Volume 4 · Issue 2 · September 2014

Geoinformatica - An International Journal (GIIJ)

ISSN : 2180-1231

Number of issues per year: 6

CSC PUBLISHERS
http://www.cscjournals.org

Geoinformatica – An International Journal (GIIJ)

VOLUME 4, ISSUE 2, 2014

EDITED BY
DR. NABEEL TAHIR

ISSN (Online): 2180-124X

Geoinformatica – An International Journal (GIIJ) is published both in traditional paper form and in Internet. This journal is published at the website http://www.cscjournals.org, maintained by Computer Science Journals (CSC Journals), Malaysia.

GIIJ Journal is a part of CSC Publishers
Computer Science Journals
http://www.cscjournals.org

Geoinformatica – An International Journal (GIIJ)

Book: Volume 4, Issue 2, September 2014
Publishing Date: 15-09-2014
ISSN (Online): 2180-124X

GIIJ Journal is a part of CSC Publishers
http://www.cscjournals.org

© GIIJ Journal
Published in Malaysia

Typesetting: Camera-ready by author, data conversation by CSC Publishing Services – CSC Journals, Malaysia

CSC Publishers, 2014

EDITORIAL PREFACE

Geoinformatica − An International Journal (GIIJ) is an effective medium for interchange of high quality theoretical and applied research in Geoinformatica domain from theoretical research to application development. This is the *Second* Issue of Volume *Four* of GIIJ. The Journal is published bi-monthly, with papers being peer reviewed to high international standards. GIIJ emphasizes on efficient and effective geomatic sciences, and provides a central for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of Geoinformatica. Some of the important topics are spatial ontologies, computational geometry and visualization for geographic information systems, geostatistics and spatial statistics, spatial analysis, interoperability, and innovative applications of geotechnologies etc.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Started with Volume 4, 2014, GIIJ appear with more focused issues. Besides normal publications, GIIJ intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

GIIJ give an opportunity to scientists, researchers, and vendors from different disciplines of Geoinformatica to share the ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development. This journal is helpful for the researchers and R&D engineers, scientists all those persons who are involve in Geoinformatics in any shape.

Highly professional scholars give their efforts, valuable time, expertise and motivation to GIIJ as Editorial board members. All submissions are evaluated by the International Editorial Board. The International Editorial Board ensures that significant developments in geotechnologies from around the world are reflected in the GIIJ publications.

GIIJ editors understand that how much it is important for authors and researchers to have their work published with a minimum delay after submission of their papers. They also strongly believe that the direct communication between the editors and authors are important for the welfare, quality and wellbeing of the Journal and its readers. Therefore, all activities from paper submission to paper publication are controlled through electronic systems that include electronic submission, editorial panel and review system that ensures rapid decision with least delays in the publication processes.

To build its international reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for GIIJ. We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. GIIJ provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

Editorial Board Members
Geoinformatica – An International Journal (GIIJ)

EDITORIAL BOARD

EDITORIAL BOARD MEMBERS (EBMs)

TABLE OF CONTENTS

Volume 4, Issue 2, September 2014

Pages

User Category Based Estimation of Location Popularity using the Road GPS Trajectory Databases

Shivendra Tiwari *shivendra@cse.iitd.ac.in*
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
New Delhi, 110016, India

Saroj Kaushik *saroj@cse.iitd.ac.in*
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
New Delhi, 110016, India

Abstract

The mining of the user GPS trajectories and identifying the interesting places have been well studied based on the visitor's frequency. However, every user is given the same importance in the majority of the trajectory mining methods. In reality, the popularity of the place also depends on the category of the visitor i.e. international vs local visitors etc. We are proposing user category based location popularity estimation using the trajectories databases. It includes mainly three steps. *First*, pre-processing – the error correction and the graph connection establishment in the road network in order to be able to carry the graph based computations. *Second*, find the stay regions where the travelers spent some time off-the-road. The visitors can be easily categorized for each POI based on the travel distance from the home location. *Finally*, normalization and popularity estimation – measure the frequency and stay time of the visitors of each category in the places in question. The weighted sum of the frequency and stay time for each category of the visitors is calculated. The final popularity of the places is computed with values of the pre-configured range. We have implemented and evaluated the proposed method using a large real road GPS trajectory of 182 users that was collected in a period of over three years by Microsoft Asia Research group.

Keywords: Trajectory Databases, Trajectory Mining, Popularity Estimation, Region of Interest.

1. INTRODUCTION

The Location Based Services (LBS) are hugely contributing to the next revolution on small computing handheld such as mobile devices through the mobile network and utilizing the ability to make use of the geographical position of the device [1]. The location based tour guide is one of the most useful LBS applications. The tour guides features include route planning, profile based route optimization, recommending Tourist Point of Interests (TPOI), constraint based tour planning, location search, map display etc [2]. Mostly the locations are recommended based on the user ratings. The tour-guide systems require the ratings and popularity measures for the regions as well in order to recommend the interesting regions known as Region of Interest (ROI). An ROI can be an individual Point of Interests (POI) at the lowest level or a group of POIs, or an administrative region such as city, district, villages etc. Such information can help users understanding surrounding locations, and would enable a better travel guidance experience. The ROIs are created using different methods as proposed in [1]. It is important to estimate the popularity of the places in order to disseminate the appropriate location content in such applications. There are various ways of assessing the popularity of the geospatial locations, such as user ratings, frequency of the user check-in etc. The average user rating sometimes may not lead to the estimation of the accurate popularity. It depends on the number of visitors who participate on rating process, and also deeply depends on their technology awareness, culture, education etc. The mining of user's GPS trajectory database is another way of reckoning the location popularity.

