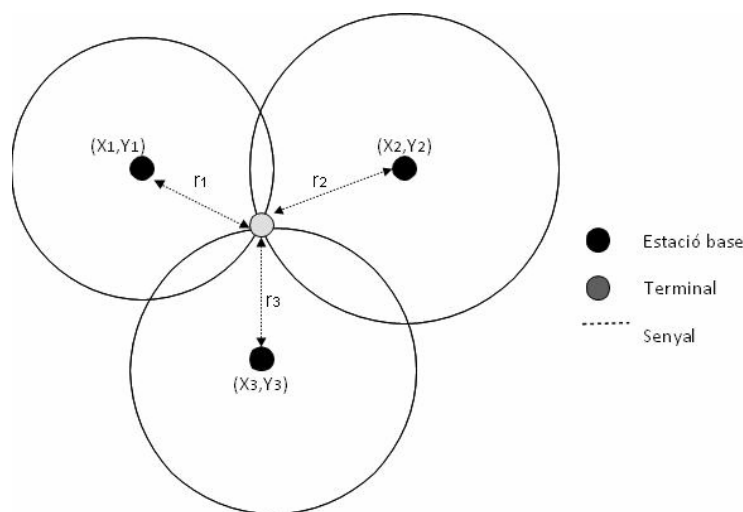


Figura 1.3. Lateració Circular

Si en el posicionament s'utilitza el temps d'arribada de les senyal (TOA), la posició es calcula per lateracions circulars. Mentre que si s'utilitza les diferències de temps d'arribada de les senyals (TDOA), la posició s'obté per lateracions hiperbòliques. A la Figura 1.3 podem veure un exemple de lateració circular en dos dimensions. A la lateració circular sabent el número d'estacions base i les respectives distàncies al terminal podem calcular la posició del terminal amb la intersecció dels cercles formats per les estacions base, on el radi de cada cercle és la distància de cada estació base respecte el terminal.

1.2.1.3 Angulació

L'angulació [20] és un altre mètode per calcular la posició d'un objectiu coneixent les coordenades de diverses estacions base. En contrast amb la lateració, el que s'observa són els angles entre l'objectiu i una sèrie d'estacions base. L'angulació també s'anomena Angle d'arribada (*Angle Of Arrival*, AOA) o direcció d'arribada (*Direction Of Arrival*, DOA).

Per poder calcular la posició es requereix la distància entre els dos punts de referència o les seves coordenades. A més a més, per obtenir aquests angles es requereix que totes les estacions base o terminals estiguin equipades amb múltiples antenes. A la Figura 1.4 es mostra el principi bàsic de l'angulació. Des del punt de vista teòric són suficients dues estacions base amb dos angles i dos distàncies per a calcular posicions en 2 dimensions.

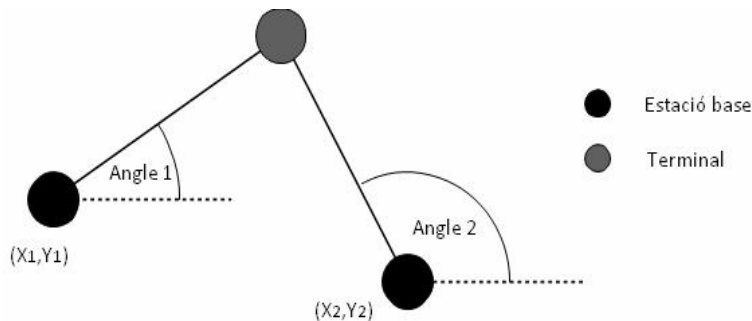


Figura 1.4. Angulació

1.2.1.4 Coincidència de patrons

El principi fonamental d'aquest mètode [20] és observar el lloc (o escena) on es vol aplicar el posicionament i extreure conclusions sobre la posició de l'objectiu a partir d'aquestes observacions. Aquests mètode es pot dividir en dos grans grups: el mètode de coincidència de patrons òptics i el mètode de coincidència de patrons no òptics.

En el mètode de coincidència de patrons òptics, també conegut com anàlisi de l'escena, hi ha una càmera que genera imatges visuals de l'escena i són comparades entre elles. Al mètode de coincidència de patrons no òptics es tenen en compte altres factors físics. Un dels mètodes més freqüents, detecta la posició de l'objectiu amb les característiques de la propagació de les senyals de ràdio. Aquest mètode també es coneix com *fingerprinting* (empremtes digitals), i s'executa en dues fases, la fase *off-line* i la fase *on-line*. A la fase *off-line*, el lloc d'interès està cobert per una quadricula, i per cada punt de la quadricula l'observador recull la força de la senyal rebuda (*Received Signal Strength*, RSS) de múltiples estacions base, el resultat és un vector de valors RSS per cada punt. Aquest vector s'anomena *fingerprint*. A la fase *on-line*, el terminal objectiu compon un vector de mostra de valors RSS rebuts a la seva posició actual i els envia a un servidor. El servidor després intenta fer coincidir la mostra amb els *fingerprints* generats a la fase *off-line* i fa una estimació de la posició de l'objectiu. L'algorisme més comú és el que calcula la distància Euclidiana entre la mostra RSS i cada *fingerprint*.

A la majoria de casos, el reconeixement de patrons es realitza amb l'ajuda del terminal, on el terminal realitza mesuraments dels resultats i les transferències a la xarxa per al càlcul de la posició. L'avantatge és que en alguns sistemes els terminals no requereixen cap modificació o solament petites modificacions.

1.2.1.5 Dead Reckoning

Aquest mètode de posicionament [20] és un dels primers que es va aplicar. El seu origen va ser a la marina, també és conegut com el mètode que va utilitzar Cristòfor Colom en els seus viatges per descobrir el Nou Món. Malgrat la seva llarga història, i degut al desenvolupament de la tecnologia de sensors avançats, el *Dead reckoning* ha adquirit una gran importància en els darrers anys i avui en dia s'utilitza per la navegació d'avions, vaixells i automòbils. La posició s'estima a partir de posicions anteriors, la velocitat i l'acceleració, la direcció del moviment i el temps. La posició inicial s'obté a partir d'un altre mètode de posicionament, com per exemple el GPS. La velocitat, l'acceleració i la direcció del moviment pot ser obtinguda de dues maneres diferents, ja sigui per deducció de dos o més posicions fixes que s'han mesurat anteriorment, o per sensors addicionals equipats a l'objectiu. A la Figura 1.5 es mostra com s'obté la posició de l'objectiu (x_1, y_1) per aproximació amb el mètode *Dead Reckoning*.

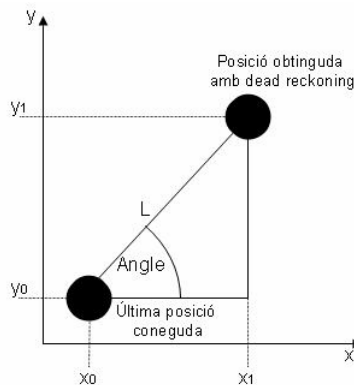


Figura 1.5. Dead Reckoning

1.2.1.6 Propostes Híbrides

És possible aplicar qualsevol combinació dels mètodes de posicionament presentats anteriorment. La principal motivació per portar-ho a terme és proporcionar una major precisió, per exemple, per aconseguir la precisió necessària per el compliment d'estàndards definits. Altres motivacions serien per exemple, augmentar la velocitat per fixar la localització i reduir el consum d'energia del terminal.

1.2.2 Classificació de les tecnologies de posicionament

El procés de localització s'ha de portar a terme independentment de si el terminal mòbil es troba a l'exterior o a l'interior d'un edifici, tal i com es mostra a la Figura 1.6. Les tècniques que s'utilitzen són diferents i depenen en gran mesura de la precisió amb que s'hagi de localitzar l'usuari. Les tecnologies de posicionament en espais oberts es classifiquen en dos. Les tecnologies de posicionament per satèl·lit i les tecnologies de posicionament en xarxes cel·lulars. A continuació es comentaran les tecnologies de localització amb aquesta classificació.

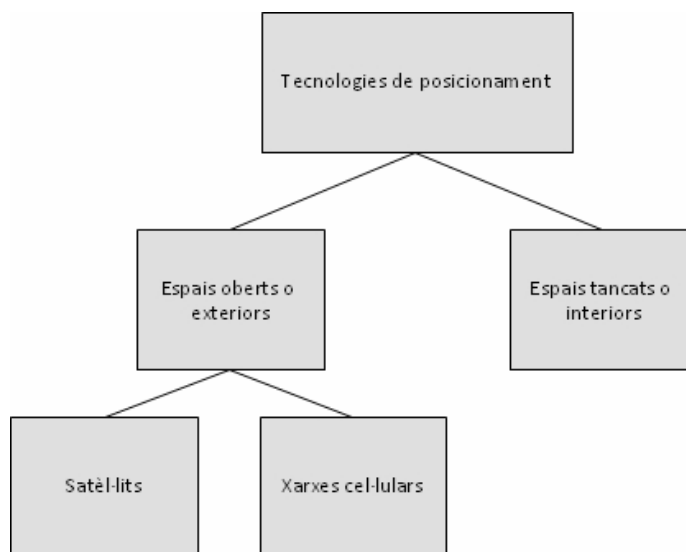


Figura 1.6. Classificació de les tecnologies de posicionament

1.2.3. Tecnologies de posicionament per satèl·lit

En aquesta secció es descriuen les tecnologies de posicionament per satèl·lit. Concretament, el sistema de posicionament global (*Global Positioning System*, GPS), el GPS diferencial, el GLONASS i el GALILEO.

1.2.3.1. Sistema de posicionament global

El Sistema de posicionament global (*Global Positioning System*, GPS) [7] és un sistema de navegació per satèl·lit (*Global Navigation Satellite Systems*, GNSS) que permet determinar en tot el món la posició d'un objecte, una persona, un vehicle, etc. Es va desenvolupar pel departament de defensa dels Estats Units al final del període de la guerra freda amb fins militars.

El sistema GPS està dividit en tres parts: l'usuari, l'espai i el control.

- La part de l'usuari són els dispositius dels usuaris mòbils (per exemple, els receptors GPS). Els receptors GPS poden estar integrats o no al dispositiu (per exemple, un telèfon mòbil o PDA). Aquests receptors cada vegada són més barats i més petits.
- La part de l'espai és una xarxa o constel·lació, coneguda amb el nom de NAVSTAR (*Navigation Satellite Timing and Ranging*) [6], de 24 satèl·lits movent-se en sis òrbites diferents amb quatre satèl·lits per òrbita (vegeu la Figura 1.7). Aquests satèl·lits es troben a una distància aproximadament de 20.200 km de la superfície de la terra. Cada satèl·lit necessita unes dotze hores per a completar la seva òrbita. Els satèl·lits es mouen d'una manera que almenys cinc satèl·lits són visibles a l'horitzó de cada punt de la superfície de la terra i onze com a màxim. Però aquest número es pot veure reduït per culpa dels obstacles, com edificis o muntanyes.
- La part de control es necessari per administrar i corregir les dades internes dels satèl·lits (per exemple, el temps del sistema i l'òrbita de cada satèl·lit). I engloba les infraestructures terrestres necessàries per al control de la constel·lació de satèl·lits.

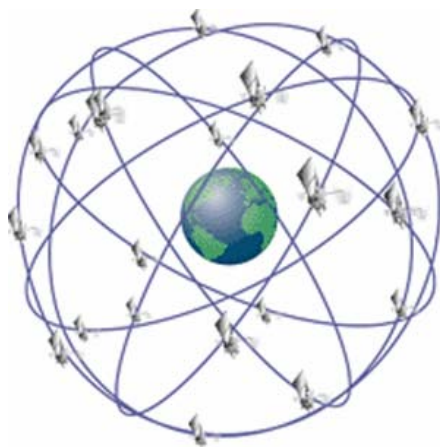


Figura 1.7. Òrbita dels satèl·lits del GPS

El posicionament amb GPS consta de tres passos: la identificació dels satèl·lits, el mesurament de les distàncies i el càlcul de la posició. La situació dels satèl·lits es coneguda pel receptor amb base en les efemèrides (5 paràmetres orbitals Keplerians), els paràmetres que són transmesos pels propis satèl·lits. La col·lecció d'efemèrides de tota la constel·lació es completa cada 12 minuts i es guarda

al receptor GPS. A continuació el receptor GPS calcula la seva distància als satèl·lits, que es mesura utilitzant lateracions circulars basats amb TOA.

Un cop el receptor GPS sap les distàncies a cada satèl·lit, calcula la seva posició per triangulació. La precisió depèn del número de satèl·lits visibles, amb vuit satèl·lits a la vista pot ser de sis a quinze metres. Encara que l'interval mitjà de precisió que ofereix està entre 5 i 20 metres.

1.2.3.2. GPS diferencial

Aquest mètode és una millora del GPS estàndard que augmenta la seva exactitud. El GPS diferencial (*Differential GPS*, D-GPS) [4], és un sistema que proporciona als receptors de GPS correccions de les dades rebudes dels satèl·lits GPS, amb la finalitat de proporcionar una major precisió de la posició calculada. El fonament radica en el fet que els errors produïts pel sistema GPS afecten per igual als receptors situats pròxims entre sí. Els errors estan fortament correlacionats amb els receptors pròxims. Un receptor GPS fix que coneix exactament la seva posició basant-se en altres tècniques, rep la posició donada pel sistema GPS, i pot calcular els errors produïts pel sistema GPS, comparant-la amb la seva posició (coneguda amb antelació). Aquest receptor transmet la correcció d'errors als receptors pròxims a ell, i així aquests poden, a la vegada, corregir també els errors produïts pel sistema dins de l'àrea de cobertura de transmissió de senyals de l'equip GPS de referència. Per a que les correccions D-GPS siguin vàlides, el receptor té que estar generalment a menys de 1000 km d'alguna estació D-GPS. La precisió aconseguida utilitzant aquest mètode és d'un a tres metres.

1.2.3.3. GLONASS

Aquest sistema global de navegació per satèl·lit es va desenvolupar a Rússia i representa la contrapartida al GPS nord-americà i al futur Galileo europeu.

Consta d'una constel·lació de 24 satèl·lits (21 actius i 3 de recanvi) situats en tres plans orbitals amb vuit satèl·lits. La constel·lació de GLONASS [30] es mou a una òrbita al voltant de la terra amb una altitud de 19.100 km (una mica més baix que el GPS) i tarda aproximadament 11 hores i 15 minuts en completar una òrbita.

La localització GLONASS és equivalent a la del GPS. La majoria dels receptors GLONASS són també equipats amb receptors GPS, això significa que la disponibilitat i la precisió de la localització GLONASS millora respecte els receptors que solament utilitzen GPS. El fet d'utilitzar els dos receptors (GLONASS i GPS) en la mateixa localització garanteix una resposta més ràpida. Això és realment molt apreciat en totes les aplicacions professionals on l'alta precisió de les posicions són requerides.

Tant el GPS i el GLONASS utilitzen dues bandes de freqüència anomenades L1 i L2. La principal diferència entre les dues tecnologies de posicionament és que el GPS utilitza un canal de 20MHz a cada freqüència, mentre que GLONASS utilitza 25 canals de 562.5KHz a cada freqüència.

1.2.3.4. Galileo

Galileo [29] és un sistema global de navegació per satèl·lits desenvolupat per la Unió Europea, amb l'objectiu d'evitar la dependència dels sistemes GPS i GLONASS, i al contrari que aquests dos, serà únicament d'ús civil. Estarà operatiu al 2010.

Les característiques bàsiques d'aquest sistema són la cobertura a tot el món, cinc serveis de posicionament per als diferents grups d'usuari, una major exactitud en comparació amb el GPS i la interoperabilitat amb GPS i GLONASS. El cinc serveis que oferirà són:

1. Servei Obert (*Open Service, OS*). Estarà dirigit a aplicacions per al públic en general. Proporciona informació de posicionament de forma gratuïta. Els clients tenen els mateixos propòsits que els usuaris del GPS, s'estima que la majoria dels receptors utilitzin senyals conjuntes de Galileo i GPS.
2. Serveis per aplicacions crítiques (*Safety Of Life service, SOL*). Estarà adreçat a persones involucrades en activitats de seguretat. S'utilitza per a la majoria d'aplicacions de transport a on la vida humana podria posar-se en perill si la prestació dels serveis del sistema de radionavegació es veiés degradada sense notificació en temps real. Aquest servei proporciona la mateixa precisió que el servei obert. La diferència principal és l'alt nivell d'integritat de cobertura mundial per les aplicacions a on la seguretat es crítica, com per exemple la navegació aèria i les aplicacions ferroviàries a on la precisió garantida és essencial.
3. Servei comercial (*Commercial Service, CS*). Estarà orientat a aplicacions de mercat que requereixen un nivell superior de prestacions que les que ofereix el servei obert. A canvi d'aquest serveis afegits és pagarà un cànon.
4. Servei públic regulat (*Public Regulated Service, PRS*). Està adreçat a aplicacions de govern: exèrcit, policia, etc. Aquest és un servei d'alta disponibilitat. Una característica que destaca aquest tipus de serveis és la robustesa de la senyal, la qual la protegeix contra els efectes de les interferències i dels intents d'emissió d'una senyal modificada de manera intencionada.
5. Servei de cerca i salvament (*Search and Rescue Service, SAR*). Aquest servei millorarà la qualitat dels serveis disponible als sistemes de cerca i salvament. Per exemple, la recepció quasi en temps real de missatges de socors transmesos des de qualsevol punt de la Terra, la localització precisa d'alertes i la detecció per múltiples satèl·lits per evitar el bloqueig en condicions de poca visibilitat dels satèl·lits. En definitiva, aquests serveis garanteixen major rapidesa de resposta i una major precisió.

1.2.4. Tecnologies de posicionament basades en xarxes cel·lulars

La instal·lació dels sistemes de posicionament és sovint una important inversió. Per reduir els costos en els serveis de posicionament s'utilitzen xarxes sense fils. Les xarxes cel·lulars són adequades per aquest propòsit perquè la identificació de cel·la ja transporta la localització aproximada. A més a

més, els mecanismes com el mesurament del temps d'arribada de les senyals (*Time Of Arrival*, TOA) o el mesurament dels angles, anomenat Angle d'arribada (*Angle Of Arrival*, AOA) permeten una delimitació més exacta de la posició.

Els dos sistemes de telecomunicacions més utilitzats amb telèfons mòbils són els protocols GSM i UMTS. GSM és el sistema global per a les comunicacions mòbils (*Groupe Spécial Mobile*, GSM), és un sistema estàndard per les comunicacions mitjançant telèfons mòbils. UMTS és la successora de GSM i una de les tecnologies utilitzades pels mòbils de tercera generació (3G).

Els protocols GSM i UMTS estan basats en arquitectura de xarxa anomenada cel·lular. En una arquitectura cel·lular es cobreix una zona geogràfica amb antenes o estacions base de transmissió (*base transceiver stations*, BTS) que emeten una senyal per a ser rebuda pels telèfons mòbils. Cada antena cobreix una àrea anomenada cel·la.

A les zones urbanes, les cel·les estan a prop les unes de les altres i cobreixen una àrea petita, el diàmetre està entre els 100 metres i un kilòmetre. A les zones rurals, les cel·les cobreixen una àrea bastant més gran amb un diàmetre que pot arribar fins al 60 kilòmetres.

A continuació veurem els diferents mètodes de posicionament utilitzats als protocols GSM i UMTS.

1.2.4.1. GSM

S'han desenvolupat diferents mètodes que difereixen amb el grau de precisió de la posició de l'objectiu. Els mètodes que s'utilitzen per la localització GSM són els següents:

- ID de cel·la combinat amb l'avanç temporal (*Cell ID Combined with timing advance*).
- Diferència de temps observat perfeccionat (*Enhanced Observed Time Difference*, E-OTD).
- Diferència en el temps d'arribada (*Uplink Time Difference Of Arrival*, U-TDOA).
- GPS assistit (*Assisted GPS*, A-GPS).

1.2.4.1.1. ID de cel·la combinat amb l'avanç temporal (*Cell ID Combined with timing advance*)

El ID de cel·la està basat en sensors de proximitat. Per millorar la precisió de la posició del terminal, *Cell ID* pot combinar-se amb l'avanç temporal (*Timing Advance*, TA) que mesura el temps de propagació del dispositiu i tres estacions base. A la Figura 1.8 podem veure el guany de precisió del mètode ID de cel·la amb TA. Aquest mètode és una de les tècniques d'identitat cel·lular perfeccionada (*Enhanced Cell-ID*). La precisió d'aquest mètode és d'uns 500 metres.

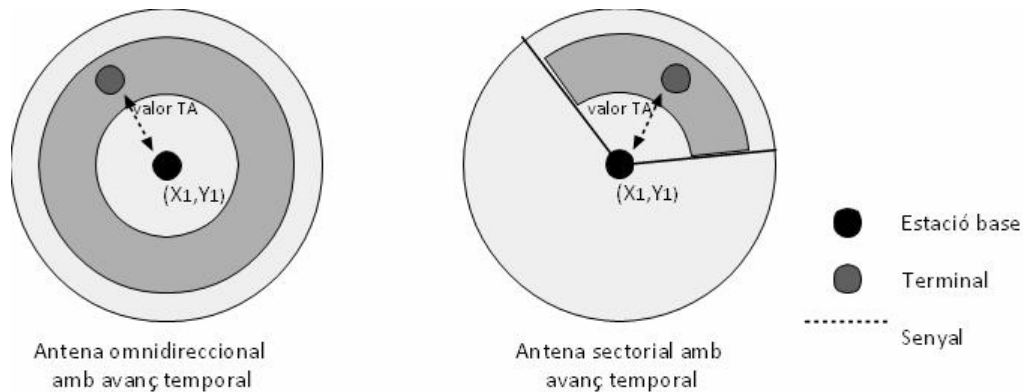


Figura 1.8. ID de cel·la + TA

1.2.4.1.2. Diferència de temps observat perfeccionat (*Enhanced Observed Time Difference, E-OTD*)

Aquest mètode està basat en lateracions hiperbòliques i s'aplica a l'enllaç de baixada (*downlink*), és a dir, el terminal observa les senyals emeses per una sèrie d'estacions base i calcula la seva posició a partir d'aquests mesuraments. Aquest sistema necessita que s'instal·lin unitats de mesurament de posició (*Location Measurement Units, LMU*) en punts dispersos geogràficament. La densitat de LMUs determinarà la precisió del sistema, i per això normalment és necessari instal·lar en tota la xarxa una LMU per cada una o dos estacions base. Aquests receptors i els terminals mòbils realitzen mesures de les senyals procedents de tres o més estacions base periòdicament. Les diferències temporals d'arribada de la senyal als dos punts (LMU i terminal) es combinen per produir línies hiperbòliques que es creuen a la posició on està el terminal mòbil, oferint d'aquesta manera la localització en dos dimensions. La precisió d'aquest mètode està en el interval de 50 a 200 metres.

1.2.4.1.3. Diferència en el temps d'arribada (*Uplink Time Difference Of Arrival, U-TDOA*)

U-TDOA també està basat en lateracions hiperbòliques, però s'utilitza l'enllaç de pujada (*uplink*), és a dir, les senyals emeses pel terminal són observades per la xarxa (el proveïdor), d'on es calcula la posició del terminal. El principal avantatge d'aquesta tècnica és que pot funcionar inclús sense visió directa, ja que la diferència de temps cancel·la possibles errors per reflexions. En entorns rurals es pot combinar amb el mètode bàsic de posicionament AOA per proporcionar major precisió.

1.2.4.1.4. GPS assistit (*Assisted GPS, A-GPS*)

Aquest mètode va ser creat per solucionar els problemes del GPS (l'engegada més ràpida, cobertura en interiors, consum de bateria, etc.), fusionant el GPS i els mètodes de posicionament cel·lulars. En aquest mètode el posicionament segueix el principi del GPS diferencial (D-GPS). S'afegeix un receptor GPS al terminal i així s'ofereixen dades d'assistència addicionals a la xarxa. Això permet reduir el temps d'adquisició i augmentar la precisió. La diferència d'aquest mètode respecte el GPS tradicional radica amb l'ús de receptors de referència. Aquests receptors recullen informació de navegació i dades de correcció diferencial per als satèl·lits GPS que estan a la zona de cobertura del servidor de localització. A partir de la informació obtinguda, el servidor de localització facilita baix

demanda, dades d'interès als terminals mòbils, principalment una llista amb els satèl·lits visibles pel terminal. També es poden fer localitzacions en tres dimensions, si el servidor de localització té accés a una base de dades d'elevacions del terreny.

1.2.4.2. UMTS

Els mètodes que s'utilitzen per la localització amb UMTS són els següents:

- ID de cel·la + RTT: La tècnica de localització ID de cel·la + RTT (*Round Trip Time*) és l'equivalent a la tècnica ID de cel·la + TA utilitzada en xarxes GSM. RTT fa una estimació del temps d'anada i tornada de l'emissor al receptor. Utilitzant ID de cel·la i mesuraments RTT podem arribar a tenir una precisió d'uns 50 metres, molt millor comparada amb la de GSM (aproximadament 500 metres).
- Diferència del temps d'arribada observat amb IPDL (OTDOA + IPDL): El mètode OTDOA és l'equivalent al E-OTD de GSM, basat en lateracions hiperbòliques. La gran diferència és la utilització dels períodes d'inactivitat en l'enllaç de baixada (*Idle Periods in Down Link*, IPDL) per a solucionar el problema conegut com "hearability problem" (problema de capacitat d'escolta). Aquest problema és degut quan un receptor escolta simultàniament diferents transmissors que tenen la mateixa potència. El receptor rep més potència del transmissor que té més a prop, això fa que sigui difícil escoltar els transmissors més llunyans. En definitiva, el problema "hearability problem" tracta d'un problema de detectar o filtrar una senyal més feble entre senyals més fortes.
- GPS assistit (*Assisted GPS*, A-GPS): Aquest mètode també es suporta amb UMTS, ja explicat als mètodes de posicionament de GSM a la secció 2.1.

1.2.4.3 Posicionament en xarxes CDMA

En aquesta secció es descriu un mètode exclusiu per xarxes CDMA (Code Division Multiple Access): la Trilateració avançada d'enllaç cap a davant (Advanced Forward Link Trilateration, A-FLT). El mètode A-FLT és exclusiu per xarxes CDMA. Aquest mètode és molt similar al TDOA, consisteix en efectuar la mesura del retard de fase entre senyals enviades a un parell d'estacions base i comparar-les amb la mesura de l'altre parell. Les dades procedents de tres estacions base permeten localitzar un terminal mòbil. També existeix una altra tècnica millorada amb els mateixos fonaments que AFLT, s'anomena EFLT (*Enhanced Forward Link Trilateration*).

1.2.5 Tecnologies de posicionament en espais interiors

Les tecnologies de posicionament per satèl·lit i les basades en xarxes cel·lulars són adequades per entorns exteriors amb visió directa dels satèl·lits o de les estacions base. No obstant això, quan aquestes tecnologies s'utilitzen a l'interior d'un edifici, presenten molts problemes deguts a l'atenuació de les senyals rebudes i als multitrajectes provocats pels fenòmens de la reflexió, la difracció i la dispersió de la senyal. Aquestes tecnologies no ofereixen la suficient precisió per saber

si una persona està dins o fora d'un edifici, per no esmentar que no és possible localitzar amb la granularitat d'habitacions o plantes. Per tant, en aquest escenari s'utilitza tecnologies de posicionament en espais interiors. A continuació es mostren aquestes tecnologies.

1.2.5.1 Bluetooth

Bluetooth [8] és una tecnologia de senyals de ràdio de curt abast, de baix cost i de baixa potència, desenvolupada per a substituir la connexió entre dispositius amb cable, per exemple, telèfons mòbils, auriculars i ordinadors portàtils. És va crear un estàndard per a comunicar-se amb aquests accessoris anomenat, Xarxa d'àrea personal (*Personal Area Network*, PAN).

Aquest mètode fa servir la trilateració per determinar la posició del dispositiu mòbil. Per a realitzar el posicionament es necessita almenys tres receptors Bluetooth instal·lats en posicions conegudes. La precisió és de 1,7 metres aproximadament. Com més receptors s'instal·lin major precisió aconseguirem. Cada receptor cobreix una distància màxima de 100 metres.

1.2.5.2 WLAN

En els últims anys les tecnologies de xarxes d'àrea local sense fils (*Wireless Local Area Network*, WLAN) van tenir un boom d'implantació. Aquestes tecnologies també suporten la localització de dispositius mòbils [21].

La majoria dels sistemes de posicionament WLAN que s'han desenvolupat estan basats amb els mesuraments de la robustesa de la senyal rebuda (*Received Signal Strength*, RSS), amb la relació de senyal / soroll (*Signal Noise Ratio*, SNR) o amb els sensors de proximitat. No es té en compte els mesuraments del temps, ja que és molt difícil la sincronització del temps de manera precisa amb WLAN. Les observacions es basen amb les anomenades balises (*beacons*). Es transmeten tant per l'enllaç de pujada (*uplink*) com per l'enllaç de baixada (*dowlink*). La precisió és de 2 a 6 metres.

1.2.5.3 RFID

La identificació per radiofreqüència (*Radio Frequency Identification*, RFID) és un sistema d'emmagatzematge i recuperació de dades remotes. Aquesta tecnologia s'utilitza principalment en aplicacions de gestió, de control d'accés, d'identificació de productes, d'automatització de fàbriques i pagament en peatges.

Un sistema RFID consta d'etiquetes i un lector amb una antena. Les etiquetes contenen informació i els lectors serveixen per llegir aquesta informació. Aquests sistemes es basen en senyals ràdio que s'intercanvien entre un lector RFID i etiquetes RFID.

Els sistemes de posicionament RFID [16] estan basats amb el mètode de posicionament bàsic, sensors de proximitat. Les etiquetes RFID es col·loquen a diferents posicions, proporcionant als usuaris mòbils (dispositius mòbils) amb informació rellevant segons la seva ubicació (per exemple, una etiqueta col·locada a l'entrada d'una sala d'un museu podria proporcionar informació als visitants sobre els autors de les pintures).

El rang de lectura és d'uns 6 metres aproximadament (normalment el rang de lectura depèn de l'entorn del lector i l'etiqueta utilitzada).

1.2.5.4. Sistemes basats en Infrarojos

La majoria de mètodes de posicionament amb senyals infrarojos estan basats amb sensors de proximitat. A diferència de les senyals de ràdio, les senyals d'infrarojos tenen l'avantatge que els seus emissors envien les senyals a pocs metres de distància (15 metres de rang de cobertura) i no penetren les parets. Aquestes propietats fa que els mètodes de posicionament amb senyals d'infrarojos s'utilitzin en aplicacions on es requereix que la posició de l'objectiu sigui com a mínim amb granularitat d'habitacions dins d'un edifici. Un petit inconvenient és que no pot haver-hi cap obstacle entre l'emissor i el receptor, si no és que la senyal rebota a la paret.

Un exemple de sistema basat en infrarojos, és el 'Active Badge' [28], desenvolupat en una investigació de Olivetti. Cada usuari porta visible un petit transmissor d'infrarojos, l'Active Badge, que cada 15 segons envia una senyal d'infrarojos. Aquesta senyal transporta un codi que especifica la identitat de l'usuari. Una xarxa de sensors d'infrarojos instal·lats a l'interior de l'edifici reben les senyals i, aquestes són processades per un servidor central.

1.2.5.5. Sistemes basats en ultrasons

Una alternativa als mètodes de posicionament amb senyals de ràdio o infraroges es utilitzar ultrasons. El major avantatge de les senyals d'ultrasons és la seva velocitat de propagació de 1.243 km/h, que és molt més baixa en comparació a les senyals ràdio o infraroges (aproximadament 300.000 km/s). En conseqüència d'això, els mètodes de posicionament amb senyals d'ultrasons utilitzen lateracions sense la necessitat de complicats i cars mecanismes de sincronització.

Les senyals d'ultrasons no penetren les parets i no requereixen que hi hagi una línia de visió entre l'emissor i el receptor. Dins dels edificis, els mètodes de posicionament amb ultrasons pot arribar a una precisió de centímetres.

Un exemple de sistema basat en ultrasons, és el 'Active Bat' [18]. A l'Active Bat es calcula la posició utilitzant el mètode de posicionament, lateracions circulars. L'usuari porta un dispositiu, anomenat 'Bat', que envia un impuls d'ultrasons breu al servidor. El servidor transmet la petició enviada per ràdio. El servidor sempre selecciona un Bat específic pel posicionament, per tant, els senyals ultrasons de diferents Bat no poden coincidir. Quan el receptor rep un senyal, immediatament passen aquesta informació al servidor de localització a través d'una xarxa sense fils. El servidor té totes les dades necessàries sobre la ubicació corresponent del usuari. A continuació, es crea i es soluciona un sistema d'equacions utilitzant el temps d'execució de les senyals d'ultrasò. El càlcul es simplifica ja que les senyal de ràdio tenen el temps d'execució molt més curt comparat amb les senyals d'ultrasò. Aquests sistemes són els que tenen la precisió més gran, d'uns 9 centímetres aproximadament.

