Exploring Human Movement Behaviour Based on Mobility Association Rule Mining of Trajectory Traces

Shreya Ghosh(✉) and Soumya K. Ghosh(✉)

Department of Computer Science and Engineering,
Indian Institute of Technology Kharagpur, Kharagpur, India
shreya.cst@gmail.com, skg@iitkgp.ac.in

Abstract. With the emergence of location sensing technologies there is a growing interest to explore spatio-temporal GPS (Global Positioning System) traces collected from various moving agents (ex: mobile-users, GPS-equipped vehicles etc.) to facilitate location-aware applications. This paper, therefore focuses on finding meaningful patterns from spatio-temporal data (GPS log) of human movement history and measures the interestingness of the extracted patterns. An experimental evaluation on GPS data-set of an academic campus demonstrates the efficacy of the system and its potential to extract meaningful rules from real-life dataset. AQ1

Keywords: Trajectory · Mobility · GPS traces · Association rule Transactional database

1 Introduction

Owing to the pervasiveness of mobile phones and development of sensor-technologies, wireless networks, the availability of mobility traces including personal-GPS trajectories, taxi-traces etc. have opened up the possibility of interpreting human mobility behaviour in space and time. This myriad of mobility data fuses interesting and challenging problems namely, determining classical travel sequences and top k interesting locations, traffic monitoring, defense applications [7], mobile-user categorization [3] etc. Obviously the major challenge is capturing the inherent knowledge so that it can be used effectively in several location based services or personalized recommendation systems. The core of any mobility-behaviour analysis task is *human moves with an intent* [2] and thus people follow a highly reproducible and meaningful patterns [6] in their daily movement. Therefore, to utilize the mobility traces of people for various services, it is crucial to perceive how location, time effects their mobility patterns. AQ2

Association rule mining finds application in several domains [1] including business analysis, clinical databases, stock market analysis etc. - where inter-relation among objects contribute in the knowledge-base. With the research

A. Abraham et al. (Eds.): ISDA 2017, AISC 736, pp. 1–13, 2018.
https://doi.org/10.1007/978-3-319-76348-4_44

advancement, several new paradigms have evolved, namely *inter-sequence* patterns, *intertransactional patterns*, where association relations are discovered among attributes from different transaction records [9]. Discovering frequent or co-related patterns in heterogeneous databases is one of the challenging and most important facets in data mining research. Since the introduction of *Association Rule Mining* problem and Apriori algorithms [1], significant research efforts have been made in the direction of dynamic dataset mining, appending additional semantics, such as time, space, ontologies etc. to discover temporal or spatial association rules, sequential pattern mining, Bayesian association rule mining etc. Temporal association rule mining uncovers a wide spectrum of paradigms for knowledge extraction [10] from time series data. Sequential pattern mining is a type of temporal pattern mining which is used in web-usage mining, discovering rules from medical databases [8], classification approaches etc. A key aspect of discovering meaningful rules is interestingness measures to extract and rank patterns according to the applications and potential interest to the users [5].

Motivation and Objectives: Discovering intertransaction multidimensional (space, time) rules from human mobility traces is a novel proposition. In this work, we aim to analyze human mobility behavioral patterns from the probabilistic graphical model of historical GPS log and discover the time-featured and categorical rules based on the spatio-temporal features. For example, the proposed system should be able to extract rules like, ***R1:*** *People who visits health-care center (say, gym, playground) regularly, mostly visits point-of-interests (POIs) like hospital, medicine-shop less frequently*, or ***R2:*** *Students are more likely to visit POIs like library, academicBuilding in weekdays while in weekdends frequency of GPS footprints are higher in POIs like cafe, movieComplex etc.* Rule $R1$ depicts time-feature based correlations among different movement patterns (or transaction in GPS trace database) while $R2$ represents categorical movement behaviour in an acedemic region-of-interest (ROI).

Contributions: The contributions of this work can be summarized as: (i) Proposing the structure of *GPS transactional log* using *time-series discretization* and a hash-based data structure to efficiently represent the spatio-temporal attributes of movement data in different resolution; (ii) Extracting the *Time-featured* and *User-categorical* rules to analyze the interdependencies of places (or stay-points) and time features of trajectory data and user mobility behavioral rules respectively; (iii) Finding the interestingness of the extracted rules using two proposed relevance-measures.

In our previous work [2], we identify the *mobility-association rule* mining task. To the best of our knowledge, no other existing work has undertaken the association rule mining problem from human mobility traces.

The rest of the paper is organized as follows: Sect. 2 presents the proposed framework. Experimental evaluation of the framework is shown in Sect. 3 through a case-study in the academic campus. We conclude in Sect. 4 with the future directions of the work.

