Augmenting Mobile Localization with Activities and Common Sense Knowledge

Nicola Bicocchi[1], Gabriella Castelli[2], Marco Mamei[2], and Franco Zambonelli[2]

[1] Dipartimento di Ingegneria dell'Informazione, Universitá di Modena e Reggio Emilia
[2] Dipartimento di Scienze e Metodi dell'Ingegneria, Universitá di Modena e Reggio Emilia
name.surname@unimore.it

Abstract. Location is a key element for ambient intelligence services. Due to GPS inaccuracies, inferring high level information (i.e., being at home, at work, in a restaurant) from geographic coordinates in still non trivial. In this paper we use information about activities being performed by the user to improve location recognition accuracy. Unlike traditional methods, relations between locations and activities are not extracted from training data but from an external commonsense knowledge base. Our approach maps location and activity labels to concepts organized within the ConceptNet network. Then, it verifies their commonsense proximity by implementing a bio-inspired greedy algorithm. Experimental results show a sharp increase in localization accuracy.

1 Introduction

Ambient intelligence services ubiquitously need to perceive and understand their operating environment to achieve adaptable and flexible behaviors. The recent introduction of powerful mobile platforms allowed a plethora of innovative user-centric services [18,12] to be developed. These services, considered as a whole, aim to improve users' experience by making use of contextual information gathered from the environment.

Location is a key aspect concurring in context definition. The embedding of GPS receivers in modern mobile devices allows millions of people to acquire their position is terms of geographical coordinates. However, this representation is not actually informative. Applications, and consequently users, make greater use of higher level representations such as being at *home*, at *work*, in a *grocery shop* or in a *bank*.

A straightforward way to achieve this result is to feed a geocoding service with a GPS signal. For example, given a point and a search radius, Google Maps returns a list of locations of interest contained within the searched area. Unfortunately, to avoid false negatives (i.e., locations outside the search area) and deal with GPS inaccuracies, the search radius must range around 250m [15]. Due to this, especially in dense urban settings, the number of retrieved locations is often not negligible. Furthermore, urban environments are extremely patchy. Even a small (few meters) physical movement might imply logically different contexts (opposite sides of the wall separating the coffee shop and the bank). To improve localization accuracy (i.e., filtering irrelevant results), Ofstad et al. argued to perform localization across both *physical* and *logical* domains [16].

In this paper we propose to use recent activity recognition results [5] paired with commonsense reasoning to augment localization. Specifically, locations are filtered on

D. Keyson et al. (Eds.): AmI 2011, LNCS 7040, pp. 72–81, 2011.

the basis of their commonsense proximity with the activities performed by the user. How to collect relations between activities and locations? A well-known solution is based on collecting large bodies of training data. Then, eventual correlations among different features can be observed. However, this approach is inefficient and time-consuming. Alternatively, we propose to extract well-know relations from an external commonsense knowledge base. Specifically, a bio-inspired algorithm has been used to compute proximity among activities and locations within ConceptNet.

This paper presents the following contributions and insights: *(i)* it proposes to augment localization with activity recognition capabilities; *(ii)* it proposes to use external commonsense knowledge sources to extract relations between activities and locations; and *(iii)* it discusses experimental results collected from a real-world case study.

Accordingly, the rest of the paper is organized as follows: Section 2 discusses other works concerning augmented localization. Section 3 illustrates both localization and activity recognition modules that have been implemented. Section 4 describes how commonsense reasoning can be used to augment localization with user activities. Section 5 details experimental results under different configurations. Finally, Section 6 concludes the paper.

2 Related Work

The spreading of location-based applications for mobile phones made localization an exciting research topic during the last decade. Since early 2000s, due to its high energy requirements and poor accuracy in indoor and dense urban areas, GPS localization has been augmented with WiFi and/or GSM data [17,7].

During the same years, the conceptual switch from "positions" to "places" was discussed. The idea has been initially proposed in [10], while later works realized probabilistic engines to identify visited places on the basis of temporal patterns [13,4].

