ONTHEWAY: A PREDICTION SYSTEM FOR SPATIAL LOCATIONS

Juan Antonio Álvarez[1], Juan Antonio Ortega[1]
[1]Departamento de Lenguajes y Sistemas Informáticos,University of Seville, Av. Reina Mercedes s/n. 41012 Sevilla, Spain.

Luis González[2], Francisco Velasco[2], Francisco Javier Cuberos[3]
[2]Departamento de Economía Aplicada I, University of Seville. Ramón y Cajal 1. 41018 Sevilla, Spain.

[3]Departamento de Planificación. Radio Televisión de Andalucía. José Gálvez s/n. 41092 Sevilla, Spain

Keywords: Ubiquitous Computing, context-awareness, location-enhanced.

Abstract: In ubiquitous computing we need to know the present context in order to interact properly with the nearby smart elements. When we are moving outdoors, mobile devices take a very important role because they provide us with a link between the world outside and ourselves through means of intelligent interfaces. There are a lot of situations in which it would be very useful to know or foresee the future context, i.e. as a geographic environment, where we could find ourselves in a near future, and at the same time being able to use that information from our devices. Therefore we must preview this location with enough precision and time and be able to use this information from our mobile device. In our "OnTheWay" system, we used GPS technology and databases made of past paths taken by a person, in order to predict the next location, once we had begun a new course, comparing the new one with those ones stored. The results were amazing: from the data collected about paths travelled during a month and five days, we got the actual destination in 98% of cases, when we have only made a 30,35% of the total path. Therefore, including statistic and semantic information will allow us to upgrade our results, due to the sedentary human behaviour, the small number of frequently visited locations and the fact that the paths used to arrive to these locations are usually the same.

1 INTRODUCTION

The information provided by indoor or outdoors positioning systems is very valuable; however it is not fully exploited. The most spread applications only provide accurate positioning in time and space of the object or person we would like to locate on earth. Some programs offer geocoded services that allow looking for buildings, roads or even taxis.

Nowadays, research works are oriented towards technical aspects, like the improvement of features of performance (i.e. acquisition times, positioning accuracy) or the reduction of energy consumed by receptors. The Federal Communication Committee and the European recommendation E112 have made these researches to grow a lot allowing the wireless providers to locate accurately within a dozens of meters, users who make emergency calls E911/E112. However the applications using this information are in an early stage of development.

Obtaining semantic information from wild life is being used incipiently with new applications, as in some works about tracking animals' lives, but there are several problems, like privacy or intimacy when research is focused on human life. However, these difficulties should be a technical challenge and not a handicap. Novel ideas must be planned and the use of positioning should become an instrument, not a problem.

This article explains our efforts to predict the future location of people. The work is based on their routine activities. We can realise that the principle of locality or locality of reference in computer science is valid for the human behaviour. In the most developed societies, life is lived in a sedentary and comfort ambient with moves that follow a pattern. Displacements are usually done to known places like

Antonio Álvarez J., Antonio Ortega J., González L., Velasco F. and Javier Cuberos F. (2006).
ONTHEWAY: A PREDICTION SYSTEM FOR SPATIAL LOCATIONS.
In *Proceedings of the International Conference on Wireless Information Networks and Systems*, pages 298-303

our house, work office, the house of our family, favourite cinema or the fashionable shopping centre. These facts are repeated periodically and to get to our final destination we use different means of transport. The autonomous recognition of the destinies we go to, without the interaction of people, will open our minds to new applications with which we will know the present context and the future context.

The known places and the ritual behaviour will be considered facts and our work will focus on frequently taken paths. The OnTheWay system will obtain an accurate and will foretell or foresee the place where we are going, knowing several points frequently visited and using the history of the paths taken by a person.

Possible scenarios and current work about localisation and mobile data management are explained in the next sections. In the development point we will describe the methodology used, called OnTheWay, the problems reported and the solutions created to avoid them. Once analysed the features of the system and the obtained results, we will illustrate the future context with some scenarios. Finally, conclusions and future work will be explained.