1.2.5.6. Banda ultra ampla (Ultrawideband, UWB)

Aquesta tecnologia va néixer durant la dècada de 1960. Es va desenvolupar per a radar, localització i aplicacions de comunicacions. La capacitat de UWB d'operar per davall del nivell de soroll evitava que les comunicacions poguessin ser interceptades.

UWB [25] utilitza ràfegues de potència mil vegades més baix que les d'un telèfon mòbil, amb duració de pico-segons, a un espectre de freqüència ampli (3.1-10.6GHz). Sobre les ràfegues es possible transferir dades a velocitats de centenars de megabits per segon. A més, la senyal és relativament immune a la cancel·lació multitrajecte, ja que degut a la seva curta duració la senyal directa, va i torna abans de que les senyals reflectides als obstacles arribin al receptor. És un sistema de baixa complexitat i cost reduït.

Degut a les seves característiques, aquesta tecnologia permet localitzar els terminals mòbils amb un error molt baix. UWB està basada amb polsos ultracurts, de tal manera que el receptor pot determinar el temps d'arribada amb precisió de pico-segons i, per tant, estimar la posició amb precisió de centímetres. La distància al mòbil es calcula mesurant el retard d'un pols des de que s'emet pel transmissor fins que arriba al receptor. A continuació, utilitzant triangulació es determina amb molta exactitud la posició del terminal. Si es realitzen mesures respecte a quatre receptors diferents és possible saber amb precisió l'altura a la que està l'usuari.

1.3 Bibliografia del capítol

[1] CORDIS: Community Research and Development Information Service, <http://cordis.europa.eu/home-es.html>.

[2] FIAT Research Center. PEPTRAN project: PEdestrian and Public TRAnsport Navigator project. <http://www.crfproject-eu.org/>.

[3] Geodan Mobile Solutions, the webpage of the WebPark project, <http://www.webparkservices.info/>.

[4] Information on the Differential Global Positioning System (DGPS), <http://en.wikipedia.org/wiki/Differential-GPS>.

[5] I. Frech and B. Koch. Multimedia Geoinformation in Rural Areas with Eco-Tourism: the ReGeo-System. In: Information and Communication Technologies in Tourism 2003. pages 421-429, 2003. Springer.

[6] Official website of NAVSTAR, <http://www.navstarinc.com>.

[7] The Official Web Site of GPS, <http://www.gps.gov/>.

[8] The official website for the Bluetooth short range wireless connectivity standard, <http://www.bluetooth.com/bluetooth/>.

- [9] The Webpage of the AMBIESENSE project, <http://www.ambiesense.com>.
- [10] The Webpage of the ELBA project, <http://www.e-lba.com/>. 2001.
- [11] The Webpage of the EUREAUWEB project, <http://eureauweb.com/>.
- [12] The Webpage of the IM@GINE IT project, <http://www.imagineit-eu.com/>.
- [13] The Webpage of the M-GUIDE project, <http://www.exodus.gr/m-guide/index.htm>.
- [14] The Webpage of the REGEO project, <http://www.felis.uni-freiburg.de/regeo/English/>.
- [15] A. Goker and H. Myrhaug. User context and personalization. In: Proceedings of the European Conference on Case Based Reasoning (ECCBR 2002) - Workshop on Personalized CaseBased Reasoning.
- [16] H. D. Chon and S. Jun and H. Jung and S. Won. Using RFID for Accurate Positioning. 2004.
- [17] H. Kirchner and B. Mahleko and M. Kelly and R. Krummenacher and Z. Wang. eureauweb: An Architecture for a European Waterways Networked Information System. 2004.
- [18] A. Harter and A. Hopper and P. Steggles and A. Ward and P. Webster. The anatomy of a Context-Aware Application, Wireless Networks. 2002.
- [19] M. Karagiozidis. The Webpage of the LOVEUS project, <http://loveus.intranet.gr/>. 2002.
- [20] A. Kupper. *Location-based services: fundamentals and operation*, volume 1. JOHN WILEY & SONS, LTD, 2005.
- [21] B. Li and J. Salter and A. G. Dempster and C. Rizos. Indoor Positioning Techniques Based on Wireless LAN. 2006.
- [22] Queen Mary and Westfield College, University of London. The Webpage of the CRUMPET Project. <http://www.elec.qmul.ac.uk/crumpet/>.
- [23] M. Megliola and L. Barbieri. INTEGRATING AGENT AND WIRELESS TECHNOLOGIES FOR LOCATION-BASED SERVICES IN CULTURAL HERITAGE. 2008.
- [24] B. Mennecke and T. Strader. Mobile Commerce: Technology, Theory and Applications. 1999.
- [25] Z. Sahinoglu and S. Gezici and I. Güvenc. *Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols*. Cambridge University Press; illustrated edition edition (4 Sep 2008), 2008.
- [26] C. Schlieder and A. Werner. Interpretation of Intentional Behavior in Spatial Partonomies. Volume 2685/2003:401, 2003.
- [27] B. Schmidt-Belz and H. Laamanen and S. Poslad and A. Zipf. Location-based mobile tourist services - first user experiences. 2002.

[28] R. Want and A. Hopper. Active Badge and personal interactive computing objects. 1992.

[29] Website with information on GALILEO positioning technology, [http://en.wikipedia.org/wiki/Galileo-\(satellite-navigation\)](http://en.wikipedia.org/wiki/Galileo-(satellite-navigation)). 2009.

2. Estudi dels diferents tipus de serveis que es poden oferir a partir de la localització i/o seguiment de turistes

2.1 Estudi de l'oferta de serveis de turistes segons la seva localització i/o seguiment.

S'ha realitzat un estudi dels de projectes i bibliografia relacionada amb la provisió de diferents tipus de serveis a partir d'informació relacionada amb la ubicació de l'usuari.

Abans de presentar els resultats, volem constatar que aquesta àrea d'investigació que es troba en una fase experimental, com veurem, això ha condicionat el desenvolupament de l'estudi i ha obligat a recopilar informació molt diversa, la qual s'ha hagut d'agrupar temàticament en blocs per tal de fer més pràctica la gestió de la informació.

S'ha creat una plantilla (base de dades *Access*) per tal de facilitar la catalogació dels articles que s'han considerat rellevants per aquest estudi. Les fitxes a on s'han registrat contenen diferents variables que els caracteritzen i els analitzen amb la finalitat de conèixer els punts que interessin pel present estudi. Així doncs, gràcies a aquest sistema, s'han pogut comparar i extreure conclusions generals, que es presentaran en el capítol 5. El nombre d'articles inventariats amb aquesta metodologia és de 26. La taula es troba al final d'aquest capítol. Els criteris amb els quals s'ha valorat cada article els trobem explicats a la taula 2.1.

Tal com s'ha dit, els estudis sobre serveis avançats per a turistes en funció de la seva localització i/o seguiment es troben en una fase experimental. Això es constata veient que dels 26 articles identificats, un total de 16 han estat publicats en els darrers 5 anys, tot i això hi ha algunes primeres referències a aquest tema a l'any 1996 i 1997.

A la taula 2.1 també es pot observar que els articles formen un conjunt força heterogeni. Tot i que no es perd de vista el fil conductor de la investigació, els enfocaments són diversos en molts sentits. El que determina aquesta diversitat és, sense cap mena de dubte, **els objectius**. Hi ha estudis que única i exclusivament volen saber la influència de les campanyes publicitàries turístiques d'una destinació en concret [17] o quines rutes han seguit els turistes i què han fet [20], també s'han llegit articles a on els investigadors s'han situat en un node (un aeroport, per exemple) preguntant als viatgers la direcció del seu viatge [14], etc. Però, per sobre de tot, hi ha un objectiu comú i general que, o bé es fa visible o es troba intrínsec en totes les recerques observades i és que, al cap i a la fi, tot estudi sobre turisme pretén millorar la gestió sobre aquest, tot i que l'objectiu concret sigui experimentar sobre un mètode de seguiment o sobre qualsevol sistema que aparentment no tracti sobre turisme.

Camp	Descripció del camp
Refworks	Identificació al programa bibliogràfic interactiu.
Autor	Investigadors/es principals.
Títol	Nom de l'article/investigació.
Any	Any de publicació de l'article.
Àrea	Nom de la zona a on s'ha realitzat la investigació.
País	Estat a on s'ha realitzat la investigació.
Tipus zona	Tipologia de la zona a on s'ha realitzat la investigació.
Superfície	Dimensions de la zona a on s'ha realitzat la investigació.
Escala	Escala emprada per a la representació en SIG de la zona a on s'ha realitzat la investigació.
Desplaçament	Mode de desplaçament (Ex: "a peu").
Software	Software emprat per a desenvolupar la investigació.
Hardware	Hardware emprat per a desenvolupar la investigació.
Objectius del "tracking"	Finalitat del seguiment als turistes.
Metodologia	Com s'ha desenvolupat la investigació.
Tipus de dades	Dades emprades per a desenvolupar la investigació (Ex: "qüestionari").
Resum	Petit resum de tot l'article.

Taula 2.1. Criteris

Un altre tret distintiu que es relaciona amb els objectius és l'**àmbit d'estudi**. Hi ha articles que localitzen la recerca en un casc antic [10], altres en una reserva natural [26] i, fins i tot, s'han trobat uns pocs que ho fan en una regió [24]. Ara bé, si s'observen els estats la qüestió passa a ser més homogènia: Austràlia, Israel, Alemanya, EEUU, Anglaterra i Hong Kong són els països que acaparen la majoria de recerques. Això en el cas de l'observació del comportament turístic. Si el que es vol és treballar el **seguiment**, trobem dues línies d'investigació: per una banda els estudis que únicament volen experimentar diferents hardware [22] i, en l'altre costat, els que experimenten amb els propis turistes per a conèixer les seves activitats [20]. Per tant, en aquesta temàtica es fa imprescindible la prèvia acotació del territori el qual, per a poder ser representat, es requereix

establir l'**escala** més idònia. S'ha de tenir en compte que l'escala amb la que es representarà la ruta seguida pel visitant vindrà condicionada per les dimensions de la zona per on s'ha mogut i el conjunt que es pren com a referència.

Per a la gestió de recintes i/o zones turístiques, un dels factors més estudiats és el **desplaçament**. Però com s'ha anat veient successivament, no es pot analitzar un criteri amb independència d'uns altres. És a dir, si estem parlant d'un casc antic o una reserva natural, és obvi que el desplaçament es realitzarà a peu [5]. En canvi, en una ciutat o regió el transport motoritzat serà el protagonista [7]. Finalment, en viatges transnacionals, l'avió es fa imprescindible [14]. Sempre, depenent dels objectius, les variables que es tindran en compte seran unes o altres. Tot i així, el mode de desplaçament és essencial en quant a la gestió dels espais turístics.

El **software** i el **hardware** no són comentats gairebé mai. Únicament, si s'escau, els anomenen però no se'ls hi dóna més importància. S'ha trobat un article el qual sí ocupa les seves línies en el software i el hardware ja que el seu objectiu és experimentar amb ells [22]. Les altres aparicions són residuals i anecdòtiques. Cal comentar algunes investigacions que, encara que no tractin de turisme estrictament, treballen amb tècniques de simulació de moviment i presa de decisió i altres tecnologies aplicables al turisme [11].

Per tal d'arribar a unes conclusions, tots els articles han treballat amb unes **dades** concretes. La majoria, al tractar-se d'un fenomen social, s'han basat en uns qüestionaris a on el visitant havia de respondre certes preguntes com la seva ruta, els llocs visitats, el mode de transport, acompanyants, etc. Si bé és cert que molts s'han recolzat en entrevistes i qüestionaris (sobretot els estudis de *tracking*), a alguns articles no s'han inclòs les preguntes realitzades fent difícil conèixer quins criteris van interessar per a la investigació en qüestió.

En canvi, els estudis més tècnics centrats en la metodologia o en l'experimentació de hardware i software han obviat la part social per valorar-la com a prescindible.

2.2 Identificació i classificació dels serveis que s'estan oferint a turistes i proposta de possibles nous serveis.

De manera general, l'objectiu d'aquests projectes, tots ells experimentals, és el d'analitzar el comportament turístic, els fluxos, les aglomeracions, el mode de transport, etc., amb la finalitat d'establir tipologies de turistes per tal de millorar la gestió de les destinacions. Els Sistemes d'Informació Geogràfica (SIG) combinats amb altres hardware i software són les eines més emprades fins aleshores.

Ara bé, hi ha dos àmbits d'aplicació dels SIG: els orientats a la planificació del turisme i els orientats al propi turista. Els primers estan més desenvolupats degut a què s'està en la fase de trobar la millor utilitat per a la gestió dels espais turístics, fet que provoca que els serveis al turista quedin, de moment, en un segon pla. Tot i així, hi ha un avenç notable en una àrea d'estudi en concret en relació als SIG que fa referència al visitant: la presa de decisions.

En quant a la **planificació del turisme**, s'ha comprovat l'extensa utilitat del SIG per a moltes funcions. Algunes d'aquestes són: (a) la realització d'inventaris de recursos turístics, instal·lacions i serveis; (b) l'anàlisi de forma integrada de dades turístiques de destinacions i la gestió d'infraestructures; (c) la determinació d'àrees que compleixin determinats requisits com, per exemple, àrees òptimes per a la conservació degut a la seva alta qualitat mediambiental o per a la instal·lació d'equipaments; (d) l'anàlisi d'aptituds per a l'activitat turística o (e) l'avaluació de l'impacte del turisme mitjançant la comparació de la situació existent abans i després de la instal·lació de l'activitat turística. Com es pot observar, les variables són moltes, les conclusions de la seva utilitat estan establertes en la seva majoria i, la discussió en aquest cas, està instal·lada en com dur-ho a terme i amb quin equipament informàtic.

En l'altre costat, **els serveis a turistes**, tot i haver-hi un bon inventari de possibilitats, estaran força immòbils fins que la gestió turística no estableixi unes metodologies concretes prèviament. En cas contrari, arribar al segon pas, que és els serveis destinats als turistes, serà difícil abordar-lo amb garanties i determinació. La majoria, per no dir tots, estan basats en els Sistemes d'Ajuda a la Presa de Decisió Espacial (*SDSS, Spatial Decision Support System*) que, com el seu nom indica, orienten al visitant a escollir les diferents ofertes que la destinació ofereix aportant-li tota la informació necessària per a que aquest prengui les seves decisions al llarg de la seva estada. Un dels exemples és el presentat per [26] els quals valoren la influència del servei més generalitzat en una zona turística natural: les marques del territori (*landmark*), o el que és el mateix, les senyalitzacions per tal de seguir una ruta (*wayfinding*). És un dels serveis més senzills que es pot oferir avui dia però molt necessari per a estructurar un circuit interessant pels visitants. Hom pot veure en la majoria d'entorns a on s'ha apostat pel turisme rural basat en el senderisme, marques als arbres, roques, o senyalitzacions verticals que, potser mitjançant un color o una explicació, indiquen el camí a seguir i informen sobre les característiques biòtiques de la zona, respectivament.

Si es parla de sistemes més tecnificats a on es requereix un hardware i un software (com els SIGs, per exemple) ja s'entraria en les Guies Turístiques Portàtils (*m-ToGuide, Mobile Tourist Guides*) sumades amb els Serveis Basats en la Localització (*Located Based Services*) o les WebGIS (*Web based Geographic Information Systems*). Les Guies Turístiques Portàtils contenen una base cartogràfica que permet al turista ubicar-se i, a la vegada, obtenir informació descriptiva (sobre els recursos i equipaments turístics) i logística (horaris, preus, etc.) a més de, si l'usuari així ho requereix, realitzar operacions de reserves o comerç electrònic (*booking* o *e-commerce*). Aquestes guies es troben en un terminal portàtil com una PDA o un telèfon mòbil que son utilitzats per a intercanviar informació en temps real entre la m-ToGuide i el visitant a través de la connexió a Internet. Una altra de les aplicacions possibles correspon a les WebGIS que són portals d'Internet que proporcionen tota la informació necessària pel turista a partir de plataformes de base espacial. Els WebGIS doten al turista d'una eina d'ús fàcil que els dona la oportunitat treballar amb mapes. L'aplicatiu permet, per un costat, operacions bàsiques com són les ampliacions, els desplaçaments,

o activar i desactivar determinades capes d'informació. I, per l'altre costat, feines més complexes com la mesura de distàncies, la recerca de determinats recursos en funció de la seva localització espacial o de les seves característiques temàtiques com, per exemple, restaurants vegetarians; o buscar l'itinerari més curt entre dos punts com podria ésser el camí a seguir des d'un hotel fins a un monument en concret.

2.3 Bibliografia del capítol

[1] AHAS, R.; AASA, A.; MARK, Ü.; PAE, T.; KULL, A. (2007). Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management*, 28(3), 898-910.

[2] AHAS, R.; AASA, A.; ROOSE, A.; MARK, Ü.; SILM, S. (2008). Evaluating passive mobilepositioning data for tourism surveys: An Estonian case study. *Tourism Management*, 29(3), 469-486.

[3] ASAKURA, Y.; IRYO, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice*, 41(7), 684-690.

[4] CHARDONNEL, S.; VAN DER KNAAP, W. (2004). Managing tourist time-space movements in recreational areas. A comparative study of a protected natural park in the French Alps and the « De Hoge Veluwe ». Dutch national park using the same methodology. *Revue De Géographie Alpine*, 90(1), 37-48.

[5] Determining hiking experiences in nature-based tourist destinations. *Tourism Management*, 25(1), 31-43.

[6] DONGGEN, W.; TAO, C. (2001). A spatio-temporal data model for activity-based transport demand modelling. *International Journal of Geographical Information Science*, Volume 15, Issue 6. A Spatio-Temporal Data Model for Activity-Based Transport Demand Modelling. January 2001 , Pages 561 - 585, 15(6), 561.

[7] A tourist space-time budget in the Shetland Islands. *Annals of Tourism Research*, 23(4), 811-829.

[8] FRIHIDA, A.; MARCEAU, D. J.; THÉRIAULT, M. (2004). Development of a temporal extension to query travel behavior time paths using an object-oriented GIS. *GeoInformatica*, 8(3), 211.

[9] ITAMI, R. M. (2003). RBSim3: Agent-based simulations of human behaviour in GIS environments using hierarchical spatial reasoning. In *Proceedings of the ModSim 2003 conference*.

- [10] Tracking the Salzburg tourist. *Annals of Tourism Research*, 24(4), 1008-1012.
- [11] GABRIEL: Gis Activity-Based tRavel sImuLator. Activity scheduling in the presence of real-time information. *GeoInformatica*, 10(4), 469.
- [12] LAM, T.; HSU, C. H. C. (2006). Predicting behavioral intention of choosing a travel destination. *Tourism Management*, 27(4), 589-599.
- [13] LAU, G.; MCKERCHER, B. (2007). Understanding tourist movement patterns in a destination: A GIS approach. *Tourism and Hospitality Research*, 7(1), 39-49.
- [14] Trip destinations, gateways and itineraries: The example of Hong Kong. *Tourism Management*, 23(6), 609-621.
- [15] LEW, A.; MCKERCHER, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of Tourism Research*, 33(2), 403-423.
- [16] LOITERTON, D.; BISHOP, I. (2005). Virtual environments and location-based questioning for understanding visitor movement in urban parks and gardens. In *Trends in Real-Time Landscape Visualization and Participation (Conference on Information and Technologies in Landscape Architecture)*, 144.
- [17] An expanded framework for measuring the effectiveness of destination advertising. *Tourism Management*, 18(3), 127.
- [18] MODSCHING, M.; KRAMER, R.; GRETZEL, U.; TEN HAGEN, K. (2006). Capturing the beaten paths: A novel method for analysing tourists spatial behaviour at an urban destination. In *Information and Communication Technologies in Tourism 2006 Proceedings of the International Conference in Lausanne, Switzerland, 2006*. 75.
- [19] MUKAI, N.; HONMA, H.; KOSUGI, M. (2007). Generation of dynamic moving path for autonomous human model. In *MODSIM07: Land, Water and Environmental Management: Integrated Systems for Sustainability*.
- [20] Geo-temporal tracking and analysis of tourist movement. *Mathematics and Computers in Simulation*, 69(1-2), 135-150.
- [21] SHOVAL, N. (2008). Tracking technologies and urban analysis. *Cities*, 25(1), 21-28.
- [22] Tracking tourists in the digital age. *Annals of Tourism Research*, 34(1), 141-159.
- [23] THORNTON, P. R.; SHAW, G.; WILLIAMS, A. M. (1997). Tourist group holiday decision-making and behaviour: The influence of children. *Tourism Management*, 18(5), 287-297.
- [24] Research report: GIS-oriented analysis of tourist time-space patterns to support sustainable tourism development. *Tourism Geographies*, 1 (1), p. 56-69. Research Report: GIS-Oriented Analysis of Tourist Time-Space Patterns to Support Sustainable Tourist Development, 1(1), 56.

[25] XIA, J.; ARROWSMITH, C. (2005). Managing scale issues in tourist spatio-temporal movement modelling. In Zerger, A. and Argent, R.M. (Eds) MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 170.

[26] The wayfinding process relationships between decision-making and landmark utility. *Tourism Management*, 29(3), 445-457.

<i>ID</i>	<i>rAutor</i>	<i>Titol</i>	<i>Any</i>	<i>Area</i>	<i>Pais</i>	<i>Tipus</i>	<i>z</i>	<i>Sup</i>	<i>Escala</i>	<i>Desplacament</i>	<i>Software</i>	<i>Hardware</i>	<i>Objectius</i>
24	VAN DER KNAAP, Wim G.M.	Research report: GIS - oriented analysis of tourist time - space patterns to support sustainable tourism development.	1999	1: Nette Tal region 2: Euregion Maas-Rijn	Alemanya, Holanda i Bèlgica	Regió europea					Cartographic Data Visualization system for time-space analysis (CVD-TS)		Dissenyar un tractament en dos passos de dades turístiques específiques per a analitzar el comportament dels visitants basat en el temps i en l'espai.
15	LEW, Alan; MCKERCHER,	Modeling tourist movements. A local destination analysis.	2006	Destinació turística	Indiferent	Turística (preferiblement)	Indiferent			De tot tipus			Identificar els factors que influeixen els patrons de moviment dels turistes dintre d'una mateixa destinació.
14	LEW, Alan; MCKERCHER,	Trip destinations, gateways and itineraries: the example of Hong	2002	Ciutat	Hong Kong	Urbana				Aeri			Saber el tipus de destinació que és Hong Kong en funció de la gent enquestada.
20	O'CONNOR, A.; ZERGER, A.; ITAMI, B.	Geo-temporal tracking and analysis of tourist movement.	2005	Twelve Apostles National Park (Victoria, Australia)	Austràlia	Parc Nacional	1:2400			A peu	Recreation Behaviour Simulator (RBSim), basat en l'Alge Timing System	Unobtrusive ankle transmitters	Simular el moviment usual i previst dels turistes per tal de determinar l'impacte en l'experiència del visitant, així com la capacitat de càrrega de la infraestructura.
22	SHOVAL, Noam; ISAACSON,	Tracking tourists in the digital age.	2007	1: Heidelberg (Alemanya) 2: Old city of Jerusalem 3: Nordoest d'Israel	Alemanya i Israel	1: Ciutat 2: Casc antic 3: Regió	1: 1/43478 2: 1/33333 3: 1/76923 1/176470; 1/789473; 1/12500	1: A peu, en cotxe i amb autobús 2: A peu i amb cotxe 3: A peu i amb autobús			1: GPS Tracking 2: Cellular triangulation 3: Land Based TDOA Tracking	1: Emtac Cruxll BlueTooth GPS receiver and Pocket PC 2: Motorola i860 AGPS. 3: Petit dispositiu de	Experimentar amb els diversos software per tal de veure els seus punts forts i febles a l'hora de realitzar el seguiment als turistes.

<i>ID</i>	<i>r</i>	<i>Autor</i>	<i>Títol</i>	<i>Any</i>	<i>Àrea</i>	<i>País</i>	<i>Tipus</i>	<i>z0</i>	<i>Sup</i>	<i>Escala</i>	<i>Desplacament</i>	<i>Software</i>	<i>Hardware</i>	<i>Objectius</i>
8		FRIHIDA, Ali; MARCEAU, Danielle J.; THÉRIAULT, Marius	Development of a Temporal Extension to Query Travel Behavior Time Paths Using an Object-Oriented GIS.	2004	Québec	Canadà	Regió					GIS		Cear un prototip que permeti simular el moviment dels turistes. Per a fer això, es vol implementar un SIG en format de prototip, amb l'objectiu que es pugui emprar en un futur com una eina de simulació.
1		Ahas, Rein; Aasa, Anto; Mark, Ülar; Pae, Taavi; Kull,	Seasonal tourism spaces in Estonia: Case study with mobile positioning data	2007	Estonia	Estonia		45.21	5	km2			Telèfons mòbils	Es vol analitzar l'estacionalitat dels espais turístics d'Estonia fent un seguiment als turistes estrangers. Les dades obtingudes s'avaluaran prenent crear un mètode amb l'objectiu de saber com emprar les dades de posicionament mòbil.
6		Donggen, Wang; Tao, Cheng	A spatio-temporal data model for activity-based transport demand modelling.	2001	Ciutat	Hong Kong	Urbana					SIG: ArcView	GPS	Desenvolupar un model de dades espacial - temporals que pugui representar el comportament dinàmic de l'activitat i amb el suport d'un SIG.

<i>ID</i>	<i>r Autor</i>	<i>Titol</i>	<i>Any</i>	<i>Area</i>	<i>Pais</i>	<i>Tipus</i>	<i>zo Sup</i>	<i>Escala</i>	<i>Desplacament</i>	<i>Software</i>	<i>Hardware</i>	<i>Objectius</i>
26	XIA, Jianhong; ARROWSMITH, Colin; JACKSON, Mervyn;	The wayfinding process relationships between decision - making and landmark utility.	2008	Koala Conservation Center (KCC) (Phillip Island, Victoria)	Austràlia	Reserva animal	15,5h	1:9090	A peu	GPS	GPS receivers	Entendre els diferents mètodes de "wayfinding" triats pels turistes per tal de millorar la gestió dels parcs, aportant ajudes complementàries als visitants alhora de triar les rutes a seguir.
3	ASAKURA, Yasuo; IRYO,	Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument.	2006	Kobi Kitano	Japó	Ciutat	300m	x300m	A peu	GPS	Wearable data logger named personal activity monitor (PEAMON)	Analitzar el comportament turístic a partir de la construcció d'un índex de rutes topològiques.
13	Lau, Gigi; McKercher, Bob	Understanding tourist movement patterns in a destination: A GIS approach	2007	Hong Kong	Hong Kong	Urbana				GIS		Estudiar els patrons de moviment del turistes individuals.
4	CHARDONNEL, Sonia; VAN DER KNAAP, Wim G.M.	Managing Tourist Time-Space Movements in Recreational Areas. A comparative study of a protected natural park in the French Alps and the « De Hoge Veluwe ». Dutch National Park using the same	2004	1: Ardèche region and the Northern Alps 2: De Hoge Veluwe	1: França 2: Holanda	1: Àrea natural protegida 2: Parc Nacional			1: A peu 2: En cotxe, bicicleta, transport públic, a peu	Cartographic Data Visualizer for Time-Space data		Es pretén avaluar el CDV-TS com a sistema útil pels gestors d'un espai turístic i si aquests són capaços de fer-lo servir i interpretar els resultats

<i>ID</i>	<i>r Autor</i>	<i>Titol</i>	<i>Any</i>	<i>Area</i>	<i>Pais</i>	<i>Tipus</i>	<i>zo Sup</i>	<i>Escala</i>	<i>Desplacament</i>	<i>Software</i>	<i>Hardware</i>	<i>Objectius</i>
2	AHASA, Rein; AASAA, Anto; ROOSEA, Antti; MARKB, Ular; SILMA, Siiri	Evaluating passive mobile positioning data for tourism surveys: An Estonian case study	2007	Estonia	Estonia		45.21 5 km2				Telèfons mòbils	Introduir i avaluar l'aplicabilitat de les "passive mobile positioning data" alhora de realitzar estudis turístics.
21	Shoval, Noam	Tracking technologies and urban analysis.	2008	Old City of Akko	Israel	Ciutat		1:6500		SIG	Emtac Cruxll BlueTooth GPS receivers and Pocket PC.	Saber l'impacte dels visitants a les ciutats i posar en relleu les possibilitats i les dificultats de la utilització de la tecnologia GPS als projectes d'investigació urbana.
5	CHHETRI, Prem; ARROWSMITH, Colin; JACKSON, Mervyn	Determining hiking experiences in nature-based tourist	2004	Pinnacle (Grampians National Park, Victoria)	Austràlia	Parc Nacional		1:25.000	A peu			Determinar l'impacte sobre els visitants dels paisatges naturals.
25	Xia, J.; Arrowsmith, C.	Managing scale issues in tourist spatio-temporal movement modelling	2005	1: Phillip Island 2: Koala Conservation Centre	Austràlia	1: Illa 2: Reserva animal		1:150000 // 1:30000				Analitzar les qüestions que fan referència a l'escala temporal i espacial dels turistes comparant i analitzant els diferents mètodes emprats fins ara.

<i>ID</i>	<i>r Autor</i>	<i>Títol</i>	<i>Any</i>	<i>Area</i>	<i>País</i>	<i>Tipus</i>	<i>z0 Sup</i>	<i>Escala</i>	<i>Desplacament</i>	<i>Software</i>	<i>Hardware</i>	<i>Objectius</i>
9	ITAMI, Robert M.	RBSim3: Agent-based simulations of human behaviour in GIS environments using hierarchical spatial reasoning.	2003	Loch Ard Gorge - Port Campbell National Park (Victoria)	Austràlia	Parc Nacional		1:5.000		SIG		Aprofundir en la millora del Recreation Behaviour Simulator (RBSim) treballant en les jerarquies espacials i probabilístiques (comportament i presa de decisió) per a la simulació dels viatges.
19	Mukai, N.; Honma, H.; Kosugi, M.	Generation of Dynamic Moving Path for Autonomous Human Model	2007						A peu			Establir un mètode de generació automàtica de moviment per a visualitzar el comportament d'humans autònoms
7	FENNELL, David A.	A tourist space-time budget in the Shetland Islands	1996	Shetland Islands	Gran Bretanya	Arxipèlag		1,468 km2			Cotxe, ferry, autobús, etc.	Es pretén analitzar les pressions turístiques sobre les illes Shetland a partir d'establir una tipologia de turistes segons les semblances espacials i temporals en l'activitat realitzada.

<i>ID</i>	<i>r Autor</i>	<i>Titol</i>	<i>Any</i>	<i>Area</i>	<i>Pais</i>	<i>Tipus</i>	<i>zo Sup</i>	<i>Escala</i>	<i>Desplacament</i>	<i>Software</i>	<i>Hardware</i>	<i>Objectius</i>	
16	LOITERTON, Daniel; BISHOP, Ian	Virtual Environments and Location-Based Questioning for Understanding Visitor Movement in Urban Parks and Gardens.	2005	Royal Botanic Gardens (Melbourne)	Australia	Parc	35	ha.			PDA - GPS	Es pretén saber si es pot modelar el comportament dels visitants en un entorn específic.	
17	McWilliams EG; Crompton JL	An expanded framework for measuring the effectiveness of destination advertising.	1997	Estat de Texas	EUA	Regió						Es pretén avaluar l'impacte i la rendibilitat de la publicitat turística en els viatgers.	
10	KEUL,Alexander; KÜHBERGER,Anton	Tracking the Salzburg tourist	1997	Old town of Salzburg	Austria	Ciutat		1:5.000	A peu			Recopilar dades bàsiques de comportament turístic a les ciutats.	
18	Modsching, Marko; Kramer, Ronny; Gretzel, Ulrike; Ten Hagen, Klaus	Capturing the Beaten Paths: A Novel Method for Analysing Tourists Spatial Behaviour at an Urban Destination	2006	Görlitz	Alemanya	Ciutat					SIG	Mobile Digital Assistant (MDA) i Bluetooth GPS	Establir una metodologia per analitzar i estudiar el comportament espacial dels turistes basat en un sistema de dades de tracking.
24	Thornton,Paul R.; Shaw,Gareth; Williams,Allan M.	Tourist group holiday decision-making and behaviour: the influence of children	1997	Cornwall (Newquay and the Bodmin area)	Anglaterra	Regió		1:1000000	A peu, en cotxe o en autobús.			Analitzar la influència dels nens en l'activitat turística en grup.	
11	KWAN, Mei-Po; CASAS, Irene	GABRIEL: Gis Activity-Based tRavel sImuLator. Activity Scheduling in the Presence of Real-Time	2006	Columbus (Ohio)	EEUU	Ciutat					GIS (ArcView // ArcInfo) - GABRIEL	Presentar el disseny i la implementació del simulador de viatges "GABRIEL".	

3. Estudi de les tècniques per garantir la privadesa dels turistes

La disciplina que s'ocupa d'estudiar com preservar la privadesa dels individus en les dades s'anomena Control d'Inferència Estadística (Statistical Disclosure Control - SDC). L'objectiu és desenvolupar mètodes que permetin la publicació de dades (estadístiques) amb la seguretat de que aquestes dades no pondran vincular-se a dades confidencials de subjectes que les han proporcionat.