Considering "places" instead of "positions" quickly grounded embryonic ideas about context-aware localization; that is, augmenting localization with contextual information gathered from alternative sensors. In 2004, Bao and Intille showed how to detect 9 everyday activities using a couple of biaxial accelerometers that could have been easily embedded in a mobile phone [3]. After that, numerous localization techniques using contextual data have been developed. For example, [1,2] discussed the use of light and sound sensors and delineated the concept of ambient fingerprinting; while [16,11] illustrated how to embed information coming from cameras and accelerometers. However, integrating alternative sensors within the localization process posed the problem of extracting relations between different domains (e.g., accelerations and geographical coordinates). The majority of existing works solved this problem by searching relations within training data. This approach, showed to be prone to over-fitting, is sensitive to application domain and requires more data to initialize the system [8].

Recently, few attempts have been made to overcome these problems. In particular, instead of extracting relations among different domains from the data itself, some authors tried to use pre-existing, external, commonsense knowledge bases. In [15], for example, Mamei et al. used Cyc to improve automatic place identification considering historical data about the user.

3 Sensor Setup

To develop our activity-aided localization system, a GPS-based localization sensor and an accelerometer-based activity sensor have been implemented. While the former has been designed to natively run on Symbian smart phones, the second requires dedicated hardware (i.e. SunSpot nodes). However, our results abstract from specific implementation details and could be reproduced with alternative systems with a comparable accuracy. Both sensors are described below in this section.

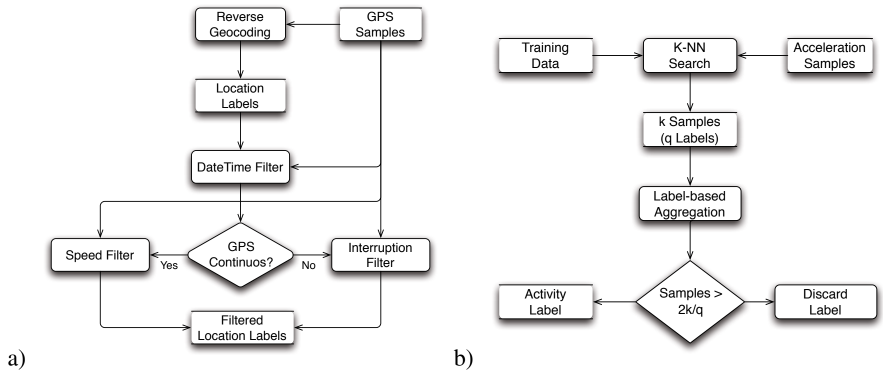

Fig. 1. (a) *Location classification process*; reverse geocoded labels are filtered on the basis of time, date, speed and eventual interruptions. (b) *Activity classification process*; for each sample, k nearest neighbors (associated to q classes, $k = 64$, $q <= k$) are identified. The sample is then associated to all the classes (at most 3) associated to al least $k/2q$ training samples.

3.1 Location Recognition

To classify user's location we implemented a tool [15,9] for Symbian OS. It samples GPS coordinates and classifies user's location by querying Google Maps. Specifically, Google Maps takes in input a couple of geographic coordinates and a radius, returning a list of locations of interest associated with a label coming from a predefine set (i.e., road, square, park, shop, cinema, mall, restaurant, gym). Unfortunately several practical drawbacks affect this process:

1. Smart phones are not equipped with high-precision GPS receivers. Under normal operating conditions this error is smaller than 100m [15], however the error can reach 200m.
2. Google Maps database is not perfect. Although we do not have accurate statistics, we noticed that a portion of locations is still missing and coordinates might be unprecise. Furthermore, Google Maps does not provide information about locations' geometry. Due to this, especially with large-sized instances (e.g., parks, squares) locations can be misclassified.

To mitigate these problems and avoid false negatives, the system has been setup to use a search radius of 250m. Clearly, the number of reverse geo-coded locations is proportional to the search radius. The bigger the radius, the more the returned location labels. Because of this, especially in densely populated areas, the system might produce numerous false positives. To reduce them, we implemented three filters acting on different dimensions of the GPS signal (see Figure 1(a)).

DateTime filter acts on the assumption that each label is more likely to be visited during defined portions of the week. For example, banks are closed during the night, while cinemas are unlikely to be visited during the morning. Thus, we associated to each label a probability distribution (i.e., 24-7) describing how likely that category of places is going to be visited.

Speed filter works on continuous GPS signals (suggesting an outdoor location). A common misclassification happens when a user moving on a street is associated to all the locations she goes by. To avoid this, the filter analyzes user's speed. If the user is moving, only *road*, *park* and *square* are allowed.