2 POSSIBLE SCENARIOS

To illustrate the power of OnTheWay system, possible scenarios will be shown where the prediction of destinations is helpful.

Tourist information: Saving the frequent tourist's journeys, the optimisation of new visits could be possible: Information about the best path, the fastest transport to use or the timetable of museums where the tourist wants to go could be notified when the destination has been predicted. Providing the same help to civil service could be helpful for citizens.

Future interest zones: Predicting that our destination is the place C, our system could track the buildings associated to our to-do-list. It could give us notice of the routes to reach marketplaces, chemist's or civil service buildings before going to C.

Prediction of traffic jams: If the probable destination of a set of motorists is actually the same and their paths pass by the same point, then calculating the number of cars that can pass in an hour, the OnTheWay system applied to all the GPS of these cars could obtain a prediction of traffic jams and notice the drivers other possible paths to avoid them.

Meetings prediction: A knowledge network could provide the share of information of relatives or friends tracking routes. Analyzing this information the probable places where some of the members of the community could meet will be possible.

Management of alerts: The possibility to track person movements can be very useful (e.g. in the case of Alzheimer or Schizophrenic) when someone gets lost in an external environment. It is possible to track and compare usual routes with current ones and decide to send an alarm before the situation might put the person in risky situations.

3 RELATED WORK

Research about enhanced localisation is until now focused to identify frequently visited places. In (Ashbrook and Starner, 2003), (Hightower, 2005), (Kang, 2004) and (Marmasse, 2000) authors describe methods based in different technologies (GPS, GSM, WiFi) to obtain algorithms that recognise a reached place during the second and successive arrivals. The object of these works is to make easy the use of context in ubiquitous computing. If we know where we are and what smart devices surround us, we could interact with the environment.

The prediction of possible destinations is made when the person begins a new journey. Moreover the possibility to foretell in advance and accurately the places he goes are in-depth approached.

In our research we have assumed the work of detecting the frequently visited places labelling them in a map through its coordinates and a textual representation of the place like we know it, i.e. we are familiar with the term "home", but we don't know that it corresponds to the coordinates 37°22'55.98"N, 5°58'14.20"W.

On the other hand, when tracking wild animal life projects like (POST project, 2002), (Pei Zhang, 2004), (VAFALCONS, 2002) and (Puma Project, 2004) positioning technology to store the journeys and extract information about animal behaviour is used. Concretely in ZebraNet project (Pei Zhang, 2004), unknown information until now has been obtained. Biologists know now that zebras explore more wooded areas and gullies at night.

Although it is not intended to track human life, it is obvious that their way of moving allow the creation of interesting and useful applications which could manage future situations obtained from predicted contexts.

Another important issue in this kind of applications is the mobile data management. In (Waluyo 2005), (Kayan, Ulusoy, 1999) and (Perich 2004), location-dependent information services, real-time data access requirements and new data management challenges are analysed.

4 DEVELOPMENT

Firstly, scope and used terms will be defined; next the used methodology there will be described.

4.1 Scope of the Problem

OnTheWay proposal is based on the similarity of journeys to foretell the place where we are going when we begin a new path. This problem depends on the spatial dimension of the course so temporal dimension will be put away because it will increase the difficulty of our problem (e.g. a journey by car compared temporally to the same path by bicycle will be temporally very different). Elected option for avoid this, was a sampling used to acquire the spatial data with a temporal frequency of one second. On this way, graphical representation will be simplified too from three to two dimensions (longitude and latitude).

4.2 Terms Definition

Concepts will be simplified on defining the next terms:

Point: Tuple composed by their geographical values of longitude and latitude.

p=(long, lat)

Place: Tuple composed by a point and a radius. Both define a circumference centred on the origin of the place. The third element is a label representing the semantic information about this place ("home", "work", etc.)