Quan un conjunt de dades proporcionades per individus es publica, els qui veuen les dades poden intentar deduir informació confidencial sobre un individu a partir de la correlació d'estadístiques publicades sobre grups d'individus. A aquestes accions (malicioses) se les denomina atacs. Els atacs es classifiquen en dos tipus:

Revelació d'identitat: associar la identitat d'un enquestat amb registres de dades publicats que contenen informació confidencial. En el cas de trajectòries seria descobrir qui fa la trajectòria.

Revelació d'atribut: associar el valor d'un atribut en les dades publicades d'un enquestat. En el cas de trajectòries seria, per exemple, descobrir la destinació d'un individu.

La protecció que ofereixen els mètodes SDC es basa en un cert grau de modificació de les dades originals que minimitzi el risc de re-identificació dels individus. Aquesta modificació es troba en un punt intermedi entre:

- No modificació (dades útils però no segures)
- Xifratge (dades intractables però segures)

El principal repte del SDC es modificar les dades de tal forma que s'obtingui la seguretat suficient amb la mínima pèrdua d'informació possible.

La subdisciplina en la qual ens centrem és la protecció de microdades. És la més jove, i esta en plena expansió. Consisteix en la protecció de les dades estàtiques i individuals, que són les que es denominen microdades. Per exemple:

Edat	Sexe	Estat Civil	Fills	Ingressos
65	Masculí	Solter	2	70.000
38	Femení	Casat	3	95.000
65	Masculí	Vidu	2	25.000

Taula 3.1 Exemple Microdades

En aquest exemple es pot apreciar com el fet d'eliminar el nom i el DNI no és suficient. Els habitants de la població on s'ha fet l'enquesta poden saber els sous de la resta a partir de les dades publicades, si coneixen altres dades dels individus. Si al poble només hi ha un senyor solter de 65 anys i amb 2 fills, es pot saber quina han estat els seus ingressos.

A continuació formalitzem una mica el problema de protecció de privadesa en microdades.

Definició: Un conjunt de microdades V pot ser vist com un fitxer d' n files on cada fila conté m atributs.

Cada atribut pot ser classificat en quatre categories (no disjundes):

- **Identificadors:** Són atributs que identifiquen de forma no ambigua a l'individu al qual fan referència. (per exemple, el DNI, passaport, N° seguretat social).
- **Quasi-identificadors:** Atributs que identifiquen a l'individu al qual fan referència amb un cert grau d'ambigüitat. Una combinació d'aquests atributs pot portar a la identificació, sense ambigüitats, de l'individu. (per exemple combinacions com Edat + Gènere + Estat Civil).
- **Confidencials (Sensibles):** Contenen informació sensible sobre l'individu. (per exemple, salari, religió, malalties que ha tingut, llocs on ha estat).
- **No confidencials:** Atributs que no encaixen en cap de les anteriors categories.

Objectiu: L'objectiu de la protecció de microdades es oferir un conjunt de microdades protegit V' a partir d'un conjunt de microdades original V de tal forma que:

- el risc d'inferència sigui baix
- l'anàlisi estadístic d'aquest nou conjunt V' doni els mateixos (o similars) resultats.

En aquest capítol es presenten els mètodes protecció de la privadesa i control d'inferència estadística en el cas d'ubicacions en sistemes LBS i de dades de trajectòries en estudis de mobilitat de les persones.

3.1 Estudi de la privadesa en Sistemes Basats en la Localització

Un servei basat en la localització (*Location Based Service*, LBS) és un sistema d'informació que empra la localització (posició) dels usuaris per proporcionar serveis personalitzats.

Els LBS ja fa molt de temps que es fan servir. Al 1973 el Departament de Defensa del Estats Units va posar en funcionament el sistema de posicionament global (*Global Positioning System*, GPS), un sistema global de posicionament per satèl·lit (*Global Navigation Satellite Systems*, GNSS) que permet determinar en tot el món la posició d'un objecte, una persona, un vehicle, etc. Inicialment, el propòsit del GPS era per a fins militars, però al 1983 el govern del Estats Units va anunciar que el GPS estaria disponible per a tothom. Això va provocar que l'indústria utilitzés aquesta tecnologia per a millorar els seus productes i serveis. Un exemple a l'indústria de l'automòbil són els sistemes de navegació. En els sistemes de posicionament tradicionals, la informació de localització era proporcionada per un dispositiu amb l'ajuda d'un receptor GPS. Les noves oportunitats de negoci dels LBS han arribat amb els avanços de les tecnologies de posicionament i comunicació sense fils que fan que els LBS estiguin disponibles en qualsevol lloc i en qualsevol moment.

Els LBS es poden classificar en diferents categories basant-nos amb la seva funcionalitat. Com a exemples d'aplicacions LBS tenim: les de navegació (orientació i control del trànsit), d'informació (guies turístiques), de seguiment (de persones, vehicles i productes), d'emergència (policia i ambulància), de publicitat (alertes d'anuncis), facturació (peatges) i xarxes socials (localització d'amics, missatgeria instantània).

La capacitat de localitzar la posició dels usuaris tenen avantatges i inconvenients. Per una banda, té una gran repercussió en la societat i l'economia ja que s'obté una bona xarxa de comunicació, com per exemple, un dispositiu de localització d'amics, però per altra banda la divulgació de la informació de la localització, o els interessos dels usuaris atempta contra la privadesa personal o la seva seguretat. Per tant, la gestió inadequada de les dades personals dels usuaris és una qüestió de considerable interès públic que pot desaccelerar el desplegament dels LBS.

Diversos països han pres iniciativa legal per fer front als problemes de privadesa en comunicacions electròniques. Per exemple a Europa, la directiva Europea sobre protecció i privadesa de les dades [19] va acordar un conjunt de mesures per assegurar la privadesa dels usuaris. Totes aquestes mesures regulen models de negoci ben establerts, però la seva aplicació es complexa.

Els actors involucrats en el procés de provisió de serveis basats en la localització (LBS) són:

- L'usuari.
- El proveïdor de LBS.
- Tercera part de confiança (*Trusted Third Parties*, TTP).

3.1.1 Classificació dels mètodes per a la privadesa en LBS

La forma més senzilla de comunicació entre un usuari del LBS (U) i un proveïdor del LBS (P), l'usuari envia una consulta (C) que conté un identificador, la seva localització (L) i una petició d'informació (I) que ell vol recuperar de P. Per tant, una consulta que s'envia de U a P es pot representar com $C = ID_U, L, I = ID_U, x_U, y_U$, "On és el restaurant més proper?" (vegeu la Figura 3.1). Mitjançant l'enviament de la seva ubicació a P, l'usuari assumeix que P gestiona les seves dades amb honestetat i s'absté de qualsevol ús indegut. Com que el proveïdor del LBS no sempre serà de confiança, es necessiten esquemes de comunicació més complexes.

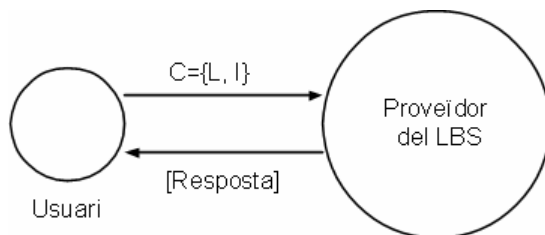


Figura 3.1. Esquema de comunicació simple entre un usuari LBS i un proveïdor LBS.

La majoria de les solucions proposades per fer front al problema de la privadesa de la localització són basades en TTP, és a dir, entitats que garanteixen plenament la privadesa dels seus usuaris. Encara que aquest enfocament és àmpliament acceptat, simplement es mou la confiança dels usuaris dels proveïdors dels LBS a unes entitats intermèdies. En aquest cas els proveïdors LBS ja no són conscients de les veritables identitats i ubicacions reals dels usuaris, la confiança i, per extensió, el poder passa en mans d'entitats intermèdies tal com *brokers*, pseudoanonimitzadors o anonimitzadors. El problema és que els usuaris no estan completament satisfets amb les entitats o proveïdors de confiança, ja que hi han hagut escàndols en els que s'ha divulgat una quantitat important de dades personals en aquest tipus d'entitats de confiança (vegeu la Figura 3.2).

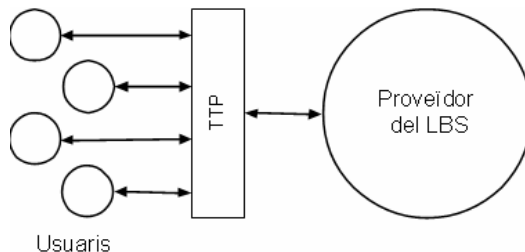


Figura 3.2. Esquema de comunicació entre un usuari LBS, una entitat intermèdia de confiança i un proveïdor LBS.

La principal diferència entre l'esquema de comunicació simple i l'esquema basat en TTP, és que en el segon el conjunt d'entitats intermèdies s'espera que sigui més petit que el nombre de proveïdors de serveis. Per tant, les entitats intermèdies poden ser molt conegudes i el risc de confiar en una entitat deshonest es redueix. Tanmateix, a causa dels escàndols que han hagut de divulgació de

dades personals, molts usuaris prefereixen no confiar en ningú, el que porta als esquemes sense TTP. Aquests representen un canvi substancial de paradigma (vegeu la Figura 3.3). En lloc de confiar en una tercera part, els usuaris col·laboren per a protegir la seva privacitat, sense la necessitat de confiar entre ells.

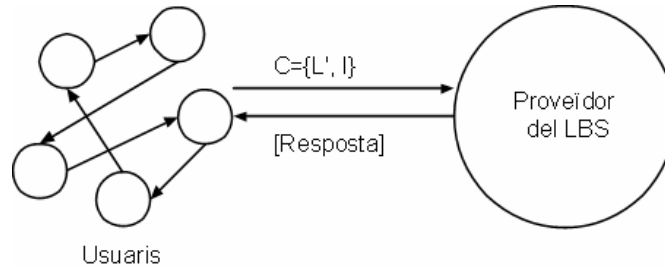


Figura 3.3. Esquema de comunicació entre un conjunt d'usuaris LBS col·laboratius i un proveïdor LBS de no confiança.

Moltes de les classificacions dels mètodes per a la privadesa en LBS proposades a la literatura utilitzen com a única variable l'ús o no d'una TTP. Per raons de claredat, s'ha classificat aquests mètodes d'acord amb dues variables: (i) la utilització o no d'una TTP i (ii) la pertorbació o no de la ubicació. A continuació, utilitzant aquestes dues variables s'ha classificat aquests mètodes en quatre categories:

1. Mètodes no pertorbatius basats amb TTP.
2. Mètodes pertorbatius basats amb TTP.
3. Mètodes no pertorbatius sense TTP.
4. Mètodes pertorbatius sense TTP.

3.1.2 Privadesa en sistemes basats amb TTP no pertorbatius

Els sistemes basats amb TTP no pertorbatius són molt comuns perquè són conceptualment simples, i generalment, ofereixen un raonable equilibri entre eficiència, privadesa i exactitud. Els sistemes que detallarem sobre aquesta categoria són: (i) els sistemes basats en polítiques, i (ii) els Pseudoanonimitzadors centralitzats.

3.1.2.1 Sistemes basats en polítiques

Els sistemes basats en polítiques van ser el primer pas per obtenir privadesa en LBS. Els proveïdors afegeixen un conjunt de polítiques de privadesa conegudes pels usuaris. Aquests sistemes permeten a l'usuari definir la seva pròpia política, és a dir la forma en que altres usuaris poden accedir a la

seva informació d'ubicació. L'usuari també pot definir el conjunt de llocs i la seva granularitat a partir de la qual aquesta informació es donarà a conèixer, per exemple, a més de retornar la ubicació dels usuaris a nivell d'habitació, retornar-la a nivell d'edifici. Si els proveïdors no segueixen adequadament les polítiques de privadesa, els usuaris tenen dret a demanar una indemnització i/o prendre accions legals contra els proveïdors.

Aquestes polítiques de privadesa serveixen com un acord contractual que especifiquen el que poden fer els proveïdors de serveis amb la informació personal dels usuaris. Les polítiques de privadesa són publicades pels proveïdors de serveis, i els usuaris decideixen si les accepten o no. Els usuaris arriben a un acord amb els proveïdors sobre quines dades poden ser recollides, per quin propòsit es poden utilitzar i com es poden distribuir a tercers. En aquest tipus de sistemes, la privadesa s'entén com la capacitat de les persones a decidir quan, què i com la informació sobre ells es divulga als altres. Idealment, els usuaris poden triar entre una varietat de polítiques.

El primer treball que utilitzava les polítiques de privadesa per protegir la privadesa espacial va ser presentat a [21]. El concepte clau en la metodologia proposada en aquest treball és l'observació de l'objecte a localitzar. Una observació típica inclou la ubicació, la identitat de l'objecte, el moment en el qual es va realitzar l'observació i la velocitat de l'objecte. La idea clau darrere d'aquesta metodologia és que una persona ha de ser capaç d'ajustar l'exactitud en que les observacions són alliberades segons uns paràmetres, com per exemple, la utilització prevista i la identitat del destinatari.

Aquests sistemes s'utilitzen molt a Internet, per exemple en comerç electrònic es defineixen les seves polítiques de privadesa, per exemple amb la Plataforma de Preferències de Privadesa (*Platform for Privacy Preferences*, P3P) [28]. És un protocol que permet als llocs Web declarar l'ús de la informació recopilada dels usuaris que la visiten. Un estudi recent sobre l'ús de polítiques i tècniques de control d'accés es pot trobar a [13]. En aquest treball es descriu la falta d'avaluació pràctica amb usuaris reals de tots els sistemes i llenguatges de control d'accés flexible. Els autors presenten un estudi amb un grup d'usuaris sobre les polítiques d'accés ideals desitjades per a la seguretat física en un entorn d'oficina.

3.1.2.2 Pseudoanonimitzadors centralitzats

Els Pseudoanonimitzadors centralitzats són una entitat intermèdia entre un usuari i un proveïdor LBS. La seva funció es ocultar la identitat real dels usuaris als proveïdors. L'usuari a més d'enviar directament les consultes al proveïdor LBS, les enviarà al Pseudoanonimitzador centralitzat que substituirà la verdadera identitat de l'usuari per una falsa (és a dir, un pseudònim). Un cop ocultada la identitat de l'usuari, s'enviarà la consulta al proveïdor LBS.

Utilitzant aquests tipus de sistema, ja no fa falta confiar directament amb el proveïdor del LBS, però si amb el Pseudoanonimitzador, ja que emmagatzema les identitats reals i els seus corresponents pseudònims amb la finalitat de donar les respostes dels proveïdors als usuaris.

El problema principal d'aquesta tècnica és que un atacant (per exemple, el mateix proveïdor del LBS) pot deduir la veritable identitat de l'usuari mitjançant la vinculació de la ubicació de l'usuari, per exemple, amb una guia telefònica pública (per exemple, utilitzant l'atac d'identificació en espai restringit (*Restricted Space Identification*, RSI) o l'atac d'identificació per observació (*Observation Identification*, OI) [11]. L'atac RSI consisteix en associar llocs a les identitats mitjançant l'ús de les consultes que es presenten en un espai restringit (per exemple, si un usuari envia les preguntes des d'una casa de les afores d'una ciutat, és fàcil vincular aquestes consultes a la seva identitat). De la mateixa manera, l'atac OI vincula les consultes a les identitats mitjançant l'observació (és a dir, l'atacant sap la ubicació de l'usuari ja que el veu en aquell moment) i correlaciona aquesta informació amb la ubicació que figura en les seves consultes.

3.1.3 Privadesa en sistemes basats amb TTP pertorbatius

Els mètodes basats amb TTP no pertorbatius no proporcionen el nivell de privadesa requerit per alguns usuaris. Per aquest motiu, es van proposar solucions que incrementen el nivell de privadesa pertorbant la ubicació real de l'usuari.

3.1.3.1 Anonimitzadors

En lloc d'ocupar-se de les polítiques o de les identifications dels usuaris, els Anonimitzadors assumeixen que les comunicacions són anònimes, és a dir, els proveïdors dels LBS no requereixen la identificació dels usuaris per respondre a les consultes. Els Anonimitzadors tenen la finalitat d'amagar la veritable identitat dels usuaris respecte la informació d'ubicació emesa. Una forma molt comuna d'ocultar la ubicació real dels usuaris del proveïdor del LBS és utilitzant la propietat de k-anonimat.

El k-anonimat és un enfocament interessant per afrontar el conflicte entre la pèrdua d'informació i el risc de divulgació, suggerida per Samarati i Sweeney [17, 18, 25, 26]. El k-anonimat és un concepte ben establert en la publicació de microdades (per exemple, als registres d'hospitals) que s'ha adaptat a la privadesa dels LBS.

En un sistema LBS, un missatge enviat per un usuari serà k-anònim respecte la seva informació temporal i espacial si és indistingible de la informació temporal i espacial d'altres $k - 1$ usuaris. La idea principal darrere dels k-anonimitzadors es reemplaçar la ubicació real de l'usuari per àrees d'encobriment (cloaking areas) en la qual es troben almenys k usuaris.

Els Anonimitzadors redueixen la resolució de la localització transformant la localització (x,y) en el temps t per $([x1,x2], [y1,y2], [t1,t2])$ on $([x1,x2], [y1,y2])$ és l'àrea en dos dimensions que conté (x,y) entre el temps $t1$ i $t2$ tal que t pertany al interval $[t1,t2]$. D'aquesta manera, els proveïdors dels LBS no poden determinar fàcilment quin dels k usuaris dins l'àrea d'encobriment és realment el que fa la consulta.

El risc de reidentificació és $1/k$, és a dir, la probabilitat d'identificar la font de la consulta és com a màxim $1/k$, fins i tot si l'atacant té el coneixement exacte de les ubicacions dels usuaris.

Molts exemples d'aquest tipus de propostes i d'altres de similars que estan basades en encobriment es poden trobar a la literatura [1, 8, 9, 11, 16]. Aquestes propostes són resistents contra atacs d'identificació com ara l'atac RSI i OI.

No obstant això, tenen alguns inconvenients importants, tals com: (i) la informació obtinguda no es precisa degut a la distorsió de la ubicació, i (ii) l'arquitectura és centralitzada, cosa que el fa vulnerable a atacs de denegació de servei (*Denial of Service*, DoS) i a la creació de colls d'ampolla.

A més, es podria donar el cas que un servidor intermediari en un moment determinat tingués una mostra reduïda d'usuaris actius (dins d'una àrea geogràfica concreta), això podria no ser suficient per satisfer el grau desitjat d'anonimitat d'un usuari que realitza una sol·licitud d'informació en una determinada regió. Si el servidor d'anonimitat retarda l'execució d'una sol·licitud fins que es compleixi la condició de k -anonimitat, llavors aquest retard pot arribar a ser inacceptable per l'usuari.

3.1.4 Privadesa en sistemes sense TTP no pertorbatius

Com hem vist a la secció 2, els mètodes que estan basats amb TTP tenen l'inconvenient de que es necessita confiar en una entitat intermèdia. A continuació veurem mètodes on no fa falta confiar amb una tercera part sense modificar la posició real de l'usuari.

3.1.4.1 Pseudoanonimitzadors distribuïts

Els Pseudoanonimitzadors distribuïts és una proposta que no està basada en TTP i no pertorba les ubicacions dels usuaris [14]. L'objectiu d'aquest mètode es preservar les avantatges dels mètodes pertorbatius tal com els Pseudoanonimitzadors centralitzats (és a dir, la senzillesa i l'exactitud), i evitar les desavantatges del mètodes basats amb TTP (és a dir, la baixa escalabilitat i la falta de privadesa). En aquest mètode es vol substituir el concepte clàssic de Pseudoanonimitzador, entès com una TTP, per un Pseudoanonimitzador distribuït que consisteix d'un conjunt d'usuaris col·laboratius. Els usuaris construeixen una cadena perquè els seus missatges segueixin un nombre de salts a l'atzar fins que el missatge s'envia a un proveïdor. En fer-ho, el proveïdor rep missatges que contenen les identifications dels usuaris i les seves ubicacions que no poden ser assignats, ja que corresponen a diferents usuaris.

Encara que aquest mètode representa una millora considerable respecte el Pseudoanonimitzadors centralitzats, no és resistent als atacs RSI i OI.

3.1.4.2 Dummy locations

Kido et al. [12] van proposar una tècnica per aconseguir la privadesa de la ubicació utilitzant localitzacions falses. La idea principal és que l'usuari envia un conjunt d'ubicacions falses anomenades *dummies* juntament amb la ubicació real al proveïdor del serveis. A continuació, el proveïdor de serveis processa totes les sol·licituds i envia totes les respostes a l'usuari. Llavors l'usuari tria la resposta que vol.

El desavantatge d'aquesta proposta és que el proveïdor de serveis gasta molts recursos de càlcul per processar consultes falses i un atacant podria detectar la ubicació real de l'usuari observant l'historial de les sol·licituds.

3.1.4.3 SpaceTwist

SpaceTwist [27] genera un àncora (és a dir, un punt fals) que s'utilitza per a recuperar la informació sobre els k punts d'interès més propers del proveïdor del LBS. L'àncora està inicialment definit en un ubicació aleatòria per l'usuari. Després de les successives consultes al proveïdor del LBS, SpaceTwist és capaç de determinar el punt d'interès més proper de la ubicació real del proveïdor del LBS, mentre que el proveïdor no pot obtenir la ubicació real de l'usuari.

El principal problema d'aquesta proposta és que requereix que l'usuari enviï diverses consultes al proveïdor i, a causa de la manca de col·laboració, aquest mètode no és capaç d'aconseguir k -anonimat.

3.1.4.4 PIR

L'any passat va aparèixer una nova proposta sobre privadesa en LBS basada amb la recuperació d'informació privada (*Private Information Retrieval*, PIR) introduïda a [10].

Els protocols PIR permeten a un client recuperar privadament la informació d'una base de dades, sense que el servidor aprengui la informació que el client ha sol·licitat. La majoria de tècniques són expressades des d'un punt de vista teòric, on la base de dades és una cadena binària X de n bits. El client vol trobar el valor de l'enèsim bit de X (p. e., X_i). Per preservar la privadesa, el client envia una petició xifrada $q(i)$ al servidor. El servidor respon amb un valor $r(X, q(i))$, que permet al client poder calcular X_i .

Encara que la idea d'utilitzar les tècniques de PIR per garantir la privadesa de la ubicació es prometedor, el sistema proposat requereix que el proveïdor del LBS cooperi amb els usuaris per complir el protocol PIR, això impedeix la utilització d'aquest mètode en entorns reals, en la qual els proveïdors simplement responen les consultes que contenen una localització o una àrea sense tenir en compte la privadesa de la localització.

Tanmateix, si aquest problema es resolts sense grans conseqüències de càlcul i rendiment, la utilització dels protocols PIR de col·laboració entre iguals (és a dir, els usuaris) podria ser realment prometedora per una futura línia de recerca.

Utilitzant un sistema PIR els usuaris estan protegits contra qualsevol atac, ja que la informació sobre la ubicació de l'usuari no és alliberada. Per tant, no es pot realitzar l'associació entre l'usuari i la informació sensible. Fins i tot en el cas que l'atacant sapigui la ubicació exacta de l'usuari (a partir d'una font externa). En aquest mètode es garanteix la privadesa en consultes contínues (és a dir, el seguiment dels usuaris).

3.1.5 Privadesa en sistemes sense TTP pertorbatius

En aquesta categoria es consideren totes les propostes que distorsionen la ubicació real dels usuaris sense utilitzar una TTP. Es divideixen aquestes propostes en dos grups: (i) mètodes col·laboratius, i (ii) mètodes individuals.

3.1.5.1 Mètodes col·laboratius

A [4], Domingo-Ferrer va proposar que els usuaris pertorbessin les seves ubicacions mitjançant soroll Gaussià. Els usuaris difonen les seves ubicacions pertorbades i demanen als veïns que tornin les versions pertorbades de les seves ubicacions. Després d'això, cada un dels usuaris seleccionen la ubicació de $k-1$ usuaris que estan dins d'una zona determinada. Finalment, cada usuari envia al LBS el centroid del grup de les k ubicacions pertorbades.

Aquest mètode no aconsegueix k -anonimat perquè el centroid només és utilitzat per un sol usuari per identificar-se. A més, a causa de la cancel·lació de soroll, els usuaris no poden utilitzar aquest mètode en diverses ocasions sense canviar-se d'ubicació.

Chi-Yin Chow et al. [3] van presentar una proposta P2P *peer-to-peer* similar pel que fa a la privadesa de la localització. El mètode proposat aprofita la cerca anònima P2P per construir les àrees d'encobriment. Els usuaris han de trobar a altres usuaris en el seu rang de cobertura, i compartir la seva informació sobre la ubicació. Una vegada que aquesta informació es coneix, els usuaris poden enviar les seves consultes als proveïdors dels LBS utilitzant l'àrea d'encobriment en comptes de la seva ubicació real.

Aquesta proposta té dos grans desavantatges. Primerament, la formació del grup i la selecció de l'agent podria ser un repte perquè no tots els dispositius mòbils es subscriuen al mateix proveïdor de serveis. Un altre inconvenient, és que els usuaris han de confiar amb els altres usuaris perquè ells intercanvien la seva ubicació real. Per tant, un usuari maliciós pot obtenir i publicar la ubicació d'altres usuaris.

A [24], els autors proposen un mètode basat en l'addició de soroll Gaussià per calcular una ubicació falsa que és comparteix amb els k usuaris. Per tant, tots els k usuaris tenen la mateixa ubicació falsa

i el proveïdor del LBS no és capaç de distingir un usuari de la resta, de manera que la seva ubicació esdevé k-anònima. Aquest mètode es va estendre per donar suport a les comunicacions no centralitzades [23].

3.1.5.2 Mètodes individuals

Aquestes mètodes, també es coneixen com mètodes d'ofuscació, són l'alternativa sense TTP als mètodes basats en la col·laboració. L'ofuscació pot ser entès com el procés de degradació de la qualitat de la informació sobre la ubicació dels usuaris, amb l'objectiu de protegir la privadesa de l'usuari [6].

A [5] es va presentar un mètode d'ofuscació basat amb la imprecisió. L'espai es modela com un graf on els vèrtexs són les ubicacions i les arestes indiquen l'adjacència. Per tant, amb la finalitat d'obtenir una ubicació imprecisa, l'usuari envia un conjunt de vèrtexs en lloc d'un únic vèrtex en la qual ell es troba ubicat. El proveïdor del LBS no pot distingir quin dels vèrtexs és el real. L'article proposa la negociació d'algorismes que permeten als usuaris augmentar la qualitat de servei (*Quality of Service*, QoS) mantenint la seva privadesa.

El principal problema d'aquesta tècnica és que els usuaris i els proveïdors han de compartir el graf que modela l'espai [7]. A [2] podeu trobar altres mètodes d'ofuscació proposats, on la ubicació real dels usuaris del LBS es substitueix per àrees circulars de centre i radi variable.

3.1.6 Anonimització segons l'origen de les dades

Avui en dia existeixen moltes tecnologies que permeten localitzar i fer el seguiment de les persones. Per exemple, per al correcte funcionament del telèfon mòbil, l'operadora requereix detectar constantment la proximitat del client a una antena específica (cel·la). Aquesta característica permet el seguiment dels telèfons mòbils.

A continuació veurem dos estudis on es recullen dades dels turistes per analitzar el seu comportament utilitzant diferents tecnologies de posicionament. Al primer estudi es va utilitzar el telèfon mòbil per a localitzar les ubicacions dels turistes i al segon el GPS.

Ahas et al. [15] van presentar un estudi geogràfic del turisme a Estònia utilitzant el telèfon mòbil per a obtenir les dades de posicionament dels turistes. Aquestes dades eren guardades automàticament en arxius dels operadors mòbils. Per a l'estudi utilitzaven la base de dades de les localitzacions de les trucades d'itinerància (telèfons estrangers). Per cada trucada s'emmagatzemava aquesta informació: la ubicació, l'hora, l'identificador aleatori i el país d'origen de la trucada de telèfon. Els resultats van demostrar que les dades de posicionament mòbil tenen valor per als estudis geogràfics.

Noam Shoval [20] va fer un estudi del comportament dels turistes a la ciutat de Akko (Israel). Aquesta ciutat va servir com un laboratori per provar la implementació d'eines de recopilació de

dades espacials. Els visitants que visiten aquesta ciutat, poques vegades hi passen la nit, la majoria trien passar un parell d'hores visitant llocs de la ciutat. Això no és molt bo per l'economia de la ciutat, però va resultar ser de gran avantatge pel que fa a la recollida de dades, ja que significava que era possible registrar la ubicació dels visitants que van accedir a prendre part de l'estudi en una molta alta freqüència (una vegada per segon) durant tota la seva breu visita a la ciutat. A cada participant se li va repartir un equip de localització (PDA + receptor GPS). El receptor GPS passiu registrava les posicions dels participants i al mateix instant, la informació es transferia a través de la tecnologia Bluetooth al PDA. Amb les dades obtingudes van analitzar de manera global el comportament espacial i temporal dels turistes, ignorant les dades individuals.

Amb relació la privadesa, en el primer estudi han utilitzat un identificador aleatori per "intentar" ocultar la identitat del turista. En el segon, la utilització d'un receptor GPS passiu. Però això no és suficient per garantir la privadesa dels turistes.

3.2 Bibliografia sobre privadesa en sistemes LBS

- [1] B. Bamba and L. Liu and P. Pesti and T. Wang. Supporting anonymous location queries in mobile environments with privacygrid. In: International World Wide Web Conference WWW. :237-246, 2008.
- [2] C. A. Ardagna and M. Cremonini and E. Damiani and P. Samarati. Location privacy protection through obfuscation-based techniques. In Baker, S., Ahn, G., eds.: Data and Applications Security. Volume 4602 of LNCS., IFIP:47-60, 2007.
- [3] C. Chow and M. F. Mokbel and X. Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based services. In: GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems. :171-178, 2008.
- [4] J. Domingo-Ferrer. Microaggregation for database and location privacy. In Etzion, O., Kuflick, T., Motro, A., eds.: Next Generation Information Technologies and Systems-NGITS. Volume 4032 of LNCS., Springer Berlin / Heidelberg:106-116, 2006.
- [5] M. Duckham and L. Kulit. A formal model of obfuscation and negotiation for location privacy. In: Pervasive Computing. Volume 3468 of LNCS., Springer Berlin / Heidelberg:152-170, 2005.
- [6] M. Duckham and L. Kulit. Location Privacy and Location-Aware Computing. Number 3. In: Dynamic and Mobile GIS: Investigating Changes in Space and Time CRC Press. :35-52, 2007.
- [7] M. Duckham and K. Mason and J. Stell and M. Workboys. A formal approach to imperfection in geographic information. Computers, Environment and Urban Systems. :89-103, 2001.
- [8] B. Gedik and L. Liu. A customizable k-anonymity model for protecting location privacy. In: Proceedings of the IEEE International conference on Distributed Computing Systems (ICDS'05). :620-629, 2005.

- [9] B. Gedik and L. Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*. :1-18, 2008.
- [10] G. Ghinita and P. Khoshgozaran Kalnis A. and C. Shahabi and K. Tan. Private queries in location based services: Anonymizers are not necessary. In: *SIGMOD'08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, Vancouver, BC, Canada, ACM*. pages 121-132, 2008.
- [11] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In: *Proceedings of MobiSys 2003: The First International Conference on Mobile Systems, Applications, and Services.*, San Francisco, CA, USA, USENIX Association, ACM, Sigmobile, ACM. :31-42, 2003.
- [12] H. Kido and Y. Yanagisawa and T. Satoh. An Anonymous Communication Technique using Dummies for Location-based Services. *Proceedings of IEEE International Conference on Pervasive Services*. 88-97, 2005.
- [13] L. Bauer and L. F. Cranor and R. W. Reeder and M. K. Reiter and K. Vaniea. A user study of policy creation in a flexible access-control system. In Czerwinski, M., Lund, A.M., Tan, D.S., eds.: *Proceedings of the 2008 Conference on Human Factors in Computing Systems, ACM*. pages 543-552, 2008.
- [14] P. A. Pérez and A. Solanas and A. Martínez-Ballesté. Location privacy through users' collaboration: A distributed pseudonymizer. In: *Manuscript to appear*. 2009.
- [15] R. Ahas and A. Aasa and A. Roose and U. Mark and S. Silm. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. 29:469-486, 2008.
- [16] R. Cheng and Y. Zhang and E. Bertino and S. Prabhakar. Preserving user location privacy in mobile data management infrastructures. In: *6th Workshop on Privacy Enhancing Technologies (PET)*. Volume 4258 of *Lecture Notes in Computer Science*:393-412, 2006.
- [17] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*. :1010-1027, 2001.
- [18] P. Samarati and L. Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, SRI Internacional. 1998.
- [19] The European Parliament and the Council: Directive 2002/58/EC on privacy and electronic communications. *Official Journal of the European Communities*:37-47, 2002.
- [20] Shoal N. Tracking technologies and urban analysis. 25:21-28, 2008.
- [21] E. Sneekenes. Concepts for personal location privacy policies. In: *ACM Conference on Electronic Commerce*. pages 48-57, 2001.