Finally, Interruption filter works on discontinuous GPS signals indicating, with high probability, an indoor location. It works on the assumption that each category of places is fairly characterized by the duration of the visit. Thus, we defined for each category a probability distribution of durations.

3.2 Activity Recognition

To classify user's activities we made use of the system detailed in [5]. Here we provide only an informal introduction. It collects data from 3-axis accelerometers, sampling at 10Hz, positioned in 3 body locations (i.e., wrist, hip, ankle) and classifies activities using instance-based algorithms. To initialize the system, each user is required to collect a certain amount (i.e., 250-500) of training samples for each activity. Furthermore, considering that human activities have a minimum duration, it aggregates classification results over a sliding window and performs majority voting on that window. Each window is associated with the most frequent label. For the sake of experimentation, we modified this module in two ways:

1. First, we implemented both training and classification modules on Sun Spot nodes. Instance-based algorithm perfectly suit this need in that they support on-line classification and training and can be implemented on resource-constrained devices. Client nodes send their samplings to a master node which classifies them and stores the result. This way, it is possible to discard raw samplings and store only high-level activity labels, allowing the execution of 4+ hours experiments without using heavy and obtrusive equipment.
2. Second, we modified it to deal with uncertainties. Instead of producing a single label for each sensor sampling, we implemented a mechanism to produce multiple labels associated with a degree of confidence. Specifically, for each sample to be classified, k nearest neighbors (associated to q classes, $k = 64$, $q <= k$) are identified. The sample is then associated to all the classes (at most 3) associated to al least $k/2q$ training samples.

The architecture of this classification process has been summarized in Figure 1(b).

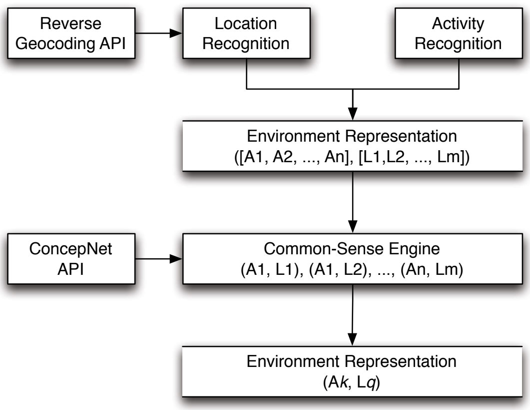

Fig. 2. Location and activity labels concur to define a representation of the environment. Each and every couple *(location, activity)* is ranked on a commonsense basis using ConceptNet assertions.

4 Augmenting Localization

The two sensors described in Section 3, periodically output classification labels. Specifically, the localization sensor outputs a tuple of candidate locations (i.e., road, square) , while the activity sensor outputs a tuple of candidate activities (i.e., walk, read).

Our approach for augmenting mobile localization is based on the assumption that locations and activities are semantically tied and commonsense knowledge can be used to measure proximity among them. The more an activity and a location are proximate, the more it is likely they have been recognized within the same situation. In other words, given a set of possible locations and activities, our goal is to rank each and every couple *(location, activity)* on a commonsense basis (see Figure 2).

To measure the commonsense proximity between locations and activities there is the need for *(i)* a knowledge base containing relations between locations and activities have to be selected; and *(ii)* a method for computing proximity among these concepts.

4.1 ConceptNet

First, since we are dealing with a situation recognition problem, the ideal knowledge base should: *(i)* include a vocabulary covering a wide scope of topics, and *(ii)* incorporate tricky relations hard to be discovered in an automatic way. ConceptNet best suits these requirement. It is a semantic network designed for commonsense reasoning. It has been built from a collection of 700,000 sentences provided by thousands of people. It is organized as a massive directed and labelled graph made of about 0.3 million nodes and 1.6 million edges, corresponding to concepts and relations between them, respectively. Most nodes represent common activities or chores given as phrases (e.g., “drive a car” or “buy food”). Its structure is uneven, with a group of highly connected nodes, and “person” being the most connected, having in- degree of about 30,000 and out- degree of over 50,000. There are over 86,000 leaf nodes and approximately 25,000 root nodes. The average degree of the network is approximately 4.7.