L=(p, radius, label)

Current path or current journey: Set of points obtained sampling the geographic coordinates using the positioning system. In our case was GPS (Global Positioning System). These points are validated by this system, satisfying the correct number of satellites that allow obtain an accurate position. The points of the current path are separated by not less than 30 meters to avoid redundant information. So, let

$X=\{ p_0, p_1, \ldots, p_n \}$ be the set representing a path.

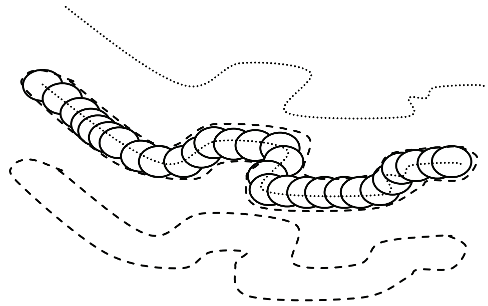

Figure 1: Generating the scope of a path.

Class: Given a place origin A and another destination B. A class will be defined as the pair A-B or B-A. Although the path from A to B will probably be different, we will suppose the same class to reduce the number of journeys needed to train the system.

Labelled path: Given a finished path, it will be labelled with its class.

$X_{A-B}=$
$X|A=(P,radius1,origin)\wedge B=(Q,radius2,destination)$

Distance between points: Although in the cities the distances are considered using the Manhattan distance, OnTheWay system uses the great-circle distance based on the spherical geometry and very important for finding the shortest distance between points on the surface of the Earth. To express it between two points it will be notated by d(pi,qj).

Scope of a path: Given a labelled path $X_{A-B} =\{ p_0, p_1, \ldots, p_n\}$ and a point *q*, this point will belong to the scope of the path if there exists some point of the path which distance to *q* is less than a given value of δ: $q \in Scope(X_{A-B})$, if $\exists\, j \mid d(p_i,q) \leq \delta$.

On our work we have determined that δ=85 meters is an adequate distance to consider. A point will be on the scope of a path if it belongs to the region generated on the figure 1.

Canonical path or representative: Given a set of labelled paths with the same class, we will choose one of more paths from this set to represent it in an accurate manner the others. The reliability will be given by the percentage of points that belongs to the scope of this path. The more paths from the class are enclosed by a canonical path, the better representative is it.

Because there could be different paths used to do a journey from A to B, more than one canonical will be accepted to represent each class.

Similarity of paths: A path matches with other if it has a percentage of points from their total that belong to the scope of the second one. The similarity level will be expressed like this:

Given $X_{A-B} =\{ p_0, p_1, \ldots, p_n \} \wedge Y_{C-D} =\{ q_0, q_1, \ldots, q_m\}$

