International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-3, Issue-2)

Research Article | **February 2014**

An Enhanced Prediction of Subsequent Mobile User Behavior in Location Based Service Environments

Mrs Maalini D
PG Scholar of CSE Department
RVS Faculty of Engineering, Coimbatore, India

Mrs Subashini S
Professor of CSE Department
RVS Faculty of Engineering, Coimbatore, India

Abstract: ***Researches on Location-Based Service have been emerging in recent years due to a wide range of potential applications. One of active topics is the mining and prediction of mobile movements and associated transactions. A large amount of existing studies focus on discovering mobile patterns from the whole logs. However, this class of patterns may not be precise enough for predictions as the differentiated mobile behaviors amongst users and temporal periods are not considered. In this paper, we advise a new algorithm that is Cluster-based Temporal Mobile Sequential Pattern Mine, to realize the Cluster-based Temporal Mobile Sequential Patterns. Moreover, a prediction strategy is proposed to guess the subsequent mobile behaviors. In CTMSP-Mine, user clusters are construct by a novel algorithm named Cluster- Object based Smart Cluster Affinity Search Technique and similarities between users are appraise by the proposed measure, Location-Based Service Alignment. Meanwhile, a time segmentation approach is presented to find segmenting time intervals where similar mobile characteristics exist. To our best knowledge, this is the initial work on mining and prediction of mobile behaviors with considerations of user relatives and temporal property simultaneously. Through experimental valuation under various simulated conditions, the proposed methods are revealed to deliver excellent performance.***

***Keywords:** Data mining, Transportation, Mining methods and algorithms, Mobile environments.*

I INTRODUCTION

The advancement of wireless communication techniques and the position of mobile devices such as the mobile phones, cellular phones, have contribute to a novel business model. Mobile users can call for services through their mobile devices through Information Service and Application Provider (ISAP) from anywhere at any time. This business model is known as Mobile Commerce that provide Location-Based Services through mobile phones. MC is likely to be as popular as E-Commerce in the prospect and it is based on the cellular network collected of several base stations. The communication coverage of every base station is called a cell as a location area. The average distance among two base stations is hundreds of meters and the number of base stations is usually other than ten thousand in a city. When users move inside the mobile network, their location and service desires are store in a centralized mobile transaction database. Fig. 1 shows a MC scenario, where a user moves in the mobile network and desires services in the corresponding cell through the mobile devices. A moving sequence of a user, where cells are underline if services are request there. The record of service transactions, where the service S1 was request when this user moved to the location A at time 5. In fact, there exist insightful information in this data, such as movement and transaction behaviors of mobile users. Removal of mobile transaction data can give insights for different applications, such as data perfect and service recommendations. A mobile transaction database is complicated since a vast amount of mobile transaction logs are produced based on the user's mobile behaviors. Data mining is a widely used method for discovering valuable information in a complex data set [12] and a number of studies have discussed the copy of mobile behavior mining. The main variation between these literatures is the involved information of proposed patterns. Tseng addressed the problem of mining associated service patterns in mobile web networks. Mobile behavior will be more exact if we can find the corresponding mobile patterns in every user cluster and time interval.