- [23] A. Solanas and A. Martínez-Ballesté. A TTP-free protocol for location privacy in location-based services. *Computer Communications*. :1181-1191, 2008.
- [24] A. Solanas and A. Martínez-Ballesté. Privacy protection in location-based services through a public-key privacy homomorphism. In: *Fourth European PKI Workshop: theory and practice*. Lecture Notes in Computer Science, Springer Berlin / Heidelberg. 362-368, 2007.
- [25] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems*. :557-570, 2002.
- [26] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems*. :571-588, 2002.
- [27] M. L. Yiu and C. S. Jensen and X. Huang and H. Lu. Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile service. In: *IEEE 24th International Conference on Data Engineering ICDE'08*. :366-375, 2008.
- [28] W3C: Platform for privacy preferences (P3P) project. Webpage (October 2007) <http://www.w3.org/P3P/>.

3.3 Privadesa en dades sobre trajectòries

Recentment estan sorgint estudis de privadesa sobre dades de espaciotemporals (trajectòries). Fins ara era difícil aconseguir aquest tipus de dades, però actualment, gràcies a les noves tecnologies, cada vegada hi ha més empreses i institucions que recullen aquest tipus d'informació seqüencial sobre les activitats dels seus clients. Per exemple, les companyies de telecomunicacions emmagatzemen dades espaciotemporals diàriament, aquestes seqüències contenen informació detallada sobre el comportament personal o vehicular, que pot permetre trobar patrons interessants per ser utilitzat en moltes aplicacions diferents, com el control de trànsit. De la mateixa manera, les persones naveguen per Internet. Aquesta és una altra font de gran potencial de les seqüències d'accions dels usuaris (per exemple, pàgines web visitades). L'estudi del comportament en Internet també pot conduir a aplicacions interessants, com la detecció d'intrusions. Hi ha altres àmbits que també produeixen seqüències temporals: les seqüències de proteïnes que descriuen la composició d'aminoàcids de les proteïnes i representen l'estructura i funció de les proteïnes, la informació genètica (ADN) que codifica la composició genètica, els registres de salut electrònics que emmagatzemen la història clínica dels pacients, etc

Tot i això, després de fer una cerca exhaustiva en les publicacions científiques, aquest àmbit encara és molt minoritari, i no s'ha aportat cap solució prou bona per protegir les dades de trajectòries. En aquest capítol es presentaran breument les diferents aproximacions que existeixen actualment.

Primer, reformularem la definició de protecció de privadesa en microdades, per aquest problema en concret.

3.3.1. Definició

L'objectiu de la protecció de privadesa en dades seqüencials consisteix en pertorbar les dades de les trajectòries per a prevenir l'associació d'individus a trajectòries. Aquests mètodes distorsionen la localització de les mostres en cada instant de temps.

Més formalment, podem dir que el problema d'anonimització d'una base dades de trajectòries T consisteix en trobar una modificació de les seqüències espaciotemporals, per a convertir-la en la base de dades segura T' que satisfaci els requeriments de privadesa com següents:

- No conèixer exactament seqüència en que s'han visitat els llocs.
- No identificar llocs on ha estat l'individu i que no vol que es coneguin (llocs sensibles).
- No poder usar la informació de les seqüències per re-identificar l'individu concret i així poder associar-la amb altra informació addicional de la base de dades que sigui sensible.
- Maximitzar la utilitat de les dades publicades T' , de forma que els anàlisis que es facin posteriorment en les trajectòries de T' arribin a les mateixes conclusions que si es fessin amb les dades originals de T .

L'exemple de la figura 3.4 mostra un fitxer amb dades de trajectòries T i la seva anonimització (a la dreta) T' .

id	trajectory
t_1	$a_1 \rightarrow b_1 \rightarrow a_2$
t_2	$a_1 \rightarrow b_1 \rightarrow a_2 \rightarrow b_3$
t_3	$a_1 \rightarrow b_2 \rightarrow a_2$
t_4	$a_1 \rightarrow a_2 \rightarrow b_2$
t_5	$a_1 \rightarrow a_3 \rightarrow b_1$
t_6	$a_3 \rightarrow b_1$
t_7	$a_3 \rightarrow b_2$
t_8	$a_3 \rightarrow b_2 \rightarrow b_3$

id	trajectory
t'_1	$a_1 \rightarrow b_1 \rightarrow a_2$
t'_2	$a_1 \rightarrow b_1 \rightarrow a_2$
t'_3	$a_1 \rightarrow b_2 \rightarrow a_2$
t'_4	$a_1 \rightarrow a_2 \rightarrow b_2$
t'_5	$a_3 \rightarrow b_1$
t'_6	$a_3 \rightarrow b_1$
t'_7	$a_3 \rightarrow b_2$
t'_8	$a_3 \rightarrow b_2$

Figura 3.4. Comparació de base de dades de Trajectòries, abans i després de l'anonimització[5].

Les tècniques de preservació de la privadesa en la publicació de trajectòries poden classificar-se en dues categories [1]:

Mètodes que publiquen mostres de localitzacions dels usuaris de forma independent i puntual. L'objectiu d'aquestes tècniques és prevenir a un atacant reconstruir trajectòries basades en obtenir conjunts d'aquestes localitzacions independents d'un mateix individu. Publicar mostres de localitzacions independents és usual en aplicacions que calculen informació agregada, tals com monitoratge de tràfic (per exemple, per saber quants individus passen per un cert lloc es pot demanar informació dels individus dels que es té informació en aquest lloc).

Mètodes que publiquen trajectòries individuals. En altres àmbits d'aplicació, els patrons de moviment i la relació de causalitat entre certes localitzacions d'origen i destinació són l'objecte

d'interès en l'estudi. En aquests casos és necessari publicar trajectòries completes, en comptes de mostres independents de localització.

3.3.2. Mètodes d'anonimització de trajectòries

Després de fer un estudi exhaustiu de la literatura, s'han trobat 4 tipus d'aproximacions al problema de la protecció de trajectòries:

- Confusió de trajectòries
- Ofuscació de localitzacions
- Modificació de projeccions
- Generalització

Cal destacar que la majoria de mètodes formulen el problema com un problema d'optimització no lineal amb restriccions. L'objectiu és maximitzar una certa funció de privadesa sota certes limitacions, com que la màxima distorsió per a cada localització publicada no excedeixi el llindar.

3.3.2.1 Confusió de trajectòries

Les tècniques de confusió de trajectòries són mecanismes que prevenen a un adversari seguir una trajectòria completa d'un usuari. La idea principal darrere aquestes tècniques consisteix en modificar les trajectòries dels usuaris o generar una o més trajectòries falses per a confondre l'adversari i impedir que esbrini les rutes verdaderes dels usuaris.

3.3.2.1.1. Generació de trajectòries falses

Una forma de confondre les trajectòries consisteix en barrejar la trajectòria real dels usuaris amb altres trajectòries "falsificades", generades per usuaris no reals anomenats *dummies* [10-11]. D'aquesta manera s'intenta reduir el risc d'exposar la trajectòria verdadera de l'usuari a un adversari. Per a la generació dels *dummies* hi ha tres metodologies diferents, que es basen en representar les trajectòries en un graf de localitzacions V unit per vèrtexs E . La posició inicial del *dummy* en aquest graf es tria a l'atzar dins d'un radi específic al voltant de la posició inicial de la trajectòria real del usuari.

En la primera metodologia, anomenada *random walk*, quan un *dummy* arriba a un node, la decisió sobre el vèrtex en el qual es mourà es realitza aleatòriament. El desavantatge d'aquesta metodologia és que els adversaris poden distingir fàcilment a l'usuari real dels *dummies*, ja que la trajectòria generada pel *dummy* no sembla natural.

En la segona metodologia, anomenada *direction control*, quan un *dummy* arriba a un node, la decisió sobre la nova localització es fa usant un conjunt de probabilitats predefinides, que intenten representar les adreces potencials a partir d'un cert punt. Per exemple, quan un usuari real arriba a una intersecció, ell o ella va sovint recte, a l'esquerra, o la dreta, però rarament torna enrere. En aquest mètode cada *dummy* es mou independentment de les trajectòries dels usuaris real.

La tercera metodologia, anomenada *collaborative direction control*, aborda aquesta qüestió fent que el moviment dels *dummies* i l'usuari real es creuin freqüentment entre ells. Aquest mètode amplia l'anterior agregant dues condicions: (a) un *dummy* intenta predir el node en el qual un usuari real es traslladarà, quan selecciona la seva pròpia direcció, i (b) un *dummy* esperarà en un node si detecta que l'usuari real s'acosta a aquest node.

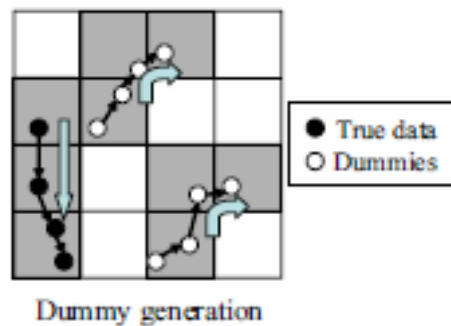


Figura 3.5. L'usuari genera 2 dummies que es poden moure en diferents direccions a partir de les dades de posició verdaderes [10]

3.3.2.1.2. Confusió de trajectòries reals

A [2] es presenta un algoritme de pertorbació de trajectòries que introdueix errors tolerables en els punts de la trajectòria original mantenint un cert nivell específic de qualitat en el servei. Concretament, El problema de confondre a l'atacant es pot formular com un problema de optimització no lineal amb restriccions.

Definim el grau de privadesa de la localització com l'exactitud amb la qual una font no fiable pot localitzar a un usuari en específic. Per a confondre a l'atacant, la ubicació de les mostres individuals és distorsionada, amb la finalitat de minimitzar la probabilitat d'identificar la ubicació precisa de l'usuari.

La idea principal del següent algoritme de privadesa és el concepte de confusió de trajectes. Cada vegada que el trajecte de dos usuaris es troba (es defineix trobar com estar prop un d'un altre) hi ha una probabilitat que l'adversari confongui els trajectes i segueixi a l'usuari equivocat. Un algoritme de privadesa pot explotar això pertorbant la informació de la localització en certes àrees de trobada per a incrementar les probabilitats de confusió.

Distorsionant les mostres de localització introdueix inexactitud de les dades, i pot tenir repercussions en l'anàlisi posterior de les dades. Per tal de poder trobar un compromís entre privadesa i inexactitud ha d'aconseguir-se, els autors de [2] proposen una mesura per a quantificar ambdues, privadesa i inexactitud. Específicament la privadesa és mesurada a través de la "Expectativa de l'error de distància", que captura que tan exacte un adversari pot igualar les localitzacions amb les seqüències. La inexactitud de les dades es mesura amb la "Qualitat del servei" (QoS).

L'ús d'aquest algoritme de pertorbació de seqüències s'il·lustra en un escenari simple on dos usuaris viatgen per trajectes aproximadament paral·lels (Figura 3.6).

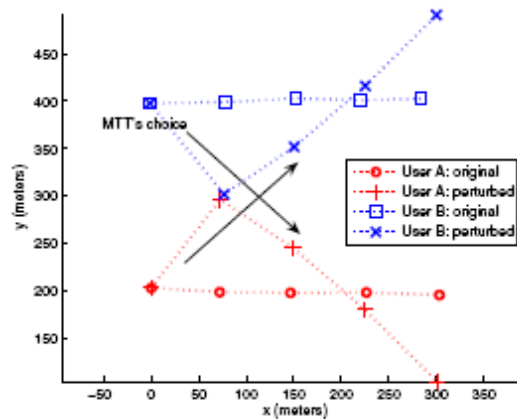


Figura 3.6. Dos usuaris movent-se en paral·lel, pertorbació d'un segment paral·lel a un segment creuat(Confusió de trajectòries) [2]

Els cercles vermells són localitzacions de la trajectòria de l'usuari A mentre que els quadrats blaus són punts de la trajectòria de l'usuari B. Les creus i les marques x són mostres pertorbades de l'usuari A i B respectivament. Ambdós usuaris es van moure d'esquerra a dreta començant amb prop de 200 metres de separació amb una velocitat horitzontal de 15m/s. El mètode, per tal de maximitzar l'expectativa de l'error de distància, converteix els trajectes paral·lels en trajectes que es creuen. Les fletxes en la imatge mostren el resultat d'aplicar l'algoritme de protecció, les noves seqüències tenen els punts més propers, per facilitar que un atacant confongui als dos usuaris i segueixi el camí incorrecte a partir del punt d'intersecció.

Un inconvenient d'aquest mètode és que la fase de pertorbació necessita considerar totes les permutacions de totes les ubicacions de les trajectòries, per tant el cost computacional és molt alt. Per a disminuir el temps de còmput, es pot realitzar una fase de segmentació de la seqüència abans de la pertorbació [2]. Donat que l'algoritme de pertorbació de trajectòries treballa millor en trajectes paral·lels curts, l'algoritme de segmentació intenta trobar aquests segments a priori.

Les limitacions d'aquest mètode són dues. D'una banda el resultat no és suficient per a protegir la privadesa dels usuaris en àrees poc denses. D'altra banda, la probabilitat d'èxit de l'atacant s'incrementa amb la longitud de la trajectòria divulgada, tot i les modificacions proposades.

3.3.2.2 Ofuscació de localitzacions

L'ofuscació es refereix a la pràctica on un individu pot degradar qualitat de la informació sobre la seva localització per a protegir la seva privadesa. A continuació es presenten 3 mètodes.

3.3.2.2.1 Ofuscació per imprecisió de coordenades

Segons [7] L'ofuscació es refereix a la pràctica on un individu pot degradar qualitat de la informació sobre la seva localització per a protegir el seu aïllament. La negociació, d'una banda, s'utilitza per a proporcionar la millor qualitat de servei, mentre que revela tant poca informació com sigui possible sobre la localització veritable de l'usuari. L'espai es representa com gràf on es modelen les localitzacions com un conjunt de vèrtexs V i la connectivitat entre els parells de localitzacions és representat per un conjunt de costats E . La localització d'un usuari es representa com un vèrtex $I \in V$. Una ofuscació de la localització d'un usuari es representa com un conjunt O de vèrtex, tals que $I \in O$. D'aquesta manera, el conjunt O proporciona una representació imprecisa de la localització d'un usuari, ja que un adversari no té cap informació sobre quin, entre els varis vèrtexs representats en O , correspon a la localització real de l'usuari.

3.3.2.2.2 Ofuscació amb translacions i votacions

En [4] es presenta un mètode de Anonimització adequat per a detecció de proximitat i separació. Pot ser classificat com Anonimització basada en ofuscació de dades, de manera que la anonimitat es defineix com "l'estat de ser no identificable dintre d'una conjunt d'objectes, el conjunt anònim". La meta és protegir la identitat dels objectius revelats per un proveïdor LBS i el proveïdor de localització (LP) respectivament. Aquesta aproximació està basada en comunicació pseudònima.

El mètode consisteix de dues parts:

- Primer totes les coordenades són transformades per una transformació global independent del temps, consisteix en una rotació seguit d'un trasllat.
- En el segon pas, el moviment local dels objectius és traslladat depenent del temps.

La motivació per al primer pas és la ofuscació de la referència global de les coordenades en fi d'evitar atacs basats a saber el parador d'un objectiu. Per al segon pas, els atacs basats en el coneixement del parador de l'objectiu, patrons de mobilitat i patrons de camí han de ser evitats.

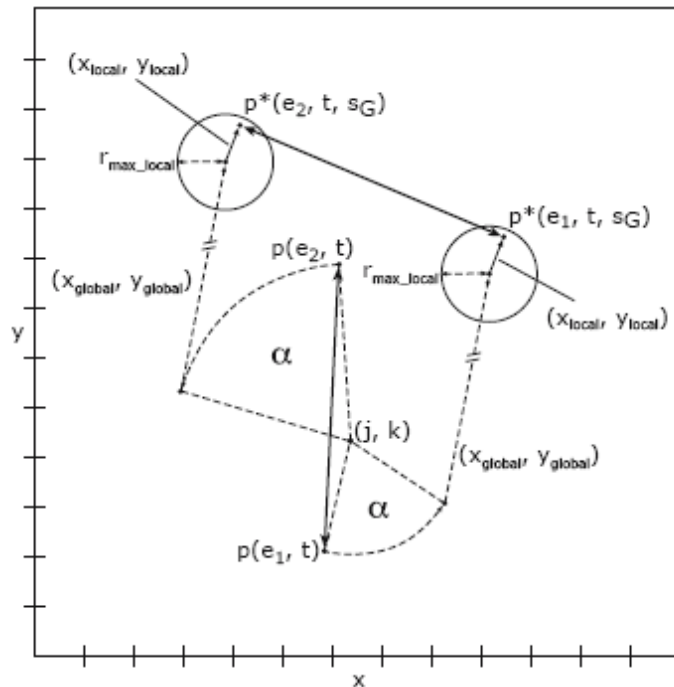


Figura 3.7. Distància preservant la ofuscatió de dos posicions $p(e_1, t)$ i $p(e_2, t)$ [4].

La figura 3.7 il·lustra aquest mètode. Primer, totes les coordenades són transformades per una transformació global independent del temps, consistint en una rotació amb angle α al voltant de $(j, k) \in \mathbb{R}^2$ seguit per la translació $(x_{global}, y_{global}) \in \mathbb{R}^2$. En un segon pas, el moviment local dels objectius és modificat afegint un vector de translació dependent del temps $v = (x_{local}, y_{local}) \in \mathbb{R}^2$. Els paràmetres α, j, k, x_{global} i y_{global} depenen d'una clau secreta entre el proveïdor de les dades i el client, i és per tant secret. El vector local v de l'ofuscatió també depèn de la clau secreta així com del temps actual.

3.3.2.2.3 Ofuscatió en la Minería de Dades per Trajectòries d'Objectes en Moviment

A [6] es proposen altres mètodes per modificar les dades aleatòriament o controladament. Per el cas de les trajectòries, l'atribut a discretitzar són les coordenades dels punts de la trajectòria. Les següents tècniques de pertorbació de dades han estat investigades:

- Value-Class Membership (també anomenat rounding o randomization)
- Value Distortion (també anomenat noise o adding noise)

En el primer mètode (*Value-Class Membership*) els valors d'un atribut són repartits en un conjunt de classes disjunctes exclusives mútuament. Un cas especial és la discretització on els valors dels atributs són repartits en intervals consecutius. No és necessari que tots els intervals siguin de la mateixa amplitud. Un cop s'ha fet la discretització, els valors d'un atribut es donen en intervals en el qual el valor original pertany.

El mètode *Value Distortion* retorna un valor $x_i + r$ en comptes de x_i , on r és un valor aleatori pertanyent a alguna distribució. És a dir, s'afegeix soroll a les dades numèriques. Es solen considerar les següents dues distribucions aleatòries:

- **Uniforme:** La variable aleatòria té una distribució uniforme, entre $[-\alpha, +\alpha]$ la mitja de la variable aleatòria és 0.
- **Gaussiana:** La variable aleatòria té una distribució normal, amb mitja $\mu=0$ i desviació estàndard σ .

Els exemples de la Figura 3.9 mostra els efectes de la distorsió per soroll uniforme i la Figura 3.10 per soroll Gausià.

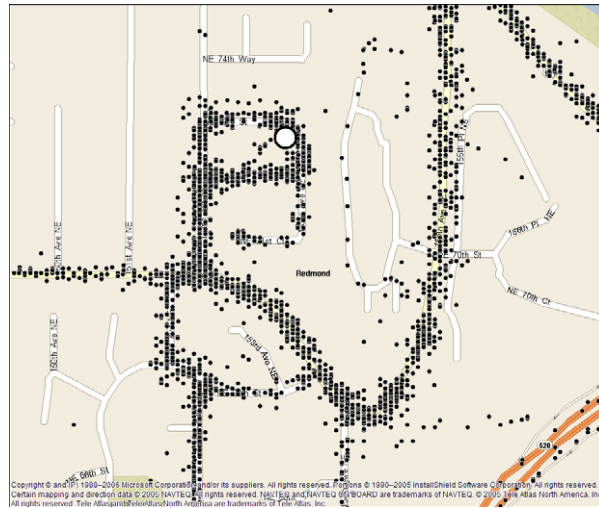


Figura 3.8. GPS no corromput [12]

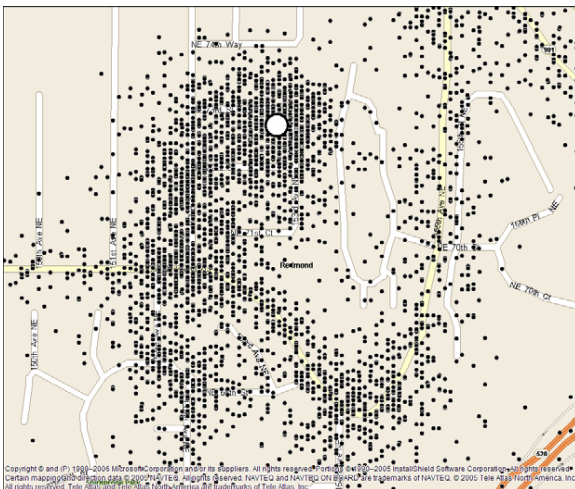


Figura 3.9. Value Distortion(Noise): afegit gaussian noise amb una desviació estandar de 50 metres [12].



Figura 3.10. Value-Class Membership(Rounding): cada punt es trenca al punt més proper en un una reixa de 50mx50m [12].

3.3.2.3 Anonimització per generalització

En privadesa es diu que un fitxer de dades compleix la propietat de k-anonimitat[15] si qualsevol adversari que conegui una clau quasi-identificadora, la relacioni com a mínim amb k registres (individus) del fitxer. D'aquesta manera, l'atacant no pot diferenciar entre els k registres i no pot re-identificar a cap individu concret.

Aquesta propietat es pot redefinir pel cas de les seqüències espacio-temporals. Assumint que els adversaris coneixen tots o alguns dels punts espacio-temporals d'un individu, aquest conjunt de tots els punts corresponents a la trajectòria poden ser usats com quasi-identificadors per tal d'intentar descobrir la resta de la seqüència. En aquest context, la propietat de k-anonimitat requereix que la porció de seqüència en el millor dels casos pugui ser vinculada almenys a k trajectòries del conjunt de dades anonimitzades.

En [3] es proposa que la privadesa de les trajectòries pot ser aconseguida aplicant el concepte de k-anonimitat mitjançant algorismes de generalització per agrupament. Aquests algorismes s'utilitzen també per taules de dades independents, tanmateix, una aproximació d'agrupació es mostra més adequada per a la anonimització d'estructures complexes, ja que treballen directament amb entitats conceptuals, com les trajectòries en aquest cas. Així doncs, els algorismes d'agrupament construeixen grups de trajectòries semblants, aleshores les trajectòries pertanyents al mateix grup són generalitzades de manera que siguin indistintes unes d'unes altres. L'agrupament de trajectòries implica la generalització de les coordenades temporals i espacials, i això requereix definir mètodes especials per aquest cas, com els que es presenten a continuació.

3.3.2.3.1 K-anonimitat per a bases de dades de trajectòries

En [3] s'estén el concepte de k-anonimitat a trajectòries sense confiar en la distorsió de dades i incertesa. En comptes d'això s'elimina la informació de les dades fent ús de l'espai i generalitzadors de temps, alineació de punts en espai i en temps, supressió de punts i trajectòries. La mètrica de cost utilitzada és obtinguda per estadística i captura el temps i la sensibilitat espacial per abordar diverses aplicacions. A més els treballs previs semblen no mesurar el nivell de distorsió a causa de la anonimització en el context d'aplicacions de mineria de trajectòries, la qual cosa es considera una de les metes principals en la publicació de trajectòries.

A [3] l'anonimització de trajectòries es realitza en tres etapes. Primer, l'algoritme selecciona les trajectòries que pertanyen a cada grup. Aquesta fase és realitzada a través d'una heurística que és similar a la comparació de cadenes de text. Després una fase de anonimització és realitzada, on es decideix quines trajectòries seran alineades amb mostres d'altres trajectòries. Per últim es generalitzen les trajectòries alineades, substituint les dades originals. A la Figura 3.11 es mostra un exemple de generalització de 3 trajectòries. Els punts alineats es substitueixen per zones (indicades en òvals).

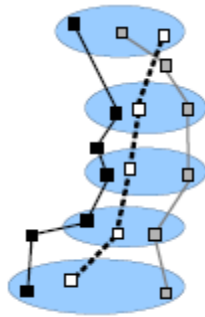


Figura 3.11. Aparellament de punts entre trajectòries [3]

L'algoritme d'agrupament de trajectòries funciona de la següent manera. Un cop els grups estan formats, es necessita especificar l'aparellament de punts òptim que minimitzi la pèrdua d'informació. A [3] es defineix una mesura de pèrdua d'informació anomenada LCM (Log Cost Metric) que quantifica la inexactitud de les trajectòries en espai i temps, i és calcula sumant l'ampliació (àrea) requerida sobre cadascun dels punts de localització publicats.

Trobar el punt òptim d'aparellament entre dues trajectòries és fàcil. El veritable repte és trobar el punt òptim d'alineament entre més de dues trajectòries. L'algoritme proposat a [3] usa funcions heurístiques per a tractar d'obtenir una possible alineació de punts. En cada pas troba l'òptim aparellament entre els punts d'una trajectòria no marcada i l'actual trajectòria anonimitzada. Les supressions de punts i les generalitzacions són aplicades d'acord amb l'aparellament, tal com es mostra a la Figura 3.12.

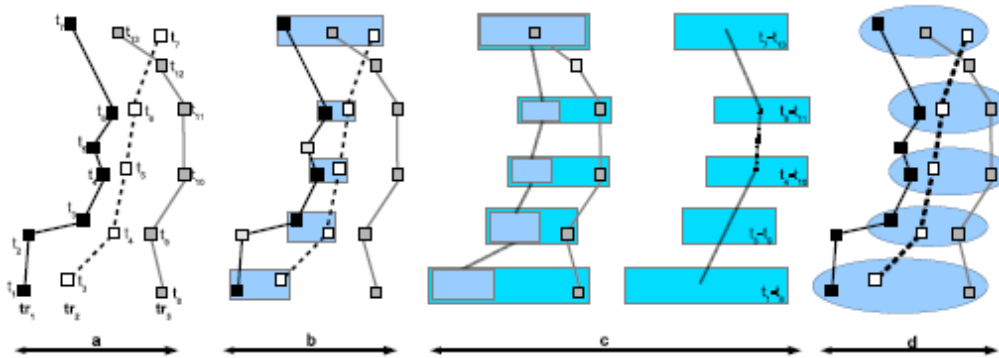


Figura 3.12. Procés d'anonimització [3]

3.3.2.3.2 Generalització per agrupació de seqüències de símbols

A [16] es proposa un mètode d'agrupació basat en les tècniques de microagregació que s'usen per dades no seqüencials. En concret es proposen nous algoritmes per les tres fases de l'anonimització per agrupament:

1. Seleccionar les trajectòries que pertanyen a cada grup: similitud OSS [11]
2. Creació dels grups: mètode KSHC [14]
3. Càlcul dels prototips: mètode OESP [13]

La funció de similitud OSS (Ordered-based Sequence Similarity) compara els símbols de dues seqüències tenint en compte el nombre de símbols comuns i no comuns, així com la diferència de posicions entre els elements comuns. A més símbols compartits i en posicions més similars, major serà la semblança entre les trajectòries.

Pel que fa a la creació de grups, l'algorisme que es proposa és una adaptació dels mètode tradicionals de classificació ascendent jeràrquica que s'utilitzen en mineria de [17]. Es proposa una adaptació d'aquests algorismes per al cas de microagregació, que asseguri obtenir grups de grandària limitada entre k i $2k-1$, de forma que compleixin la propietat de k -anonimitat. A aquest algorisme se l'ha denominat K-Sized Hierarchical Clustering (KSHC)[14].

Finalment es creen els centroides o prototips per cadascun dels grups obtinguts en el pas anterior. Aquests s'utilitzen per a emascarar les dades originals i generar una nova versió que no permeti la re-identificació dels individus. A [13] es proposa un mètode per a generar el prototip d'un conjunt seqüències de valors categòrics, denominat *Ordered Element Scoring Prototyping* (OESP). Aquest mètode es basa en el càlcul d'una taula de puntuacions per cada símbol que apareix a la seqüència, que dona punts diferents segons la posició en la que apareix el símbol. Els símbols que apareixen a l'inici de la seqüència reben més puntuació que els seus successors, de forma que els símbols que apareixen al final reben molt poca puntuació. Cada símbol va acumulant punts per cada seqüència del grup. Quan s'han comptabilitzat totes les puntuacions del grup, es procedeix a la construcció de la seqüència prototipus, que va agafant els símbols segons l'ordre decreixent de punts i els va concatenant fins arribar a la llargada desitjada.

3.3.2.3.3. Generalització segons el model del terreny

A [8] es presenta un mètode basat en rutes sobre un terreny per anonimitzar seqüències transformant les peticions originals en anònimes oferint k -anonimitat a la trajectòria. Les trajectòries considerades contenen coordenades (x,y) de l'individu. Es defineix el concepte de “ruta freqüent” com una ruta que apareix entre les trajectòries del mateix usuari un número de vegades que és més gran que el llindar mínim de freqüència. D'altra banda, es defineix com a “ruta insegura” una ruta d'un usuari u que és freqüent per u i infreqüent per un número definit pel sistema d'altres usuaris.

El mètode de generalització té dues fases. La primera fase és responsable de la computació de cada ruta insegura per cada usuari del sistema.

Per a derivar les rutes “insegures”, es cobreix una graella 3D sobre la superfície total coberta per les trajectòries. En aquesta graella s'estudia el moviment de l'usuari en un sistema de regions o de cel·les espaciotemporals. Cada trajectòria es representa a partir de les cel·les de la graella tal com es mostra a la figura 3.13 (dreta). Amb aquest nou model de cel·les es busquen les rutes insegures.

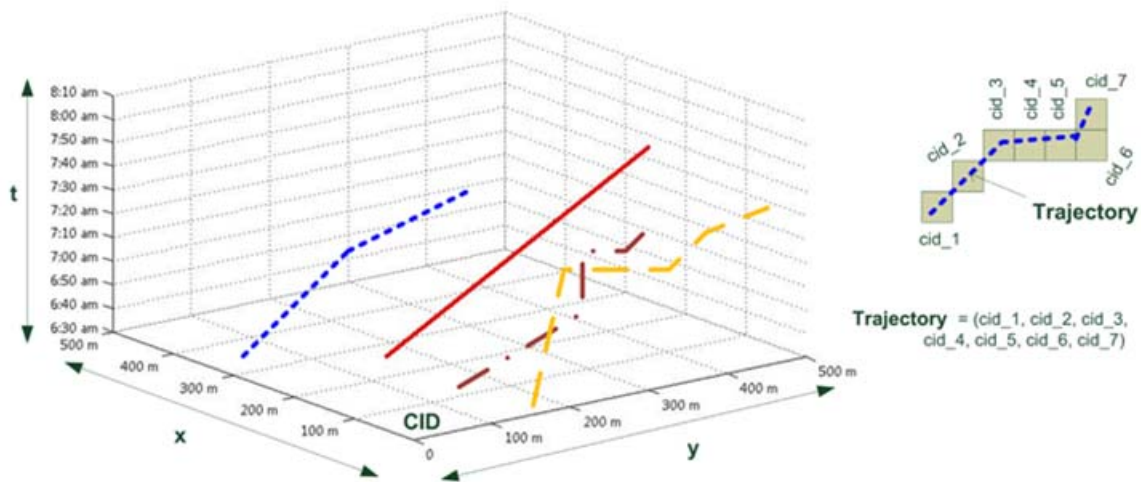


Figura 3.13. Free terrain [8]

La segona fase modifica les rutes insegures per obtenir una trajectòria segura. El primer pas examina el potencial de aparellament de la localització del sol·licitant respecte de les rutes “insegures”. Si existeix tal aparellament, es procedeix a anonimitzar la trajectòria per tal de que el sol·licitant no pugui re-identificar l’individu.