$Similarity(X_{A-B}, Y_{C-D})=$

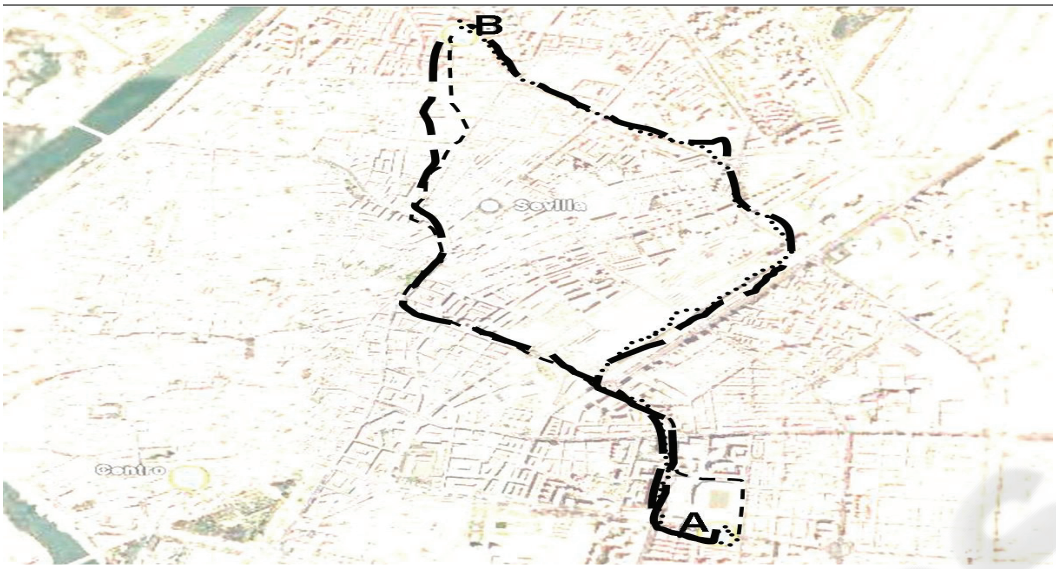

Figure 2: Four paths from A to B.

100*(Number of qi $\in$ Scope(X_{A-B}))/m

Hence, given $X_{A-B} = \{ p_0, p_1, \ldots, p_n \} \wedge Y = \{ q_0, q_1, \ldots, q_m \}$

Y will be identical to X_{A-B} if $\forall i$, $q_i \in$ Scope(X_{A-B}) and Y is labelled with the class A-B so it is a labelled path: Y_{A-B}

Figure 2 shows an example of four journeys on Seville, X^1_{A-B}, X^2_{A-B} (both on the left of image), X^3_{B-A} and X^4_{A-B} (both on the right).

All of the paths are from the class A-B. Although the places origin and destination of all are the same, for a correct tagged, OnTheWay should take two canonical paths, the first one for represent two on the right and the other for represent the two on the left. Section called selection of canonical will describe the manner to take the canonical paths.

4.3 Methodology

4.3.1 Data Retrieval

Data retrieval was done by the authors during the period from December 10th, 2005 and January 14th, 2006. The journeys were tracked in the cities of Sevilla (31 days), Almeria (3 days) and Granada (1 day). Data obtained from Granada was insufficient to any prediction but the tracked paths on Sevilla and Almeria were very useful. 103 tracked paths were obtained. To achieve this set of routes, a Royaltek Bluetooth GPS was used. Due to the impossibility of storing these tracked paths on the device, a Dell Axim X30 PDA was used to store the routes on NMEA format.

The vehicles used were only bicycles and cars.

4.3.2 Labelled

When the destination was reached, the NMEA file was renamed including the date and time of the journey, the origin and the destination. Moreover, the file was transformed to KML format, used by the GoogleEarth software to obtain an easy and fast representation. On the conversion, unreliable or erroneous points were filtered.

4.3.3 Data Cleaning and Bizarre Paths

Two kinds of problems were distinguished on the Data retrieval process: some derived from the GPS technology and others derived from the real experience.

GPS technology:

- First valid GPS point: The validation system of GPS takes some time to obtain the first correct point. If the receiver is still, the duration is about 1 minute. If the receiver is in movement the duration could be much more. On the collection of data, the first valid point sometimes was very far on space and time from the real origin of the journey. Despite this, the path was tagged with the real origin (not the first validated) and destination. These paths would not be a canonical path because the canonical ones must represent with fidelity the real path. So, we need a rule to select it checking that initial and finals points are close to the real places of origin and destination.
- "Urban canyon" effect: The global positioning satellite signal cannot be received by the

receiver if it is surrounded by tall buildings or hills. It causes interruptions of validated positions during a variable time. So it reduces the number of correct points on the track file. Although the cities where the Data retrieval did not have skyscrapers or lots of tunnels, the effect was observed on narrow roads. Because of this, we add a new rule to discriminate canonical paths from non canonical ones: Each couple of points composed by one point and the next one, have to be near. If the distance between them is more than the advanced in a car the points will not be close. Data sampling interval is 1 second in our Royaltek GPS and it was considered the possibility of having one point not validated between two NMEA validated, so due to the speed limit is 120 k.p.h. in most countries, the distance considered was 66,7 meters.