A trajectory is a sequence of sampled locations and time stamps along the route of a moving object. The analysis of such trajectory data is a critical component in a wide range of research and decision-making fields. However, it is a challenging problem to analyze and understand patterns in massive movement data, which can easily have millions of GPS point locations and trajectory segments. Sometimes the GPS location is far from the actual road and hence it is important to project them on the road or on the POIs in order to carry about accurate computations. We have proposed to use map matching techniques to preprocess the GPS trajectory before actually estimating the popularity of the locations. Map Matching, is the process of projecting the GPS fixes on the road network graph G = (V, E) [6]. In this paper, we have used map matching in order to project the inaccurate trajectories on the road network. However, the trajectories where the GPS points are already map-matched, this step can be skipped. Figure 1 shows the places popularity and their overlap in their popularity index.

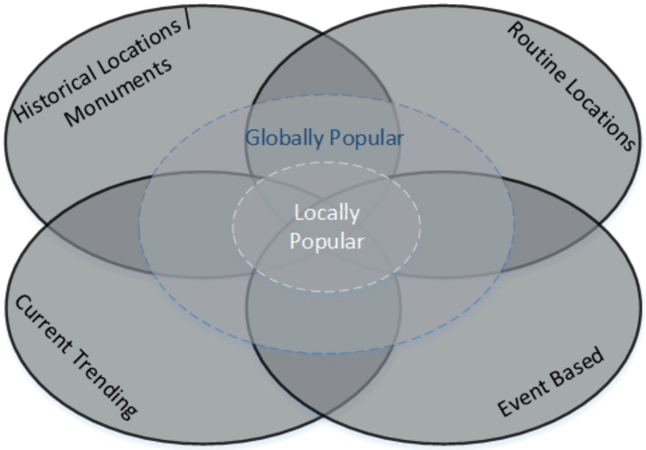

FIGURE 1: The place's popularity depending on different scenarios and factors.

The advanced LBS applications requiring the highly accurate information are the main source of motivation of this work. The ROI based tour guides need to the deliver the appropriately granulized infotainment data depending on desired level of details. The tourists prefer visiting a location with several attractions. From the tourist point of view, most of the time the whole region is considered while planning instead of just one POI. The current popularity estimation systems mainly evaluate the POIs; and hence, the ratings of the ROIs are unavailable. The popularity of ROIs is important to be estimated correctly as the granularity of the geographical information, and dissemination of the content with the desired level of details needs to be assured. There are multiple challenges of estimating the popularity based on the trajectory database. *First*, *inaccurate trajectories* – the points are sometimes inconsistent, off-the-road and unequal sampled. The trajectory segments need to be accurate in order to compute the travel distances, hence they need to be processed in order to reduce the error level. The locations where there is no digitized map available, therefore, it requires a graph generation process that considers the average of the multiple trajectories. *Second*, establishing the relationship between popularity of the ROI and POIs based on the trajectory visiting trend. The POI visit is not mutually exclusive with the ROI that encloses a POI. Hence, the POI visit also contributes to the enclosing region's popularity. *Third*, *popularity estimation* – other than the ratings of the places in order to estimate the likings, the user visiting frequency also contributes to the trend estimation. We need to consider the kind of people visiting the underlying while computing their popularity indices.

Our contribution in this paper includes – *First*, proposed a trajectory analysis method to estimate the popularity of the geospatial locations. *Second*, introduce the notion of "who" in order to categorize the visiting travelers to a location. *Finally*, we demonstrate and evaluate the method with the real dataset Geolife provided by Microsoft [8]. Apart from this section as introduction, section 2 talks about the related work in the trajectory mining area. The data modeling and pre-processing of the trajectory databases, and determination of the stay points have been discussed in section 3. Section 4 and 5 establish the algorithms estimating the popularity and experiments by implementations separately. Finally, the section 6 has concluded the paper.

2. RELATED WORK

Many different methods have been developed for trajectory and movement analysis. In general, most trajectory analysis methods involve the two steps, *first* – simplify and generalize each trajectory; *second* – compare and group trajectories to find general patterns [5]. Zheng et al in 2009 [9] proposed hierarchical graph based method for mining the interesting locations and travel sequences from GPS trajectory databases. They model multiple individuals' location histories with a tree-based hierarchical graph (TBHG). Based on the TBHG, they defined a HITS (Hypertext Induced Topic Search) – based inference model, which infers the interest of a location by taking into account three factors i.e. users travel experience; mutual reinforcement relationship between travel experience and location interest. *Finally*, the method mined the classical travel sequences among locations considering the interests of these locations and users' travel experiences. Later in 2010 Zheng et al suggested supervised learning based approach to infer people's motion modes from their GPS logs [10]. They also introduced a social networking service, called GeoLife, which aims to understand trajectories, locations and users, and mine the correlation between users and locations in terms of user-generated GPS trajectories [11].

Kang et al in 2010 [12] suggested method to mine the spatio-temporal pattern in the trajectory data – it first finds meaningful regions and extracts frequent patterns based on a prefix-projection approach from the sequences of these regions. They experimentally proved that the proposed method improves mining performance and derive more intuitive patterns. Lee et al in 2007 [13, 14] proposed a trajectory clustering method based on the partition and group framework. They established the importance of discovering the common sub-trajectories in many applications, especially if we have regions of special interest for analysis. The new framework partitions a trajectory into a set of line segments, and then, group's similar line segments together into a cluster. The primary advantage of the framework is to discover common sub-trajectories from a trajectory database. Based on the partition-and-group framework, they developed a trajectory clustering algorithm that consists mainly two phases: partitioning and grouping. The main advantage of the algorithm is the discovery of common sub-trajectories from a trajectory database.

