Processing Location Data for Ambient Intelligence Applications

Samuel del Bello[1], Jared Hawkey[2], Sofia Oliveira[2], Olivier Perriquet[2], and Nuno Correia[1]

[1] CITI and DI/FCT/UNL, Quinta da Torre, 2829 - 516, Caparica, Portugal
sdelbello@gmail.com, nmc@di.fct.unl.pt
[2] CADA, Ed. Interpress, Rua Luz Soriano, 67, 3º, Sala 43, 1200 - 246, Lisbon, Portugal
jaredhawkey@gmail.com, sofiaoliveira@cada1.net, olivier@perriquet.net

Abstract. The paper presents contributions in the area of location data processing for pattern discovery. This work forms part of a project which explores an ambient intelligence application designed to present individual users with an overview of their time usage patterns. The application uses location data to build interfaces and visualizations which highlight changes in personal routines, with the aim of stimulating reflection. Data is processed to extract significant places and temporal information about them. The paper presents the questions that can be answered by a data processing layer and the strategy to handle the different types of queries. Location data is processed to identify significant locations, discover patterns and predict future behavior.

Keywords: Ambient Intelligence, Location Data, Clustering, Visualization.

1 Introduction

Mobile devices can capture different types of sensor data, including location and motion, which may be used in context aware applications. Since such devices are carried almost continuously by their users extended data sets maybe readily acquired. The Time Machine project explores personal location data within an artistic context. The goal is to provide each user with an overview of their time usage, identifying routines and extraordinary events. It aims to provide a means for reflection and attempts to present ones habits and personal uses of time. A major component of the project is the construction of mobile interfaces with advanced visualization capabilities. Input data, based on GPS logs, has to be processed to extract relevant and meaningful patterns [1]. This paper focuses on this data processing layer, an essential component for presenting information in meaningful ways. The data processing layer includes a combination of techniques chosen in accordance with the visualization and interface needs, identified as a set of questions the system should be able to answer. A key concern is that the system should run on a mobile device, without resorting to external processing or storage servers, thereby introducing additional constraints on tools used. The paper is organized as follows. The next section presents the

S. Gabrielli, D. Elias, and K. Kahol (Eds.): AMBI-SYS 2011, LNICST 70, pp. 66–69, 2011.

requirements for a data processing layer as a set of questions/queries the system should be able to handle. Section 3 describes the specific technique being implemented, including the identification of significant locations, the discovery of patterns and the prediction of future behavior. The paper ends with conclusions and directions for future work, regarding its integration within a mobile system.

2 Data Processing Requirements

The architecture of the application, developed within the Time Machine project, relies upon a data processing layer that handles location data and provides services to the upper visualization and interaction layers. The requirements for this layer were defined as a set of questions/queries and visualization requirements. This section presents a summary of the processing needs identified. The questions are high level, closer to what a user might request, may contain ambiguities and sometimes need to be decomposed across several data processing functions.

Different time/places/speed usage show different patterns. Here a major concern is to show "different", or extraordinary time usage behavior. The project aims to calculate and represent personal cycles of time, and incorporate what might be considered the natural/human limits of time. Thus, the system will collect data in an ambient manner. The concerns above and the desired functionality to show the routine in terms of the variables being processed: time, speed and places visited, may be further specified and used to provide answers to the following queries, divided in four groups:

1. Is this a new place? Is this a significant place, home/work/other? At which places do I spend more time? Have I travelled a lot today? Which places do I frequent more often?
2. Is this my (usual) pattern for the time of day? Is this my (usual) pattern for this place? Is this sequence of places usual?
3. Is this day different from the norm? Was this a calm/busy day (morning/afternoon)? Was this a long/short day?
4. Where will I go next? Where will I be tomorrow/next week/next month? When will I return to this place?

Each of the above groups will need a different set of tools for location data processing, as described in the next section.

