

GeoLife: Managing and Understanding Your Past Life over Maps

Yu Zheng, Longhao Wang, Ruochi Zhang, Xing Xie, Wei-Ying Ma

Microsoft Research Asia, 4F Sigma Building, No. 49, Zhichun Road, Beijing, 100080, P.R. China

{yuzheng, v-lowang, v-ruzhan, xingx, wyma}@microsoft.com

ABSTRACT

The increasing popularity of GPS device has boosted many applications where more and more GPS logs have been accumulating continuously. Managing and understanding the collected GPS data are two important issues for these applications. On one hand, by indexing the increasing GPS data, we can provide effective retrieval method for users to find the corresponding GPS data interests them. On the other hand, by understanding user's GPS data, we are more likely to enable novel services which would stimulate people's passion on contributing GPS data in turn. However, so far, GPS data are still used directly without much understanding. In our project, referred to as GeoLife, we focus on visualization, organization, fast retrieval, and effective understanding of GPS track logs for both personal and public use. It not only provides a powerful platform for people to effectively manage their GPS data but also help them well understand a person's past experience from GPS data.

1. INTRODUCTION

With decreasing price and increasing locality accuracy, GPS devices such as GPS phones have become prevalent in modern life. Hence, as never before, lots of GPS log data have been accumulated both continuously and unobtrusively. The large volumes of GPS data have given rise to a generation of novel applications [1, 2] on the Web. These applications allow users to upload, share and browse GPS track logs and associated multimedia content over Web maps. So far, however, GPS data has been used directly without much understanding in these applications. Moreover, existing search method by tags, like activity tags and region tags [1, 2], offered by these Web applications cannot satisfy users' needs either. To address these issues, we start project GeoLife [3, 4] where GPS data are well managed and understood. As compared to current technologies, the novelties of our work lie in that:

- Visualizing GPS data over digital maps and display the past events lively.
- Indexing the GPS trajectories submitted by users based on their uploading behavior.
- Searching GPS trajectories over digital maps using a spatial range and/or a temporal interval as query.
- Understand people's life pattern from raw GPS data.

2. FRAMEWORK

As shown Figure 1, in GeoLife, users can upload their GPS logs as well as associated multimedia data to the system for personal and/or public use depending on their own will. After parsing the received files, we tag multimedia content with corresponding GPS coordinates where they are taken. Then, based on user behavior of uploading GPS trajectories, we build spatial-temporal index over the parsed GPS data for fast retrieving GPS tracks over maps. I.e. given a spatial range over maps and/or temporal interval that a user is interested in, our system will retrieve all the GPS tracks across the spatial range and/or temporal interval. (Refer to Figure 3)

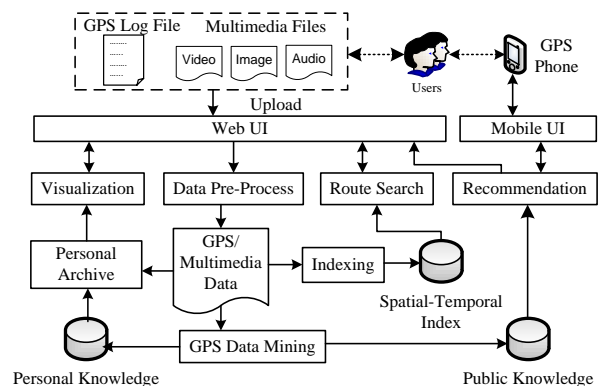


Figure 1. Architecture of GeoLife

For personal use, we help each user archive his/her own historical data from which we can mine a lot of knowledge such as personally transportation routines, significant places, life pattern etc. Further, the knowledge is leveraged to help users summarize their own travel/sports experience and acquire healthy habit in daily life. From the public data, we learn the classic sports routes, popular travel routes, hot places and traffic condition of different routes in different time etc. The mined knowledge can be recommended to users via Web or mobile user interface (UI) when they need suggestions.

3. PROTOTYPE

Figure 2 depicts the Web user interface of GeoLife prototype. Compared to traditional text-based description, visualizing users' GPS log over Web maps can provide a

more fancy, concise and explicit approach to express their experience. Hence, users can connect to their personal past better, and obtain more information from other people's experience when they browse a GPS track.