Real experience:

- In Data retrieval and the labelling process, problematic situations that could alter the selection of canonical paths were found. The main problem occurred when it was difficult to find a place to park the car during a journey. That caused the driver to become stressed-out and the increase of points collected on the surroundings of the destination place. The selection of this kind of journeys and their associated paths as canonical paths would be an error because another path would not be similar to this one even if their origins and destinations were the same. So we add another rule that specifies that if there are two paths that satisfy the last two rules, we will take the one that has fewer points.

After having verified the tagged paths, 7 were discarded because their results were totally erroneous or nonsense. So it rested 96 with checked information including the ones that had the features explained.

4.3.4 Canonical Paths Selection

Given a set of labelled paths from the place *origin=(o, r1,A)* to the place *destination=(d,r2,B)*, $X^0_{A-B}\ X^1_{A-B} \ldots X^n_{A-B}$ the selection of a canonical path $X^c_{A-B} = \{ p_0, p_1, \ldots, p_m\} \mid 0 \le c \le n$ will be done following the next rules:

- **Rule 1**. Nearby to origin and destination:

$\forall\ i \in [0,k],\ dist(p_i, o) \le r1\ \wedge\ \forall\ i \in [m-k,m],\ dist(p_i, d) \le r2$

- **Rule 2**. Uniform distribution:

$\forall\ i \in [0,n),\ dist(p_i, p_{i+1}) \le 66{,}7$ meters.

- **Rule 3.** Less number of points:

If $X^i_{A-B}\ X^j_{A-B}$ satisfy the rules 1 and 2 and *similarity*$(X^i_{A-B}\ X^j_{A-B})>80$, then it would be selected the one with less points.

- **Rule 4.** More than one canonical:

If X^c_{A-B} is a canonical path and we found another path $X^i_{A-B} \mid$ *similarity*$(X^c_{A-B}\ X^i_{A-B})<80$, then we will consider X^i_{A-B} also as a representative.

After applying this rules to the learning set of paths, we obtained 44 canonical paths and 52 non canonical.

4.3.5 Off-line Results

Once the set of canonical paths were obtained, legitimacy of the methodology was checked comparing the sets of representative and non-canonical paths. The goal was to predict the destination where the non-canonical path went to. So for a non-canonical path, his similarity index was obtained for all the canonicals, and the one that produces the greater index was selected. We compare the destination from the selected canonical path with the real destination of the non-canonical path. The results were correct in 51 of 52 cases. The incorrect case was because a canonical path was not selected when it should. These results showed us the strength of the methodology and the importance of the selection of the canonical paths. Due to the good results obtained in the off-line mode, we faced the task of obtaining the point where the correct prediction begins.

4.3.6 On-line Results

Having checked the similarity index with all the canonicals over all the points of the non-canonical path, the predicted destination in each point was obtained. The point when the real destination was predicted not being altered until the end of the path, was called detection point. Results were coherent with common sense: Until the itinerary did not enter in a not common area with other paths, it was impossible to distinguish where we go. Nonetheless, we also observed that although there were overstrike paths, often small changes taken on the paths allowed the OnTheWay system to distinguish the correct destination. The results were very hopeful because the remaining distance to the destination measured in a straight line, on average was the 69,65% of the total path and the remaining time to reach the destination was more than the 70% of the total (when the journey finally stops). It was considered that a new detection point that remained stable for some meters, could be defined as a decision point. The spatial pertinent value seemed to