Yan in 2009 [15] proposed semantic trajectory analysis based on the statistical computation and semantic concepts. It involves three major perspectives, i.e. trajectory modelling, trajectory computing, and trajectory pattern discovery. The early stage of the work has surveyed three types of modelling requirements for comprehensively explaining trajectories, in terms of geometric knowledge, geographical knowledge, and application domain knowledge. Zenger in 2009 [16] proposed the POI recommendation techniques based on the GPS trajectory databases. Zenger presented a new framework for trajectory-based POI recommendation. The method constructs a k-truncated generalized suffix tree to represent a historical trajectory database, and use it to execute exact matching recommendation queries. Two variants are developed, allowing for the execution of fuzzy matching and order-flexible queries.

3. PROPOSED SOLUTION

Figure 2 shows the high-level component and flow of the trajectory data processing the popularity estimation including the three major components i.e. preprocessing, determination of the stay regions, and popularity estimation. The preprocessing module involves improving the trajectory to make it usable – this step, however can be skipped in case the underlying trajectory has high quality GPS sequences. The stay region determination step finds out all the regions where the traveler spent more than a minimum amount of time off-the-road. This step simply produces a list of rectangular regions that could be POIs as well and hence these regions are useful for popularity computation in the later stage of the procedure. The popularity estimation step includes identifying the visiting locations and making a check-in entry into the place registry as discussed in section 4. In this section we discuss the preprocessing of the trajectory data and the stay regions determination.

3.1 Data Modeling and Preprocessing

The road GPS trajectory is sometimes inaccurate and has irregular behavior in terms of the time and the location. Also, the trajectories might have irregular time interval of logging the location. The other problem is that the trajectories do not consider the connectivity of the roads as a graph. It is hence difficult to calculate the travel distance only based on the trajectories. In this section, the preprocessing of the trajectory data is carried out based on the existing *map matching* and *interpolation* techniques [5, 6]. The trajectory is converted into the road network graph which is aware of the connectivity of the roads and hence it becomes more accurate to compute the reachability among consecutive points in the trajectory.

Map Matching – The map matching problem is characterized by two objectives – identify the link traversed by the traveler and find the actual location within that link. Road network map and GPS data are often enough for post processing map matching. The shortest path algorithm can be appropriately used for post-processing map-matching. The map matching plays an important role in putting the user either on the road or in a region based on the users location as discussed in section 3.2. It is possible that the user stays at the POIs for some time and hence there might not be the road at all. In such cases, the map matching algorithm has to put the user on the POI so that it can be concluded that the user indeed visited the POI. So the output of the map matching is not only the road network graph, but also a list of regions where the user stayed. The stay point finding method includes the task of map matching in case there is inaccuracy; however, it also finds the stay regions where the traveler has spent some time. The input to the map marching is the road network, trajectory database; however the output is the accurate GPS sequence that guarantees the point being exactly on the road if moving or in any geographical location (if there is a stay). For the sake of completion, we have discussed the map matching in short, that is inspired from the work in [6].

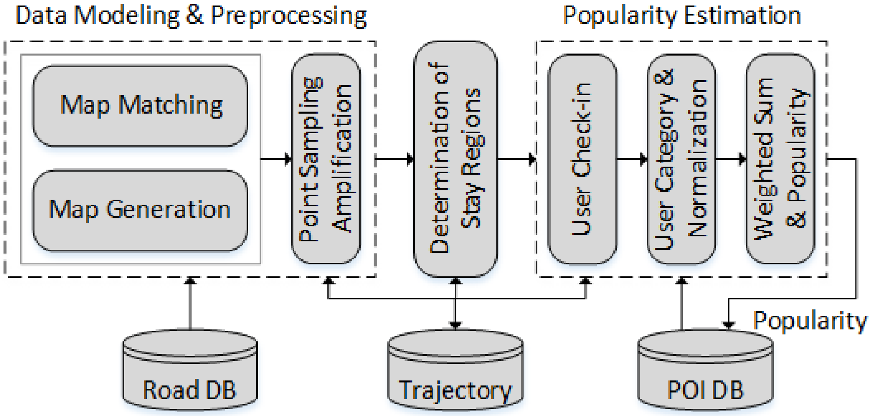
Data Modeling & Preprocessing
Popularity Estimation
Map Matching
Map Generation
Point Sampling Amplification
Determination of Stay Regions
User Check-in
User Category & Normalization
Weighted Sum & Popularity
Popularity
Road DB
Trajectory
POI DB

FIGURE 2: Overall component diagram for proposed popularity estimation.

Map Generation – In case the digitized geographical map is not available in the area, we need to generate the map based on the all the trajectory paths. Map generation is a process of creating the road network based on the trajectory databases. Considering the inherent inaccuracy in GPS measurements, a circular window is used to smooth/aggregate GPS points and to extract a much smaller number of representative points for a group of GPS points. A circle with a fixed radius is placed at each GPS point, whose location will be changed to the average of all the GPS points covered by the circle. The smoothing process brings the point closer to the road median. If a GPS point does not have any other point within the specified distance, it will remain in the original location. We choose a smaller set of new locations as representatives of the original GPS points to reduce data redundancy and size. If there is no other GPS point within distance criteria to a GPS point p_i, then p_i will represent itself. The representative point based map generation method is inspired by the graph based trajectory analysis method in [5].