3 Data Processing Tools

This section describes the processing techniques currently being used and the results so far obtained. Some of the questions above, e.g., "Is this a new place?" result from simple data processing, while others like "Is this my (usual) pattern for this place?" require more complex techniques. Preliminary processing aims to find the significant places from the noisy data captured by GPS. Here, raw location data, containing all the data collected by the user, is parsed to extract the approximate semantic notations of location. Fig1 (a), shows such data from a user log as place marks in space.

GPS logs are first parsed into stay points as explained in [3]. This step eliminates non-relevant places and defines those places where the user actually spends time. Fig1(b) shows the stay points parsed from the raw locations. To find the meaningful places for the user, the stay points are clustered using an approximation to the density-based algorithm, DJ-Cluster [4]. Significant places from the previous stay points are shown in Fig1(c). Information about the sequence of places visited is also kept.

Fig. 1. Two weeks of an user location history: (a) all GPS points from the logs; (b) parsed stay points; (c) meaningful places

Simple statistical data processing using the different dimensions measured provides promising results. The data was processed as described below:

- Daily activity: A time based representation showing whether the individual is moving or stopped. This allows the recognition of daily time patterns, including major changes between states (sleep/awake), and the identification of activity periods.
- Duration and number of visits associated with a location: to identify the most relevant places.
- Features associated with a day: here the idea is to experiment with and study different features of individual days in order to define their signatures, as a first step to cluster and classify days. Currently, the features being tested are the number of visits average/standard deviation, the duration average/standard deviation, stopped time, distance, speed average/standard deviation, and number of visited locations.
- Features associated with an hour: The user behavior can be compared over several days for the same period of the day. The features being tested are distance, speed average, and number of locations visited.

This statistical processing is used to directly support some of the queries mentioned above but also as input for further clustering and classification of the data. Initial approaches to classify and cluster the data have been completed. In concrete, three experiments took place: classify days as weekend/weekday, cluster days and cluster locations. Days were characterized by the durations and frequency of visits to the most relevant locations for that day. These days, labelled as weekends or weekdays were, in some cases, correctly classified in 95% of the cases using cross validation.

The Naïve Bayes classifier was used in this experiment. The X-means algorithm was used for the clustering experiments, due to its efficiency and the advantage of not needing number of clusters as an input; durations and number of visits where also used to characterize the days and the locations. Results show that the days were clustered as expected in many cases but the signature for clusters is still inefficient and further data and tests are required. All experiments were done using the Weka[1] platform.

4 Conclusions and Future Work

The paper presents ongoing work on data processing for a mobile system to show individual time usage patterns. The requirements were identified as a set of queries that the system should answer and current work focuses on testing the processing algorithms. Most of the techniques were identified and tested but further tests with extended sets of data are required. Currently, the project has one dataset for one user with over a year of daily logs, and four shorter data sets each with a few weeks of data. Next, the integration of final versions of the chosen techniques will be developed for the mobile device. With this data processing layer in place, visualizations and interaction options that go beyond the initial proposal [2] will be developed. Regarding additional processing, work is also underway to predict future behavior, using Markov models.

Acknowledgments. We thank DGArtes (The Directorate-General for the Arts of the Portuguese Ministry of Culture) and FCT/MCTES for funding this research as part of the Time Machine project (PTDC/EAT-AVP/105384/2008).

References

1. Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across multiple users. Personal Ubiquitous Computing 7, 5 (2003)
2. Correia, N., Rodrigues, A., Hawkey, J., Oliveira, S., Amorim, T.: A Mobile System to Visualize Patterns of Everyday Life. In: ISAmI 2011 - International Symposium on Ambient Intelligence, Salamanca, Spain (2011)
3. Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., Ma, W.: Mining User Similarity Based on Location History. In: Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 1–10. ACM, New York (2008)
4. Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L.: Discovering Personal Gazetteers: An Interactive Clustering Approach. In: Proceedings of the 12th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, New York (2004)

[1] `http://www.cs.waikato.ac.nz/ml/weka`