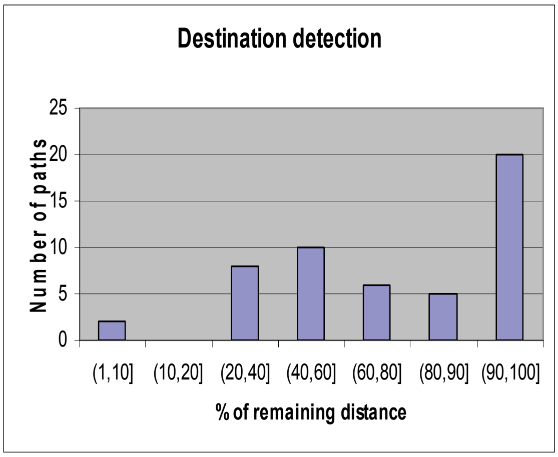

Figure 3: On-line results.

be 120 meters. This value adds these amount of time and space to the decision point, being specially harmful for the paths in which the decision arrived very late, as it causes the arrival to the destination comes before we can predict it. However, the capacity to predict future context is real and very helpful.

5 CONCLUSION AND FUTURE WORK

This work opens a new research line on the enhanced localization systems, founded on the ritually behaviour of people. Although the habits of movements of people are not the same, knowing some of the last and past routes, we can intuit the future paths. This paper exploits this intuition to show in a concrete experience, how to predict in advance the places where we will go in a nearby future.

The exercise of tracking the set of journeys which will be the canonical paths suppose that the efficiency of the methodology followed in OnTheWay will be better with more tracked journeys, but it will include some obstacles as the close places or the overstrike paths that difficult the prediction and produce the detection of decision points very close to the final destination. To avoid these disadvantages, we are planning to include semantic information and statistic models to our system. It will permit to distinguish the most probable place between some of them using the spatial, temporal, semantic and statistical information of the user. We are including also the diary book to obtain the semantic data.

Short paths done in home and indoor buildings will be researched to use as base for handicapped persons who moves slowly or in a wheelchair. In this way, we will pass to technologies used in indoor: RFID or Bluetooth. Finally, connection between indoor and outdoor positioning systems will be studied for a better use the future context. Concretely, in (Shun-Yuan, 2005) is shown a novel system to locate persons in indoor and outdoor environments founded on the relative position of shoes when someone walks.

REFERENCES

Ashbrook, D. and Starner, T. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Computing, (7):275–286, 2003*.

Census of Marine Life. (Oct. 2002) POST: Pacific Ocean Salmon Tracking Project. *http://www.postcoml.org/*, 2003. (ASPLOS-X).

Hightower, J. et Al. (Sep. 2005) Learning and Recognizing the Places We Go, In *Proceedings of the Seventh International Conference on Ubiquitous Computing (Ubicomp 2005), pp. 159-176.*

Kang, J.H., Welbourne, W., Stewart, B., Borriello, G. (2004): Extracting places from traces of locations. In *Proceedings of the Second ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH 2004), Philadelphia, PA, ACM Press 110–118*

Kayan E, Ulusoy O. (1999) An evaluation of real-time transaction management issues in mobile database systems. *Computer Journal 42 (6): 501-510*

Marmasse, N., Schmandt, C. (2000) Location-aware information delivery with commotion. In *Proceedings of the Second International Symposium on Handheld and Ubiquitous Computing (HUC). Volume 1927., Springer-Verlag 151–171*

Pei Zhang et Al., (2004) Hardware Design Experiences in ZebraNet. Department of Electrical Engineering Princeton University, Sensys 2004

Perich F et Al, (2004) On data management in pervasive computing environments. *IEEE Transactions on Knowledge and Data Engineering 16 (5): 621-634 May 2004*

Shun-Yuan Yeh et Al. (2005) Geta sandals: knowing where you walk to. *Ubicomp 2005*

The Center for Conservation Biology. (2002)VAFALCONS. *http://ccb-wm.org/vafalcons/falconhome.cfm*

UC Davis Wildlife Health Center. Southern California (2004) Puma Project. *http://www.vetmed.ucdavis.edu/whc/*

Waluyo AB et Al. (2005) Research on location-dependent queries in mobile databases. *Computer Systems Science and Engineering 20 (2): 79-95 Mar 2005*