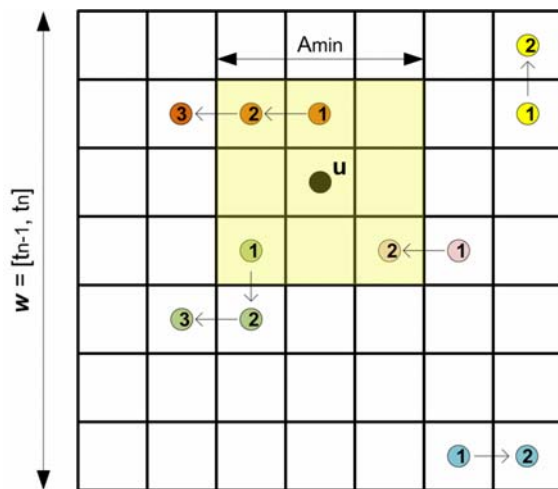


Figura 3.14. localització dels usuaris [8], on Amin és l’àrea mínima on els k-1 usuaris haurien de ser trobats

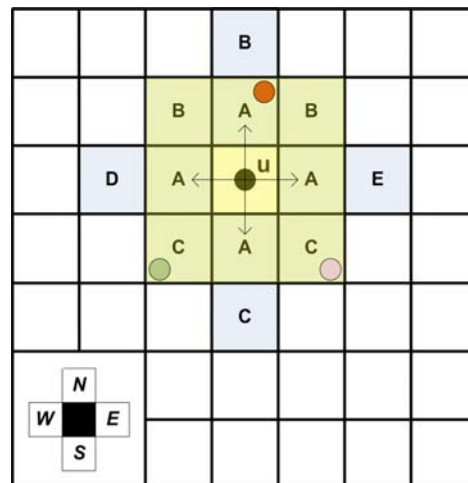


Figura 3.15. Estratègia de cerca.

La Figura 3.14 mostra la trajectòria de cinc usuaris durant un interval de temps i la cel·la demanada per un usuari u . Basant-se en l’exactitud de la localització que és necessària per al servei demanat, el sistema crea la corresponent projecció en una graella. Els nombres en els punts acolorits en la graella de l’esquerra representen les trajectòries. La graella de la dreta mostra estratègia de cerca de rutes insegures a partir de u . Inicialment es miren les caselles indicades per A, després les B, C, etc. Si es troben rutes insegures, s’anonimitzen.

Per a fer això, es segueix la següent estratègia:

A cada actualització d'ubicació, el servidor de confiança comprova si els veïns dels usuaris prèviament calculats han canviat de cel·la i per als que tenen, si la seva nova cel·la es troba dins de la regió d'anonimitat anteriors. Si és així, s'ajusta la informació temporal de la sol·licitud i es manté la mateixa àrea d'anonimat. Si alguns dels $K - 1$ usuaris han sortit de la regió, es busquen el mateix nombre d'usuaris entre els altres usuaris en el sistema. Si el nombre requerit d'usuaris que manquen no es pot trobar en aquesta regió, es torna a computar una nova regió.

3.3.2.4. *Modificació de projeccions*

Fins ara s'han presentat mètodes d'anonimització sobre trajectòries on els punts estan donats per coordenades (x,y) numèriques, tanmateix, també es poden obtenir seqüències de localitzacions on els punts vinguin representats per símbols, és a dir, on els punts siguin llocs identificats per una paraula (que pot ser el seu nom o algun tipus de codificació). En el treball [5] es proposa un mètode de protecció de la privadesa en aquest tipus de seqüències de símbols.

En concret, es considera que l'atacant ja coneix alguns fragments de la trajectòria i la identitat dels usuaris corresponents a aquests fragments, i vol descobrir la resta de la informació de la seqüència. Per exemple, una companyia P (publicador) comercialitza targetes de pagament que poden ser usades per a pagar el transport, així com per fer les compres del dia a dia. En un temps P (que és de confiança per a tots els usuaris de targetes) reunirà grans quantitats de dades de trajectòries. Suposem que hi ha una companyia de distribució A que té accés a totes les compres d'un cert usuari en els seus establiments, i que a més coneix la identitat de l'usuari degut a un servei de fidelitat de clients. Així doncs, si A té accés a un subconjunt de trajectòries seguides per l'usuari, podria mirar d'usar les dades públiques de P per identificar la resta de llocs on l'usuari ha usat la targeta. Aquest cas s'il·lustra en la Figura 3.16. En aquest exemple A pot identificar que u_1 correspon a la trajectòria t_1 ja que solament t_1 correspon al patró de moviment $a_1 - a_2 - a_3$ conegut per A. Per tant A, mirant les dades originals, pot inferir amb certesa que u_1 ha visitat b_1 que pot correspondre a un club nocturn. Una associació tan sensitiva és clarament una violació a la privadesa.

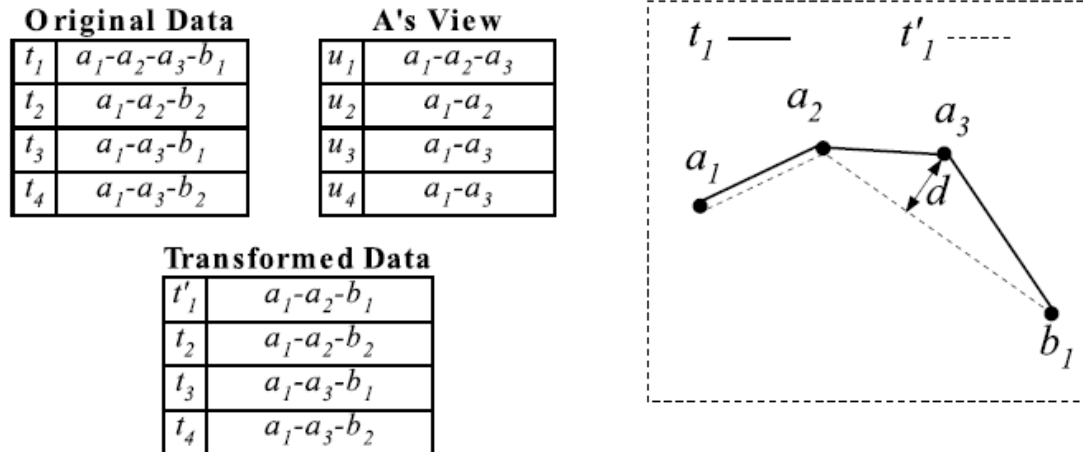


Figura 3.16. Anonimització de dos trajectòries, protegint dels intrusos que coneixen sub-seqüències de les trajectòries [4]

A [5] es proposa un algorisme d'anonimització basat en transformar seqüències llargues i detallades en seqüències més petites i simples. Per decidir quins punts de la trajectòria s'eliminen es miren les projeccions dels punts en un espai 2D, desviant-se de la trajectòria original, tal com es mostra a la dreta de la Figura 3.16. Fent això es dificulta als atacants inferir amb alta certesa si les trajectòries inclouen certs punts.

Cal esmentar que trobar el conjunt de punts òptim per a esborrar de T en ordre per a obtenir una nova versió segura T' i aconseguir la mínima perduda de la informació, és més complicat que assegurar la propietat de k-anonimitat en bases de dades relacionals.

3.4. Bibliografia sobre anonimització de trajectòries

[1] Ghinita Gabriel (2009), Private Queries and Trajectory Anonymization: a Dual Perspective on Location Privacy. Transactions on Data Privacy Vol.2, No. 1 3-19.

[2] Hoh Baik & Gruteser Marco (2005), *Protecting Location Privacy Through Path Confusion*, Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM 2005).

[3] Nergiz Mehmet., Atzori Maurizio, Saygin Yucel (2009) *Towards Trajectory Anonymization: a Generalization-Based Approach*. Transactions on Data Privacy Vol.2, No.1. 47-75.

[4] Ruppel Peter, Treu George, Küpper Axel, Linnhoff Claudia (2006), *Anonymous User Tracking for Location Based Community Services*. LoCA 2006, LNCS 3987,116-133, Springer-Verlag Berlin.

[5] Terrovitis Manolis, Mamoulis Nikos (2008), *Privacy Preservation in the Publication of trajectories*. The Ninth International Conference on Mobile Data Management.

- [6] R. Agrawal and S. Ramakrishnan. Privacy-preserving data mining. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 439-450, 2000
- [7] M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In Proceedings of the Third International Conference on Pervasive Computing (Pervasive'05), pp. 152-170, 2005.
- [8] Gkoulalas-Divanis, Aris, AND S.Verykios, Vassilios. A Free Terrain Model for Trajectory K-Anonymity 2008
- [9] H. Kido, Y. Yanagisawa, and T. Satoh. An anonymous communication technique using dummies for location-based services. In Proceedings of the Third International Conference on Pervasive Computing (Pervasive'05), pp. 88-97, 2005.
- [10] H. Kido, Y. Yanagisawa, and T. Satoh. Protection of location privacy using dummies for location-based services. In Proceedings of the 21st International Conference on Data Engineering (ICDE'05), pp. 118-122, 2005.
- [11] Valls, A., Nin, J., Torra, V.: On the use of Aggregation Operators for Location Privacy. IFSA-EUSFLAT Conference, (2009) 489-494
- [12] Krumm, J.: Inference attacks on location tracks. In: Anonymous : Pervasive Computing, Vol. Volume 4480/2007. Springer, Berlin / Heidelberg (2007) 127-143
- [13] Valls, A., Gómez-Alonso, C., Torra, V., Generation of prototypes for masking sequences of events, 3rd . Int. Workshop on Advances in Information Security (WAIS), Proc. 4th Int Conf on Availability, Reliability and Security (ARES), pp. 947-952, IEEE Computer Society, Fukuoka, Japan, March, 2009.
- [14] Valls, A., Gómez-Alonso, C., El algoritmo de microagregación KSHC para anonimización de secuencias de datos categóricos, Reporte de investigación de la Universitat Rovira i Virgili, DEIM-RR-09-001, 2009.
- [15] Sweeney, L.: K-Anonymity: A Model for Protecting Privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10 (2002) 557-570
- [16] Gómez-Alonso, Cristina. Protección de Datos Categóricos que siguen Patrones Secuenciales, Master Thesis in Artificial Intelligence, Escola Tècnica Superior d'Enginyeria, Universitat Rovira i Virgili, Tarragona, Spain, 2008.
- [17] Han, J., & Kamber, M.: Cluster analysis. In: Morgan Kaufmann (ed.): Data Mining: Concepts and Techniques. Second edition, Cerra, D., San Francisco (2006) 383-464.

4. Estudi de les tècniques d'anàlisi de dades per modelar el comportament de turistes

Aquest capítol està dividit en dues parts. Primer es fa una introducció als mètodes d'anàlisi intel·ligent de dades que permeten extreure coneixement estructurat a partir d'un conjunt de dades d'entrada. En aquest cas, la secció 4.1 es centra en l'estudi dels mètodes que existeixen per dades organitzades en forma de seqüències espacio-temporals (és a dir, trajectòries de moviments) i amb l'objectiu de generar perfils de trajectòries habituals. La segona part del capítol està dedicada a l'estudi de quin altre tipus d'informació es sol considerar quan es fan estudis sobre comportament mòbil dels turistes, concretament, s'ha tractat d'identificar els criteris no espacials que es consideren en els estudis.

4.1. Generació de perfils d'usuari amb dades espacio-temporals

En aquesta secció ens centrem en explicar les particularitats dels processos d'extracció de coneixement de dades espacio-temporals amb tècniques d'anàlisi intel·ligent de dades (*Intel·ligència Artificial*).

Des del punt de vista dels mètodes a aplicar, aquestes dades tenen unes característiques especials que no es troben en els conjunts de dades que habitualment s'estudien. En aquest cas, les dades són elements d'una seqüència i tenen una relació temporal entre ells.

Cal diferenciar els estudis de sèries temporals de la mineria de dades seqüències espacio-temporals. En l'anàlisi de sèries temporals l'instant precís en què es produeix cada acció és molt important, mentre que en les dades seqüencials l'instant no és el més important, sinó l'ordre en què es produeixen[31].

En l'àmbit de l'anàlisi de seqüències s'han fet alguns intents per estendre les tècniques clàssiques de mineria de dades a aquest tipus de dades. Tanmateix, gran part de l'esforç es centra en adaptar els algorismes existents però no s'ha proposat cap marc teòric específic per l'anàlisi de dades espacio-temporals, la qual cosa suposa un repte a nivell científic [31].

A continuació s'enunciaran diferents tipus de problemes que ens pot interessar abordar usant tècniques de mineria de dades sobre seqüències espacio-temporals. Després es presentaran els mètodes existents, distingint 3 tipus: mètodes per agrupar les seqüències segons la seva similitud, mètodes d'extracció de patrons (models) de trajectòries i mètodes de predicció de moviments.

4.1.1 Tipus de Problemes

Quan es parla d'objecte en moviment, o dades de trajectòries, es poden identificar diferents objectius de la cerca de coneixement en les dades, és a dir, de l'objectiu del procés de mineria de dades. Seguint l'exemple de [31], suposarem que tenim dades de les trajectòries de vehicles dins de la ciutat italiana de Pisa, i en base a això identificarem un conjunt tasques interessants.

4.1.1.1 Identificar patrons concrets

Cercar tots els embussos de trànsit a Pisa entre 7 i 9 del matí.

Els embussos de trànsit poden ser definits en termes de la densitat i la velocitat del trànsit, per tant, hi ha una clara relació amb els mètodes d'agrupació o classificació (*clustering*). Com que hi ha moltes formes de mesurar de similitud entre les trajectòries es poden trobar moltes variacions de l'agrupació són diferents criteris, això implica una dificultat addicional a l'hora de definir quina és la semàntica de la similitud de trajectòries en un cert context. També és important destacar que la identificació de patrons es pot fer sobre determinats fragments de les dades, en aquest cas cal algun mecanisme per identificar quin és el conjunt de dades d'interès.

4.1.1.2 Reconeixement d'una situació segons una variable

Trobar tres grups d'objectes que tenen velocitat similar (lent, mitjà i ràpid).

Les propietats físiques de les trajectòries, com la velocitat, l'acceleració i la longitud, es pot esperar que ocupin un paper important en el descobriment de coneixement en objectes en moviment. Típic dels objectes que es mouen és que tenen velocitat, i l'agrupació es pot dirigir de manera similar per detectar objectes en moviment ràpid. Així doncs, en aquest exemple es vol fer una agrupació dels vehicles segons la velocitat (p.ex. identificant cotxes, bicicletes i vianants).

4.1.1.3 Regles d'associació

En molts casos, els diferents embussos de trànsit estan relacionats temporalment. Relacions espaciotemporals entre els fenòmens poden expressar-se mitjançant regles d'associació, com per exemple.

Embús de trànsit (Pisa, 7.30) \Rightarrow embús (Lucca, 8:30)

Això significa que cada vegada que el primer esdeveniment es produeix (un embús de trànsit a Pisa a les 7.30 AM), generalment és seguit pel segon (un tràfic embús a Lucca a les 8.30 AM). Un versió més general d'aquesta regla podria ser:

Embús (Pisa, t) \Rightarrow embús (Lucca, t +1 h)

aquests casos sol ser interessant descobrir regles més generals, per exemple una generalització dels dos exemples anteriors seria la següent:

Embús (Pisa) \Rightarrow embús (Lucca)

4.1.1.4 Patrons de comportament

En el àmbit més general, es pot descobrir quins són els patrons de comportament més freqüents donat un conjunt de trajectòries. Per exemple, suposant que els esdeveniments associats les trajectòries s'identifiquen pel nom dels llocs on passen (per exemple, A, B, C, ...), ens pot interessar trobar patrons de transicions habituals com:

$A \rightarrow B \rightarrow C$

Aquest patró indica que A, B i C apareixen en aquest ordre temporal.

$A \rightarrow_3 B \rightarrow_7 C$

Aquest segon patró s'afegeix una informació temporal dels esdeveniments, on el subíndex i indica una demora temporal de i minuts.

Respecte als patrons de comportament, podem identificar patrons dins d'un cert interval de temps, o patrons que es repeteixen per períodes, per exemple cada dia, com els embussos de trànsit.

4.1.1.5 Tendència espaciotemporal

Una altra classe de problemes són els que busquen un model de tendència espaciotemporal. Un exemple d'una tendència és:

La velocitat dels objectes augmenta a mesura que s'allunyen de Pisa.

4.1.1.6 Predicció

Una darrera categoria de les tasques es refereix a l'extrapolació de trajectòries de dades. Un exemple és la pregunta:

Quines trajectòries creuaran Pisa matí a les 5:00 pm?

4.1.2 Mètodes d'agrupació (clustering)

Els mètodes d'agrupació (*clustering*) consisteixen en organitzar un conjunt de dades en grups diferents, de tal manera que els objectes en cada un dels grups siguin més semblants entre ells que no pas respecte als objectes d'altres grups. En aquesta secció, ens centrem en explicar els mètodes d'agrupament en el context d'objectes en moviment, per tant, mètodes que l'agrupació està basada en les trajectòries espaciotemporals que descriuen els seus moviments dels individus. En aquest context, l'agrupació essencialment intenta identificar grups de persones que mostren comportaments similars.

4.1.2.1 Classificació de tècniques de clustering

Existeixen molts algorismes de clustering diferents, però se solen classificar en dues tipologies [10]:

Mètodes Particionals: Donada una base de dades de n objectes o registres de dades, aquest mètode construeix una partició de les dades en k grups (*clústers*) diferents, tenint en compte la condició $k \leq n$. Els grups satisfan les següents condicions:

1. Cada grup ha de contenir com a mínim un objecte.
2. Cada objecte ha de pertànyer a un sol grup (aquesta condició no es compleix estrictament en algunes tècniques de partició difusa - *fuzzy*).

Donada k , el nombre de grups a construir, els mètodes de partició creen una partició inicial. Llavors, mitjançant una tècnica iterativa de reubicació van movent objectes d'un grup a un altre fins a trobar la partició millor. El criteri general per a determinar una partició correcte és que els objectes en el mateix clúster són "propers" o molt semblants, de la mateixa manera, els objectes de diferents clústers són "llunyans" o molt diferents.

Per arribar a obtenir una optimització global s'hauria d'estudiar l'enumeració exhaustiva de totes les possibles particions, com que això no és viable computacionalment, els mètodes intel·ligents usen funcions heurístiques per avaluar la bondat de les particions obtingudes. Un cop obtinguda la partició, es pot generar un representant (prototipus) del grup, per exemple:

1. L'algoritme *k-means*, cada clúster es representat per el valor mitjà dels objectes en el clúster.
2. L'algoritme *k-medoid*, cada clúster es representat per un dels objectes ubicat a prop del centre del clúster.

Mètodes Jeràrquics: Els mètodes jeràrquics creen una descomposició jeràrquica del conjunt de dades donat. Un mètode jeràrquic pot seguir diferents estratègies, identificant-se dues tipologies basades en com es forma la descomposició jeràrquica: algorismes aglomeració o de divisió.

L'enfocament d'aglomeració, comença amb cada objecte formant un grup separat, i successivament uneix els objectes o grups que estan més a prop un de l'altre, fins que tots els grups estiguin units en un (l'arrel de la jerarquia), o fins que hi hagi una condició de finalització.

L'enfocament de divisió, comença amb tots els objectes al mateix clúster. En cada iteració, un clúster es divideix en petits clústers, fins que cada objecte està en un clúster o fins que hi hagi una condició de finalització.

Els mètodes jeràrquics tenen l'inconvenient de que un cop un pas (unir o dividir) és fet, ja no es pot desfer. Aquesta rigidesa és útil ja que porta a un cost computacional petit ja que no s'ha de preocupar d'un nombre combinatori per a les possibles opcions. Tot i això, aquestes tècniques no poden solucionar decisions errònies. Hi ha 2 enfocaments per provar la qualitat del clustering jeràrquic [10]:

1. Realitzar un anàlisi acurat dels objectes "units" a cada partició jeràrquica.
2. Integrar aglomeració jeràrquica i altres enfocaments utilitzant primer l'algorisme d'aglomeració jeràrquic per agrupar objectes en *microclusters*, i llavors realitzar *macroclustering* en els *microclusters* utilitzant un altre mètode de clustering com ara reubicació iterativa.

Cal tenir en compte que l'algorisme de construcció dels grups determina les característiques finals d'aquests. Per exemple, algorismes basats en el centre com el k-means produiran grups esfèrics i compactes; mètodes jeràrquics organitzaran grups en una estructura multi-nivell de grups i sub-grups, mentre que altres enfocaments com els mètodes de classificació per densitat, formen grups segons la densitat dels objectes, per tant, poden no limitar la mida del grup.

Un cop presentades les aproximacions més habituals al problema de l'agrupament, cal estudiar si els mètodes tradicionals poden utilitzar-se en el cas de manejar dades de seqüències espaciotemporals. En la literatura es distingeixen dos enfocaments:

L'aplicació de mètodes tradicionals d'agrupació utilitzant una nova definició d'una mesura de distància entre les trajectòries. En aquest cas, la semàntica de la trajectòria de dades està completament encapsulada en funció de la distància.

Definir nous algorismes ad hoc adaptats al cas d'aquest tipus de dades espaciotemporals.

En la resta del capítol es presenten es donen detalls d'ambdós casos.

4.1.2.2 Agrupació tradicional basada en distàncies de trajectòries.

Aquest enfocament del problema d'agrupació es centra en la definició adequada d'una mesura de distància o similitud entre trajectòries. Un cop definida la manera de mesurar la distància es poden aplicar els mètodes tradicionals d'agrupament de dades simples.

Cal tenir en compte que d'aquesta manera, la definició de la funció de distància entre els objectes implícitament determina quins objectes que han de ser part del mateix grup i, per tant, determina també quin tipus de grups anem a descobrir.

Per identificar quins individus es mouen junts, un criteri bàsic per definir una distància és considerar els parells similars d'objectes que segueixen aproximadament la mateixa trajectòria espai-temporal, és a dir, en cada moment instantani estan aproximadament al mateix lloc [31].

Cal notar que hi pot haver trajectòries que segueixen similars camins però amb diferents velocitats. En la figura 4.1 els individus del primer grup comencen a avançar a una velocitat ràpida i van disminuint la velocitat durant el trajecte, mentre que els del segon grup comencen més lentament i després van accelerant. Una manera senzilla de modelar aquesta comparació és representar les trajectòries com a vectors de longitud fixa segons el nombre de punts rellevants del trajecte i després comparar aquests vectors per mitjà d'alguna mesura de distància estàndard usada en sèries de temps, com la Distància Euclidiana (la més comuna).

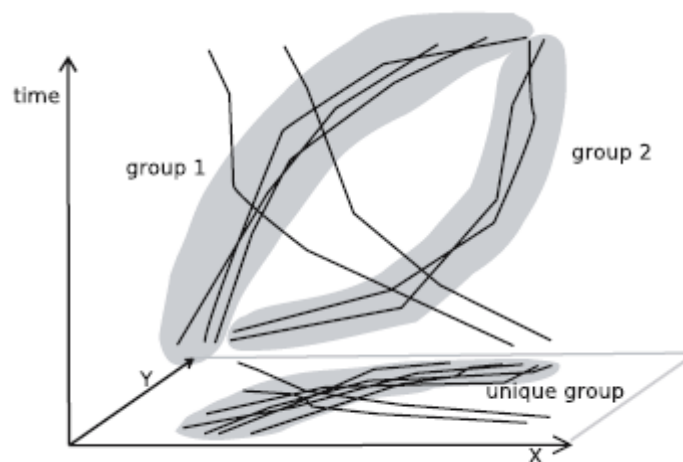


Figura 4.1. Trajectòries en coordenades espaciotemporals [31]

Una solució alternativa es dona en [23], on a cada instant de temps es calcula la distància espacial entre dos objectes i els resultats s'agreguen per obtenir la distància total, per exemple, calculant el valor mitjà, mínim o el màxim. Tanmateix, analitzar la seqüència per intervals iguals de temps a vegades és massa restrictiu per descobrir informació útil i, per tant, la limitació temporal es podria eliminar. En aquests casos, podríem veure els grups d'objectes que segueixen el mateix camí independentment de l'instant de temps. En aquest cas es fa una projecció de les dades en el component espacial, que podria servir per trobar grups de persones que es desplacen al llarg de les mateixos camins. Per exemple, vaixells seguint el mateix itinerari per creuar un mar, o cotxes seguint el mateix camí des del domicili fins al lloc de treball i de tornada, etc.

A la part inferior de la figura 4.1 , on les trajectòries són espacialment projectades en el pla XY es mostra un exemple d'això. En aquest exemple, el resultat és un únic grup d'objectes que segueixen el mateix camí, encara que les dades inicials indiquen diferents moments i velocitats.

Una variant d'aquest mètode es la que es presenta a [31]. Aquest mètode es basa en la comparació de parelles de sèries de temps, permetent (dinàmicament) transformacions no lineals de la informació temporal. D'aquesta manera l'ordre d'aparició dels llocs de la sèrie es manté, però possiblement comprimint / expandint els temps de moviment. Un altre mètode, proposat en [1], consisteix a calcular la distància com la "longitud de la menor sub-seqüència comú" (LCSS) de les dues sèries. En aquest sentit es pot treballar amb seqüències des d'un punt de vista d'edició de cadenes de caràcters.

Un altre aspecte en la flexibilització de les restriccions imposades a les agrupacions consisteix a no exigir una estricta co-ubicació de trajectòries i rutes, sinó només identificar grups d'objectes que realitzin moviments similars, com anar en la mateixa direcció o realitzar de la mateixa un circuit (és a dir, fixar-se en el moviment relatiu i no l'absolut). El primer exemple, pot ser modelat definint com semblant qualsevol parella d'objectes que segueixen aproximadament el mateix camí, permeten fer translacions espacial, tal com es proposa en [25]. Un pas més s'aconsegueix amb una distància que és també rotació-invariant.

D'altra banda, cal esmentar l'existència d'altres mètodes d'agrupació basades en definir nous criteris que descriguin a les dades espaciotemporals a partir de característiques extrems de la sèrie, en lloc de comparar les seqüències en si mateixes [27]. Per exemple, podríem extreure tots els parells de valors consecutius en cada sèrie (en el nostre context, llocs consecutius dins de cada trajectòria), i després comptar el nombre de parelles que comparteixen les dues sèries comparades, com es proposa en [1].

Finalment, s'hauria de tenir en compte la incertesa present en les dades de trajectòria, que generalment s'obtenen a través d'una sèrie d'observacions GPS de GSM. A [16] presenten una funció de distància que té en compte la incertesa utilitzant la noció de *prismes espaciotemporals*. Aquests prismes són una forma gràfica de representar les trajectòries en 3 dimensions (X,Y i temps). Així doncs, es calcula la distància entre les cadenes de prismes corresponents a dues trajectòries.

4.1.2.3 Agrupament basat de Trajectòries específiques

Els mètodes anteriors, basats en la funció de distància entre seqüències, imposen algunes limitacions a dos nivells: el nivell d'expressivitat i el nivell d'acompliment. Això és principalment a causa de la forta separació entre els criteris de similitud i l'esquema d'agrupament. Per exemple, qualsevol mètode basat en una idea de centre o, en general, representant d'un grup, ha de calcular els prototipus d'una manera que sigui coherent amb la funció de la distància adoptada, i això sovint no es té senzill.

L'exemple més destacat és l'algoritme K-means, on el representant és generalment calculat com l'objecte (possiblement nou, que encara no està presents en les dades d'entrada) que minimitza la

distància mitjana entre ell i tots els objectes de l'agrupació. En altres casos, el model de clúster demanat no es basa en cap idea de distància en absolut o, almenys, no tota la distància entre trajectòries. Segons[2] s'utilitza per agrupar les categories d'atributs en l'espai reduït. Aquest mètode no es limita a dues o tres dimensions.

En l'aproximació d'agrupament basat directament en la trajectòria completa, es plantegen algoritmes dissenyats especialment per aquest tipus de dades, enlloc de reaprofitar els tradicionals. Això és especialment interessant quan el model de clúster que es busca no es basa en cap idea de distància o, almenys, no només en la distància entre trajectòries. Segons[2] s'utilitza per agrupar segons un conjunt reduït dels atributs dels objectes, per exemple, quan la finalitat és l'anàlisi de la segmentació del mercat. Un altre punt fort d'aquest enfocament és que permet l'ús de mostres de mida gran, la qual cosa és molt interessant per a l'estudi de grans trajectòries. S'ha utilitzat amb força freqüència en la investigació de mercat. A continuació es presenten alguns mètodes utilitzats en la generació de grups de trajectòries.

En [21] un algoritme de clustering basat en densitats anomenat OPTICS s'utilitza per agrupar jeràrquicament punts en regions geogràfiques de manera divisiva. En comparació amb un mètode d'aglomeració com K-Means, l'enfocament basat en la densitat és capaç de detectar els clústers amb estructures irregulars, que pot representar a un conjunt de restaurants propers o punts de viatge.

Suposant que la desviació entre les trajectòries en un clúster és pot interpretar com soroll, a [6] es proposa una barreja d'un model basat en mètodes d'agrupació contínua de trajectòries, que agrupa els objectes que són probablement generats des d'un nucli comú de trajectòria afegint soroll Gaussià.

A [6] es proposa la utilització de models de Markov en què el representant d'un grup no és una trajectòria sinó un model de Markov que intenta explicar la transició entre una posició i la següent, generant posicions discretes a priori. Més específicament, els models ocults de Markov (HMMs) s'utilitzen per a modelar les agrupacions (clústers) i una barreja de model d'enfocament, i l'algorisme EM (Expectation-maximization algorithm)[32], en particular, és adoptada per el paràmetre d'estimació de tasques.

En[30] consideren informació local i global en dues etapes de l'agregació. En el cas d'aplicació al turisme, aquesta aproximació ens permet realitzar l'estudi sobre característiques de mobilitat a diferents nivells. Primer amb zones més precises i anar pujant fins obtenir una classificació a nivell més global. Aquestes tasques també es poden realitzar en paral·lel. A [30] utilitzen el "weka" que és un kit d'eines per mineria de dades per implementar aquest mètode.

En [29] també utilitzen el "weka" per a estudiar el comportament turístic a les Filipines. En concret han utilitzat l'algoritme EM (incorporat a l'eina weka) ja que accepta atributs tant numèrics com categòrics.

Un enfocament alternatiu es basa en la recerca de subsectors de trajectòries que encaixen prou. A [12], les trajectòries es representen com una peça de manera lineal, possiblement amb els segments que falten (per exemple, a causa de la desconexió d'un telèfon de la seva xarxa cel·lular). A continuació, es defineix un breu interval de temps com l'interval màxim, tal que tots els objectes són parells propers un de l'altre en aquest interval. Els grups de trajectòries tenen associat un pes que

expressa la proporció del temps en què les trajectòries estan a prop i, a continuació, el problema de la mineria és trobar tots els grups de trajectòria amb un pes més enllà d'un determinat llindar. A continuació, el mètode intenta descobrir la mida màxima i la màxima extensió temporal dels grups de segments de prop de la trajectòria. Des d'aquest punt de vista, un semblant però simplificat objectiu es persegueix en [24]. Aquí, una extensió d'una micro-agrupació per a objectes en moviment es proposa, que agrupa els segments rectilinis de les trajectòries dels que s'estableixen dins d'un rectangle de tamany determinat en un interval de temps. Fins i tot en aquest cas, la proximitat espacial es decideix a través de llinars (la mida del rectangle), mentre que la mida del grup i l'extensió temporal és maximitzada, en aquest cas la restricció a considerar només són els intervals de temps. Finalment, un enfocament diferent a un problema similar s'ha proposat recentment a [20]. Les trajectòries es representen com a seqüències de punts sense informació explícita temporal i una simplificació heurística s'aplica a cada partició de la trajectòria en un conjunt quasi-lineal de segments. Després, tots els segments s'agrupen per mitjà de la densitat d'un mètode basat en l'agrupació, i al final un representant de la trajectòria es calcula per a cada grup resultant.

Un equilibri entre els enfocaments basats en la distància i la trajectòria específica es produeix quan la distància no ha de comparar les trajectòries senceres sinó només algunes parts d'elles, per exemple, centrant-se en un interval de temps. A [31] es proposa un mètode d'agrupament genèric basat en densitats, on la distància adoptada és la mitjana de distància espacial entre les trajectòries dins d'un determinat interval de temps, que és un paràmetre de la distància. Llavors, per a cada interval de temps T , l'algorisme es pot executar centrant-se en segments de la trajectòria on els esdeveniments tenen lloc en el temps T . En aquest treball es proposa un mètode en què l'objectiu final és descobrir en quins intervals de temps els grups es diferencien millor. La figura 4.2 representa un conjunt de trajectòries que formen tres grups (a més d'una mica de soroll) i mostra l'interval de temps òptim (on els grups estiguin clars) com segments de la trajectòria més obscurs.

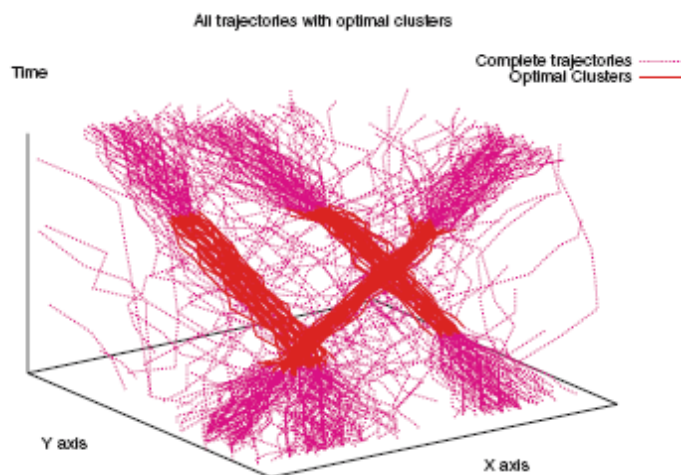


Figura 4.2. Agrupacions sobre un interval de temps [31]

Un objectiu similar es persegueix en [13], però des d'una perspectiva diferent. Aquí, els autors consideren els objectes en moviment com una associació amb una posició espacial d'un conjunt de línies de temps. El problema que afronten és el de descobrir agrupacions basades en la densitat espacial que aproximïn persistentment al llarg de certes línies de temps contigües, com es mostra en la figura 4.3 on una agrupació persistent en tres unitats de temps és trobada.