Sampling Amplification – GPS points collectively can reveal the road network as we discussed in the earlier subsection. However, the trajectory also needs to include enough sampling of the points. If the GPS is retrieved in a long time interval, then it can be heavily inaccurate while computing the other attributes such as travel speed, distance, stay time etc. Therefore, we need to insert additional sampling of the points in order to make the trajectory well sampled. The trajectory interpolation is a method of estimating such points within the trajectory line segments. The challenge is that this is not a linear interpolation since a straight-line trajectory segment should be interpolated to follow the curves and turns of the 'road'. The interpolation achieves important outcomes i.e. improves the resolution and accuracy; enables accurate location-based summary statistics; and establishes the topological relations between trajectories.

3.2 Determining Stay Points

The stay points are the regions where the user has spent some time off-the-road. If it is on the road, it could be due to heavy traffic and hence it needs to be carefully avoided. This process takes two inputs i.e. road network graph G = (V, E); and user GPS trajectory database $T = \{T_1, T_2, T_3 \ldots T_m\}$; where $T_i = \{<x_1, y_1, t_1>, <x_2, y_2, t_2>, <x_3, y_3, t_3> \ldots <x_k, y_k, j_k>\}$. It generates a list of regions $R = \{R_1, R_2, R_3, \ldots R_k\}$ where user spent some time off the road network for more than a minimum threshold. The region is a rectangular area with a list of points within it. It will help us further knowing the POIs user stayed in the later part of the solution.

The extraction of a stay-point depends on two scale parameters, a time threshold (T_{thresh}) and a distance threshold (D_{thresh}). For the points $\{p_5, p_6, p_7 \ldots p_{17}\}$, a single stay-point s can be regarded as a virtual location characterized by a group of consecutive GPS points $P = \{p_m, p_{m+1}, \ldots, p_n\}$, where $\forall m < i \leq n$, $Distance(p_m, p_i) \leq D_{thresh}$ and $|pn.T - pm.T| \geq T_{thresh}$. Formally, conditioned by P, D_{thresh} and T_{thresh}, a stay point s=(Region, Latitude, Longitude, $T_{arrival}$, $T_{departure}$), where –

$$\text{s.Latitude} = \frac{\sum_{i=m}^{n} \text{p}_i.\mathit{Latitude}}{|P|} \tag{1}$$

$$\text{s.Longitude} = \frac{\sum_{i=m}^{n} \text{p}_i.\mathit{Longitude}}{|P|} \tag{2}$$

The region s.Region is the rectangle with center as (s.Latitude, s.Longitude). Equation (1) and (2) respectively stand for the average latitude and longitude of the collection P. However, $s.T_{arrival} = p_m.T$ and $s.T_{departure} = p_n.T$ represent a user's arrival and departure times on stay point s included in rectangular region. These stay points occur where an individual remains stationary exceeding a time threshold. In most cases, this status happens when people enter a building and lose satellite signal over a time interval until coming back outdoors. The other situation is when a user wanders around within a certain geospatial range for a period. In most cases, this situation occurs when people travel and are attracted by the surrounding environment. As compared to a raw GPS point, each stay point carries a particular semantic meaning, such as the shopping malls we accessed and the restaurants we visited, etc.

4. POPULARITY ESTIMATION

The popularity is measured based on the different attributes such as visiting frequency, visitor category, and stay time. The process includes three steps, i.e. determining the user's 'home location' and checking-in into the locations in questions. It is interesting to note that our solution is generic to a single point location i.e. POI as well as a geographic region i.e. ROI. The database in question may contain both kinds of records which can be evaluated in this section. In order to keep terminology clear, we use the term 'location' for both the POI and ROI.

4.1 User Check-in

The user enters in a location, and spends sometime is known as user check-in. When a user leaves the location, it is marked as check-out. The check-in location, frequency and the stay-time creates a pattern that would be helpful determining if it is a 'home', 'routine' or an 'interesting' locations of the other users. A home location for a user is a location that is visited and spent time there on a regular basis. The visiting pattern such as leaving the place in the morning and coming back to the same location in the evening in a regular pattern, then it is marked as the home location. In order to determine the home location, we need to analyze the regions set R for the individual users. Intuitively, the stay time between two GPS points is the most important attribute in the trajectory to identify the home location. The POIs under evaluation are stored in the spatial grid so that the algorithm only search the appropriate grids while looking for the check-in locations. Only those points are evaluated for the POI check-in only if the user has spent more than a specified minimum time. The POIs are searched using the adjacent grid based search (AGBS) method [4]. The stay location where there is no registered place, add the reverse-geocoded [17] address as the new place. The newly discovered place can be used as potential POI that is yet to be digitized or it could be a home location for the user that is normally not digitized as a POI.

```
Algorithm 1. User-Check-in
Input: Set of POIs to be evaluated denoted by Ŋ; and the stay regions
   set R = {$R_1$, $R_2$, $R_3$,… $R_k$}
Output: Ѣ =Set of user's home locations; updated check-in information
   for the set of POIs Ŋ;
Steps:
1. Repeat through each region $R_i \in R$
   a) Repeat through each point $p_i$ in region $R_i$
     i)   Search the nearest POI ∈ Ŋ from the point $p_i$ using the adjacent
          grid search method.
    ii)   If a valid POI exists within the region $R_i$, then
           - Go to step (iv);
   iii)   Else
           - Create new POI = call reverse-geocode($p_i$);
           - Add the POI place into POI database Ŋ.
    iv)   Update registry
           - Add an entry <user-id, arrival-time, departure-time> in the
             POI registry.
           - Add the POI into the user's registry.
2. Evaluate POI database Ŋ:
     i)   Mark the highest frequency visiting locations and the highest
          average stay-time as 'home' or 'routine' point for this user.
    ii)   Remove the current user from the POI registry (the user's home
          visit is discounted from the popularity computation)
   iii)   Remove the home location(s) from the user registry.
3. Update travel distance (from the nearest home location) for the
   current user in all the POIs in the POI database registry to
   maintain <user-id, arrival-time, departure-time, travel-distance>
4. End.
```