4.2 Measuring Semantic Proximity

A preliminary round of experiments with ConceptNet led us to identify the following principles:

1. Proximity increases with the number of unique paths. However, this is not a reliable indicator given that even completely unrelated concepts might be connected through long paths or highly connected nodes.
2. Proximity decreases with the length of the shortest path; nodes connected directly or through some niche edges are in a short distance, hence they are proximate;
3. Connections going through highly connected nodes increase ambiguity, therefore proximity should be inversely proportional to the degrees of visited nodes;
4. ConceptNet has been created from natural-language assertions. Thus, errors are frequent and algorithms have to be noise-tolerant;

Majewski et al. recently proposed an interesting algorithm for commonsense text categorization inspired to similar observations [14]. Despite it has been conceived for a different problem, it can be applied to localization as well. The algorithm is based on the assumption that proximity among concepts (i.e., in our case locations and activities) is proportional the to amount of some substance s that reaches the destination node v as a result of injection to node u. The procedure has been built around two key biological paradigms such as *diffusion* and *evaporation* and works as follow:

1. a given amount of substance s is injected to a node u;
2. at every node, a fraction α of the substance evaporates and leaves the node;
3. at every node, the substance diffuses into smaller flows proportional to the out degree of the node;
4. nodes never overflow. If multiple paths visit the same node, the previous amount of substance s can be incremented;
5. target nodes are ranked according with the amount of substance s received.

Figure 3 exemplifyies the algorithm in action. A certain amount (i.e., 128) of substance s is injected into a node representing a candidate activity (i.e., *Run*). Then, the substance diffuses over the graph and halves by evaporation at each node it visits. The overall amounts of s that reach nodes *Park* and *Road* respectively are 36 and 16. *Park* is considered more proximate than *Road* to *Run*.

Finally, it is interesting to note how this bio-inspired algorithm matches with the principles we deduced from our preliminary studies on ConceptNet. In fact: *(i)* the evaporation process assures that short paths imply high proximity; while *(ii)* the diffusion process takes into account the total amount of connections among two concepts while diminishing the relevance of highly-connected paths.

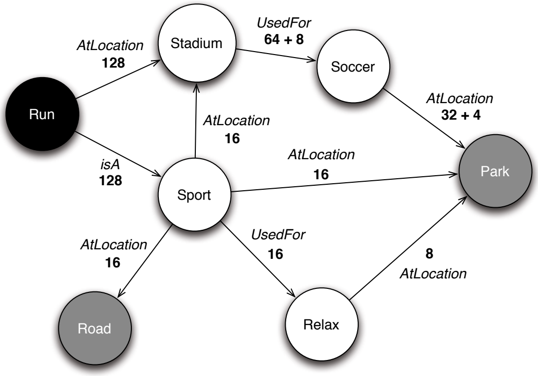

Fig. 3. An example of the proposed algorithm in action. Nodes and edges are associated with their respective weights.

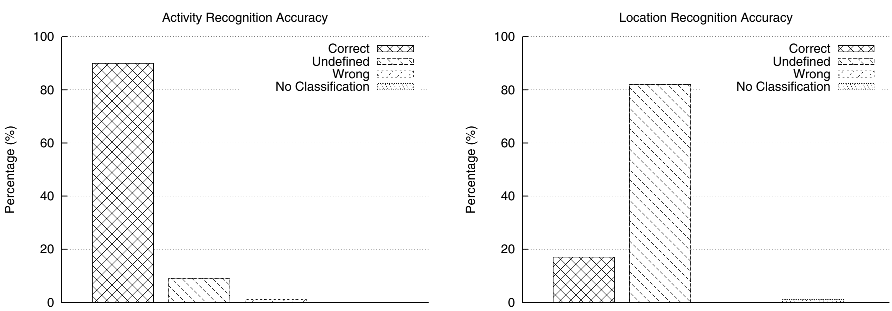

Fig. 4. Experimental results of the Activity Recognition and Location Recognition module stand alone respectively. While the Activity Recognition (a) provides reliable classifications, the Location Recognition (b) can rarely provide a correct classification.

5 Experimental Results

In this section we discuss the experimental results obtained. A volunteer equipped with the sensing system collected data while going about his normal life and manually annotating ground truth data.

The activity recognition module has been trained to recognize 8 activities (i.e., climb, use stairs, drive, walk, read, run, use computer, stand still, drink). For each class, 300 training samples have been collected. The location recognition module, instead, has been run on a Nokia N95 smartphone. GPS traces has been collected using a sampling period of 30 seconds and has been associated with a label coming a predefined set (i.e., road, square, park, shop, cinema, mall, restaurant, gym).