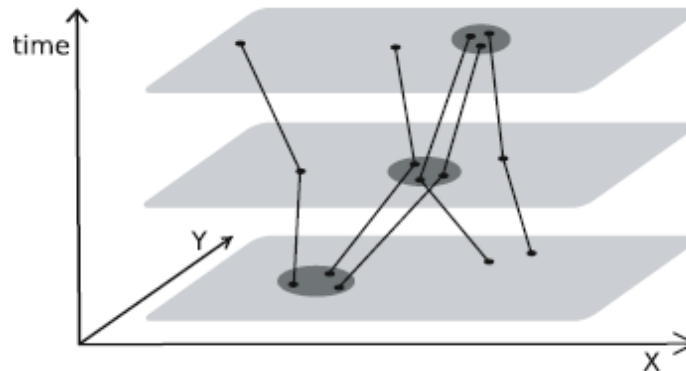


Figura 4.3. Una agrupació en moviment [31]

La persistència d'una agrupació significa que els individus continguts en un grup en una certa línia de temps són aproximadament els mateixos que apareixen en el grup en la següent línia de temps. Cal tenir en compte que la primera línia d'una agrupació en moviment i la última podrien compartir només alguns objectes (en el nostre exemple, només un), o fins i tot cap objecte en absolut, ja que gradualment i per etapes, les migracions en els grups són permeses, i cap verificació global es porta a terme. Amb aquest mètode, l'enfocament és a l'espai de les regions cobertes per grups i no en les trajectòries que contenen, per aquesta raó, aquest enfocament es pot considerar un cas límit entre l'agrupació (amb vista a trobar una classificació dels objectes) i la detecció de patrons freqüents (per trobar regularitats que impliquen a moltes persones). Així, les agrupacions identifiquen segments semblants de les trajectòries implicades i els patrons busquen les localitzacions espaciotemporals on les els segments es troben.

Tots els enfocaments descrits anteriorment comparteixen una propietat comuna: intenten cercar grups d'objectes que es mouen junts en relació amb diferents variants. Relaxant requisits temporals, podem buscar grups d'objectes només buscant en els camins que segueixen.

4.1.2.4 Classificació de Trajectòries en Grups Pre-Definits

A partir del nostre estudi de la literatura, no hem trobat cap mètode per classificar trajectòries en grups ja definits. No obstant això, hi ha problemes similars en l'àrea d'anàlisi de sèries temporals. A [15] utilitzen la representació de sèries de forma lineal segons el factor temps i associen un pes cada

un dels segments en funció de la seva importància. A [7], les sèries de temps es classifiquen a partir de l'aplicació d'unes pautes com a criteris de prova en arbres de decisió. Cada patró del que correspon a un model temporal porta associat un valor que pot representar un concepte relacionat, com per exemple, la velocitat d'un objecte. En general, les trajectòries es poden classificar usant el veí més proper utilitzant algorismes sempre que una adequada funció de la distància sigui donada. No obstant això, la definició d'una funció de distància depèn de la tasca de classificació, com s'ha vist també en els mètodes d'agrupació presentats en aquesta secció. A més, no és fàcil determinar l'escala, el trasllat i els efectes de soroll, elements que s'han de tenir en compte.

4.1.3 Patrons Locals Espaciotemporals

El segon tipus de mètodes de mineria de dades que es presenten són els que tenen com a objectiu buscar patrons de comportaments interessants dels objectes en moviment o grups d'objectes. Els patrons són representacions concises dels grups d'objectes. En particular, en aquesta secció estem interessats en la mineria de patrons locals, és a dir, els patrons que tenen per objecte caracteritzar petites porcions de les dades espacials, com ara sub-conjunts d'individus, petits intervals de temps limitat o de regions de l'espai.

El tipus de patrons de comportaments interessants que volem descobrir depèn estrictament del context. En general, el context s'especifica per la selecció d'un sub-conjunt de tots els patrons possibles i, possiblement, l'aplicació de certes restriccions sobre com es produeixen els patrons en les dades.

El problema de cercar patrons es pot plantejar de dues maneres diferents. Mitjançant una cerca directa, on podem especificar un espai de cerca molt gran i tractar d'identificar tots els patrons freqüents en aquest espai. Alternativament, es pot especificar un conjunt petit de patrons interessants (o hipòtesis) amb antelació, i trobar totes les aparicions que coincideixin amb aquests patrons inicials en les dades. Una característica addicional d'aquests dos tipus de problemes és que els patrons en cerques directes són generalment (però no necessàriament sempre) molt simples i inclouen individus simples, a més d'algunes restriccions sobre el nombre de successos. Per contra, les cerques inverses en els patrons són generalment bastant complexes i tenen limitacions en la mida o la composició de les seqüències.

Els algorismes d'anàlisi de dades aplicats sempre dependran de les dades objecte de la cerca. La informació en els patrons pot incloure conjunts o seqüències d'esdeveniments (com els llocs visitats per la trajectòria), o esdeveniments espaciotemporals (com ara les maniobres realitzades: Voltes a U, aturar-se, acceleracions extremes, etc.). En alguns casos, podem derivar informació que descriu a la població, per exemple embussos de trànsit a la ciutat. També es poden trobar associacions entre els patrons, per exemple podem descobrir si un embús de trànsit aquí hi ha normalment seguit d'un accident de trànsit més endavant, o si algú que està aturat en el trànsit, normalment, després té un accident. A continuació es donen més detalls dels mètodes de cerca de patrons.

4.1.3.1 Extraient patrons freqüents

Els patrons freqüents són un element bàsic de la mineria de dades, una aproximació simple i molt comú en la mineria de patrons freqüents espaciotemporals consisteix en buscar patrons de freqüència en base a l'extracció de característiques rellevants o en l'anàlisi estadística de certes característiques. L'alternativa oposada davant de l'enfocament basat en extracció de característiques consisteix en l'anàlisi directe de trajectòries, per exemple, per descobrir els camins sovint seguits pels turistes al centre de la ciutat, les maniobres més freqüents dels habitants de la ciutat.

4.1.3.1.1 Extracció de característiques rellevants

En primer lloc, es deriven conjunts de característiques de les dades, generant esdeveniments, valors d'atributs o, més en general, predicats espaciotemporals que descriuen cada trajectòria. A continuació, els algorismes de mineria de dades tradicionals s'apliquen sobre la nova representació de les dades, com per exemple l'aprenentatge de regles d'associació, la detecció de la freqüència de les seqüències de característiques. Amb aquesta aproximació es poden aprendre regles de la forma:

Longitud (trajectòria) > 50 quilòmetres \Rightarrow velocitat mitjana (trajectòria) > 60 quilòmetres

Cal tenir en compte que seguint aquest enfocament, la informació de les dades espaciotemporals queda resumida en les característiques extretes durant l'etapa de tractament previ i, per tant, ja no es pot usar en la fase de la mineria. Òbviament la selecció dels atributs (característiques) a extreure és un aspecte crucial en el procés d'aprenentatge, atès que defineix el patró a cercar. Una família bàsica de característiques per trajectòries d'objectes en moviment consisteix en característiques dels individus, aquelles que descriuen el comportament de cada objecte separat dels altres, per exemple podem tenir:

- Agregats espacials i / o temporals (la longitud del trajecte cobert, la quantitat de temps gastat en el centre de la ciutat, el mínim i el màxim mitjana de velocitat, la direcció més freqüent seguida, etc.)
- Esdeveniments espacials (visitant algunes regions predefinides o visitant dues vegades el mateix lloc).
- Esdeveniments espaciotemporals (maniobres localitzades temporalment com realitzar una volta en U, parades sobtades, acceleracions inesperades o comportaments de llarga durada com cobrir cert segment del camí en algun moment i cobrir-lo novament més tard en la direcció oposada).

Això combina esdeveniments espacials (visitar una regió donada) amb comportaments temporals simples (parades i voltes en U).

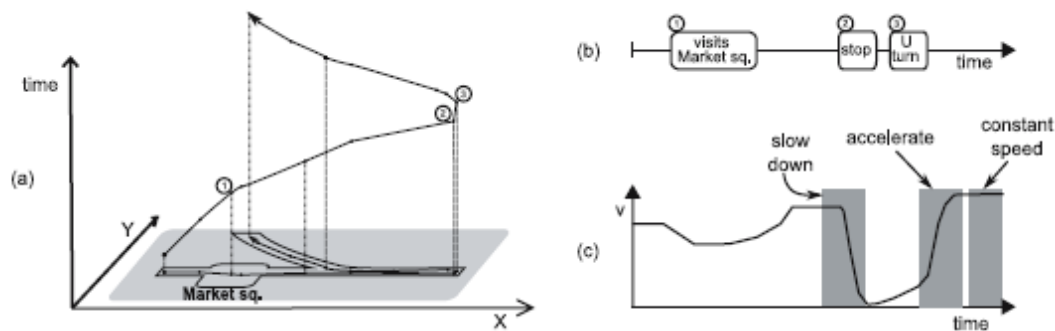


Figura 4.4. Simple trajectòria (a), Esdeveniments derivats espai temporal (b), Variació de la velocitat (c) [31]

En alguns contextos, es requereixen característiques més complexes per descriure el comportament de grups en moviment. Per exemple característiques que incloguin incertesa espacial (per exemple, les localitzacions no són punts sinó límits circulars que contenen la posició real) o temporal (en un interval de temps).

En [29] utilitzen la funció de tabulació creuada aportada per l'eina STATICA¹, aquesta funció és una tècnica d'anàlisi estadístic per a taules de freqüència multivariant, que comprova els efectes de creuar diverses variables. En [28] es presenta un algorisme per a descobrir patrons temporals híbrids de seqüències d'esdeveniments híbrides. L'algorisme anomenat HTPM (Hybrid Temporal Pattern Mining) descobreix patrons d'esdeveniments puntuals, d'esdeveniments basats en intervals o ambdós.

Encara que els patrons de seqüències poden dir quins ítems han d'anar junts i en quin ordre, ens pot interessar també que tinguin informació sobre l'interval de temps entre els ítems. En [11] defineixen una nova variant de patrons seqüencials d'interval de temps, la qual pot revelar els intervals de temps entre un parell d'ítems en un patró, tal com es mostra a les figures 4.5, 4.6 i 4.7.

Patró de seqüències tradicional:



Figura 4.5

¹ <http://www.statsoft.com/>

Patrò seqüencial d'interval de temps simple:

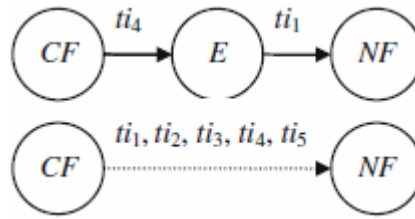


Figura 4.6

Patrò seqüencial d'interval de temps múltiple:

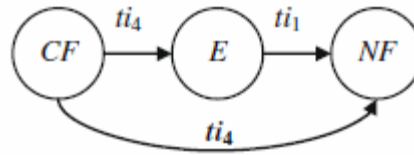


Figura 4.7

4.1.3.1.2 Exploració espacial estadística

Els mètodes d'anàlisi estadístic també es poden usar per extreure característiques de les seqüències espaciotemporals. Per exemple es pot definir una característica basada en la concurrència de diversos esdeveniments simples en el mateix lloc al mateix temps. A [17] s'utilitza l'exploració espacial estadística per cercar regions concorregudes en d'espaitemps, concretament cerca cilindres espaciotemporals (és a dir, regions circulars dins d'un interval de temps) on la densitat d'esdeveniments és més alta que fora. El model pot ser ampliat afegint informació temporal a les característiques extretes, per exemple la data i hora dels esdeveniments espaciotemporals. També es poden extreure atributs dinàmics, és a dir, els atributs que tinguin valors dependents del temps. Finalment, els segells de temps permeten extreure patrons més detallats que descriuen les relacions temporals entre esdeveniments. Per exemple, seqüències amb els temps de transició entre les característiques, esdeveniments consecutius, o informació de les limitacions temporals entre esdeveniments.

A partir d'aquest tipus de característiques, de manera similar a la presentada anteriorment, es poden obtenir automàticament regles d'associació entre els esdeveniments de la sèrie.

4.1.3.1.3 Anàlisi directe de trajectòries

En aquest tipus de mètodes d'obtenció de patrons o regles, no hi ha una discretització prèvia ni qualsevol altra forma de tractament previ de la informació espaciotemporal. Per tant, la informació que hi ha en les dades espaciotemporals s'utilitza durant tot el procés de mineria de dades. Una primera conseqüència d'aquest escenari és que el concepte de patró freqüent que s'utilitza habitualment, és a dir, un patró que exactament passa diverses vegades en les dades, en general no es pot aplicar. De fet, si les dades de les coordenades i del temps són numèriques, la continuïtat de l'espai i del temps fa que sigui gairebé impossible trobar una configuració que ocorri més d'una vegada a la perfecció de la mateixa manera. La solució consisteix en admetre algun tipus de tolerància a les petites perturbacions de la posició i el temps.

Aquest problema de la continuïtat pot ser abordat almenys de dues formes complementàries:

Tenint en compte les pautes que es presenten en segments de la trajectòria, considerant una certa incertesa en l'aproximació de les dades.

Tenint en compte els patrons que es troben en la forma de moure les regions dels intervals de temps, com els cilindres o tubs espaciotemporals que, d'alguna manera, representen un segment d'una trajectòria amb una certa aproximació o incertesa. Aleshores, els successos de la trajectòria són segments plenament continguts en el moviment de regions.

El treball en [5] proporciona un exemple del primer cas: s'aproxima una trajectòria per mitjà d'una successió de segments espacials obtinguts a través d'un procés de simplificació de les dades. A continuació els patrons són extrets en forma de seqüències de segments contigus, en particular, cada element de la seqüència ha de ser semblant a diversos segments de la trajectòria d'entrada. La similitud entre segments es defineix en base tres paràmetres: la proximitat espacial, la longitud i angle. Els patrons es construeixen com una successió de rectangles de tal manera que la seva amplada quantifica la distància mitjana entre cada segment i els punts de la trajectòria que abasta. La figura 4.8 mostra un model simple d'aquest tipus, format per dos segments i els seus rectangles corresponents.

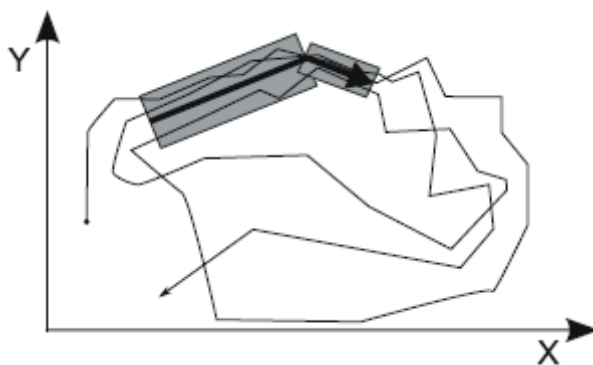


Figura 4.8. Exemple bàsic de patrons espai-temporals [31]

El segon enfocament, basat en les regions en moviment, és seguit per [13]. En aquest cas s'estudia el descobriment de la densitat espacial basat en grups d'individus que persisteix al llarg de diverses fraccions de temps. D'una forma similar, però centrada en models cíclics, es poden definir patrons espacio-temporal periòdics, [13] i proposa un algoritme ràpid i eficient per a la recuperació de patrons diaris. En aquest mètode el temps es suposa que és discret i les localitzacions espacials són discretitzades dinàmicament cada vegada que identifica un patró periòdic en el que totes les regions en el patró són denses i, per tant, significatives.

Algunes variants dels models esmentats s'han proposat i estudiat en els darrers anys. Un exemple és el manegament de la incertesa en la localització dels objectes des d'un punt de vista probabilístic, de forma que la posició de cada un dels objectes té associada una probabilitat de distribució sobre l'espai, al voltant del punt de localització donat [31]. Aquí, els patrons candidats són construïts sobre una discretització pre-definida de l'espai i temps, i el suport d'un patró es calcula a partir de la distribució esperada dels objectes d'entrada en aquesta graella.

4.1.3.2 Recuperació d'ocurrència

Contrari a l'extracció de patrons freqüents de les dades, un usuari pot ja tenir alguns patrons específics en ment i preguntar per totes les ocurrències d'aquests en un determinat conjunt de dades. Hi ha dos tipus de consultes que poden distingir: elementals i sinòptiques. Mentre que les consultes elementals representen el comportament del moviment d'entitats individuals, les consultes sinòptiques representen patrons del comportament de moviments col·lectius.

Les consultes elementals són sobre patrons que poden ser resolts analitzant les trajectòries de forma individual, per exemple la consulta:

Trobar totes les trajectòries que passin per la ubicació A en els temps t_1 i t_2

Això pot generar moltes trajectòries, però cada trajectòria per si mateixa és suficient per decidir si el patró s'ha complert o no. En la consulta anterior, la localització és especificada explícitament mentre que la constant temporal correspon a un rang de cerca. Noteu que el model no implica la informació seqüencial. Per incloure informació seqüencial a la consulta, es podria demanar que després d'una ubicació A l'individu hagi anat a B . A [9] aquest tipus de consulta que es denomina patró espaciotemporal de consulta (STP) i es defineix com una seqüència ordenada de punts del territori, ja sigui considerant l'ordre temporal absolut o relatiu entre els esdeveniments.

Les consultes sinòpticament inverses identifiquen objectes que conformen un específic comportament col·lectiu. Aquests patrons tenen com a objectiu estudiar moviments simultanis i també la interacció entre els objectes. Es refereixen a grup de patrons i poden incloure informació derivada relativa a tot el grup d'objectes (per exemple, la velocitat mitjana). Intuïtivament, un grup és format per un nombre d'objectes que estan propers en espai en un període de temps significatiu.

A part de la definició general de la proximitat espaciotemporal, un grup pot ser especificat per algunes característiques de la seva estructura interna. Per exemple, un grup podria estar encapçalat

per algun individu que preveu el moviment del grup. Aquest patró es diu lideratge i es va introduir per primera vegada per [19] en el marc del concepte general de moviment relatiu (REM). Un altre grup de models espai-temporals bàsiques de REM són ramat, la convergència i divergència. Un ramat correspon a un grup mòbil de patrons que s'ha definit anteriorment, mentre que la convergència i la divergència simultània descriuen el moviment d'objectes cap o des d'algun punt en l'espai (veure il·lustració 18, mitjà i dreta). És fàcil pensar en una extensió d'aquests patrons per tal d'incloure noves característiques de moviment. Per exemple, la velocitat dels cotxes podria ser necessari que augmenti a mesura que s'allunyen de Pisa com en l'exemple (8), amb la qual cosa es cobrirà en certa mesura el problema de la detecció de tendències. No obstant això, és important assenyalar que no tots els patrons, segons l'especificat per REM i relacionats en la literatura, són derivats directament de les trajectòries. La recuperació dels successos de convergència o de trobada de patrons es basa normalment en un sol instant en el temps i assumeix una velocitat constant i l'orientació dels objectes.

Fins ara, tots els patrons es basen en un grup estable dels objectes. No obstant això, un patró pot continuar amb el pas del temps, encara que canviïn els membres del grup. Per exemple, un embús de trànsit pot prevaler durant diverses hores, mentre que cotxes nous contínuament arriben a un extrem i escapin a l'altre extrem. Aquest fenomen és anomenat **grup en moviment** i es refereix a un grup que manté la seva densitat (i altres propietats similars, com la mida de clúster o el diàmetre), encara que diferents objectes participen en el grup durant la seva durada.

4.1.4 Predicció

En l'era de la logística "*just-in-time*", la gestió del trànsit en temps real, els serveis basats en la localització i de navegació GPS, la predicció sobre el futur de la posició o el destí d'un individu o objecte té un paper central. Anticipar el moviment de persones o grups d'objectes d'aquests sistemes permet prendre les accions preventives, per exemple per tal d'evitar aglomeracions o per lliurar informació útil en el temps desitjat. Per tant, les dades espaciotemporals són una font d'informació molt important des de la perspectiva de les tasques de predicció. A partir d'aquest tipus de dades es pot fer predicció de les trajectòries, de la densitat, de l'abast o d'esdeveniments futurs, entre altres.

4.1.4.1 Predicció d'ubicació i Trajectòries

Durant els darrers anys, la predicció fiable de la futura ubicació d'objectes en moviment ha estat d'interès principalment en dues àrees de recerca: sistemes de bases de dades i xarxes de comunicació sense fils. Les bases de dades d'objectes geo-referenciades suporten consultes sobre moviment i ubicacions futures dels objectes. A les xarxes sense fils, la previsió de moviment futur és important per permetre una assignació eficient dels recursos de la xarxa.

En la literatura de bases de dades que suporten aquest tipus de preguntes requereixen sofisticades estructures d'indexació de les futures posicions dels objectes en moviment. Les preguntes es basen en la previsió de la indexació de les posicions actuals d'estructures i vectors de moviment. Donada la ubicació actual l_c i el vector de velocitat v_c d'un objecte, la futura posició després de temps Δt pot ser calculat com $l_f = l_c + v_c \Delta t$. Amb aquesta finalitat, s'ha proposat una estructura de dades en arbre, anomenada arbre TPR [24], que s'ha desenvolupat per manejar les preguntes de predicció de rang, consultes de veí més proper o sobre el futur de les posicions dels objectes en moviment [3]. La hipòtesi subjacent de totes les tècniques és que els objectes que participen continuen el seu moviment amb el vector de velocitat donat fins l'hora de finalització de l'interval de la consulta.

Aquest supòsit val per al moviment lineal en els espais lliures d'obstacles, com per exemple, els vaixells, avions o fenòmens meteorològics. No obstant això, no és raonable per a xarxes de carrers on els objectes canvien de direcció i de velocitat en intervals curts de temps. Aquestes condicions inestables es compleixen en les xarxes de comunicacions sense fils. En aquest entorn, en primer lloc, els recursos han de ser assignats per garantir una transició de serveis suau si un usuari es mou d'un mòbil a un altre. Segon, quan una trucada entrant arriba, la xarxa ha de trucar el mínim de mòbils possible en una determinada àrea. Ambdues tasques requereixen anticipar el moviment dels usuaris en el futur proper.

A més de la predicció de la ubicació en el futur proper, una part important de l'esforç en recerca en aquest tema s'ha dedicat a estudiar com anticipar la ruta més probable i el destí d'un objecte en moviment. Això obre moltes possibilitats als proveïdors de serveis. Per exemple, serveis basats en la localització poden oferir serveis més sofisticats si saben per quins llocs passarà l'usuari o si està en el camí al treball o al supermercat. La premissa general darrere de la predicció de les rutes i destinacions és que la gent segueix rutines diàries o setmanals. Normalment, la gent visita només uns pocs llocs amb freqüència, com ara la seva casa, lloc de treball o restaurant favorit.

A [14] proposen un mètode basat en matrius de transició adaptades a les preferències personals i, per tant, són capaços de predir la ruta més probable i el destí d'una sola persona amb un segment de temps donat. Mentre que Karimi i Liu, basen les seves prediccions únicament en informació d'enrutament, [18] incorpora els temps de residència en el seu model. L'autor detecta els llocs en què un usuari gasta una quantitat comparativament més gran de temps. Aquests llocs formen el conjunt de totes les destinacions possibles i delimiten les rutes. Laasonen agrupa rutes històriques i compara els tipus obtinguts amb la trajectòria actual. Les previsions de destinació pertanyen al tipus de trajectòria més similar i opcionalment, pot ser condicionada a l'hora del dia i dia de la setmana.

4.1.4.2 Predicció de Densitat

La densitat d'una zona es defineix com el nombre d'objectes dins de la zona en proporció a la mida de l'àrea en un punt donat en el temps. És una característica que emergeix a través de la interacció d'una sèrie d'objectes i canvis en el temps. La predicció de densitats promet molts beneficis, especialment en gestió de fluxos de visitants. Per exemple, un sistema de gestió del fluxos que és capaç d'identificar les regions denses i els colls d'ampolla ha de contrarestar aquests efectes en el

temps. El concepte de densitat s'estén a un interval de temps on l'esmentat conté el nombre mínim d'objectes que estan a la vegada a l'àrea donada. Per calcular la densitat es sol treballar amb graelles espacio-temporal on cada cel·la conté la densitat d'una àrea donada durant algun moment o interval de temps (eix z). Per a la predicció de la densitat [8] assumeix un moviment lineal dels objectes i computa futures densitats per extrapolació.

4.1.4.3 Predicció de Abast

L'abast del moviment és una mesura dependent del temps. Per exemple, Imaginem que s'inaugura un nou restaurant al centre de la ciutat, després d'1 dia el 20% dels habitants hauran notat la nova ubicació, després d'una setmana l'abast s'haurà incrementat a un 60%, si el restaurant hagués obert als afores de la ciutat, només el 40% dels habitants haurien passat després d'una setmana.

L'abast no està limitat a una sola ubicació sinó que pot abastar una xarxa de llocs. En aquest cas es defineix com la proporció de la població que passa al menys per una de les ubicacions de la xarxa en un cert període de temps donat.

Donades les trajectòries d'un GPS d'un grup de persones sobre diversos dies, el nombre de contactes amb una xarxa donada pot ser fàcilment calculat. El repte està en la extrapolació de les dades d'aquesta mostra de les trajectòries. Si les persones en l'enquesta no són representatives de tota la població, per exemple si viuen la majoria en una sola part de la ciutat, les mostres de dades necessiten ser estratificats per proveir un abast imparcial. A més, les trajectòries incompletes que s'originen per defectes dels dispositius GPS, oblit o abandonaments de l'enquesta, constitueixen un greu problema en els mesuraments de les seqüències de dies consecutius.

Hi ha pocs treballs de recerca en aquest problema. A [31] es proposa la utilització de mètodes d'anàlisi de supervivència aplicats a la predicció de l'abast dels moviments d'individus.

4.1.4.4 Predicció d'Esdeveniments

El treball en [4] estudia el problema de predir esdeveniments espaciotemporals que estan associats amb certes característiques predefinides, per exemple, quina probabilitat hi ha que algun crim sigui comès donada certa regió i interval de temps basat en les localitzacions, els temps i les característiques soci econòmiques d'incidents passats. La justificació subjacent és la definició del model de densitat de transició, que prediu la probabilitat de densitat en espai i temps donades les dades històriques, juntament amb el procés d'estimació de densitat per descobrir importants conjunts de característiques i ubicacions d'espai i temps.

4.1.5 El paper de la incertesa en la Minería de Dades Espai-Temporals

La incertesa és una característica inherent en les dades espaciotemporals que es recullen automàticament. Sorgeix a causa de les limitacions físiques i tècniques durant la recollida de dades i emmagatzematge. Normalment la informació sobre els instants de temps és pot aconseguir amb gran precisió, ara bé, la incertesa es troba en la ubicació, que varia en funció de la tecnologia aplicada entre uns pocs metres (GPS) i quilòmetres (GSM).

A més, la taxa de mostreig té una gran influència en la precisió. Com més ràpid es mou un objecte, més freqüentment la ubicació d'un objecte ha de ser reportada per sostenir un nivell de donat d'incertesa espacial. El coneixement general del domini, així com certs supòsits sobre el comportament del moviment ajuden a reduir la incertesa en els dades. Per exemple, pel que fa el seguiment d'un vehicle, un pot estar segur que tots els moviments estan restringits a la xarxa de carrers, és poc probable que els cotxes puguin desplaçar-se a través dels edificis. Un altre supòsit és el fet de moviment lineal entre dues posicions, en general, tenint en compte dues posicions P_1 i P_2 en moments t_1 i t_2 i una velocitat màxima, la posició d'un objecte en cada moment en el temps $t \in [t_1, t_2]$ es limita a algunes àrees, si no es dona més informació, una distribució uniforme dels objectes dins d'aquesta zona pot ser assumida.

4.1.6 Software per a la Minería de Dades

En aquesta secció es presenten els paquets de programari que es poden utilitzar per desenvolupar algorismes de minería de dades.

Open Source

- **Weka²**: és una col·lecció d'algorismes de machine learning per a tasques de minería de dades. Els algorismes poden ser aplicats directament a un conjunt de dades o cridats des del propi codi Java. Weka conté eines per a les dades de pre-processament, classificació, regressió, clustering, regles d'associació, i la visualització. També és molt adequat per al desenvolupament de nous esquemes de machine learning.
- **RapidMiner³**: (abans anomenat YELA) És un entorn per a machine learning i minería de dades, RapidMiner i els seus plugins proporcionen més de 400 operadors de tots els aspectes de la Minería de Dades. RapidMiner s'utilitza tant per a la investigació com per a dades reals.

² <http://www.cs.waikato.ac.nz/ml/weka/>

³ <http://rapid-i.com>

- **Rattle de R⁴**: és un kit d'eines per a la mineria de dades utilitzat per analitzar conjunts de dades molt grans. Rattle presenta resums tant visuals com estadístics de dades, transforma les dades en formes que puguin ser fàcilment modelades, es basa tant en els models de supervisió com de no supervisió de les dades.

Rattle també proporciona una entrada sofisticada a la mineria de dades utilitzant el llenguatge R.

- **Orange⁵**: És una llibreria per a Python per a realitzar mineria de dades i anàlisis predictiu. També permet visualització de dades. Conté extensions per bioinformàtica.

Propietari

- **Enterprise Miner de SAS Institute⁶**: És un programari per agilitzar el procés de mineria de dades per crear models predictius i descriptius d'alta precisió basats en l'anàlisi de grans quantitats de dades de tota l'empresa. Les organitzacions estan utilitzant el programari de mineria de dades de SAS per detectar el frau, minimitzar el risc de crèdit, anticipar la demanda de recursos, les taxes de resposta de l'augment de les campanyes de màrqueting i frenar la pèrdua de clients.
- **PASW Modeler⁷** (abans anomenat Clementine) És un programari per a la mineria de dades. L'usuari mitjançant la interfície envia al servidor el codi d'una comanda (no visible) per a que el servidor el processi. El codi serà compilat en SQL i processat en el magatzem de dades.
- **Pentaho⁸**: La Suite Pentaho BI ofereix una gamma completa de Business Intelligence (BI), incloent les capacitats de consulta i presentació d'informes, anàlisi interactiu, quadres de comandament, integració de dades ETL, mineria de dades i una plataforma per a l'empresa, BI Suite.
- **Intelligent Miner de IBM⁹**: Les capacitats de mineria de dades d'IBM ajuden a detectar el frau, el segment dels seus clients i simplificar l'anàlisi de la cistella de mercat. Les capacitats de mineria de bases de dades s'integren amb els sistemes existents per oferir escalabilitat i alt rendiment d'anàlisi predictiu sense moure les dades en les plataformes d'extracció de dades de propietat. Es pot utilitzar SQL Server, serveis web, o Java per accedir a les capacitats de mineria de dades de DB2 de les seves pròpies aplicacions o eines d'intel·ligència empresarial dels socis de negoci d'IBM.

⁴ <http://rattle.togaware.com/>

⁵ <http://www.ailab.si/orange/>

⁶ <http://www.sas.com/technologies/analytics/datamining/miner/>

⁷ <http://www.spss.com/software/modeling/modeler/>

⁸ <http://www.pentaho.com>

⁹ <http://www-01.ibm.com/software/data/iminer/>

Llicència Pròpia

- **Knime¹⁰**: (o Konstanz Information Miner) és una plataforma de mineria de dades que permet el desenvolupament en un entorn visual. Està construït sota la plataforma Eclipse. La versió base KNIME ja inclou més de 100 nodes de processament de dades I / O, pre-processament i la neteja, el modelatge, anàlisi i mineria de dades, així com diverses visites interactives, com ara diagrames de dispersió, coordenades paral·leles i altres. Integra tots els mòduls d'anàlisi del conegut entorn de dades WEKA la mineria i els plugins addicionals permeten R-seqüències de comandaments a executar, que ofereix accés a una vasta biblioteca de rutines d'estadística. Utilitza una llicència dual(free public software)¹¹

4.1.7 Bibliografia sobre Minería de dades

[1] Agrawal, Rakesh, Lin, King-ip, S. Sawhney, Harpreet and Shim, Kyuseok. Fast similarity search in the presence of noise, scaling and translation in time series databases 1995.

[2] Arimond, George, Elfessi, Abdulaziz. A Clustering Method for Categorical Data in Tourism Market Segmentation Research 2001.