2 Proposed Framework

In this section, we discuss our proposed framework [Fig. 1] to discover the interesting rules from mobile-user GPS traces. The framework consists of three modules, namely, *Mobility Data Pre-processing*, *Mobility Rule Generation* and *Significant Rule Learning*. Before describing the modules, few basic concepts and the problem definitions have been presented.

2.1 Preliminaries

1. **Labelled User GPS log (G):**
   User GPS log (G) is a sequence of time-stamped geo-tagged latitude, longitude points of an individual.
   $G = (C, U, < lat_1, lon_1, p_1, t_1 >, \ldots, < lat_n, lon_n, p_n, t_n >)$, where C is the category of mobile-user (say, student, faculty of an academic ROI), U represents the unique user-id and p_i stores the geo-tagged place (say, university, cafe etc.) of i^{th} point in the sequence.
2. **User Movement Summary (UMS):**
   UMS of an individual depicts the graphical representation of probabilistic relationship among the stay-points of the trajectory traces. $UMS = (N, \theta)$, where $N = (V, E)$ is the directed graph consisting V, stay-points and connecting edges E among the stay-points. The probability distribution among the stay-points or variables is quantified by θ. The detailed study of representing movement summary as probabilistic graphical model has been discussed in [3].
3. **Stay Point (S) and Point-of-interest Taxonomy (T):**
   S is defined as $S =< lat, lon, P, T_s, T_d >$ where within a radius of $d > D_{thresh}$ distance, an individual spends $T_d > T_{thresh}$ time at T_s timestamp. P represents the geo-tagging information of the stay-point. Point-of-interest taxonomy T is generated to represent the geo-tagged information in a hierarchical manner [3].
4. **GPS Transaction record (T):**
   A GPS transaction record, $T =< S, T_d, T_s, E_T, E_D >$ consists of staypoint information (S: Stay-point POI, T_d: Time-duration at S, T_s: Time-interval visited at S) along with the edge traversal information (E_T: Distance travelled from the previous stay-point, E_D: Time-duration to travel from the previous stay-point) between two consecutive stay-points.

In our proposed framework, the input dataset is a tuple of *latitude*, *longitude* and *timestamp* along with some application-specific information, namely, user-category, types of place visited etc. Therefore, there are several types of data available unlike transactional database. For a discrete (POIs, namely *ResidentialBuilding*, *Cafe* etc.) or categorical attribute (user category: *student*, *professor*), all the possible categories are mapped to a set of integers and continuous attribute (time) is discretized into several intervals. Further, a *day* is partitioned into 4 non-overlapping time-ranges allowing the continuous time-series to be mapped to different time-slots or buckets. Consequently, each data-item in the dataset can be represented as *attribute, integer-value*.

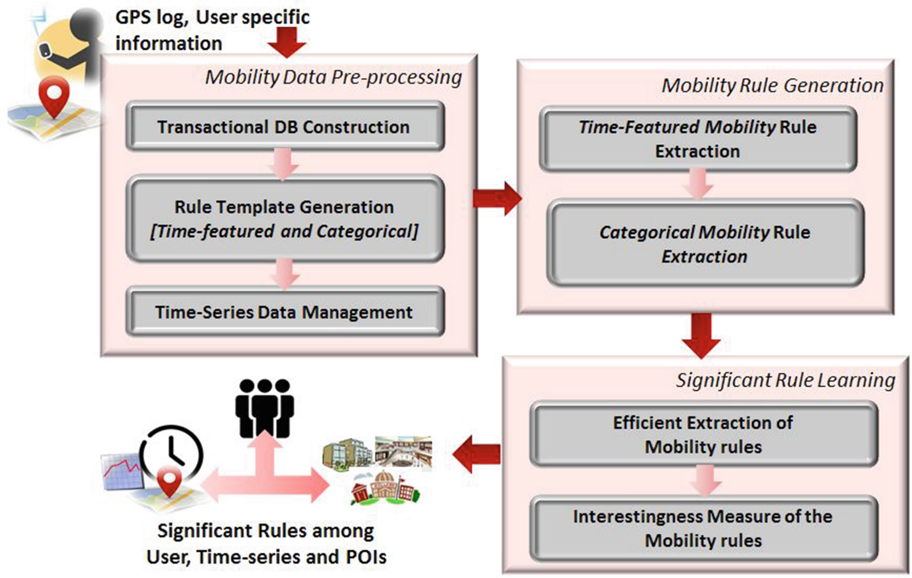


Fig. 1. Architecture of the mobility-rule mining framework

2.2 Problem Definition

In association rule mining algorithm, data record is taken as attribute-value pairs and association rules related to certain features of the attributes are extracted. In our GPS trace dataset, a number of features each having finite number of possible values are present which represent the objects or items in this domain. Example of transaction databases are shown in Figs. 2a and 3 where dimensions are user-category, place-information and time-features. We aim to extract two types of inter-transactional rules from the GPS trajectory of users.