Experimental results have been analyzed considering to the following four categories: *Correct Classification* for the situations in which both modules provide a single and correct result that can be easily put together; *Undefined Classification* for accounting situations in which one or both the module provide more than one result, in this case there is the need for an additional mechanism to choose the most proper activity and/or location; *Wrong Classification* for situations in which either a classifier or both of them provided a wrong classification; *Missing Data* for accounting situations in which the GPS or the accelerometers didn't provide data to be classified.

We first discuss the results obtained independently by activity recognition and the location recognition systems. The activity recognition produce reliable classifications (see Figure 4 (a)) . Indeed around 90% of the samples are correctly classified, while 9% are undefined (undefined results may occur for actions that generate alike patterns of accelerations) and only the remaining 1% classifications are wrong.

On the other hand, the location recognition module doesn't exhibit a comparable level of reliability. The results of this experiment are reported in Figure 4 (b), and they actually show that in the majority of cases (82,4%), the algorithm provides an undefined classification because many locations are ranked equally likely. Analyzing undefined data points, it emerges that the the module can filter out few categories.

The combined classifier uses these data and combines them with the use of ConceptNet in order to improve the classification accuracy. It classifies data collected and processed from acceleration sensors and GPS off-line and works with two separate files that must be synchronized and streamed in parallel in order to get corresponding feature vectors and places in a given time window. Four situations can occur: *(i)* both data about the action and the place are available for the given time slot; *(ii)* only the data about the action is available for the given time slot; *(iii)* only the data about the place is available for the given time slot; *(iv)* no data available for the given time slot.

The first situation make use of ConceptNet to combine data coming from sensors and improve the context recognition result with the common sense exploitation. In both the second and the third situation only a data source is available. In these cases the common sense can be used to identify a possible place or action respectively in order to complete the action-place couple. Finally in the last situation, no data is available and different processing based on predictions can be performed. In the reported experiments we focus on the first situation, that is the most common ones.

Figure 5 (c,d) shows the results of the combined classifier and compares them with a basic system composed by the action recognition and the location recognition modules but without the Common Sense glue. In the basic system (see Figure 5 (c)) the majority of results are in the Unclassified category, mainly due to the location recognition module's output. This proves that the is there is the need for a mechanism to effectively put together data coming from diverse classifiers. The combined system (see Figure 5 (d)), instead, shows a significant improvement with 75% of data correctly classified and 25% of wrong classifications. It is worth noticing that the Undefined Classification category is lowered to zero meaning that ConceptNet is always capable of providing a ranking of action-place couples. Also the Missing data category is lowered to zero, in fact one of the advantage of the use of ConceptNet is to provide missing data. Please note that in our experiment we never experienced the concurrent lack of both sensorial data, that should have called for different strategies similar to activity and location prediction, such as bayesian networks [4]. Overall, these preliminary experiments prove that:

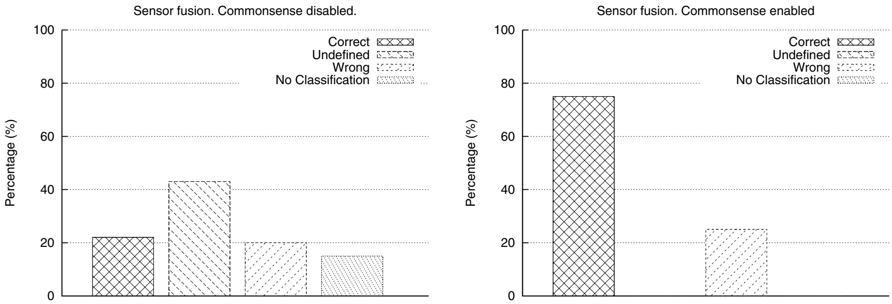

Fig. 5. Experimental results combining the Activity Recognition and Location Recognition systems. In particular, (c) shows that there is the need for some intelligent mechanism to combine the data, and (d) proves that CommonSense can be effectively used in this application.

1. Common Sense can provide an effective basis to combine contextual data coming from different classifiers
2. Common Sense is useful in a number of likely to happen situations, such as either when classifiers outputs more than a classification label or when a label is missing
3. The better the accuracy of the classifications module combined together through common sense reasoning, the better the overall results.