Lack contribute to a novel business model. Mobile users can call for services through their mobile devices through Information Service and Application Provider (ISAP) from anywhere at any time. This business model is known as Mobile Commerce that provide Location-Based Services through mobile phones. MC is likely to be as popular as E-Commerce in the prospect and it is based on the cellular network collected of several base stations. The communication coverage of every base station is called a cell as a location area. The average distance among two base stations is hundreds of meters and the number of base stations is usually other than ten thousand in a city. When users move inside the mobile network, their locations and service requests are stored in a centralized mobile transaction database. Fig. 1 shows a MC scenario, where a user moves in the mobile network and desires services in the corresponding cell through the mobile devices. Fig. 1(a) shows a moving sequence of a user, where cells are underlined if services are requested there. Fig. 1(b) shows the record of service transactions, where the service S1 was request when this user moved to the location A at time 5. In fact, there exist insightful information in this data, such as movement and transaction behaviors of mobile users. Mining mobile transaction data can provide insights for different applications, such as data perfect and service recommendations. A mobile transaction database is complicated since a vast amount of mobile transaction logs are produced based on the user's mobile behaviors. Data mining is a widely used method for discovering valuable information in a complex data set [12] and a number of studies have discussed the copy of mobile behavior mining. The main difference between these literatures is the involved information of proposed patterns. Tseng addressed the problem of mining associated service patterns in mobile web networks. Mobile behavior will be more exact if we can find the corresponding mobile patterns in every user cluster and time interval. To provide precise location-based services for users, effective mobile behavior mining systems are required pressingly. Tseng also proposed SMAP-Mine to efficiently mine users' sequential mobile access patterns, based on the FP-Tree [13]. Chen [8] proposed the path traversal patterns for mining mobile web user behaviors. Yun proposed a novel method of mining mobile sequential patterns. To increase the accuracy of predictions, the moving path was taken into consideration in the above studies. However, mobile behaviors vary among different user clusters or at various time intervals. Prediction of mobile behavior will be more precise if we can find the corresponding mobile patterns in each user cluster and time interval. To provide precise location-based services for users, effective mobile behavior mining systems are required pressingly.

Clustering mobile transaction data helps in the find of social groups, which are used in application such as targeted advertising, public data allocation and personalization of pleased services. In previous studies, users are classically clustered according to their special profiles. However, in actual applications of mobile environments, it is frequently complex to obtain users' profiles. That is, we can have only right to use to users' mobile transaction data. To get the goal of user clustering lacking user profiles, we need to estimation the similarity of mobile transaction sequences. Although an amount of clustering algorithms have been planned in the rich literature, they are not official in the LBS scenario in thought of the following issues: (1) most clustering method like [9] can only course data with spatial match measures, while clustering methods with non-spatial parallel measures are essential for LBS environments. Request the users to set up a few parameters. However, in true applications, it is complex to establish the right parameters manually for the clustering tasks. Hence, a preset clustering method is essential. Although there exist many non-spatial similarity measures like [4], [6], [7] most of them are used to evaluate the string comparison. However, the mobile transaction sequences discussed in this paper include multiple and heterogeneous information likes time, location, and services.

Therefore, the existing measures are not applicable directly for measuring the similarity of mobile transaction sequences. The point interval segmentation technique helps us locate various user behaviors in different time intervals. For example, users may call for different services at different time even in the similar location. If the time interval factor is not in use into account, some behaviors may be missed during exact time intervals. To find whole mobile behavior pattern, a time space table is required. even if some studies used a pre-defined time distance table to mine mobile patterns [11] the data characteristic and data distribution vary in real mobile applications. Therefore, it is difficult to pre-define a suitable interval table by users. Automatic time segmentation. We offer a new data mining algorithm name Cluster-based Temporal Mobile Sequential Pattern Mine to capably mine the Cluster based Temporal Mobile Sequential Patterns of users. Then, new prediction strategy are proposed to successfully predict the user's subsequent behaviors using the exposed CTMSPs. To mine CTMSPs, a transaction clustering algorithm named Cluster Object based Smart Cluster Affinity Search Technique that build a cluster model for mobile transactions based on the proposed Location-Based Service Alignment similarity get advantage of the Genetic Algorithm to make more fitting time interval table. Based on the shaped user clusters and the time interval table, all CTMSPs can be exposed by the proposed method. To our finest knowledge, this is the first work on mining and prediction of mobile sequential patterns by allowing for user clusters and temporal relations in LBS environments concurrently. Firstly, through experimental evaluation on various replicated conditions, the proposed method is shown to deliver luminous performance in terms of exactness, recall and F-measure. The major contributions of this work are that we propose not only a new algorithm for mining CTMSPs but also two non-parametric techniques for growing the predictive precision of the mobile users' behaviors. Moreover, the proposed CTMSPs offer information including both user clusters and temporal relations. Now, user profiles like personal information are not desired for the clustering method and time segmentation technique proposed in this revise.