Time-featured Mobility Rule: Given a transactional database of GPS or trajectory traces of moving agents, where each transaction consists of user-id, stay-point information (i.e., T_d: duration, T_s: Time-interval, P: place-information), edge-traversal information (i.e., E_d: distance and E_t: travel-duration to reach the stay-point), discover all the rules of the form $A \Rightarrow B$, where $A \in \{T_s, T_d, E_d\}$ and $B \in \{P, T_d, E_d, D_{AB}\}$ and $A \cap B = \phi$ and $Support(A \Rightarrow B) > thresh$. D_{AB} depicts derived attributes of the transactions within a sliding-window w. For example, D_{AB} may be frequency of visit at a stay-point or number of unique places visited within a time-interval.

Based on this definition, a rule might be "if the time of edge traversal is sufficiently large, then at a particular time-stamp, an individual's next stay-points belong to a certain set of POIs and stay-duration is likely to be more" and it can be expressed as $\Delta_{e_d} \wedge \Delta_{T_s} \Rightarrow \Delta_P \wedge \Delta_{T_d}$.

Category based Mobility Rules: Given a transactional database of summarized GPS or trajectory traces of different user-categories, where each transaction consists of user-id (u), category-id (C), stay-point information (S) along with the probability to follow the paths and time-interval (T), discover all rules of the form $A \Rightarrow B$, where $A \in \{C^{'}, T, S\}$ and $B \in \{S, T\}$, and $Support(A \Rightarrow B) \geq minSupp_{thresh}$. $C^{'}$ depicts derived user-categories from the transactional database. For example, 'Users visiting health-care center regularly' or 'Users having higher footprint at Library' are two examples of derived categorical attributes from the database based on movement behaviour.

2.3 Mobility Data Preparation

In this section, we briefly describe the database preparation from time-series GPS log followed by mobility rule template definitions and efficient storage of several mobility features of the GPS traces.

Generation of GPS Transactional Log from Movement Sequence: Figs. 2(a) and 3 show transactional database from typical GPS log and UMS at a particular time-instance.

Each of the stay-point related information along with the edge-traversal from the previous stay-point are recorded as a *transaction* in the database. An *item* or *literal* i in the transactional database (G_1) is an attribute value pair of the form (A_i, v), where A_i is an attribute and takes value of stay-point information (*place* and *time*) and edge-traversal information (distance covered from the previous stay-point and time duration of the travel). Any transaction in G_1 is identified by $< u_{id}, t_i >$; i.e., *user id* and the actual timestamp when the GPS point is captured.

In the next transactional database (G_2), each transaction is generated from any existing path in the UMS. Each path $< s_i, CPT_i >$ consists of a sequence of stay-points visited and the corresponding conditional probability values. Hence, there must be a corresponding entry in the transactional database to represent the movement history along the given sequence. For example, in Fig. 3, stay-points $< s_1, s_2, s_3, s_5 >$ and $< s_1, s_4, s_3, s_5 >$ depict two different sequence present in UMS and these two sequences are reflected as two transactions in G_2.

It is clear from the definitions and terminologies, some of the literals (time-duration, distance, timestamp) take continuous values while some literals (point-of-interests, user-category) form hierarchical structure or taxonomy. Figure 2(d) and 2(b) represent typical taxonomy of user-category and place-of-interests in an academic region-of-interest. In order to implement classical association rule mining techniques, we need to partition the continuous or qualitative variables into different intervals. Both of the literals *time-duration* and *distance* are partitioned into three intervals namely, *high*, *medium* and *low* and timestamp of a particular transaction falls into timeInterval-1 (0600–1200), timeInterval-2 (1200–1600), timeInterval-3 (1600–2100) and timeInterval-4 (2100-0600). We assume all the qualitative attributes belong to a non-overlapping partition or class.

