

SPATIO-TEMPORAL PATTERN MINING ON TRAJECTORY DATA USING ARM

S. Khoshahval^{a*}, M. Farnaghi^a, M. Taleai^a

^a Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran
(samira_khoshahval@yahoo.com; farnaghi@kntu.ac.ir; taleai@kntu.ac.ir)

KEY WORDS: User Trajectory, Association Rule Mining, Location-based Application, Frequent Pattern Mining, Apriori Algorithm

ABSTRACT:

Preliminary mobile was considered to be a device to make human connections easier. But today the consumption of this device has been evolved to a platform for gaming, web surfing and GPS-enabled application capabilities. Embedding GPS in handheld devices, altered them to significant trajectory data gathering facilities. Raw GPS trajectory data is a series of points which contains hidden information. For revealing hidden information in traces, trajectory data analysis is needed. One of the most beneficial concealed information in trajectory data is user activity patterns. In each pattern, there are multiple stops and moves which identifies users visited places and tasks. This paper proposes an approach to discover user daily activity patterns from GPS trajectories using association rules. Finding user patterns needs extraction of user's visited places from stops and moves of GPS trajectories. In order to locate stops and moves, we have implemented a place recognition algorithm. After extraction of visited points an advanced association rule mining algorithm, called Apriori was used to extract user activity patterns. This study outlined that there are useful patterns in each trajectory that can be emerged from raw GPS data using association rule mining techniques in order to find out about multiple users' behaviour in a system and can be utilized in various location-based applications.

1. INTRODUCTION

Traditionally, the mobile phones were a device of communication; but today thanks to the technology advancements we can use them as tracking gadgets. Mobile devices with GPS (Global Positioning System) receivers are able to capture and record traces of users' movement, including coordinates, timestamps, elevation and other attributes such as velocity and bearing. Owing to GPS-enabled mobiles, huge amounts of user traces are recorded and stored from user daily activities. The need for using hidden information in trajectories developed studies in various fields such as urban planning, transportation, surveillance and security (Gudmundsson, Laube et al. 2011).

Extracting useful information from raw GPS trajectory data is a complex task. Trajectory data is basically a series of points which is not mostly human readable and needs interpretation; hence, trajectory analysis methods are necessary to extract useful information about users' behaviour (Zheng, Zhang et al. 2009).

Users movements make a continuous number of points in each GPS trajectory. Examining trajectory points conveys stops and moves in a route (Zheng, Li et al. 2008). User's stop places make a concentration of points in a specific section of the trace. Losing GPS signal may cause outlier in a track which implies the necessity of trajectory data pre-processing. Movement analysis calculates several measures such as distance or time between each two points of a trace and uncovers stops and moves which leads to users visited points. If user visits some specific locations permanently, those locations will be considered as user's point of interest (POI) which can be determined using place recognition algorithms (Gong, Sato et al. 2015).

Association rule mining (ARM) is one of the most popular methodologies of data mining which is an admirable apparatus for information discovery in large data itemsets. The major purpose of ARM is to distinguish the most frequent groups of items transpiring together (Zhang and Zhang 2002). Diverse association rule mining algorithms have been proposed by various researchers (Agrawal and Srikant (1994), Agrawal,

Mannila et al. (1996) and Narvekar and Syed (2015) to name a few). Among multiple available algorithms we have selected Apriori which is one of the most famous rule mining algorithms. Apriori advantages are simplicity and its ability to generate rules from frequent itemsets. Furthermore, Apriori divides the problem of finding all association rules into two subproblems which are finding all frequent itemsets and generating rules (Agrawal, Mannila et al. 1996).

This paper aims to depict a solution based on Apriori algorithm to find out user behavioural patterns from user's trajectory data by generating rules. Our solution is composed of two major stages of place recognition and generating rules based on discovered visited places. Place recognition gives us coordinates of user's visited points as stops which are then going to be semantically processed and formed to visited places by adding location name, part of day and some other features to the visited points' coordinates. Rule generation is obtained by taking advantage of Apriori which is a famous ARM algorithm. Eventually, produced rules are interpreted to discover user behavioural patterns.

This paper is divided into 5 sections. Section 2 gives a brief review of the related works. The proposed solution is thoroughly described in section 3. The experimental results are discussed in section 4 and our conclusion is drawn in the final section including the future work.

2. RELATED WORKS

Recently trajectory analysis and extracting user behavioural patterns have risen interest of researchers who study moving objects activities. Spaccapietra, Parent et al. (2008) exhibited modelling approaches in order to show the role of trajectories by attaching semantic annotations to bird's migration trajectories. Baglioni, de Macêdo et al. (2009) reached an interpretation of moving objects via enriching trajectories with semantic and geographical information of an ontology.