II RELATED WORK

In this part, we evaluate previous associated studies, which can be classify into four category: mobile pattern mining method clustering technique, temporal pattern mining method and mobile behavior predictions. In new year's, a number of studies have discuss the practice of data mining method to discover valuable rules/patterns from WWW Transaction databases [1],[2], [3], [13]and mobility data [10] Mining association rules [1] are proposed to locate important items in a transaction database. In [2], Agrawal proposed the Apriori algorithm to extract the association rules. In Park et al. proposed the DHP algorithm to recover the act of association rule mining. In the author designed an algorithm named WAP- Mine to capably grasp web access patterns in web logs, using a tree-based data structure with no candidate generation. Sequential pattern mining was first introduce in [3] to discover for time-ordered patterns, famous as in order patterns within transaction databases. For the study considering the relation between place and service, Chen [8] proposed the path traversal patterns for mining web user behaviors. Tseng first proposed the problem of mining connected service patterns in mobile web environments. SMAP-Mine was proposed by Tseng for competently mining users' sequential mobile.

Lee proposed T-MAP to efficiently find the mobile users' mobile access patterns based on SMAP in distinct time intervals which are pre-defined by users. Yun proposed the Mobile Sequential Pattern (MSP) to take moving paths into consideration and add the moving path between the left hand and the right hand in the pleased of rules. Jeung future a prediction approach called Hybrid Prediction Model (HPM) for estimate an object's future locations based on its model information. This paper considered an object's movements are more complicated than what the mathematical formulas can represent. Giannotti proposed trajectory pattern [10] for moving objects. However, there is no work considers user clusters and temporal relations in the mobile pattern mining simultaneously. The clustering analysis can be generally separated into two categories. The first group is on parallel measures that may concern the final clustering results indirectly. The Euclidean distance Edit distance LCSS [4], DTW [4], ERP [6], and EDR [7] are most popular similarity measures for string series or time series data analysis. Since mobile transaction sequence are not only time series movement string but also with service sequence, it is critical to properly define the parallel between dissimilar sequences. The second group is on the clustering methods. The mainly well-known clustering technique is the k-Means algorithm which is partition based. Other partition-based methods contain k-medoid PAM etc. These methods separation the data set into k clusters, based on similarities among data items, where k is a parameter specified by the user. Hierarchical clustering methods are one more popular kind of clustering technique. For density clustering methods, Ben-Dor potential the Cluster Affinity Search Technique (CAST) [5] require a similarity threshold t, where 0<t<1.The algorithm guarantee that the average similarity in each generate cluster is higher than the threshold t.