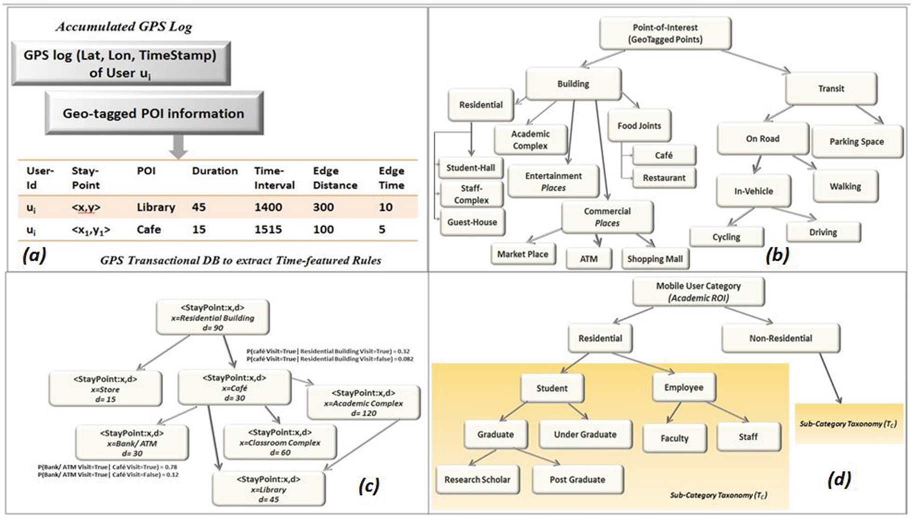
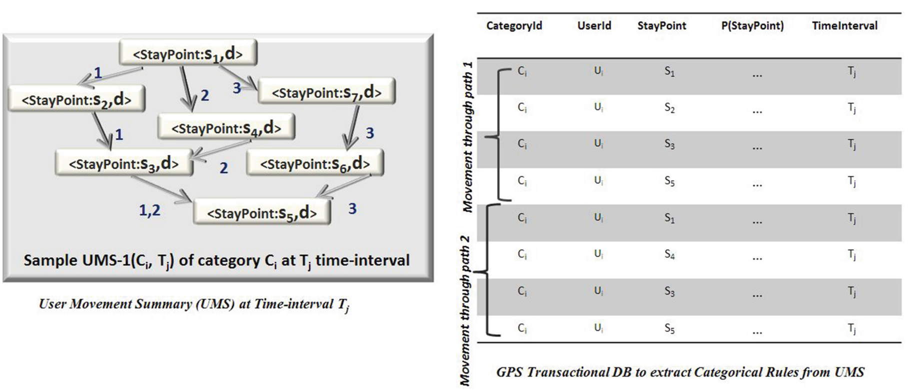


Fig. 2. (a) Sample entries on transactional DB of mobility traces from accumulated GPS log (b) Sample POI taxonomy of an academic ROI ($POI_{Taxonomy}(P)$) (c) Sample UMS (User Movement Summary): probabilistic graphical model of user's summarized mobility trace (d) User Category taxonomy (C) of an academic ROI



CategoryId	UserId	StayPoint	P(StayPoint)	TimeInterval
C_i	U_i	S_1	...	T_j
C_i	U_i	S_2	...	T_j
C_i	U_i	S_3	...	T_j
C_i	U_i	S_5	...	T_j
C_i	U_i	S_1	...	T_j
C_i	U_i	S_4	...	T_j
C_i	U_i	S_3	...	T_j
C_i	U_i	S_5	...	T_j

Fig. 3. Sample entries on transactional DB of categorical movement traces from UMS

Mobility Rule Template Generation: Clearly, *time* attribute is the key feature of UMS and acts as a principal factor to extract significant patterns or association rules from the movement summary. The key reason is "time has several meanings in movement data" and thus we need to combine all time-features to discover the inter-relationships between them. Examples of such interrelating time-feature rules are "if the edge-time T_e is more then next location stay-time T_s is likely to be more" or "a long-duration stay-time T_s is followed by a

Table 1. Mobility behaviour and corresponding potential rule-templates

T Id	Rule template mobility behaviour
T_1	$TimeStamp(t) \rightarrow stayPoint(f, P); f \in F, P \in POI$
	Time impacts footprint distribution at different POIs
T_2	$TimeStamp(t) \rightarrow StayPoint(d, P); d \in D, P \in POI$
	TimeStamp impacts on StayDuration at different POIs
T_3	$EdgeTraversal(dis, d) \rightarrow StayPoint(d); d \in D, dis \in L$
	Edge traversal information impacts on duration at next stay-point
T_4	$StayPoint(x, d_1) \wedge EdgeTraversal(dis, d_2) \rightarrow StayPoint(y, d_3)$
	$d_1, d_2, d_3 \in D, dis \in L, x < y \in S$
	Staypoint information and edge traversal influences consequent stay points
T_5	$StayPoint(x, d, t) \rightarrow is\_a(x, P)$
	$x \in S, d \in D, t \in T$
	Time-features of a particular stay-point implies the type of POI
T_6	$visits(x, A, f_1) \rightarrow visits(x, B, f_2)$
	$x \in U, A, B \in POI, f_1, f_2 \in F$
	Users visiting a POI with a particular frequency are likely to visit other POIs with certain frequency
T_7	$is\_a(x, c) \rightarrow visits(x, B, f_2)$
	$x \in U, c \in C, f_1 \in F, B \in POI$
	User category impacts on the visiting frequency of various POIs
T_8	$is\_a(x, c) \wedge TimeStamp(t) \rightarrow visits(x, B, f_2)$
	$x \in U, c \in C, f_1 \in F, B \in POI$
	User category and timestamp of the transaction influences the visiting frequency at different POIs