Algorithm 1 takes the user's trajectory road network graph and a set of POIs, and the region sets visited by the user. Once the home location and the visited POIs are identified, the distance from the nearest home location to the POI is computed and the <user id, stay time, travel distance> tuple is stored in the POI registry. This algorithm needs to be executed for each visiting user in the POI database.

4.2 User Categorization, Normalization and Popularity Estimation

Since not all the attributes have the same importance, the analytics hierarchy process can be used to find the weights for the attributes. The comprehensive decision weights for each alternative are calculated by the weight sum as suggested in [7]. The global travelers' visits are considered more significant in terms of popularity than that of the local visitor. It is important to note that the user category can be different for a user visiting the different locations depending on travel distance from the home location.

First*, *user categorization – the first step is to categorize the visitors in a POI based on their home location and travel distance. The categorization also uses the home location's administrative information in order to decide the border cases. For example, The Golden Gate Bridge in San Francisco, in USA is closer to the Mexican people living near the Tijuana region than for the people living in the rest of the states in the USA. It means that only the distance cannot decide the user category and hence we take the help from the home location's administrative information. We propose to divide the users in 5 categories (i.e. global, national, regional, local, and native) that would have their own weights in the effective weighted frequency and stay time computation. The distance criterion for each category is application and database dependent that can simply divide the

Second*, *weighted sum – for each place the check-in registry is grouped for each user category. The weighted sum of the frequency and stay time is computed i.e. the frequency F_i and stay-time T_i of i^{th} POI are defined in eq (3, 4). The frequencies (i.e. F_i) and time (i.e. T_i) values multiplied by the corresponding user category weights and are added up to have an overall weighted sum for each place. All the frequency and weights are denoted respectively, for global (i.e. F_g, W_g), national (i.e. F_n, $W_{n)}$, regional (i.e. F_r, W_r), local (i.e. F_l, W_l) and native (i.e. F_t, W_t). Here $\sum W = 1$.

$$F_i = (F_g.W_g) + (F_n.W_n) + (F_r.W_r) + (F_l.W_l) + (F_t.W_t) \tag{3}$$

$$T_i = (T_g.\ W_g) + (T_n.W_n) + (T_r.W_r) + (T_l.W_l) + (T_t.W_t) \tag{4}$$

Finally, **normalization of the results** – the stay time and the frequency values are normalized before actually used in the popularity estimation as shown in eq (5, 6). The overall popularity is the weighted-sum of the normalized frequency and stay time for the underlying place.

$$\text{F-Norm}_i = F_i / \max_i^n (F_i) \tag{5}$$

$$\text{T-Norm}_i = T_i / \max_i^n (T_i) \tag{6}$$

$$P_i = (\text{F-Norm}_i.W_{freq}) + (\text{T-Norm}_i.W_{stay\text{-}time}) \tag{7}$$

$$P_{i\text{-final}} = P_i / \max_i^n (P_i) \tag{8}$$

Here, F-Norm$_i$ and T-Norm$_i$ are the normalized accumulated frequency and stay time respectively for the i^{th} place. The frequency and the stay time have separate weights as W_{freq} and $W_{stay\text{-}time}$ to compute the weighted popularity P_i in eq (7). Finally, eq (8) computes the normalized popularity index $P_{i\text{-final}} \in [0, 1]$. The popularity $P_{i\text{-final}}$ can be finally mapped to the context dependent popularity can be further mapped to the desired range, i.e. $\rho = P_{i\text{-final}} * \rho_{max\ \wedge}\ \rho \in [0, \rho_{max}]$, ρ_{max} is the maximum popularity value in the range.

5. IMPLEMENTATION AND EVALUATION

We have used the GPS trajectory dataset that was collected in Microsoft Research Asia's Geolife project by 182 users in a period of over three years. This dataset contains 17,621 trajectories with a total distance of about 1.2 million kilometers and a total duration of 48,000+ hours. These trajectories were recorded by different GPS loggers and GPS-phones, and have a variety of sampling rates. This dataset recoded a broad range of users' outdoor movements, including not only life routines such as to go home and go to work, but also some entertainments and sports activities, such as shopping, sightseeing, dining, hiking, and cycling. The majority of the data was created in Beijing, China [4]. The total data size is approximately 1.55GB which takes around 2 hours 45 minutes to complete parsing the popularity estimation (excluding the preprocessing and sampling amplification in the trajectory database). With the experiments we have extracted 5617 places as the visited regions. The regions have roughly the rectangular size of 100x80 meters. Figure 3 shows the data distribution based on the travel distance, collection duration, and the effective travel duration.