6 Conclusions

Although pervasive services require to perceive (i.e., classify streams of data) their operating environment, current classifiers are still inaccurate and unreliable. In this paper we presented a novel approach that combines well established classifiers using the ConceptNet knowledge base. User's activities and their relations with locations have been used to improve localization accuracy. The approach has been discussed through a realistic case study and encouraging results have been presented.

Acknowledgment. Work supported by the ASCENS project (EU FP7-FET, Contract No. 257414).

References

1. Azizyan, M., Choudhury, R.R.: Surroundsense: mobile phone localization using ambient sound and light. SIGMOBILE Mob. Comput. Commun. Rev. 13, 69–72 (2009)
2. Azizyan, M., Constandache, I., Roy Choudhury, R.: Surroundsense: mobile phone localization via ambience fingerprinting. In: Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, MobiCom 2009, pp. 261–272. ACM, New York (2009)
3. Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data, pp. 1–17. Springer, Heidelberg (2004)

4. Bicocchi, N., Castelli, G., Mamei, M., Rosi, A., Zambonelli, F.: Supporting location-aware services for mobile users with the whereabouts diary. In: Proceedings of the 1st International Conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications, MOBILWARE 2008, pp. 6:1–6:6. ICST, Brussels (2008)
5. Bicocchi, N., Mamei, M., Zambonelli, F.: Detecting activities from body-worn accelerometers via instance-based algorithms. Pervasive and Mobile Computing 6(4), 482–495 (2010)
6. Brush, A.B., Karlson, A.K., Scott, J., Sarin, R., Jacobs, A., Bond, B., Murillo, O., Hunt, G., Sinclair, M., Hammil, K., Levi, S.: User experiences with activity-based navigation on mobile devices. In: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI 2010, pp. 73–82. ACM, New York (2010)
7. Chung Cheng, Y., Chawathe, Y., Lamarca, A., Krumm, J.: Accuracy characterization for metropolitan-scale wi-fi localization. In: Proceedings of Mobisys 2005, pp. 233–245 (2005)
8. Duong, T., Phung, D., Bui, H., Venkatesh, S.: Efficient duration and hierarchical modeling for human activity recognition. Artificial Intelligence 173, 830–856 (2009)
9. Ferrari, L., Mamei, M.: Discovering daily routines from google latitude with topic models. In: IEEE International Conference on Pervasive Computing and Communications, Workshop on Context Modeling and Reasoning. IEEE Computer Society, Washington, DC, USA (2011)
10. Hightower, J.: From position to place. In: Proceedings of The 2003 Workshop on Location-Aware Computing, pp. 10–12 (October 2003)
11. Jung, D., Teixeira, T., Savvides, A.: Towards cooperative localization of wearable sensors using accelerometers and cameras. In: Proceedings of the 29th Conference on Information Communications, INFOCOM 2010, pp. 2330–2338. IEEE Press, Piscataway (2010)
12. Krumm, J.: Ubiquitous Advertising: The Killer Application for the 21st Century. IEEE Pervasive Computing 10(1), 66–73 (2011)
13. Liao, L., Fox, D., Kautz, H.: Extracting places and activities from gps traces using hierarchical conditional random fields. Int. J. Rob. Res. 26, 119–134 (2007)
14. Majewski, P., Szymański, J.: Text Categorization with Semantic Commonsense Knowledge: First Results. In: Neural Information Processing, pp. 769–778. Springer, Heidelberg (2008)
15. Mamei, M.: Applying commonsense reasoning to place identification. IJHCR 1(2), 36–53 (2010)
16. Ofstad, A., Nicholas, E., Szcodronski, R., Choudhury, R.R.: Aampl: accelerometer augmented mobile phone localization. In: Proceedings of the first ACM International Workshop on Mobile Entity Localization and Tracking in GPS-Less Environments, MELT 2008, pp. 13–18. ACM, New York (2008)
17. LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G., Schilit, B.: Place Lab: Device Positioning Using Radio Beacons in the Wild. In: Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 116–133. Springer, Heidelberg (2005)
18. Zheng, Y., Zhang, L., Ma, Z., Xie, X., Ma, W.-Y.: Recommending friends and locations based on individual location history. ACM Trans. Web 5, 5:1–5:44 (2011)