Tseng planned the Smart Cluster Affinity Search method. The major ideas of the Smart-CAST are as follows. First, the technique uses the CAST as the essential clustering method. Second, the technique uses a value validation method, Hubert's Γ (gamma) statistic to find the finest clustering result. However, mainly clustering methods can only route data with spatial similarity measures. For example, k-Means, PAM, and DBSCAN [9] can only use the Euclidean preserve as similarity measure. Still, the similarity between mobile transactions cannot be exact by the Euclidean distance. Besides, most clustering methods apply for the users to set up some parameters before the clustering task. For example, DBSCAN want a density radius ε and a minimum digit of objects Min Pts to be generate. However, in real applications, it is difficult to resolve the right parameters manually for the cluster tasks. Previous studies and application judge time to be an important factor. Users have some detailed behaviors in specific time [11], [19]. Accordingly, knowing the model of each time space helps to enhance customize service. In [11], the future method that divides the data into multiple time intervals and determines user navigation in each time interval increases the prediction rate in a mobile web environment. In Lee proposed T-MAP to efficiently find the mobile users' mobile access patterns in distinct time intervals. The exposed patterns afford real-time adapted personal service for users. However, this proposed technique lacks flexibility since we should set up the start time and end time of the time interval in advance. In fact, a same cutting method is not apposite for all data. The segmenting point of the time interval controls the accuracy rate of mobile behavior prediction. Because it is not simple to locate the top segmentation points of time intervals, the inherent algorithm is generally used to solve such complex problems. The inherent algorithm [15] was future by Holland. It desires to define a fitness function to assess the value of a chromosome, and then accidentally produce a population. Through the growth process: Selection, Crossover and mutation, the chromosomes of the population repetitively generate new generations. The weakest chromosomes turn into obsolete. The mobile actions predictions can be around divided into two category. The first category is time series based calculation that can be separated into two types: (1) linear models and (2) non-linear models. The non-linear models measured the object's activities by more sophisticate regression functions. Thus, their prediction accuracies are superior to those of the linear models. Recursive Motion Function (RMF) is the mainly correct prediction method in the writing based on regression functions. The second category is model based prediction. Ishikawa derived a Markov Model (MM) that generate Markov transition probability from on cell to another for predict the next cell of the object. It means that when the idea of the pattern occurs, the outcome will also occur with chance c. However, these methods can only expect the next spatial locations of objects. SMAP Mine was first planned to learn sequential mobile access rules and predict the user's next locations and services. The form of the rule is {ri, si} {rj, sj} with a assurance c, where ri and rj are locations, and si and sj are services. It imply that a user request si in ri will have next location and service as rj and sj with c probability. In Yun planned the Mobile Sequential Pattern (MSP) to calculate the next mobile behaviors. The form of the model is {(ri, si), (r1), (r2), (r3), (rj, sj)}, where item (ri, si) indicate a user request service si at location ri. It means a user requests service si in location ri and then desires service.

III PROPOSED METHOD

In this section, we express our system plan. Four main research issue are addressed here: clustering of mobile transaction sequences, time segmentation of mobile transaction sequences, discovery of CTMSPs, and mobile behavior prediction for mobile users.

A. System Framework

The proposed system framework. Our system has an "offline" mechanism for CTMSPs mining and an "online" engine for mobile behavior prediction. When mobile users shift within the mobile network, the information which include time, locations, and service requests will be store in the mobile transaction database. It shows a model of mobile transaction database which contain 7 records. In the offline data mining device, we propose two techniques and the CTMSP- Mine algorithm to learn the knowledge. First, we propose the CAST algorithm to cluster the mobile transaction sequences. In this algorithm, we intend the LBS to assess the similarity of mobile transaction sequence. Secondly, a GA-based time segmentation algorithm to discover the most suitable time intervals. After clustering and segmentation, a user cluster table and a time interval table are generate, correspondingly. Thirdly, we suggest the CTMSP-Mine algorithm to mine the CTMSPs from the mobile transaction database according to the user cluster table and the time interval table. In the online prediction engine, we offer a behavior prediction strategy to expect the subsequent behaviors according to the mobile user's earlier mobile transaction sequences and current time. The major idea of this structure is to give mobile users a exact and resourceful mobile behavior prediction system.