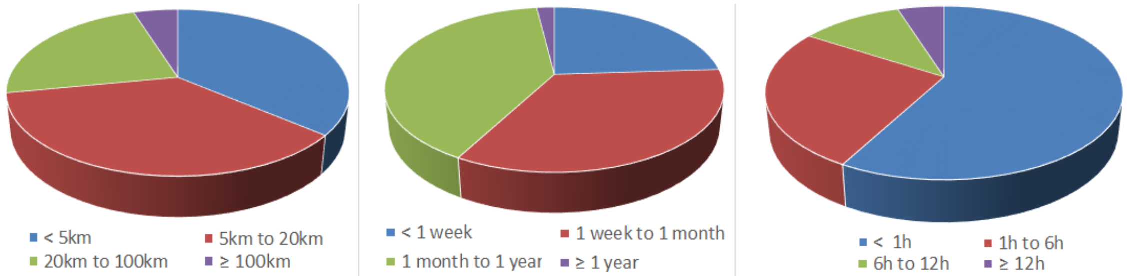

FIGURE 3. a) Trajectory distribution by distance, **b)** Data collection duration distribution, **c)** Effective travel duration distribution.

The demo implementation has been carried out using C++ as the programming language. The trajectory data have been stored in the flat files; however the spatial grid data structure has been used to store the places and their user registry. We have used user category weights as 0.39, 0.29, 0.18, 0.09, and 0.05 intuitively in order to compute the weighted sums of the frequency and the stay times. The global visitors have been assigned higher weights so that their effect is significantly seen considering that the experimental data has most of the users of the same country. However, equal weights for frequency and the stay time have been used to compute the weighted popularity.

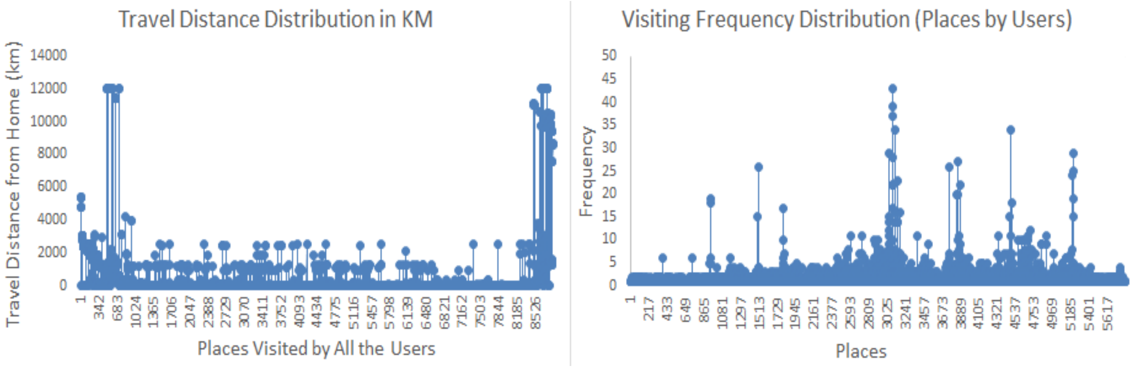

FIGURE 4: a) Travel distance to the individual places travelled by the users to visit the location. **b)** Places visit frequency distribution per user.

Figure 4 shows the travel distance distribution to the places visited by the travelers during the data collection period. We have considered all the regions where the user spent more than 10 minutes off-the-road. The first part of the graph shows the travel distance distribution by any user to visit any place. Mostly the users belong to China who travel within the country under nearly 2000 km away from the home; however, there are some users who travel from US to China back and forth; hence the travel distance has been significantly higher for some of the instances. In the second part, the places are plotted with their corresponding visiting frequencies. Although the number of places visited is good enough; but the frequency in most of the places is low as the data collection volunteers do not necessarily visit the common places repeatedly. The Figure 5 displays the places visit frequency per user in the first part and stay time distribution by any user at any place in minutes in the second part of the graph. The high stay time is either due to the home location or the devices was turned off and restarted after a few days. We have accepted a maximum of 24 hours of stay at any location.

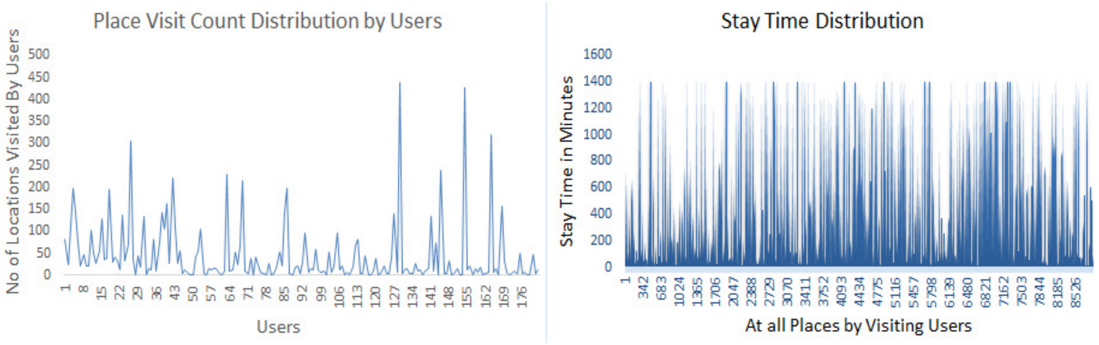

FIGURE 5: a) The distribution of overall accumulated visit frequency of the places by the individual travelers; **b)** The distribution of the travelers stay time across all the places in the <place, user> combination.