The related works in the field of user behavioural pattern discovery can be generally divided into place recognition studies

* Corresponding author

and spatial association rule mining studies. In order to discover visited points, researchers put effort on not only semantic trajectory analysis but also finding stops and moves in traces by using time and distance based algorithms. These methods can be divided to three major categories. The studies in the first category are mostly based on time and location. Zheng, Zhang et al. (2009) modelled location histories of user with a graph and proposed a HITS-based model. The studies in the second category focus on indirect parameters such as direction and acceleration which are calculated based on velocity of each GPS points in order to extract user's movement type (see Rehrl, Leitinger et al. (2010) and Bhattacharya, Kulik et al. (2012)). The last category of place recognition algorithms utilizes clustering to group data that shares identical characteristics. In the study of Ashbrook and Starner (2003), a system is designed to cluster GPS data into locations. An improved version of DBSCAN is presented in Gong, Sato et al. (2015) which uses coordinates and time of GPS trajectories to identify activity stop points.

Various approaches have been hypothesized to generate rules using ARM algorithm. Presenting a beneficial ARM algorithm was a concern of a number of researchers. Apriori, the most famous ARM algorithm, was first introduced by Agrawal and Srikant (1994). Apriori finds the frequent itemsets and generate rules based on discovered itemsets. AprioriTID determines the candidate itemsets before starting the pass in database. AprioriHybrid is a combination of Apriori and AprioriTID which transfers between algorithms in different passes which causes extra costs (Agrawal and Srikant 1994). In the following studies, attempts have been made by Han and Pei (2000) to use FP-tree concepts and they presented FP-Growth algorithm which has complexity in calculation. Moreover, other algorithms such as

Eclat and Recursive elimination was introduced by Schmidt-Thieme (2004) and Borgelt (2005), respectively.

All of the named researchers tend to develop more frequent rule mining algorithms but the gained methodologies have rarely been used for spatial pattern mining issues. Mousavi, Hunter et al. (2016) presented semantically interpretation of trajectory patterns using ontologies but researchers mostly have tended to focus on place recognition and spatial association rule mining separately but the purpose of our study is mainly on providing a methodology which uses the results of a place recognition algorithm as inputs for spatial association rule mining in order to discover user patterns. We have used a time and distance based algorithm to discover visited points and Apriori algorithm to generate rules based on extracted places.

3. METHOD

The proposed solution is composed of 4 steps of data preparation, visited place extraction, spatiotemporal analysis and association rule mining on extracted visited places (Figure 1). This solution receives trajectory data as input and generate users' behavioural patterns as output. Each step has an output which is going to be the input of next step.

In the first step OSM data and moving object data is gathered and pre-processed. Then, place recognition algorithm extracts visited points. Spatiotemporal analysis is implemented on the visited points to generate visited places' feature table. Afterwards, association rules are generated based on visited places' feature table from which users' behavioural pattern are obtained.

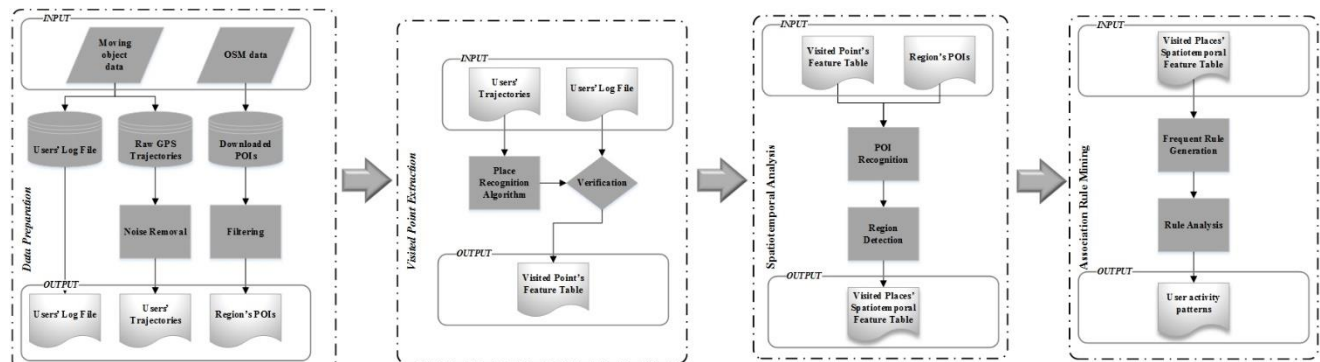


Figure 1. Proposed approach

3.1 Data preparation

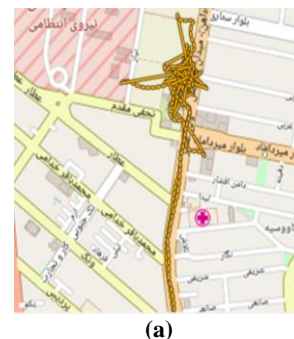
The input data of the solution is composed of three different datasets; including users' Log file, users' trajectories and regions' POIs.