[3] Benetis, R., Jensen, C., Karciuskas, G. and Saltenis, S. Nearest and reverse nearest neighbor queries for moving objects 2006

[4] Brown, D., Liu, H., Xue, Y.: Mining Preferences from Spatial-Temporal Data. The 1st International Conference on Data Mining (SDM'01), (2001)

[5] Cao, H., Mamoulis, N., Cheung, D. W.: Mining Frequent Spatio-Temporal Sequential Patterns. The Fifth IEEE International Conference on Data Mining (ICDM'05), (2005) 82-89

[6] Gaffney, S., & Smyth, P.: Trajectory Clustering with Mixture of Regression Models. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, (1999) 63-72

[7] Geurts, P.: Pattern Extraction for Time Series Classification. The 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'01), (2001) 115-127

[8] Hadjielefteriou, M., Kollios, G., Gunopulos, D. et al.: On- Line Discovery of Dense Areas in Spatio-Temporal Databases. The 8th International Symposium on Spatial and Temporal Databases (SSTD'03), (2003) 306-324

¹⁰ <http://www.knime.org>

¹¹ <http://www.knime.org/downloads/license>

- [9] Hadjieleftheriou, M., Kollios, G., Bakalov, P. et al.: Complex Spatio-Temporal Pattern Queries. The 31st conference in the series of the Very Large Data Bases conferences(VLDB'05), (2005) 877-888
- [10] Han, J., & Kamber, M.: Cluster analysis. In: Morgan Kaufmann (ed.): Data Mining: Concepts and Techniques. Second edition. Cerra, D., San Francisco (2006) 383-464
- [11] Hu, Y., Huang, T. C., Yang, H. et al.: On Mining Multi-Time-Interval Sequential Patterns. Data & Knowledge Engineering, 68 (2009) 1112-1127
- [12] Hwang, S. Y., Liu, Y. H., Chiu, J. K. et al.: Mining Mobile Group Patterns: A Trajectory Based Approach. The Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'05), (2005) 713-718
- [13] Kalnis, P., Mamoulis, N., Bakiras, S.: Discovering Moving Clusters in Spatio-Temporal Data. The 9th International Symposium on Spatial and Temporal Databases (SSTD'05), (2005) 364-381
- [14] Karimi, H., & Liu, X.: A Predictive Location Model for Location Based Services. The 11th International Symposium on Geographic Information Systems (GIS'03), (2003) 126-133
- [15] Keogh, E., & Pazzani, M.: An Enhanced Representation of Time Series which Allows Fast and Accurate Classification, Clustering and Relevance Feedback. The Fourth International Conference on Knowledge Discovery and Data Mining (KDD'98), (1998) 239-241
- [16] Kuijpers, B., Moelans, B., Othman, W. et al.: Analyzing Trajectories using Uncertainty and Background Information. SSTD '09: Proceedings of the 11th International Symposium on Advances in Spatial and Temporal Databases, (2009) 135-152
- [17] Kulldorff, M.: A Spatial Scan Statistic. Communications in Statistics - Theory and Methods, **26** (1997) 1481-1496
- [18] Laasonen, K.: Clustering and Prediction of Mobile User Routes from Cellular Data. The 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'05), (2005) 569-576
- [19] Laube, P., & Imfeld, S.: Analyzing Relative Motion within Groups of Track Able Moving Point Objects. The 10th ACM international symposium on Advances in geographic information systems (GIS'02), (2002) 132-144
- [20] Lee, J. G., Han, J., and Whang, K. Y.: Trajectory Clustering: A Partition and Group Framework. Proceedings of the 2007 ACM SIGMOD international conference on Management of data, (2007) 593-604
- [21] Li, Q., Zheng, Y., Xie, X. et al.: Mining User Similarity Based on Location History. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS '08), (2008) 298-307

- [22] Li, Y., Han, J., Yang, J.: Clustering Moving Objects. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04), (2004) 617-622
- [23] Nanni, M. Clustering Methods for Spatio-Temporal data, PhD Thesis. Dipartimento di Informatica, Università di Pisa, 2002
- [24] Tao, Y., Papadias, D., Sun, J.: The TPR* Tree: An Optimized Spatio-Temporal Access Method for Predictive Queries. The 29th International conference on Very Large Data Bases(VLDB'03), (2003) 790-801
- [25] Vlachos, M., Kollios, G., and Gunopulos, D.: Discovering Similar Multidimensional Trajectories. IEEE International Conference on Data Engineering (ICDE), (2002) 673-684
- [26] Wang, Y., Lim, S., Hwang, S.: Mining Group Patterns of Mobile Users. The 14th International Workshop on Database and Expert Systems Applications (DEXA'03), (2003) 287-296
- [27] Warrenliao, T.: Clustering of Time Series Data. Pattern Recognition, 38 (2005) 1857-1874
- [28] Wu, S., & Chen, Y.: Discovering Hybrid Temporal Patterns from Sequences Consisting of Point- and Interval-Based Events. Data & Knowledge Engineering, 68 (2009) 1309-1330
- [29] Xia, J., Ciesielski, V., Arrowsmith, C.: Data Mining of Tourists Spatio-Temporal Movement Patterns: A Case Study on Phillip Island. The Eighth International Conference on GeoComputation, (2005) 1-15
- [30] Zanda, A., Koerner, C., and Giannotti, F. et al.: Clustering of German Municipalities Based on Mobility Characteristics. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS 2008), (2008) 479-482
- [31] Giannotti, Fosca and Pedreschi, Dino (Eds.): Mobility, Data Mining and Privacy - Geographic Knowledge Discovery. Springer 2008
- [32] Witten, I.H and Frank, E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers, San Francisco (2001).

4.2. Estudi de les variables que es tenen en compte per l'anàlisi del comportament i fluxos de turistes

En aquesta secció es presenten els resultats de l'estudi de la recerca bibliogràfica sobre els criteris (i.e. variables) que es tenen en compte en els estudis de mobilitat, especialment els centrats en el sector del turisme. La quantitat d'articles recollits que treballen amb criteris espacials i aespacials són 15 del total de 26 treballs sobre oferiment de serveis a turistes, basats en la seva localització o seguiment, presentats en el capítol 2.

S'han identificat quines són aquelles variables que han estat valorades pels experts com a necessàries, amb l'objectiu de comprendre amb més exactitud el comportament turístic dels visitants en les diferents zones a on s'han realitzat les investigacions i observacions i, com a conseqüència, la interacció entre els investigadors i els turistes en forma de qüestionaris, entrevistes, etc.

Cal esmentar que en els articles analitzats, no ha estat fàcil trobar la informació sobre els criteris tinguts en compte en els estudis. En alguns casos s'han trobat directament les dades, però en altres s'han hagut de deduir degut a la poca importància que se li han donat a moltes investigacions a difondre el contingut de l'entrevista o qüestionari. En molts estudis, l'únic que importava era mostrar el resultat o el mètode, però no les dades amb les quals s'ha treballat.

Totes les variables que s'han trobat estan recollits en aquest document. S'han establert diversos blocs temàtics per a poder dissecar millor els criteris i analitzar-los. Així doncs, veient a què feien referència les variables que es consideraven en els diferents treballs, van sorgir tres divisions. La descripció i comentaris dels resultats per cadascuna d'aquestes divisions es presenta en el capítol 5.

- *Sociodemogràfica*: variables que tenen relació amb la caracterització del turista. Són dades socials i personals, no de l'activitat en si. Són: nom, edat, any de naixement, sexe, nivell educatiu, rol familiar, lloc d'origen, lloc de residència i/o nacionalitat, adreça, ingressos, professió/ocupació, lloc de feina, realització de vacances últimament o intenció de fer-les properament, lloc preferit com a destí turístic.
- *Dades del viatge*: variables corresponents al viatge i factor que poden haver influenciat en el moviment del turista. Són: acompanyants, amb nens, temps d'estada, mode de transport, hora d'arribada, hora de sortida, propòsit del viatge, influència de la publicitat, primera visita, única destinació del viatge, destinació principal, llocs visitats, activitats realitzades, temps gastat en cada activitat, opinió del lloc visitat, tipus d'allotjament, destinacions futures, percepció del lloc visitat i temps gastat en omplir el qüestionari.
- *Dades de l'itinerari*: aquestes variables només són aplicables en els casos en què s'ha estudiat el trajecte concret realitzat pel visitant. Són: recorregut seguit, mètode emprat pel visitant per decantar-se per un itinerari o un altre i distància entre l'allotjament i l'activitat.

5. Estudi de les noves tendències en la provisió de serveis avançats a turistes que ofereixen les tecnologies basades en localització i seguiment

En aquesta secció s'analitza el material explicat en les seccions anteriors. D'una banda, es comenten i comparen les tècniques existents. D'altra banda, s'identifiquen noves línies de recerca, així com els investigadors i grups principals en els diferents àmbits de l'estudi que s'ha dut a terme en aquest projecte.

5.1 Anàlisi dels projectes de recerca sobre mobilitat

Al capítol 1 s'ha fet un estudi dels projectes de localització i seguiment finançats per la Unió Europea. En total s'han trobat 29 projectes dels quals 15 eren sobre turisme mòbil, i són en els que s'ha centrat l'estudi. Es va veure que la major part d'aquests projectes es van realitzar al cinquè Programa Marc i utilitzaven el GPS per determinar la posició.

En aquesta secció es comparen les característiques més importants dels sistemes proposats en els 15 projectes de recerca finançats per la Unió Europea detallats en el capítol 1: PEPTRAN, TOURSERV, CRUMPET, WEBPARK, IMAGE, LOVEIUS, EUREAUWEB, AMBIESENS, REGEO, M-ToGuide, ELBA, M-Guide, EXPLORE, IM@GINE i ITACITUS.

En primer lloc, a la Taula 5.1, es pot observar que tots els projectes, excepte EXPLORE, tenen com a entorn d'aplicació espais a l'aire lliure. EXPLORE es va dur a terme en museus. A més, tots els projectes van realitzar la prova dels sistemes en zones limitades, ja sigui una ciutat, un parc natural o una pista d'esquí.

Taula 5.1. Entorn d'aplicació dels projectes

PEPTRAN	TOURSERV	CRUMPET	WEBPARK	IMAGE	LOVEUS	EUREAUWEB	AMBIESENS
Ciutat	Pistes d'esquí	Ciutat	Parc natural	Ciutat/ Regió rural	Ciutat	Vies navegables	Ciutat
REGEO	M-ToGuide	ELBA	M-Guide	EXPLORE	IM@GINE IT	ITACITUS	
Regió natural	Ciutat	Ciutat	Ciutat	Museus	Ciutat	Ciutat	

La Figura 5.1 mostra el període de desenvolupament dels projectes. S'observa que els anys 2002 i 2003 hi va haver un boom en la realització de projectes europeus relacionats amb els serveis basats en la localització.

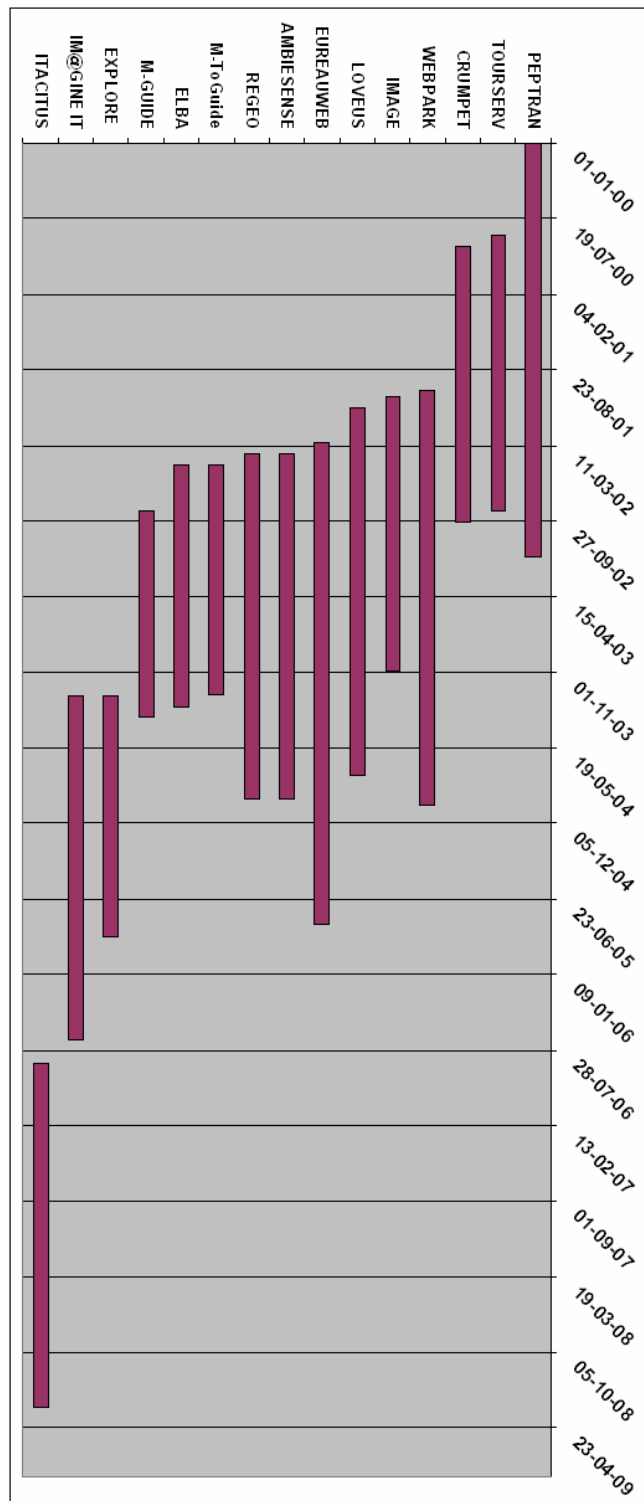


Figura 5.1. Període de desenvolupament dels projectes

A la Figura 5.2 podem veure les tecnologies de posicionament utilitzades als projectes. Com es pot apreciar la gran majoria utilitzen el GPS. El motiu és que la major part dels projectes tenen com a finalitat l'aplicació d'aquests serveis en espais oberts, i el GPS en aquest tipus d'entorns és el més apropiat, per la seva precisió, cobertura i cost.

Avui en dia no hi ha cap sistema en funcionament dels projectes de turisme mòbil estudiats degut a la manca d'estàndards.

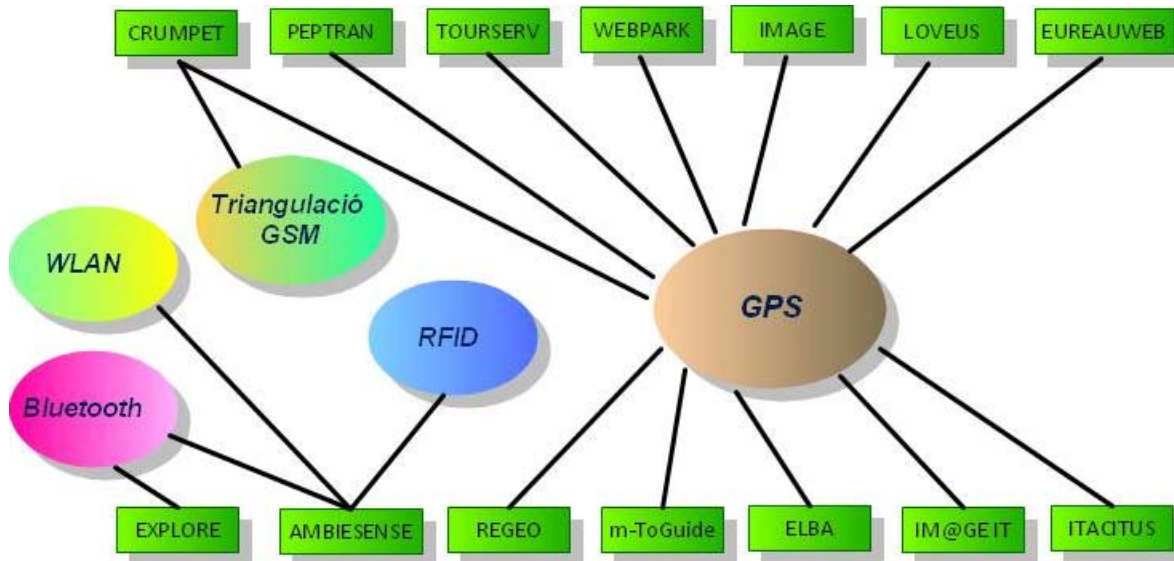


Figura 5.2. Tecnologies de posicionament utilitzades als projectes

5.1.1 Estudi de la privadesa en els projectes de turisme mòbil

Existeixen moltes maneres de proporcionar privadesa als usuaris en un LBS . En els projectes de turisme mòbil analitzats, la privadesa dels usuaris està a una fase molt inicial, ja que solament hi ha dos projectes (WEBPARK i CRUMPET) que l'han tingut en compte.

En el projecte WEBPARK utilitzen un sistema basat en polítiques que permeten que els usuaris defineixin la seva pròpia política utilitzant el protocol P3P. En el projecte CRUMPET es descriu un agent mediador que només revela informació absolutament necessària per dur a terme la comunicació, però no s'ha trobat informació de com s'ha portat a terme aquesta protecció.

5.1.2. Comparativa de la tecnologia segons el tipus de servei

Per finalitzar, el capítol 1 també presenta els resultats de l'estudi de les tecnologies que s'estan utilitzant per localitzar i fer el seguiment de persones. El tipus de tecnologia que s'utilitza en un LBS depèn de les característiques del servei. A l'hora de decantar-se per una tecnologia és necessari tenir en compte factors tan variats com:

- La precisió: el marge d'error a la posició real.
- La disponibilitat: la possibilitat d'utilització en diferents entorns.
- El consum de potència: en dispositius mòbils és important la durada de la bateria.
- Latència: el temps necessari per obtenir la posició de l'usuari.
- Els costos d'operació: l'instal·lació i manteniment de la infraestructura de localització.

Els serveis de localització imposen requisits de precisió molt diferents en funció dels seus objectius. Existeixen aplicacions que admeten un error de desenes de metres, i altres per a les quals és necessari conèixer la ubicació de l'element buscat al centímetre, i en les quals a més es necessita una resposta en temps real. Per a oferir aquests serveis existeixen diverses tecnologies que permeten conèixer la posició d'un terminal mòbil. La precisió, el cost i la dificultat d'implementació són paràmetres claus a valorar abans de decidir-se per una o altra opció.

Servei	Exemple	Tecnologia	Entorn	Precisió
20.8in Emergència	Trucades d'emergència	TOA / OTD / A-GPS	Interior / Exterior	Mitjana - alta
	Assistència a l'automòbil	TOA / OTD / A-GPS	Exterior	Mitjana
40.8in Navegació	Direccions	A-GPS	Exterior	Alta
	Gestió del trànsit	TOA / OTD / A-GPS	Exterior	Mitjana
	Navegació interior	WLAN / Bluetooth / RFID	Interior	Alta
	Gestió de grup	Cell-ID / TOA / OTD / A-GPS	Exterior	Baixa - Mitjana
30.650in Informació	Serveis de viatge	TOA / OTD / A-GPS	Exterior	Mitjana - Alta

	Pàgines grogues	TOA / OTD / A-GPS	Exterior	Mitjana
	Serveis d'informació i oci	TOA / OTD / A-GPS	Exterior	Mitjana - Alta
Publicitat	-	TOA / OTD / A-GPS	Exterior	Mitjana - Alta
40.650in Seguint	Seguiment de persones	Cell-ID / GPS	Interior / Exterior	Alta
	Seguiment de vehicles	TOA / OTD / A-GPS	Exterior	Mitjana
	Seguiment de productes	RFID / Bluetooth	Interior	Alta
	Seguiment de personal	TOA / OTD / A-GPS	Exterior	Mitjana
Facturació	Facturació sensible a la localització	Cell-ID / TOA / OTD	Interior / Exterior	Baixa - Mitjana

Taula 5.2. Comparativa de les tecnologies utilitzades segons els tipus de LBS

A la Taula 5.2 podem veure un estudi [24] de les tecnologies utilitzades segons si es tracta d'un servei d'emergència, de navegació, d'informació, de publicitat, de seguiment o de facturació. Un dels factors més importants a l'hora de triar la tecnologia és concretar el nivell de precisió que es vol tenir i si l'entorn d'aplicació és a l'aire lliure o dins d'un edifici.

Els sistema LBS és un dels desenvolupaments més interessants que sorgeixen del sector de les telecomunicacions mòbils. Però no han tingut l'èxit esperat per culpa de que els LBS tenen importants barreres a superar. La més important és la manca d'estàndards. Com en altres tantes solucions tecnològiques, la integració i la interoperabilitat sembla ser el futur dels LBS.

5.2 Serveis avançats per turistes

En aquesta secció s'analitzen mètodes utilitzats en aquests estudis per tal d'obtenir les trajectòries de moviment dels turistes (*tracking*), a partir dels resultats explicats al capítol 2. S'identifiquen els investigadors més rellevants involucrats en aquest aspecte. Finalment es presenta una classificació dels serveis que actualment s'estan oferint als turistes en base a la seva localització i/o seguiment, i s'identifiquen nous serveis que es creu interessant estudiar.

5.2.1 Investigadors que fan estudis de mobilitat de turistes

Després de l'estudi dels projectes i treballs de recerca en oferiment de serveis basats en localitzacions i seguiment dels turistes, presentat al capítol 2, en aquesta secció es comenta de forma general l'estat de la recerca en aquest àmbit.

Degut a la fase inicial en la qual es troba l'estudi de mobilitat de turistes, hi ha uns grups d'investigació determinats que han establert les seves pròpies línies d'estudi amb les respectives metodologies. Com es veurà, no hi ha gran quantitat d'investigadors, o el que és el mateix, no hi ha una equivalència numeral entre la quantitat d'articles i autors. Molts dels estudis repeteixen autors i col·laboracions.

Un grup important és el format per O'CONNOR, A. ZERGER i B. ITAMI, els quals se centren en la millora de la gestió dels recintes turístics a partir de conèixer els moviments dels visitants mitjançant el *tracking*. Els diferents articles trobats d'en Noam SHOVAL i Michal ISAACSON, (tant junts com per separat), incideixen en l'experimentació dels diferents sistemes de localització amb la finalitat de conèixer els punts forts i febles de cada metodologia. Jianhong XIA, Colin ARROWSMITH i Mervyn JACKSON treballen en les tècniques del *wayfinding* a Victoria (Austràlia). Alan LEW i Bob MCKERCHER fan unes aportacions molt interessants sobre els factors que influencien els moviments turístics dintre d'una destinació aprofitant els estudis realitzats fins aleshores. Un autor important per a la realització del present estudi ha estat Wim G.M. VAN DER KNAAP, el qual les seves successives investigacions incideixen en la necessitat de l'ús dels SIG com a eina idònia per a la gestió dels espais turístics recollint la relació entre espai – temps. Altres grups investigadors a anomenar són: Yasuo ASAKURA i Takamasa IRYO; Rein AHASA, Anto AASAA, Antti ROOSEA, Ular MARKB i Siiri SILMA; David A. FENNELL; Ali FRIHIDA, Danielle J. MARCEAU i Marius THÉRIAULT.

5.2.2 Classificació dels serveis que s'estan oferint als turistes i proposta de possibles nous serveis

A la secció 2.2, s'han identificat dos àmbits d'aplicació d'aquest tipus de serveis: els orientats a la planificació del turisme i els orientats al propi turista. S'han presentat les aplicacions existents en aquests dos àmbits. A partir d'aquesta informació, en aquesta secció es fa una llista de propostes seguint aquests dos punts de vista.

Aplicacions per la planificació turística:

- **Perfil 'espacial' del turista:** la combinació del *tracking* amb qüestionaris sociodemogràfics permet obtenir el perfil del turista. Aquest perfil no ha de ser el general ni el tradicional, sinó que ha de tenir en compte el criteri espacial, és a dir, ha d'estar determinat per a on i com s'ha mogut el visitant a la destinació. Això és necessari per tal d'adaptar l'oferta de la pròpia destinació als diferents visitants que rep.
- **Redistribució i redirecció de l'afluència:** degut a l'aportació de dades instantànies que ofereix el GPS i la tecnologia de la localització, s'hauria d'aprofitar aquest atribut per a evitar massificacions podent redistribuir la gent, per exemple, controlant i adequant els fluxos de turistes a la capacitat de càrrega de cada espai. A més, per a poder fer això, s'han de proposar diferents rutes alternatives amb les quals es puguin redirigir a altres espais.
- **Seguretat i prevenció d'accidents:** aquesta proposta va enfocada a espais tancats o semi-tancats com podrien ser els parcs naturals. A través del *tracking* es pot organitzar la seguretat localitzant els turistes amb la finalitat de prevenir accidents o pèrdues o, en cas d'emergència, poder procedir a un rescat sense perdre el temps buscant al visitant. També seria interessant poder avisar al turista en el cas que s'estigui desviant de la ruta establerta.
- **Diferenciació de la destinació:** per a poder diferenciar la destinació i dotar-li d'un caràcter d'exclusivitat, es proposa oferir eines de planificació de l'estada com les guies digitals portàtils amb serveis basats en la localització relacionades amb Internet per a poder descarregar-les de forma gratuïta. Això obliga a investigar la relació entre la Web i els LBS.
- **Senyalització de recursos especials:** hi ha rutes que és impossible senyalitzar-les, com podrien ser les aquàtiques en el cas d'una zona turística marina. En aquests casos, es proposa que el SIG les indiqui juntament amb altres elements de característiques similars.

Aplicacions pels turistes:

- **Generalitzar l'ús de guies turístiques portàtils (*mobile guides*)** per a possibilitar al turista demanar un taxi, fer una reserva al restaurant que vulgui, disposar de telèfon d'emergència, etc. Estaríem, així doncs, parlant d' **E-commerce** i d'informació descriptiva dels recursos i la logística (horaris, dies d'obertura, preus...).

- **Assistència i seguretat:** en el cas que el turista, en espais tancats o semi-tancats, estigui perdut o s'hagi accidentat, que pugui demanar ajuda a través d'un PDA o un hardware de l'estil que se li hagi proporcionat al començament de la visita. Per a fer efectiu aquest servei d'emergència, el PDA hauria de tenir l'aplicació de la localització.
- **Facilitar i adaptar les marques dels camins** per a discapacitats amb la finalitat d'obrir la destinació a tot tipus de visitants sense posar obstacles físics.
- **Compartir experiències de manera col·laborativa:** amb la Web 2.0 i el LBS es poden compartir experiències viscudes durant l'estança. L'objectiu d'això no és només explicar la pròpia experiència sinó localitzar-la incloent-hi fotografies, vídeos, etc. Aquesta proposta també afecta a la gestió ja que els visitants, indirectament, fan publicitat de la destinació gratuïtament.

Les destinacions que presenten aglomeracions en zones concretes, com poden ser els parcs temàtics, poden oferir informació consultable als turistes en temps real sobre **l'aflluència** en una determinada atracció per tal que el visitant tingui l'opció de seleccionar i establir un ordre propi de visita.

Com a conclusió, més que pensar en noves aplicacions turístiques, s'hauria de procurar millorar les actuals i intentar potenciar el seu ús. Oferir les guies digitals portàtils de forma gratuïta i a través d'Internet que siguin compatibles amb el PDA es veu com el primer pas essencial per a impulsar els serveis turístics tecnificats, diferenciar les destinacions i apropar i facilitar el territori amb els seus recursos al turista.

5.3. Anàlisi de la protecció de la privadesa dels turistes

En aquest treball s'han fet estudis que proporcionen una visió acurada dels mecanismes de recollida i anonimització de dades procedents dels turistes.

Avui en dia existeixen moltes tecnologies que permeten localitzar i fer el seguiment de les persones. Per exemple, per al correcte funcionament del telèfon mòbil, l'operadora requereix detectar constantment la proximitat del client a una antena específica (cel·la). Aquesta característica permet el seguiment dels telèfons mòbils.

A la secció 3.1 s'han presentat les dues tecnologies més habituals per recollir les dades de posicionament dels turistes: el telèfon mòbil i el GPS. L'elecció de la tecnologia de posicionament dependrà de l'escala de les dades a analitzar. Per exemple, en un entorn tancat (parc d'atraccions o una ciutat) o directament de tots els turistes que visiten un país.

Els telèfons mòbils tenen una precisió relativament baixa com a mètode de seguiment que variarà segons la geometria de la xarxa i la densitat d'estacions base (aprox. entre 150 i 1500 metres). En

canvi, els dispositius GPS són un mitjà molt eficaç per a recollir grans quantitats de dades amb alta resolució del comportament espacial i temporal dels visitants.

Sigui quina sigui la tecnologia, és necessari proporcionar mecanismes de protecció de la privadesa.

Després de presentar, en el capítol 3, els mètodes de protecció de la privadesa dels individus en quant a la informació sobre localització, es pot dir que aquest és un àmbit de recerca bastant nou i poc estudiat, en comparació amb altres àmbits. A més, fins ara, els mètodes desenvolupats són estudis teòrics que no s'han aplicat en gaires casos reals, i menys encara en el cas del turisme. En aquesta secció es fa un anàlisi dels mètodes existents, tant per serveis de localització com per anonimització de trajectòries.

5.3.1 Anàlisi dels mètodes d'anonimització en LBS

Al capítol 3.1 s'han examinat els mètodes més recents i rellevants per a la protecció de la privadesa en serveis basats en la localització (*Location Based Service*, LBS) classificats segons dues variables: (i) la utilització o no d'una TTP i (ii) la pertorbació o no de la ubicació.

Encara que els mètodes basats en TTP són els més comuns, els mètodes sense TTP són superiors en termes de privadesa a causa de les deficiències dels mètodes basats en TTP. L'inconvenient més important dels mètodes basats en TTP és que la seva arquitectura és centralitzada. Això els fa vulnerables a atacs de denegació de servei i a la creació de colls d'ampolla. A més, si la ubicació no es pertorba, és fàcil deduir la identitat de l'usuari utilitzant l'atac RSI o OI. Per fer front a aquests atacs van sorgir mètodes basats en TTP que pertorben les ubicacions dels usuaris, però tenen l'inconvenient que la informació no és precisa degut a la distorsió de la ubicació.

En general els mètodes basats en TTP són febles perquè els usuaris depenen d'una sola entitat de confiança. Per fer front a aquest problema es van proposar els mètodes sense TTP. Tot i que aquests mètodes són més segurs en protecció de la privadesa, la seva aplicació real és més difícil degut a la volatilitat de la xarxa.

En quan el treball futur, hi ha diverses línies obertes de recerca per protegir la privadesa de la ubicació. En el problema de les consultes privades, és interessant l'estudi dels mètodes que aconseguen anonimats i diversitat impedit que els usuaris puguin associar les ubicacions amb les consultes sensibles.

Una altra direcció és el desenvolupament de mètodes basats en PIR que redueixin la seva sobrecàrrega de PIR. Una possible proposta podria ser el disseny d'un mètode híbrid mitjançant la combinació d'àrees d'encobriment, seguit d'una fase de PIR en una part restringida de dades espacials.

Per últim, seria interessant desenvolupar un mètode basat amb TTPs distribuïdes. Aquest mètode resoldria els problemes de les arquitectures centralitzades, és a dir, la vulnerabilitat a atacs de denegació de servei i a la creació de colls d'ampolla.

5.3.2 Anàlisi dels mètodes d'anonimització de trajectòries

En aquesta secció es resumeixen breument els mètodes de anonimització estudiats i a continuació es fa un anàlisi dels mateixos.

Mètode		Localització Independent	Trajectòries	Espai	Temps
Confusió de Trajectòries	Generació de trajectòries falses		x	Coordenades (x,y)	No considera l'instant de temps
	Confusió de trajectòries reals		x	Coordenades (x,y)	Cada mostra correspon a un instant concret
Ofuscació de Localitzacions	Ofuscació per imprecisió de coordenades		x	Coordenades (x,y)	No considera l'instant de temps
	Ofuscació amb translacions i votacions	x		Coordenades (x,y)	No considera l'instant de temps
	Mineria de Dades per Trajectòries d'Objectes en Moviment	x	x	Coordenades (x,y)	No considera l'instant de temps
Modificació de Trajectòries			x	Esdeveniments o localitzacions concrets (dades simbòliques)	No té en compte moments de temps.
Anonimització per Generalització	K- anonimitat per a bases de dades de trajectòries			Coordenades (x,y)	No considera l'instant de temps
	Generalització per agrupació de seqüències de símbols	x		Esdeveniments o localitzacions concrets (dades simbòliques)	No considera l'instant de temps
	Generalització segons el model del terreny		x	Coordenades (x,y)	No considera l'instant de temps

Taula 5.3. Mètodes de Anonimització vs Informació Espaciotemporal

Mètode		Mesura de la qualitat del Resultat.(error)	Mesura de la qualitat de la Anonimització.	Qui fa els càlculs?
Confusió de Trajectòries	Generació de trajectòries falses			L'usuari en si
	Confusió de trajectòries reals	La inexactitud de les dades es mesura d'acord a la qualitat del servei (QoS).	La privacitat es mesura a través de la "Expectativa del error de distancia".	Varis usuaris
Ofuscació de Localitzacions	Ofuscació per imprecisió de coordenades	Mitjançant un procés de negociació aconseguix la millor qualitat		L'usuari en si
	Ofuscació amb translacions i votacions			Varis usuaris (comunicació P2P)
	Míneria de Dades per Trajectòries d'Objectes en Moviment		Es mesura en termes de tendència, precisió i consistència	
Modificació de Trajectòries			L'algoritme d'identificació de violacions a la privacitat identifica les trajectòries no segures.	
Anonimització per Generalització	K- anonimitat per a bases de dades de trajectòries	La pèrdua de informació es mesura d'acord a la equació LCM (log cost metric)		TTP
	Generalització per agrupació de seqüències de símbols	SSE(Sum of Squared Euclidean distances)		TTP
	Generalització segons el model del terreny		Comparant les rutes dels usuaris amb les rutes no segures.	TTP

Taula 5.4. Característiques de la protecció de la privadesa.