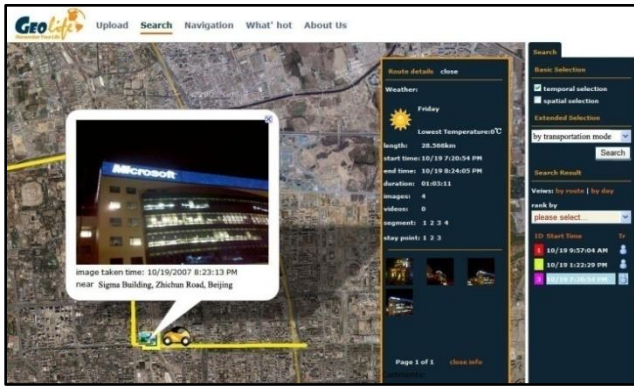
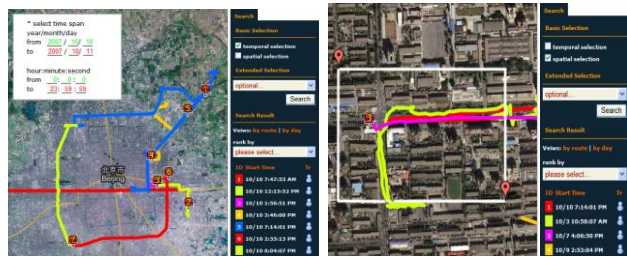


Figure 2. Prototype of GeoLife

In Figure 3, the spatio-temporal search function is powered by a flexible indexing scheme based on user behavior of uploading GPS tracks. We observe that people are more likely to upload GPS data of the near past than the distant past. Hence, we leverage B+ tree to index frequently updated groups and sorted dynamic array for rarely updated ones. Once the update frequency of a group drops below a threshold, we convert the index from B+ tree to a sorted dynamic array. Since the skewed nature of accumulative GPS tracks is taken into account, as compared to other spatio-temporal indexing methods, our scheme requires less index space and less update cost while keeping satisfactory retrieval performance.



(a) Temporal search (b) Spatial search
Figure 3. Spatio-temporal search of GeoLife

In prototype of GeoLife, we also propose an approach based on supervised learning to automatically learn the transportation modes including walking, taking bus, riding bike and driving from raw GPS data. As shown in Figure 4, when a GPS Log file comes, first, we divide the GPS track into trips and then partition each trip into segments. Then, we extract the features from each segment and send these features to the inference model. Two different ways are considered when we select the inference model to learn a user's transportation mode. In one way, we regard the segments of GPS tracks as independent instances and handle them as normal classification problem using general classifiers. After the inference, the post-processing

is performed to improve the prediction accuracy by taking the transition probability between different transportation modes into account. In the other way, conditional random field (CRF) is leveraged to perform the inference. Since the conditional probabilities between different transportation modes have been considered in the CRF, in this way, the post-processing is not performed.

The advantages of our approach are in that: 1) It can infer compound trips, which contain more than one kind of transportation modes. 2) The approach is independent of other information from maps and other sensors. 3) The model learned from the dataset of some users can be applied to infer GPS data from other people.

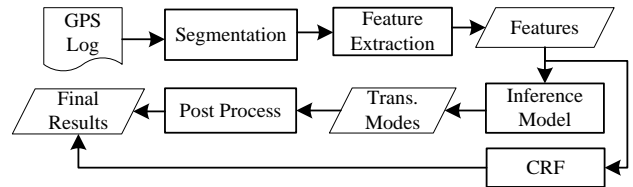
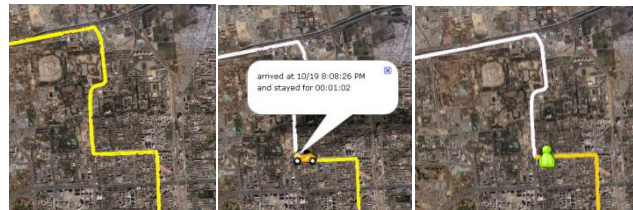


Figure 4. Procedure of inferring transportation mode

Figure 5 presents an example to distinguish the different Web experience between without and with transportation mode. Without the tag of transportation mode, the track shown in the Figure 5 (a) only provides us some basic location information. However, after we apply our inference model to this track, as depicted in Figure 5 (b) and (c), richer knowledge has been discovered over the plain track. We realize that the user first drives a car to downtown, and then at the parking lot he/she switches to walk.



(a) Before inference (b) Stop at parking lot (c) Switch to walk
Figure 5. An case of inferring transportation modes

REFERENCE

- [1] Mountain Bike. <http://www.mtb-tracks.co.uk/northyorkmoors/default.aspx>
- [2] Bikely: <http://www.bikely.com/>
- [3] Yu Zheng et al. Learning Transportation Mode from Raw GPS Data for Geographic Application on the Web. In Proc. of WWW'08, Beijing, China. 2008
- [4] Longhao Wang, Yu Zheng, Xing Xie. A Flexible Spatio-Temporal Indexing Scheme for Large-Scale GPS Track Retrieval. In Proc. of MDM'08, Beijing, China. 2008
- [5] Yu Zheng, Longhao Wang, Xing Xie. GeoLife: Memorizing Your Life over Web Maps. Submitted to Workshop on Location and Web. Beijing, China, 2008.