In an attempt to gather needed moving object data, we kept tracks of users for 60 days. Users were asked to use a defined tracking application and start tracking from early morning until night. Moreover, users had to keep a log file of places that they visited each day as a diary to validate the gained results. Losing GPS signals in some environmental conditions may cause outliers. Therefore, we preprocessed the trajectories in order to eliminate useless points.

In this study, regions' POIs and some base maps of city of Tehran in Iran were downloaded from the Open Street Map (OSM) website. Tehran has 22 districts and 134 regions. Due to public access to OSM there is a mixture of true and false registered POIs for a region and a lot of placemarks without any useful information. The filtering process removes unknown

places of the downloaded POIs. Losing GPS signal may cause outliers in trajectories and therefore noise removal is necessary to eliminate useless points. Finally, the outputs of this step are available in the format of shape files, KML and tables.

Figure 2a. is a sample trajectory gathered by a user. As it can be seen, there is a concentration of GPS points which is probably a visited place. Figure 2b. is a sample of OSM POIs restricted to a specific region.



(a)



Figure 2. (a) User trajectory; (b) Explored POIs for a restricted region of Tehran

3.2 Visited point extraction

A trajectory is a record of changes in the position of a moving object during a period of time (Spaccapietra, Parent et al. 2008). In other words, it is a series of points, in the form of equation 1, in which there are features like coordinate, timestamp and etc.

$$\{(t_0, x_0, y_0), (t_1, x_1, y_1), (t_2, x_2, y_2), \dots, (t_n, x_n, y_n)\} \quad (1)$$

where $\begin{cases} t_i, x_i, y_i \in R, i = 0, 1, 2, \dots, N \\ t_0 < t_1 < t_2 < \dots < t_N \end{cases}$

A visit point is a location where user stays in the place for a specific time duration (Lv, Chen et al. 2016). In equation 2, P is a set of N visit points that an algorithm can find in a trajectory.

$$P = \{VP_1, VP_2, \dots, VP_N\} \quad (2)$$

To determine whether the trajectory has any visited point or not, we developed a place recognition algorithm. Figure 3 presents the developed time-based algorithm adopted from Lv, Chen et al. (2016). This algorithm examines the time interval between each two GPS point. In the algorithm, T is a set of GPS trajectories and ϵ_t is the time threshold which is considered as a minimum of 5 minutes. ϵ_d is defined as the distance threshold which is set to 200 meters. In the first run of the algorithm constant thresholds are applied to parameters and the set of visited points is empty. For each trajectory, the algorithm calculates distance and time interval between points incrementally and compares it with the given thresholds. If the distance is less than the minimum threshold, the point is considered as a potential visited point (Line 2-5) but if it is more than minimum threshold the algorithm checks the time interval (Line 7-9). Eventually, the output of the algorithms is a list of discovered visited points each of which includes 5 attributes of latitude, longitude, elevation, time and day.

Algorithm 1
 Input: GPS traces T, threshold ϵ
 Output: Visited Places VP
 1: ϵ_t and ϵ_d constant, initial set of VPs = \emptyset
 2: **for** trajectory points in T **do**
 3: calculate distance interval
 4: **if** measured distance < ϵ_d **then**
 5: consider Point as a potential VP
 6: **else**
 7: calculate time duration
 8: **if** measured time > ϵ_t **then**
 9: add Point to VP set
 10: create the table
 11: **else**

12: print no VPs
 13: **end**

Figure 3. Place recognition algorithm (Lv, Chen et al. 2016)

3.3 Spatiotemporal analysis

Taking advantage of GIS tools, we exploited the exported visited points in order to append semantic meanings to the extracted visited points, this entails merging coordinates of visited points to downloaded OSM POIs using proximity tools. Then, visited places are discovered and connected to their location names.

In order to accomplish spatial analysis, we considered human activities and part of day in 8 and 7 separate fields respectively (Figure 4). Each visited place is connected to one or more activity type, therefore user daily activity types are determined. For temporal analysis, we used a division of hours of day like morning until night in the form of early/late morning, early/late afternoon, early/late evening and night. Then, exact hours of each visited place were matched to the correct time interval.