En el primer tipus de mètodes, que hem denominat "Confusió de Trajectòries", s'estudien mecanismes per a impedir a un adversari rastrejar completament una trajectòria individual. Es genera una confusió de les trajectòries. Cada vegada que el trajecte de dos usuaris es troba, hi ha una possibilitat que un atacant confongui els trajectes i pugui seguir a l'usuari equivocat, pel que un algoritme de privadesa explota aquesta possibilitat pertorbant la informació de la localització en algunes d'aquestes àrees que es troben augmentant així les probabilitats d'aquesta confusió.

La següent aproximació es basa en “Ofuscació de localitzacions”. Consisteix ofuscar les dades pertorbant la localització exacta de l’individu. Hi ha diferents maneres de fer-ho, afegint soroll a les coordenades x,y o fent moviments de rotació i translació dins uns límits establerts.

El mètode de “Modificació de projeccions” considera un escenari on les mostres de la localització són assenyalades en un conjunt. La idea principal és transformar projeccions llargues i detallades en projeccions petites i simples, amb això s’és capaç de diversificar les ubicacions que estan sent monitoritzades pels adversaris fent així impossible inferir en les dades amb certesa si les trajectòries inclouen certs punts. Per a aconseguir-ho és necessari suprimir certs punts en les trajectòries, augmentant així la privadesa.

Finalment, l’ “Anonimització per generalització” proposa que la privadesa de les trajectòries pot ser obtinguda aplicant la propietat de k-anonimitat. En aquest cas, les trajectòries de diverses línies de temps poden ser anonimitzades simultàniament. S’elimina informació de les dades fent generalitzacions espaciotemporals de forma que s’obtenen agrupacions de trajectòries semblants que després són anonimitzades.

La taula 5.3 indica el tipus d’informació espacial i temporal que tracta cadascun dels mètodes. Això ens permet veure que la majoria treballen amb coordenades numèriques (x,y), mentre que pocs són els que tenen informació simbòlica, és a dir, el nom del lloc en forma textual. En quant a l’aspecte del temps, la majoria no consideren l’instant de temps concret en què l’individu està a cada lloc, si no l’ordre relatiu entre els llocs. En aquesta taula també es pot veure que la majoria de mètodes tracten l’anonimització des del punt de vista global de la trajectòria i no lloc a lloc.

La taula 5.4 compara els mètodes des de diversos punts de vista relacionats amb la protecció de la privadesa, com són: la mesura de la qualitat del resultat (o errors introduïts), la mesura de la qualitat de la protecció de les dades i finalment, si es necessita algun intermediari per tal de dur a terme l’anonimització. Aquest darrer aspecte és important, perquè si es necessiten intermediaris (com el cas TTP), cal estudiar la confiança que es té en l’intermediari.

D’aquest estudi podem definir un conjunt de qüestions obertes:

- Com es pot incloure informació geogràfica sobre el lloc on es realitzen les trajectòries? En la secció 2 s’ha argumentat que hi ha moltes característiques espacials que influeixen en les trajectòries, com altitud, pendent, aspecte i altres característiques del terreny, accessibilitat en relació amb diverses restriccions (obstacles, la disponibilitat de camins, etc), els objectes presents en una ubicació (edificis, arbres, monuments, etc) o fins i tot la funció o manera d’ús (l’habitatge, les compres, la indústria, l’agricultura o el transport).
- Com incloure informació demogràfica dels individus? Això ajudaria a contextualitzar la trajectòria i reconèixer alguns patrons de moviment en espai i temps.
- Com tractar una altra informació temporal sobre les dades? En els mètodes presentats solament un tracta la qüestió temporal dintre de l’algoritme, és a dir, l’instant de temps en què es van prendre les dades; els altres comparen les localitzacions per instant relatiu de

temps (primera dada, segona dada, etc.). No obstant això, cap d'ells inclou intervals de temps, ni considera periodicitat en les dades, o altres de les característiques temporals esmentades en la secció 2 durant el procés d'anonimització.

- Quina és la millor manera de mesurar el grau d'anonimització i la disminució de qualitat en les dades? En els articles analitzats es proposen solament dues formes de mesurar el nivell de protecció de les dades i també el grau d'error introduït. Seria interessant abordar aquest tema en més detall i trobar unes mesures més estàndard per a aquest tipus de dades, possiblement en funció de la utilitat que hagin de tenir.

5.3.2.1 Grups d'investigació

A continuació es detallen els centres que estan treballant en anonimització de dades spatiotemporals, identificant qui són els integrants del grup especialment dedicats a aquesta línia d'investigació.

Integrants	Centre de recerca	Línia de Treball	Web
Bjaik Hoh , Marco Gruteser	ECE Department, Rutgers, The State University of New Jersey,USA.	Sistemes de comunicació sense fils. Xarxes i privadesa en xarxes. Comunicació entre vehicles.	http://www.winlab.rutgers.edu http://www.winlab.rutgers.edu/~gruteser/
Gabriel Ghinita.	Dept. of Computer Science Purdue University, USA.	Control d'accés Geo-espacial, anonimització de micro-dades, privacitat en LBS i en xarxes de gestió de dades.	http://www.cs.purdue.edu/ http://www.cs.purdue.edu/homes/gghinita/
Peter Rupel, George Treu, Axel Küpper, Claudia Linnhoff	Mobile and Distributed Systems Group, Institute for Informatics, Ludwig - Maximilian University Munich, Germany.	Serveis basats en xarxes cel·lulars (GSM, UMTS), tecnologies pro-actives per LBS, middleware per serveis sensibles al context.	http://www.mobile.ifi.lmu.de/
Manolis Terrovitis, Nikos Mamoulis.	Department of Computer Science, University of Hong Kong.	Bases de dades espaciales, multimedia i espacio-temporals. Minería de dades d'alta dimensió. Preservació de privadesa en bases de dades i xarxes sensorials.	http://www.cs.hku.hk/ http://www.cs.hku.hk/people/profile.jsp?teacher=nikos
Mehmet Nergiz, Yücel Saygin.	Sabancı University, Istanbul Turkey.	Seguretat i privadesa en dades distribuïdes. Criptografia aplicada a seguretat en bases de dades.	http://www.sabanciuniv.edu/eng/anasayfa/anasayfa.php http://www.cs.purdue.edu/homes/mnergiz/research.xml

Maurizio Atzori	KDD Lab., ISTI-CNR, Pisa, Italia.	Raonament i aprenentatge amb coneixement geogràfic. Descobriments de patrons. Mètodes d'anàlisi que preserven la privadesa.	http://www-kdd.isti.cnr.it/ http://www.di.unipi.it/~atzori/
Baris Güc	Swiss Federal Institute of Technology, Zurich, Switzerland.	Anonimització de trajectòries amb mètodes d'agrupament.	http://www.inf.ethz.ch/
Fosca Giannotti	KDD Laboratory. ISTI-CNR, Istituto di Scienze e Tecnologie dell'Informazione "A. Faedo", Pisa, Italia.	Estudis de mobilitat. Privadesa i mineria de dades en seqüències espacio-temporals.	http://www.isti.cnr.it/index.html http://www.isti.cnr.it/php-pers/iselpers.php?Giannotti+Fosca
Dino Pedreschi	KDD Laboratory Dipartimento di Informatica Università di Pisa, Pisa, Italia	Privadesa en dades geogràfiques. Mètodes de raonament espacio-temporal.	http://www-kdd.isti.cnr.it/ http://www.di.unipi.it/~pedre/
Vassilios S. Verykios	Data Mining & Privacy Group, University of Thessaly (Dept. Computer & Communication Eng) Volos, Grece	Mètodes de mineria de dades que preserven la privadesa. Privadesa en associació de registres. Computació paral·lela.	http://dmpg.inf.uth.gr/index2.html
Rakesh Agrawal	IBM Almaden Research Center	Mineria de dades. Datawarehouse. Són els autors del software Intelligent Miner.	http://www.almaden.ibm.com/cs/projects/iis/hdb/Projects/data_mining/mining.shtml
Egemen Tanin	Department of Computer Science Software Engineering, Univ Melbourne	Consultes amb bases de dades espaials.	http://www.cs.mu.oz.au/research/kd.html#SaC
Takeshi Yamada, Hiroshi Sawada	NTT Communication Science Laboratories	Visualització de dades seqüencials. Anàlisi de grans volums de dades. Extreure patrons de moviment a partir d'imatges.	http://www.kecl.ntt.co.jp/as/index.html
Antonio Moreno, Aïda Valls	Group Itaka, Department of Computer Science and Mathematics	Operadors d'agregació i fusió aplicats a la privadesa de dades. Sistemes multi-agent.	http://deim.urv.cat/~itaka/

Taula 5.5. Grups d'investigació en anàlisi i privadesa en trajectòries

5.4. Mètodes de mineria de dades

La Mineria de dades espacio-temporals i, en particular, les dades sobre les trajectòries d'individus, és un àmbit de recerca de creixent interès. Els problemes i mètodes presentats en la secció 4 s'han classificat segons els paràmetres clàssics de mineria de dades, que inclouen l'agrupament (o clustering), extracció de patrons i predicció. Les referències que figuren en les taules següents són les que es poden trobar a la secció 4.1.7.

Ref	Objectiu de l'anàlisi de dades	Mètode d'anàlisi de les dades	Tipus Clustering	Mètode existent/nou	Tipus de seqüència
[23]	Services location	Clustering and Profiling	partitioning and hierarchical (k-means, k-medoid, ward)	existent	símbols, caràcters i números
[2]	Increase visitor satisfaction	Clustering	partitioning	existent	números
[29]	Increase visitor satisfaction	Clustering and Profiling,	partitioning	existent	símbols
[21]	Recommendation systems	Clustering and Profiling	hierarchical	existent	punts
[30]	Recommendation systems	Clustering	hierarchical	nou	punts
[1]	Flow analysis	profiling			seqüències
[25]	Recreation behaviour	Clustering	hierarchical	existent	punts
[20]	Recreation behaviour	Flow analysis	partitioning	nou	punts
[22]	Recreation behaviour	Clustering	partitioning	existent	punts
[27]		Clustering	hierarchical	existent	punts i temps
[6]		Clustering	partitioning	existent/nou	punts
[12]		Profiling			punts
[13]		Clustering	partitioning	existent	punts
[17]		Profiling			regions, temps

[5]	Crowd analysis	Profiling			punts
[8]	Crowd analysis	Rule induction			punts
[9]	Crowd analysis	Rule induction			punts
[26]	flow analysis	profiling			punts
[19]	Recreation behaviour	Profiling			punts
[24]		Temporal data mining			punts
[3]		Temporal data mining			punts
[14]	Prediction	Temporal data mining			punts
[18]	Prediction	Temporal data mining			punts
[4]	Prediction	Temporal data mining			punts
[15]		profiling			punts
[7]		profiling			punts
[10]		Clustering and Profiling	tots	existent	símbols, caract, núms, seqüències
[16]		Clustering	hierarchical	existent	seqüències
[28]		profiling			seqüències
[11]		profiling			seqüències

Taula 5.6. Propietats Mètodes d'Anàlisi Intel·ligent de Dades

En la Taula 5.6 es pot veure com la majoria de mètodes són adaptacions dels mètodes de Clustering jeràrquics o particionals ja utilitzats per dades no seqüencials.

Si, per contra, ens fixem amb el tipus de dades, veiem que la majoria treballen amb coordenades conegudes (punts x,y), mentre que pocs treballen amb seqüències de símbols (i.e. noms dels llocs).

En la taula 5.7, es pot veure els diferents tipus de serveis que es consideren en els articles que han fet un anàlisi de les dades. La majoria es centren en entendre el comportament dels usuaris, més que no pas a oferir-los dinàmicament algun servei personalitzat.

	Services location	Increase visitor satisfaction	Recommender systems	Flow analysis	Recreation behaviour	Crowd analysis	Prediction
Nº of articles	1	2	2	3	4	3	3
ref	[23]	[2][29]	[21][30]	[1][20][26]	[25][20][22][19]	[5][8][9]	[14][18][4]

Taula 5.7. Nº articles

De tota manera, com ja s'ha dit, aquest és un camp encara inexplorat, com mostra la distribució dels articles per anys en la figura 5.3.

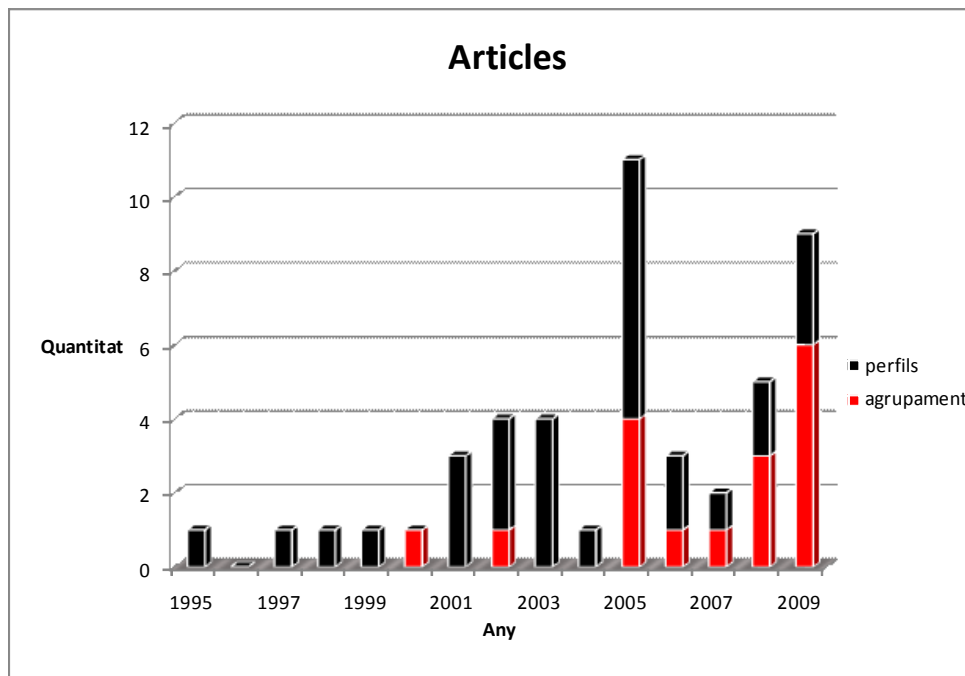


Figura 5.3. Distribució d'articles sobre anàlisi de trajectòries per anys

Després d'analitzar les dades recopilades, volem destacar les següents qüestions o reptes que encara estan per resoldre des del punt de vista general de recerca en mètodes d'anàlisi de dades de trajectòries:

Identificar quines característiques modelen millor els diferents tipus d'esdeveniments (com informació agregada, descriptors resum, co-ocurrències, etc.)

Identificar quins tipus de patrons representen millor la informació de trajectòries en cada domini en concret (punts freqüents, patrons seqüencials o regles d'associació)

Dissenyar tècniques que permetin treballar de forma adequada amb la incertesa que hi ha en les dades. Seria interessant desenvolupar mètodes que puguin treballar a nivell aproximat, per exemple, utilitzant mètodes d'anàlisi i de raonament difús.

Finalment, com pot ser usat el coneixement sobre el context de l'estudi i, en particular ontologies especialment dissenyades per a l'espai geogràfic, per extreure informació més significativa i més útil. La integració espai-temporal de l'extracció de patrons i els mètodes avançats de representació del coneixement és una línia de recerca molt interessant, però està molt poc desenvolupada.

En general, cal dissenyar i desenvolupar nous mètodes d'agrupament i predicció que donin resposta a totes les necessitats que s'han anat presentant en aquest treball. Fins ara, les solucions algorítmiques proposades adapten les tècniques de mineria de dades tradicionals i les estenen al cas de les dades seqüencials. Tal i com s'ha explicat, això simplifica el problema però alhora no explota tota la informació continguda en les dades. Per tant, el repte està en dissenyar noves tècniques específiques per aquest tipus de dades. El següent pas serà fer un marc teòric unificat per aquest domini.

5.5. Estudi de les variables a tenir en compte

En el capítol 4.2 s'han identificat quines són aquelles variables que han estat valorades pels experts com a necessàries per analitzar i comprendre el comportament turístic dels visitants en les diferents zones. En aquesta secció es classifiquen aquestes variables, es comenta la seva utilitat i es proposa una descripció i un conjunt de respostes possibles, a fi d'homogeneïtzar futurs estudis.

5.5.1. Variables en els estudis de mobilitat de turistes

Aquestes variables (o criteris) identificats s'han classificat en tres tipologies, per tal de facilitar el seu estudi: dades sociodemogràfiques, dades del viatge i dades de l'itinerari.

A la taula **Sociodemogràfica** (Taula 5.8) s'han inclòs els criteris que tenen relació amb la caracterització del turista, recollint les dades més socials, més personals i no de l'activitat en sí. Aquest conjunt de criteris pot arribar a permetre establir patrons d'activitat en funció de la tipologia del visitant. Les dades que s'han repetit més al llarg de les investigacions i/o que són més importants per a la caracterització del turista són: *Edat* (8 vegades), *Sexe* (5), *Nivell educatiu* (2), *Rol familiar* (1), *Lloc d'origen /residència i/o Nacionalitat* (5), *Ingressos* (2) i *Professió/Ocupació* (3). Aquests set criteris aporten una informació molt valuosa ja que són determinants a l'hora de la

realització de qualsevol activitat turística. L'explicació és molt senzilla: els gustos no són els mateixos el d'un nen que el d'un adult, així com el d'un enginyer o el d'un perruquer. Tampoc aniran a una mateixa destinació una família amb poc poder adquisitiu que una sola amb un alt nivell econòmic. Degut a la importància d'aquestes dades, es fa estrany que el seu ús no estigui gaire generalitzat en tots els estudis sobre el comportament turístic.

Criteri	Descripció	Respostes
Nom	Nom (i cognoms) de l'enquestat.	Ex: "Joan (García...)"
Edat	Edat de l'enquestat.	Ex: "21 anys"
Any de naixement	Any de naixement de l'enquestat.	Ex: "1987"
Sexe	Sexe de l'enquestat.	Home o dona
Nivell educatiu	Formació acadèmica de l'enquestat.	Primària, secundària, estudis superiors, etc.
Rol familiar	Posició a la família.	Pare, mare, fill, etc.
Lloc d'origen/residència i/o Nacionalitat	A on viu/d'on ve?	Estat, regió o localitat de procedència. Ex: "França"
Adreça	Direcció de la residència del país d'origen.	Ex: "Carrer X, núm. 1, 1er 1a Tarragona (Espanya)"
Ingressos	Sou (mensual/anual)	Ex: "20.000 €"
Professió/Ocupació	A què es dedica?	Ex: "Professor"
Lloc de feina	Localitat de la feina?	Ciutat. Ex: "Barcelona"
Realització de vacances últimament o intenció de fer-les properament	Ha fet vacances en l'últim any o espera fer-les?	Sí/No
A on t'agradaria viatjar?	Lloc a on es voldria passar les vacances.	Ex: "Barcelona, Catalunya"

Taula 5.8. Sociodemogràfica

La taula de *Dades del viatge* (Taula 5.9) és la que correspon a les dades recollides sobre el viatge, aquelles que pretenen recopilar tota la informació necessària per conèixer i estudiar tots els

components que han afectat, influenciat o condicionat el moviment del turista. D'aquesta manera, es pot millorar la gestió de les destinacions i adaptar-les en funció dels moviments dels visitants. Com a la taula anterior, aquí també s'observen uns criteris principals: *Acompanyant/s* (1), *Amb nens/es* (1), *Temps d'estada* (8), *Mode de transport* (6), *Propòsit del viatge* (3) i *Llocs visitats/Activitats realitzades* (6). És cert que aquestes variables també determinen l'activitat que realitzarà el turista però, sobretot, se centren en el viatge en sí i les eleccions que fa el turista durant aquest. Una vegada més, es fa difícil entendre l'absència a molts articles d'aquestes dades tan determinants.

La tercera i última taula (*Dades de l'itinerari*) (Taula 5.10) és la més concreta de totes. Només és aplicable en els casos a on s'ha estudiat l'itinerari en concret que ha seguit el visitant. Aquests casos són, bàsicament, els estudis de seguiment (*tracking*) i els de triar la ruta (*wayfinding*). Com a conseqüència, les variables que es refereixen a aquests casos seran concretes, molt específiques i necessàriament complementàries a les altres dades que s'han vist anteriorment ja que requereixen d'informació més general per a entendre el context. Només n'hi ha tres. Pel que fa a l'itinerari, hi ha diverses formes d'emmagatzemar les dades, normalment són les coordenades (x,y) dels punts per on es mou. Caldria dissenyar mètodes més eficients de recollida de dades, també per dades qualitatives (identificant el nom del lloc on est roba el visitant).

Criteri	Descripció	Respostes
Acompanyants	L'entrevistat viatja acompanyat?	Sí / No
Amb nens/es	Viatja amb nens/es?	Sí / No
Temps d'estada	Duració de l'estada en dies.	Nombre de dies
Mode de transport	Mitjà de transport utilitzat des de l'allotjament fins el lloc a visitar.	A peu, en cotxe, en bicicleta, en transport públic, etc.
Hora d'arribada	Hora d'arribada a la destinació.	Hora en concret. Ex: "16:00"
Hora de sortida	Hora de sortida de la destinació.	Hora en concret. Ex: "16:00"
Propòsit del viatge	Per què ha viatjat?	Raonament. Ex: "Perquè estic interessat per la seva cultura."
T'ha influenciat la publicitat?	Influència de la publicitat a l'hora d'escollir la destinació.	Sí/No
Primera visita o no	És la primera visita a la destinació en concret?	Sí/No

És l'única destinació del viatge?	És l'única destinació o n'hi ha més?	Sí/No
És la destinació principal?	És la destinació més important de totes a les que anirà?	Sí/No
Llocs visitats/Activitats realitzades	A on ha anat? Què ha fet?	Explicació. Ex: "He anat al Museu"
Temps gastat en cada activitat	Quant temps ha passat realitzant cada activitat?	Segons convingui: hores i minuts.
Opinió del lloc visitat	Quina opinió té?	Adjectiu qualificatiu. Ex: "bona, dolenta, etc."
Tipus d'allotjament	A on s'ha allotjat?	Hotel, apartament, càmping, etc.
Direcció/Destinacions	A on es dirigeix?	Ex: "A Taiwan"
Percepció del lloc visitat	Què li ha semblat el lloc?	Descripció qualificativa
Temps gastat en omplir el qüestionari	Quant temps ha trigat en completar el qüestionari?	Minuts

Taula 5.9. Dades de viatges

Criteri	Descripció	Respostes
Recorregut seguit	Itinerari seguit pel visitant.	Indicació de l'itinerari (descripció o realització d'un esbós).
Mètode de wayfinding	Quin mètode ha emprat el visitant per decantar-se per seguir un itinerari o un altre.	Descripció de l'elecció. Ex: "Hem seguit les indicacions als arbres".
Distància entre l'allotjament i l'activitat	Distància entre l'allotjament i la visita realitzada.	Ex: "10 quilòmetres"

Taula 5.10. Dades de l'itinerari

5.5.2 Anàlisi de les variables

S'ha observat que els criteris que es consideren en els estudis de trajectòries de turistes vénen determinats per dos factors: els objectius i l'escala de treball.

Pel que fa al primer component, **els objectius de la investigació** determinarà quines seran les variables a tenir en compte. Tot i així, algunes dades, sobretot socio-demogràfiques, acostumen a aparèixer generalment per ser imprescindibles. Per posar un exemple, la variable *Edat* és necessària per establir patrons d'activitat turística ja que no és el mateix un adult que un nen o un vell. Igualment passa amb el sexe: les dones no tenen les mateixes inquietuds que els homes i a la inversa. També es pot dir exactament igual pel que fa referència al nivell educatiu.

Si es passa de les dades socio-demogràfiques a les *Dades del viatge*, també s'observen criteris difícils d'obviar en qualsevol investigació sobre el tema en qüestió, com *Llocs visitats/Activitats realitzades*, *Temps gastat en cada activitat* i *Opinió del lloc visitat*. Si el que es vol és millorar la gestió de la destinació (objectiu principal i prioritari en tot estudi turístic), aquestes dades són essencials.

La taula de *Dades de l'itinerari* (Taula 5.10), presenta els criteris dependents de l'àmbit d'estudi i, en conseqüència, **l'escala de treball**, el segon factor que determina les variables. Mentre que les taules *Sociodemogràfica* i *Dades del viatge* fan referència a una escala més general, *Dades de l'itinerari* focalitza en una escala més petita, en rutes precises i més curtes. La majoria dels estudis llegits sobre tècniques de seguiment (*tracking*) són experimentals, és a dir, encara estan intentant establir un mètode estàndard. Això treu importància a la recopilació de dades sobre els turistes causant que només es vegin tres variables a la taula en qüestió. Per tant, fins que la metodologia no s'acabi d'establir, els objectius turístics del *tracking* seguiran sent complementaris dels anteriors.

Així doncs, prenent com a base el que s'ha dit fins ara i fent una visió més general, s'ha de dir que totes aquelles dades que influeixen al visitant a l'hora d'anar a una destinació o a una altra s'haurien de registrar sempre si es volen extreure unes conclusions reals i fidedignes sobre el comportament turístic. Potser, el debat s'instal·laria en establir quines són aquestes variables imprescindibles i en quins estudis ho són. Tot i així, es fa difícil trobar un article sobre turisme a on no importin les dades que determinen el comportament humà. A més, aquestes dades generalment són poc concretes i, com ja s'ha vist, poc repetides en els successius articles. D'altra banda, no es troben a faltar més criteris perquè els que hi ha cobreixen amb garanties la informació requerida.

Un altre punt igual o més important que l'elecció dels criteris és el seu tractament. Un bon exemple es troba en la variable *Acompanyants*, la qual admet respostes de sí o no. Si la resposta fos quantitativa (0, 1, 2, 3, etc.) es podria saber si es va acompanyar i, en el cas que sí, per quantes persones. A més, aquí es podria incloure amb qui es va acompanyar perquè anar en família, amb amics o en parella és molt diferent. Un altre exemple seria en el criteri que hi ha a continuació, el *Amb nens/es*, que podria establir, també, el nombre i l'edat, perquè les activitats realitzades pel turista canvien molt en funció de l'edat i la quantitat dels nens/es.

Com a conseqüència de tot el que s'ha dit fins ara, es proposa la creació d'uns models estàndards que estableixin els criteris a analitzar i el seu tractament amb la finalitat que serveixin com a base d'unes conclusions útils per a investigacions posteriors que segueixin la mateixa línia, a més de

possibilitar les comparacions entre els propis estudis turístics. Els models podrien agrupar projectes comuns de seguiment, de comportament, etc.

A continuació es presenten unes taules que indiquen quins treballs han utilitzat cadascun dels criteris explicats. Les referències de les publicacions són les que s'han donat al capítol 2.3.

CRITERIS SOCIODEMOGRÀFICS													
Article.	Nom	Edat	Any de	Sexe	Nivell	Rol	Lloc	Adreça	Ingressos	Professió/	Lloc de	Realització de	A on
4		x		x			x			x			
14							x						
5													
6	x					x		x	x	x	x		
7													
8		x		x	x					x			
10		x											
11		x		x	x			x	x				
12													
16													
26													
24		x	x	x			x						
13		x		x			x						
17		x					x					x	x
24		x											
TOTAL	1	8	1	5	2	1	5	2	2	3	1	1	1

DADES DEL VIATGE																		
Article. Número	Acompanyant/s	Amb nens/esd'	Temps d'estada	Mode de transport	Hora d'arribada	Hora de sortida	Propòsit del viatge	T'ha influenciat la publicitat?	Primera visita o no	És l'única destinació del viatge?	És la destinació principal?	Llocs visitats/Activitats realitzades	Temps gastat en cada activitat	Opinió del lloc visitat	Tipus d'allotjament	Direcció/ Destinacions	Percepció del lloc visitat	Temps gastat en omplir el qüestionari
4	x		x	x								x		x				
14																x		
5							x							x			x	
6			x	x								x						
7			x	x								x			x			
8			x	x			x					x						
10			x						x									
11			x									x						
12							x							x			x	
16														x			x	
26									x									
24				x	x	x												x
13			x						x	x	x							
17								x										
24		x	x	x								x	x					
TOTAL	1	1	8	6	1	1	3	1	3	1	1	6	1	4	1	1	1	1

DADES DE L'ITINERARI			
Article. Número	Recorregut seguit	Mètode de "wayfinding"	Distància entre l'allotjament i l'activitat
4	x		
14	x		
5			
6			
7			
8			
10			
11			
12			
16	x		
26		x	
24			
13			
17			
24			x
TOTAL	3	1	1

6. Conclusions i comentaris finals

Un gran nombre de dispositius que tenim a l'abast diàriament poden transmetre la nostra localització a un servidor. Això permet recollir la informació de la nostra posició al llarg del temps, de forma que es poden tenir dades dels recorreguts efectuats per una persona en un entorn tancat (un parc o una ciutat) o un entorn obert (una regió o un país).

Aquesta informació pot ser utilitzada per oferir millors serveis a l'usuari. En aquest projecte s'ha estudiat el cas concret del turisme.

Com a primera conclusió, els tres grups hem constatat que aquest és un camp de recerca bastant inexplorat, tant a nivell general com a nivell d'aplicació a l'entorn turístic.

En la documentació que teniu, s'han presentat i catalogat possibles serveis orientats tant al turista de forma individual, com a la gestió del turisme per part d'una entitat. Tots els articles, conferències, llibres i altres fonts bibliogràfiques interessants per a realitzar aquest estudi han estat catalogades a un programa d'administració bibliogràfica a la Web, anomenat *Refworks*. El material s'ha classificat segons els eixos d'estudi d'aquest projecte en diferents carpetes, que posem a la vostra disposició, prèvia petició al responsable del projecte.

D'aquest estudi podem concloure que encara hi ha pocs treballs de turisme que incloguin aquest tipus d'informació, malgrat que s'han pogut identificar nombrosos serveis interessants. En aquest estudi s'ha fet una proposta de variables o criteris que es poden considerar en aquest tipus d'estudis i que permetrien poder fer anàlisis del comportament dels turistes que puguin donar lloc a proporcionar serveis personalitzats. Creiem que aquest és un àmbit de treball molt interessant donat que avui en dia les persones estan cansades de rebre informació no desitjada. El futur passa per proporcionar la informació adequada a la persona que la necessita.

Però per oferir aquests serveis avançats a partir de dades de localització i/o seguiment, calen dues coses:

1. Disposar d'eines informàtiques que siguin capaces de recollir i analitzar grans volums de dades, tenint en compte les relacions espacio-temporals entre elles.
2. Garantir la privadesa dels usuaris d'aquest tipus de sistemes.

Aquests dos aspectes han estat tractats àmpliament en aquest projecte i els resultats s'han presentat en els capítols d'aquest document. En els dos àmbits s'han trobat estudis preliminars, que adapten tècniques ja conegudes per al cas de dades espacio-temporals. D'aquesta manera, ja hi ha disponibles alguns programes informàtics per fer anàlisis d'aquest tipus de dades. Per fer la recollida de les dades, caldria dissenyar receptors que puguin fer una recollida filtrada de les dades. En quant a l'àmbit de la privadesa, cada cop més els fòrums de discussió posen de manifest que sovint és un aspecte que no es té en compte (tal com s'ha mostrat en l'estudi dels projectes de

recerca finançats per la UE). Tanmateix, garantir la privadesa és un dret fonamental que s'ha de tenir en compte, i hem vist que hi ha iniciatives de recerca en aquest sentit.

Després d'aquest estudi creiem que seria interessant proposar un projecte de recerca finançat que tracti el tema de la provisió de serveis personalitzats als turistes garantint la privadesa. En el capítol 5 s'han plantejat diverses qüestions obertes que es poden abordar. Degut a que el camp d'aplicació concret determina tant la tecnologia com els mètodes que s'han de desenvolupar, seria recomanable centrar-se en un tipus de seguiment de turistes, decidint si s'estudiarà un entorn obert o tancat, quina serà la durada dels trajectes (quantitat de dades) i quines de les altres variables proposades es volen considerar, així com quin tipus de serveis es volen proporcionar a l'usuari (pels turistes o per planificació turística). A partir d'aquí, es podran proposar nous mètodes de recollida, anonimització i anàlisi de dades centrats en un problema més concret.