short-duration stay-time". Noticeably, all time-features of the time-series data: stay-time, duration, edge-time along need to be considered in the computation. We define potential mobility rule templates [Table 1] containing both time-featured [rule template $T_1 - T_5$] and categorical [rule template $T_6 - T_8$] mobility patterns of GPS traces. The intuition behind these rule-templates are "mobility traces (or sequence of movement) are highly dependent on timestamp values and each place-of-interest has direct correlation with several time-features which influences the movement behaviour of different categories of user."

Time-Series (GPS traces) Data Management: One of the major challenges to extract mobility rules is the huge amount of spatio-temporal traces. In order to reduce the computation load, we propose a *hash-based Place* (B_1) and *user-category* (B_2) bucket, where visit-frequency count and categorical visit-frequency count at different time-intervals are stored. A hash function is used to maintain the spatial correlation among the stay-points, i.e., two stay-points with minimum distance are stored into two consecutive buckets.

Figure 4(a) shows the visit-frequency lattice structure of the place-visit frequency where each node n_j at level j is a combination of $< H, L, M >$ or $< High, Low, Medium >$ tuple. Clearly, using a simple count function on $H(B_1, T)$ and $H(B_2, T)$, the visit-frequency lattice structure of different POIs is formed. Frequent item-set is extracted using a top-down approach on the place-frequency lattice. An item-set $< P_i, P_j >$ is *frequent* with the attribute value $< H, L >$, iff it is observed that people usually visit P_i place for a *high* stay-duration followed by *short*-duration visit at P_j.

2.4 Mobility Rule Generation

The anti-monotone property of *Apriori algorithm* presumes $\forall X, Y : (X \subseteq Y) \Rightarrow S(X) \geq S(Y)$, i.e., all superset of infrequent item-sets will be infrequent. Therefore, all the supersets of an infrequent item-set (support is less than min_{supp}) are discarded in the procedure. Our proposed algorithm utilizes the non-overlapping property of the frequency class and uses the top-down apriori algorithm. Further, we extract rules from two taxonomies, place-taxonomy and user-category taxonomy having different support values. Algorithm 1 depicts the steps of the mobility rule mining algorithm. The hash-based structures are initialized and updated when new place-visit or stay-point information is extracted from the transactional log. Using *IDDFS (iterative deepening DFS)* on the visit-frequency lattice structure, frequent item-sets are extracted. In the next run of the algorithm, derived or extended item-sets are discovered and all possible combinations of categorical and time-featured rules are extracted. Finally, all the rules having less values compared to the threshold interestingness measure value are discarded. The output of the mobility rule generation algorithm provides distinct mobility rules along with the interestingness measures.

2.5 Significant Rule Learning

We quantify the extracted rules using five measurements. While three commonly used measures *Gini Index (G)*, *Mutual Information (M)* and *J-measure (J)* along with support and confidence depict the rule-interestingness, another two proposed measures *category-relevant-measure (CRM)* and *user-relevant-measure (URM)* illustrate the applicability of a particular association rule to an user-category or group of people and an individual. In our study, we evaluate the rules using *J-measure (J), Gini Index (G)* and normalized *Mutual Information (M)* [5]. All of these measures are asymmetric measure, i.e. it implies there is a difference between $X \Rightarrow Y$ and $Y \Rightarrow X$. Clearly, in our case time-featured rules maintain a sequence of stay-points visit and there is a strong need to distinguish the strength of the implication rules for both the cases.

All of these measures captures the variance of the probability of GPS footprint distribution. It is interesting to note that, although there are several existing interestingness measures of association rules in the literature, they are not

Algorithm 1. Extracting distinct Mobility Rules from time-series data

Input: The GPS Transactional log $G = (g_1, g_2, \ldots, g_l)$, $POI_{Taxonomy}(P)$, $UserCategory_{Taxonomy}(C)$, sliding window w

Output: Set of unique time-featured mobility rules (R_T) and Categorical mobility rules (R_C);

$R_T = \{\}$; $R_C = \{\}$ ▷ *Initialize Mobility Rule-set:* R_T *and* R_C

L ← Generate *visit-frequency lattice-structure* for place-hash table

Iterative Deepening Depth First Search L to extract $i_m = \{< t_m, c_m, p_m > | (r_m\ is\ an\ item-set) \wedge (support(< r_m >\geq supp_{thresh}) | \}$ ▷ t_m*: Time-information,* p_m*: Place-information,* c_m*: category-information*

while $|i| > 0$ **do**

  $C^{'} \leftarrow C, F^{'} \leftarrow \{H, L, M\}, T \leftarrow \{1, 2, 3, 4\}$ ▷ *Derived attributes*