Finally, the output of this step is visited places spatiotemporal feature table which illustrates a list of user POIs including location, day, part of day and region.

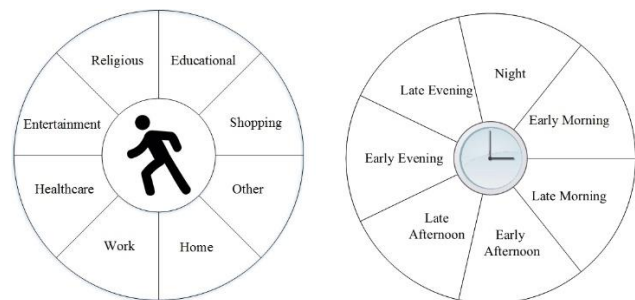


Figure 4. Human daily visited places/Part of day

3.4 Association rule mining on extracted visited places

Association rule mining is one of the most important data mining techniques and it is a well-known methodology for disclosing considerable relations and patterns among items in large datasets. Basically, association rule mining methods are also called market basket analysis because they enable us to define customers shopping behaviour (Agrawal, Mannila et al. 1996). In market basket analysis, association rules are generated for items with support and confidence larger than the user defined measures. In equation 3, I is an itemset with m items in the database of T with j transactions.

$$I = \{i_1, i_2, \dots, i_m\} \quad (3)$$

$$T = \{tr_1, tr_2, \dots, tr_j\}$$

There are two major measures of support and confidence which help us to define the frequent itemsets. The support of an association rule, as it is defined by equation 4, is the number of transactions with both items A and B over the whole number of transactions in the database.

$$\text{Support}(A \rightarrow B) = \frac{P(A \cap B)}{\text{NO. transactions with both A and B}} \approx \frac{\text{All transactions in DB}}{\text{All transactions in DB}} \quad (4)$$

The confidence of an association rule is shown by equation 5. It is the number of transactions with both items A and B over the number of transactions with A.

$$\text{Confidence}(A \rightarrow B) = P(B|A) \quad (5)$$

$$\approx \frac{\text{NO. transactions with both A and B}}{\text{Transactions with only A}}$$

Apriori firstly introduced by Agrawal, Mannila et al. (1996) is an association rule mining algorithm that finds frequent itemsets of L in database D using support and confidence. This search is level-wise and it goes up the lattice of itemsets. In the first step, the algorithm scans the whole database and eliminates items that are lower than the defined minimum support and then all items are combined together to produce subsets of items for the next scan.

For association rule mining using Apriori algorithm we considered spatial and temporal attributes exported from spatiotemporal analysis step as items of the rule mining database. Thus, we used spatial features such as location name and location category. Moreover, temporal features such as day and part of day were also added to the rule mining database. Then, the frequent items were selected based on defined ARM parameters to generate rules.

4. EXPERIMENT AND ANALYSIS

After applying the four steps of the proposed approach on the data, the activity rules were generated. Some of the rules, that were generated for a particular user who was a student with a part-time job can be seen in Tables 1 and Table 2. Table 1 and Table 2 present spatial patterns and spatio-temporal patterns, respectively. Among all generated rules, there was a significant rule by support of 1.2% and confidence of 98% which shows that if the user is at home the next location is work and if the user goes shopping after work the next location is home (Table 1). In table 2, rules with a combination of time and location are shown. For the first rule in table 2, if it is Saturday (the first day of week in Tehran) and part of the day is early in the morning, user is going to have an educational activity such as a class in university by the support of 2.2% and confidence of 87%. For this particular user with a part time job, it can be seen that the second rule (Table 2) shows a type of work activity in the early afternoon. Finally, the user's ending location is home after workplace in the late evening (the last row in Table 2).

Rules
If Home Then Work
If Work Then Shopping Then Home

Table 1. Spatial user activity patterns

Rules
If day Saturday part of day Early Morning Then Educational
If part of day Early Afternoon Then Work
If part of day Late Evening place Work Then Home

Table 2. Spatio-temporal activity patterns

According to the users' log files, the active user is mostly used to do two different kinds of activities on Fridays which are in the category of shopping and entertainment. Both of these activities happen sporadically. For instance, in one month, the user does shopping activities three times but in the next month, the activity is declined to one. Therefore, the system considers the activities on Friday generally and presents the most frequent one. According to the gained user's behavioral pattern, entertainment was the most frequent event on Fridays early evening (Table 3). Therefore, there would be some differences between our user log file and the systems' presented result for the activities if the frequency of activities were lower than the threshold. This event can be fairly alleviated by reducing the support that is not recommended. Instead of decreasing thresholds it would be better

to increase the period of trajectory gathering process in order to obtain a supplementary database.