S. No	Place category	Original User Ratings	Total frequency	Weighted frequency	Visit Frequency based rating	Weighted frequency based rating
1	Hotel	5	490	1283.161	4.1727	4.705
2	Hotel	0	307	759.8676	2.5418	2.867
3	Restaurant	0	205	558.1162	2.4835	2.980
4	Shop	0	170	424.6227	2.3625	2.972
5	Restaurant	0	131	355.2668	1.9940	2.292
6	Sights & Museums	0	122	330.9166	1.8704	2.233
7	Building	5	133	329.8117	2.0202	2.435
8	Restaurant	0	114	297.9203	1.7595	2.656
9	Electronics	0	105	283.7347	1.6374	2.548
10	Snacks	0	126	253.6427	1.9262	2.342

TABLE 1: A comparison of the place rating trend and their estimated popularity.

The most of the places found in the experiment are unrated in the popular mapping and POI data providers. The major advantage of our approach is that it offers the trend of the place visited by the overall crowd. Since the user based rating is available on the scale of 0 to 5; we have mapped the results in the same scale for the uniformity in comparisons in Table 1. It can be easily noted from the table above that the hotel and the restaurants are the most visited places. Also, only 2 out of 10 places have been rated by the users. The remaining places did not have any ratings. The frequency oriented and the weighted frequency based methods have better data mainly for the unrated places. It is clear that a trajectory based popularity method can be used as a fallback method for the unrated places or a hybrid approached can be used.

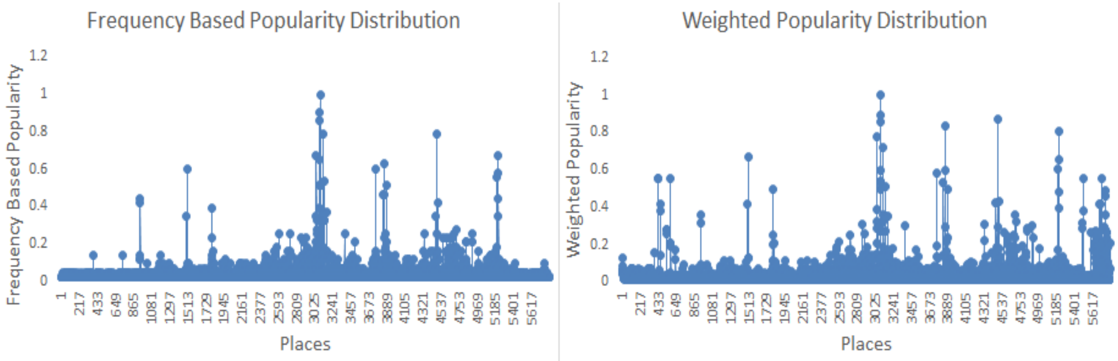

FIGURE 6. a) The distribution of overall normalized frequency based popularity of the individual places; **b)** the distribution of normalized weighted popularity.

The travel distance, visit-frequency and stay-time is finally aggregated to the weighted popularity estimation method in order to compute the final popularity index. Figure 6 shows the normalized final trend of the popularity indices of the places. We show the comparison based on the frequency based and weighted method based popularity indices. The first part of the Figure 6 shows the popularity simply based on the frequency; however the second part is the corresponding weighted popularity ranging within [0, 1]. The travel distances in the experimental data do not include large variation in the distances as only a few cities of the two countries are involved in the data collection. However, a clear difference between the popularity distributions can be seen in a few places where the travel distance was high. Since, the user category does not vary in the database significantly, the distribution trend is more identical than expected.

6. CONCLUSION AND FUTURE WORK

We have proposed and implemented the popularity estimation method that gives more importance to the travelers who is taking long way to visit the place. It also considers the fact the higher visit frequency and stay-time eventually leads to the higher popularity index. The method is effective enough in the popularity estimation using a rich trajectory database. It is mainly useful for the places with low user ratings. It can also be used as a fallback method in the recommendation systems where the initial data is not available. The proposed method has its own limitations in different aspects. This method might consider the highway side restaurants as highly popular based on the travel distance of the truck drivers. Our method is mainly focused for the tourist places; however, it can be easily extended for other kinds of scenarios by simply considering the context of the use for the locations. The proposed method might reach to incorrect conclusion in case of the long traffic jams in front of a POI. However, this problem can be solved by greater GPS accuracy measures. The method does not consider the altitude of the trajectories in order to estimate the popularity of the POIs on different floors of the same building. We have estimated the places without considering their POI categories and hence it is good for the same category places only. However, it is fairly easy to extend the system considering the specific scenarios and by handling the edge cases in the implementation without loss of generality. The trajectory data we have used for the experiments have multiple shortcomings including the small number of users, small number of visited places per user etc.

7. REFERENCES

[1] S. Tiwari, and S. Kaushik. “Modeling On-the-Spot Learning: Storage, Landmarks Weighting Heuristic and Annotation Algorithm.” In proceedings of CIIA'13 Saida, Algeria, 2013, pp. 357-366.

[2] S. Tiwari, and S. Kaushik. “Fusion of Navigation Technology and E-Learning Systems for On-the-spot Learning.” In: proceedings of ICWCA’12, Kuala Lumpur, Malaysia. 2012, pp. 72-78.

[3] G. J. Klir and B. Yuan: “From Classical (crisp) Sets to Fuzzy Sets” in Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall PTR (ECS Professional), 1997.

[4] S. Tiwari, and S. Kaushik. “Boundary Points Detection Using Adjacent Grid Block Selection (AGBS) kNN-Join Method.“ In: proceedings of MLDM’12, Berlin, Germany, 2012, pp. 113-127.