  $rule\ a \leftarrow NULL$

  for $c = 0; c \leq |C|; c++$ **do** ▷ *Candidate set Generation for Categorical Mobility Rules*

    $Generate\ Candidate - Set\ C_i^{'} = \{\{c_i \cup t_i\} \wedge \{c_i \cup p_i\}\}$

    $Prune\ candidate\ items : C_i = C_i^{'} - \{\gamma | (\gamma \in C_k^{'}) \wedge (k \subset (i-1)^{th} height\ of C\ taxonomy) \wedge (\gamma \notin c_i)\}$

    $Extract\ a_i = \{c | (c \in C_i) \wedge (support(c) \geq supp_{thresh}(\kappa))\}$

    $Extract\ pattern\ a_i\ \in L\ with\ maximum\ confidence$

    if a_i covers all instances in T **then** ▷ *Check for all time-intervals*

      $R_C = R_C \cup \{a_i\}$; $L = L - \{a\}$

    end if

  end for

end while ▷ *Repeat the same procedure to extract* R_T *using* w *time-sliding window*

$R = R_C \cup R_T$; $\kappa \leftarrow Find\ interestingness\ Measures(R)$ ▷ *Append all extracted mobility rules*

return $R, [\kappa]$ ▷ *Result: (Mobility Rule, Interestingness Measures)*

suitable for measuring associative patterns of trajectory data. There are few questions like, "Whether a rule (r_i) is prevalent in the mobility traces of an individual?" or "Whether a rule (r_j) is relevant or applicable to a particular user-category or a group of users showing higher support and confidence within the community?". Clearly, while the former rule (r_i) represents an individual's movement behaviour, the next one (r_j) is useful to capture the group-mobility behaviour. To this end, we propose two new measures, namely *category-relevant-measure (CRM)* and *user-relevant-measure (URM)*.

$$URM = 1 - \sum_{u=1}^{n} \frac{t_u(A \rightarrow B)}{\sum_{j=1}^{n} t_j(A \rightarrow B)} \log_2 \frac{t_u(A \rightarrow B)}{\sum_{j=1}^{n} t_j(A \rightarrow B)} \tag{1}$$

where u denotes an user-id and $t_j(A \rightarrow B)$ represents count of transactions containing rule $(A \rightarrow B)$ in the mobility trace of user j. Similarly, *category-relevant-measure (CRM)* is defined on user-group or categorical mobility summaries.

$$CRM = 1 - \sum_{c=1}^{|C|} \frac{t_c(A \rightarrow B)}{|c| \times t_c} \log_2 \frac{t_c(A \rightarrow B)}{|c| \times t_c} \tag{2}$$

Typically, CRM quantifies a rule applicability to a particular user-category, i.e., a higher value of CRM denotes only a few categories follow the mobility-rule, while a less CRM value depicts the rule is applicable for most of the user-categories. Similarly, a high value of URM represents only few individuals' mobility behaviour is represented by it.

3 Experimental Observations

DataSet: We demonstrate our approach using a real-life GPS traces of mobile-users of IIT Kharagpur Campus, an academic region. We collected 6 months GPS dataset from 56 volunteers (specifically 8 categories of users) from their mobile-GPS sensor and *GoogleMap Timeline*. Reverse geo-coding technique is used to extract the POIs of the region and each of the stay-points are geo-tagged.

Extracting Mobility Rules and Notations: Table 2 depicts few extracted association rules (TimeFeatured and Categorical) from the collected GPS traces. Rules are represented by following notations: $StayDuration(High, S, w = 1)$ implies stayduration of a movement-transaction is high at staypoint (S), where $w = 1$ and $w = T$ indicate same transaction and transactions within the same time-interval respectively. $TimeStamp(1)$ implies time-interval 1, i.e., [0600–1200]. $FootprintFrequency(f, P)$: Footprint frequency (f) at P POI(s). $countPlacePOI(a)$: The count of unique POIs(leaf nodes of $POI_{taxonomy}$) visited by the users in a time-interval is a. $User(f, P)$: Visit frequency (f) of users at P POI(s). $Category(c)$: c is a user-category of user-category taxonomy (C) [Fig. 2].