Rules
If day Friday part of day Early Evening Then Entertainment

Table 3. The most frequent activity

Although, the resulted if/then rules may appear to be obvious for a known user such as a friend who we are familiar with her daily routine, the rules generated by the proposed solution are gained automatically in a systematic procedure. These rules can be obtained for multiple unknown users with different attributes by the proposed methodology and the activity patterns can be exploited in a location-based service to recommend useful information to users. In such a location based service, where multiple users are producing trajectory data, obtaining behavioural rules is a complex task and it is not obvious for human and analysis is necessary.

5. CONCLUSION

In this paper, we proposed a method to reveal user behavioral patterns from GPS trajectory data using place recognition and association rule mining. We applied place recognition algorithm to extract visited points; then, spatiotemporal analysis resulted in visited places feature table as input items to discover spatial association rules. The extracted rules conveyed human readable patterns that provide useful information about user daily activities.

For future work, behavioral patterns can be used in location based services. Understanding a user activity pattern in a system can be a beneficial factor in further usage of location-based services. Having known the user behavioral patterns, a location-based service can provide more accurate recommendations to users. For instance, if the service knows about the history of user activities on Saturday, it can recommend correct places in exact time to the user. Therefore, suggestions of the service would be precise and user-directed.

REFERENCES

- Agrawal, R., H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo (1996). "Fast Discovery of Association Rules." *Advances in knowledge discovery and data mining* 12(1): 307-328.
- Agrawal, R. and R. Srikant (1994). *Fast algorithms for mining association rules*. Proc. 20th int. conf. very large data bases, VLDB.
- Ashbrook, D. and T. Starner (2003). "Using GPS to learn significant locations and predict movement across multiple users." *Personal and Ubiquitous computing* 7(5): 275-286.
- Baglioni, M., J. A. F. de Macêdo, C. Renso, R. Trasarti and M. Wachowicz (2009). Towards semantic interpretation of movement behavior. *Advances in GIScience*, Springer: 271-288.
- Bhattacharya, T., L. Kulik and J. Bailey (2012). *Extracting significant places from mobile user GPS trajectories: a bearing change based approach*. Proceedings of the 20th International Conference on Advances in Geographic Information Systems, ACM.
- Borgelt, C. (2005). *Keeping things simple: finding frequent item sets by recursive elimination*. Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations, ACM.
- Gong, L., H. Sato, T. Yamamoto, T. Miwa and T. Morikawa (2015). "Identification of activity stop locations in GPS trajectories by density-based clustering method combined with

- support vector machines." Journal of Modern Transportation **23**(3): 202-213.
- Gudmundsson, J., P. Laube and T. Wolle (2011). Computational movement analysis. Springer handbook of geographic information, Springer: 423-438.
- Han, J. and J. Pei (2000). "Mining frequent patterns by pattern-growth: methodology and implications." ACM SIGKDD explorations newsletter **2**(2): 14-20.
- Lv, M., L. Chen, Z. Xu, Y. Li and G. Chen (2016). "The discovery of personally semantic places based on trajectory data mining." Neurocomputing **173**: 1142-1153.
- Mousavi, A., A. Hunter and M. Akbari (2016). "USING ONTOLOGY BASED SEMANTIC ASSOCIATION RULE MINING IN LOCATION BASED SERVICES."
- Narvekar, M. and S. F. Syed (2015). "An Optimized Algorithm for Association Rule Mining Using FP Tree." Procedia Computer Science **45**: 101-110.
- Rehrl, K., S. Leitinger, S. Krampe and R. Stumptner (2010). An approach to semantic processing of gps traces. Proceedings of the 1st Workshop on Movement Pattern Analysis, Zurich, Switzerland.
- Schmidt-Thieme, L. (2004). Algorithmic Features of Eclat. FIMI.
- Spaccapietra, S., C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto and C. Vangenot (2008). "A conceptual view on trajectories." Data & knowledge engineering **65**(1): 126-146.
- Zhang, C. and S. Zhang (2002). Association rule mining: models and algorithms, Springer-Verlag.
- Zheng, Y., Q. Li, Y. Chen, X. Xie and W.-Y. Ma (2008). Understanding mobility based on GPS data. Proceedings of the 10th international conference on Ubiquitous computing, ACM.
- Zheng, Y., L. Zhang, X. Xie and W.-Y. Ma (2009). Mining interesting locations and travel sequences from GPS trajectories. Proceedings of the 18th international conference on World wide web, ACM.