[5] D. Guoa, S. Liua and H. Jina (2010): “A graph-based approach to vehicle trajectory analysis.“, Journal of Location Based Services, 4:3-4, 183-199

[6] R. Dalumpines and D. M. Scott (2011): “GIS-based Map-matching: Development and Demonstration of a Postprocessing Mapmatching Algorithm for Transportation Research.“ In Lecture Notes in Geoinformation and Cartography, 2011, pp 101-120.

[7] X. Gu and Q. Zhu. “Fuzzy multi-attribute decision-making method based on eigenvector of fuzzy attribute evaluation space.“ ScienceDirect Decision Support Systems, Volume 41, Issue 2, January 2006, Pages 400–410.

[8] “GeoLife GPS Trajectories – Microsoft Research“: http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/default.aspx

[9] Y. Zheng, L. Zhang, X. Xie and W. Y. Ma. “Mining interesting locations and travel sequences from GPS trajectories.“ In Proceedings of WWW’09, Madrid Spain, 2009, pp. 791-800.

[10] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Y. Ma. “Understanding Mobility Based on GPS Data.“ In proceedings of ACM UbiComp’08, Seoul, Korea, 2008, pp. 312-321.

[11] Y. Zheng, X. Xie, and W. Y. Ma. “GeoLife: A Collaborative Social Networking Service among User, location and trajectory.“ In IEEE Data Engineering Bulletin 33(2), 2010, pp. 32-40.

[12] J. Kang, and H. S. Yong. “Mining Spatio-Temporal Patterns in Trajectory Data.“ In Journal of Information Processing Systems, Vol.6, No.4; December 2010, pp. 521-536.

[13] J. Lee, J. Han, X. Li and H. Gonzalez. “TraClass:Trajectory Classification Using Hierarchical Region based and Trajectory based Clustering.“ In VLDB’08, Auckland, NZ, 2008, pp 1081-1094.

[14] J. Lee, J. Han, and K.Y. Whang. “Trajectory Clustering: A Partition-and-Group Framework.“, In proceedings of SIGMOD’07, Beijing, China, June 11–14, 2007, pp. 593-604.

[15] Z. Yan, S. Spaccapietra. “Towards Semantic Trajectory Data Analysis: A Conceptual and Computational Approach.“, In proceedings of VLDB ’09 Lyon France, August 2009, pp. 24-28.

[16] Zenger, B. (2009): “Trajectory based Point of Interest Recommendation.“ In Thesis: School of Computing Science, Simon Fraser University, Canada.

[17] OpenStreetMap (2013): http://nominatim.openstreetmap.org/reverse?lat=37.80948&&lon=-122.4773&zoom=3 Last accessed on 14-Jul-2014.

INSTRUCTIONS TO CONTRIBUTORS

Geoinformatica – An International Journal (GIIJ) aims at publishing scientific and technical developments in the diverse field of Geoinformatics. GIIJ covers all aspects and information on scientific and technical advances in the geomatic sciences. The journal is providing a platform for exploring research, development and innovative applications in geographic information science and related areas. GIIJ provides a privileged view of what is currently happening in the field of geoinformatics as well as a preview of what could be the hottest developments and research topics in the near future. Additionally, it includes recent research results on spatial databases, spatial ontologies, computational geometry and visualization for geographic information systems, geostatistics and spatial statistics, spatial analysis, interoperability, and innovative applications of geotechnologies.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for GIIJ.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Started with Volume 4, 2014, GIIJ appear with more focused issues. Besides normal publications, GIIJ intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

We are open to contributions, proposals for any topic as well as for editors and reviewers. We understand that it is through the effort of volunteers that CSC Journals continues to grow and flourish.

LIST OF TOPICS
The realm of Geoinformatica – An International Journal (GIIJ) extends, but not limited, to the following:

- Applied Geography
- Close Range and Videometric Photogrammetry
- Computational Geometry and Visualization
- Distributed GIS/GIS and the Internet
- Geodata: Capture, Sources and Standards
- Geographic Information Science
- Geospatial Web
- Geostatistics
- Guidance Systems
- Land and Geographic Information Systems
- Location-Based Services
- Mobile Maps
- Sensor Networks
- Spatial Cognition
- Spatial Ontologies and Interoperability
- Surface Modeling
- Digital Mapping
- Geo Tags
- Geographic Data
- Geographic Information
- Geoinformatics
- Geospatial Applications
- Geospatial Databases
- Geospatial Processing
- Global Positioning System
- Integrated Geodesy
- Web Services
- Map Services
- Remote Sensing
- Sensor Web
- Spatial Data Analysis
- Spatial Databases

CALL FOR PAPERS

Volume: 4 - **Issue:** 3

i. Paper Submission: October 31, 2014 **ii. Author Notification:** November 30, 2014

iii. Issue Publication: December 2014

CONTACT INFORMATION

Computer Science Journals Sdn BhD
B-5-8 Plaza Mont Kiara, Mont Kiara
50480, Kuala Lumpur, MALAYSIA

Phone: 006 03 6204 5627
Fax: 006 03 6204 5628
Email: cscpress@cscjournals.org

CSC PUBLISHERS © 2014
COMPUTER SCIENCE JOURNALS SDN BHD
B-5-8 PLAZA MONT KIARA
MONT KIARA
50480, KUALA LUMPUR
MALAYSIA

PHONE: 006 03 6204 5627
FAX: 006 03 6204 5628
EMAIL: cscpress@cscjournals.org