Results and Discussion: Table 2 depicts the *Support (S)*, *Confidence (C)*, *J-measure (J)*, *Mutual Information (M)*, *Category-relevant-measure (CRM)* and *User-relevant-measure (URM)* of each rule. Clearly, rules with higher URM are useful to capture individuals' mobility patterns and could be used for personalized mobility pattern mining and higher CRM-rules are useful for group-mobility pattern mining. It is worth noticing that none of the existing measures are capable to distinguish between these rules. Moreover, a simple support-pruning may eliminate personalized mobility-pattern rules as they might have a less support values. For example, rules r_4 has low support value as the number of participant of the category is less, however higher CRM value indicates meaningful patterns extracted from the GPS traces.

A rule is considered to be *interesting*, "if $A \rightarrow B$ is strong then $A \rightarrow \bar{B}$ must be a weak rule." The 'goodness of fit' between the rule hypothesis and data is measured by *J-measure*. While *mutual-information* depicts the average information shared by antecedent and consequent parts of the rule, quadratic entropy decrease is measured by *gini-index* [5]. From the experimental dataset, it is observed that rules R_1, R_2 have higher J values and average *G, M* values indicating higher information relative to the truth of the antecedent part.

Figure 4(b) shows a graph of running time of the algorithm (in sec) against the number of transactions. It clearly shows the improvement of running-time using the proposed data structure to capture inherent patterns. It is observed that time-featured rules are more time-intensive than categorical-rules. The reason is significant time is required to search the complete transactional log instead of only the summarized categorical patterns.

Table 2. Few examples of extracted TimeFeatured ($R_1 - R_5$) and categorical mobility-rules ($r_1 - r_5$)

Rule Id	Rule representation
R_1	EdgeDuration (High) $\wedge$ EdgeDistance (High) $\Rightarrow$ StayDuration (High,$S, w = 1$)
	S = 0.68, C = 0.85, J = 0.91, M = 0.67, G = 0.56, URM = 0.34, CRM = 0.39
R_2	StayDuration (High,S)$\wedge$ EdgeDuration (High) $\wedge$ EdgeDistance (High) $\Rightarrow$ StayDuration (Low,$S, w = T$)
	S = 0.58, C = 0.81, J = 0.92, M = 0.64, G = 0.61, URM = 0.38, CRM = 0.42
R_3	TimeStamp (1, 3) $\Rightarrow$ FootprintFrequency (High, {AcademicComplex, Transit, Department})
	S = 0.51, C = 0.75, J = 0.67, M = 0.72, G = 0.68, URM = 0.54, CRM = 0.59
R_4	TimeStamp (1, 4) $\wedge$ StayDuration (High) $\Rightarrow$ countPlacePOI (Low), countPlaceName (High)
	S = 0.52, C = 0.84, J = 0.61, M = 0.56, G = 0.52, URM = 0.56, CRM = 0.51
R_5	TimeStamp (3) $\Rightarrow$ StayDuration (Low) $\wedge$ Transit (High)
	S = 0.48, C = 0.67, J = 0.78, M = 0.66, G = 0.59, URM = 0.51, CRM = 0.56
r_1	User (High, HealthCare Center) $\Rightarrow$ User (Low, {Hospital, MedicineShop})
	S = 0.51, C = 0.87, J = 0.83, M = 0.87, G = 0.85, URM = 0.78, CRM = 0.51
r_2	User (High, {Cafe, Restaurant, Transit}) $\Rightarrow$ User (Low, {AcademicBuilding, ClassRoomComplex, Department})
	S = 0.78, C = 0.88, J = 0.78, M = 0.84, G = 0.78, URM = 0.65, CRM = 0.89
r_3	User ({High, Transit}, {Medium, Bank, ATM }) $\Rightarrow$ User (High, {MarketPlace, Store})
	S = 0.65, C = 0.84, J = 0.83, M = 0.75, G = 0.71, URM = 0.51, CRM = 0.57
r_4	TimeStamp (1,3) $\wedge$ Category (Faculty) $\Rightarrow$ User (High, {ResidentialArea, MarketPlace, Store, Transit})
	S = 0.32, C = 0.94, J = 0.51, M = 0.43, G = 0.41, URM = 0.87, CRM = 0.95
r_5	Category (Student, Residential, UnderGraduate) $\Rightarrow$
	User ({High, {Library, Cafe, HealthCareCenter, AcademicComplex}},
	{Medium, {MarketPlace, Store, CommercialPlace}})
	S = 0.42, C = 0.87, J = 0.79, M = 0.67, G = 0.64, URM = 0.78, CRM = 0.89

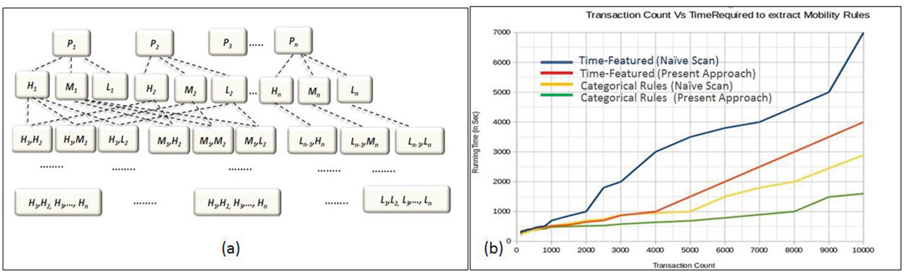


Fig. 4. (a) Lattice structure of place-visit frequency (b) Running time comparison

The experimental findings to explore human movement behaviour have following significances:

- The extracted mobility-rules (Table 2) with high *support, confidence* values demonstrate the potential of the proposed mobility rule mining framework to extract meaningful rules from accumulated GPS log. It has also been shown that although a mobility-rule may have lesser support value, it captures inherent patterns of individuals' mobility behaviour. This is a direct consequence of individuals' unique movement behaviour.
- It also turns out that mobility rules are highly dependent on timestamp values, i.e., people largely follow regular mobility patterns on the same time-intervals [Rules: R_3, r_4]. For example, footprint-frequency is high at market-place in evening and medium in morning while cafe, restaurants are crowded at evening. These may provide important insights about aggregated footprint density on different POIs and may be useful for resource-allocation.
- People also follow regular spatio-temporal patterns in their daily movement summary. Rule R_4 indicates people generally spends a long duration at specific places (POIs) in time-interval 1 [0600–1200] and time-interval 4 [2100-0600], while at time-interval 3 [1600–2100], low stay-duration and more transit-points (i.e., travelling) have been observed [Rule R_5].
- Mining mobility traces also provide interesting movement behaviour of people. For example, R_1 indicates people generally stays a longer duration in a POI after travelling a long distance. Also, it has been observed that a long-duration stay at a place is generally followed by short-duration stop within same time-interval $w = T$ [Rule R_2]. Again *high* visit at *Cafe or Hangout-spots* may be a reason of low footprints at*Acaemic-Complex, Department etc.* [Rule r_2].

In summary, our framework is capable to extract user-specific and categorical mobility-rules depending on several key factors, namely, time-intervals, POIs, stay-point duration etc. Further, data-preparation including GPS transactional log management reduces the time complexity of the procedure which is a major challenge in spatio-temporal data mining.

4 Concluding Remarks

In this work, we proposed a novel approach for mining mobility association rules of movement behaviour of people. The proposed framework extracts movement behavioural rules from accumulated GPS log of mobile users. An experimental evaluation on a real-life dataset of an academic campus demonstrates the potential of the framework to extract meaningful rules. Discovering reasonable mobility rules from GPS log is a major contribution of this work. We strongly believe the present work will act as a foundation of association rule mining framework from GPS trajectories. In the future, we also aim to assimilate other contextual information such as, weather information, traffic information and extend the present framework to extract interesting and meaningful rules from heterogeneous spatio-temporal data-set.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. ACM SIGMOD Rec. **22**(2), 207–216 (1993)
2. Ghosh, S., Ghosh, S.K.: THUMP: semantic analysis on trajectory traces to explore human movement pattern. In: Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, pp. 35-36 (2016)
3. Ghosh, S., Ghosh, S.K.: Modeling of human movement behavioral knowledge from GPS traces for categorizing mobile users. In: Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, pp. 51-58 (2017)
4. Czibula, G., Marian, Z., Czibula, I.G.: Software defect prediction using relational association rule mining. Inf. Sci. **264**, 260–278 (2014)
5. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: A survey. ACM Comput. Surv. (CSUR) **38**(3), 1–32 (2006). Art. No. 9 AQ3
6. González, M.C., Hidalgo, C.A., Barabási, A.-L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)
7. Zheng, Y.: Trajectory data mining: an overview. ACM Trans. Intell. Syst. Technol. (TIST) **6**(3), 1–41 (2015). Art. No. 29
8. Nahar, J., Imam, T., Tickle, K.S., Chen, Y.P.P.: Association rule mining to detect factors which contribute to heart disease in males and females. Expert Syst. Appl. **40**(4), 1086–1093 (2013)
9. Wang, C.-S., Lee, A.J.T.: Mining inter-sequence patterns. Expert Syst. Appl. **36**(4), 8649–8658 (2009)
10. Roddick, J.F., Spiliopoulou, M.: A survey of temporal knowledge discovery paradigms and methods. IEEE Trans. Knowl Data Eng. **14**(4), 750–767 (2002)
11. Tew, C., Giraud-Carrier, C., Tanner, K., Burton, S.: Behavior-based clustering and analysis of interestingness measures for association rule mining. Data Min. Knowl. Disc. **28**(4), 1004–1045 (2014)