B. Clustering of Mobile Transaction Database

In a mobile transaction database, users in the dissimilar user groups may have dissimilar mobile transaction behaviors. The first charge we have to attempt is to cluster mobile transaction sequences. We future a parameter less clustering algorithm CO- Smart Cluster. The entry Sij in matrix S represent the similarity of the mobile transaction sequence i and j in the database, with the degree in range of [0, 1]. A mobile transaction sequence can be view as a sequence string, where every element in the string indicate a mobile transaction. The main challenge we have to deal with is to quantify the content similarity among mobile transactions. The base similarity make is set as 0.5. Two mobile transactions can be allied if their locations are the similar Otherwise, a place penalty is generate to decrease their comparison score. Detect that the maximal number of location penalty is |s1| + |s2|. When two sequence are totally special, the similarity score of them is 0. When two mobile transactions are allied, we calculate their time penalty (TP) and service reward (SR). TP focuses on their instance distance. The farther the time distances linking them, the larger the time penalty of them. TP that is generated to decrease their parallel score is defined as, where length indicates the time extent. SR focuses on the similarity of the service desires. The more parallel the service requests of them, the larger the service prize of them. SR that is generated to enhance their similarity score is defined as the procedures of an LBS-Alignment measure. Input data include two mobile transaction sequences (line 1). Output data is the similarity between two mobile transaction sequences, with the degrees in range from 0 to 1 (line 2). Some parameters are initialized (line 4 to line 7). The bottom similarity score is set as 0.5 (line5). We use active programming to analyze Mi,j (line 8 to line 18). Mi,j indicates the rate of matrix M in column I and row j, where M is the score matrix of LBS-Alignment. In this method, if the locations of two transactions are the similar (line 10), both of the time penalty (line 11) and the service reward (line 12) are intended to compute the similarity score (line 13). Otherwise (line 14), the location penalty is generated to reduce the similarity score (line15). Finally, Ms.length,s' length is returned as the similarity attain of the two mobile transaction sequences (line 19). Example 1. Let s and s' be location penalty is 0.5 / (|s| + |s'|)= 0.05, where both |s| and |s'| are 5. Fig. 4(a) shows the detailed process of s and s' by using the dynamic programming. The similarity of s and s' is 0.405. Fig. 4(b) shows the LBS-Alignment result of s and s'. After obtaining the similarity matrix, we cluster the mobile transaction sequences by the proposed CO-Smart-CAST. Fig. 5 shows the procedure of CO-Smart-CAST. The input data is an N-by-N comparison matrix S (line 1).The output data is the clustering result (line 2). CO-Smart-CAST can frequently cluster the data according to the parallel matrix without any user input parameter. The main thoughts of CAST are as follows. First, the CAST method that takes a parameter named parallel threshold t is used as the basic clustering technique. Secondly, we use a value validation method, called Hubert's Γ Statistics, to locate the finest clustering result. Thirdly, we use a hierarchical concept to decrease the sparse clusters. For a clustering effect, we use Hubert's Γ Statistics to calculate its quality by taking the similarity matrix and the clustering effect as the input. In each clustering result, we analyze its Γobj and Γclu which characterize the clustering qualities precise by the original object similarity matrix S and the last cluster similarity matrix S', respectively. The initial values of S' and S are the similar since we let each object be an independent cluster. We use the F1 score which is the harmonic mean to join Γobj and Γclu as ΓCO. A higher value of ΓCO represents the superior clustering quality. To conclude the most suitable t, the easiest way is varying t with a permanent increment and iterating the executions of CAST to locate the best clustering result with the highest ΓCO. The main disadvantage of this way is that many iterations of multiplication are required. For this reason, we try to decrease the number of computations by eliminate unnecessary executions and then get a "near-optimal" clustering result. That is, we try to achieve a minimal number of CAST executions. The main design is to narrow down the choice of *t* effectively. A trying range *R* for setting *t* is from 0 to 1. (Line 5).By the points *P0*, *P1*, *P2*, *P3*, and *P4*, *R* is equally separated into 5 points, where *P0* < *P1* < *P2* < *P3* < *P4*. Then, the value of each *Pi* (line 8) is in sequence taken as the attraction threshold to execute the CAST algorithm (line 9), and then obtain the *ΓCO* of the clustering result of each *Pi* (line 10 to line 12). When a run of execute the clustering is complete (line 7 to line 13), the clustering at point *Pb* that produce the highest *ΓCO* is careful to be the best clustering (line 14). Then, the testing range *R* is partial within the new range [*Pb*-1, *Pb*+1] contain the point *Pb* (line 15). The above process is constant until the testing range *R* is lesser than the threshold *ε* (line 16), where *ε* is a very small value, i.e., less than 10-5. If the *ΓCO* marker produced by point *PBest* is upper than the best *ΓCO* statistic (line 17), the best cluster product is recorded (line 18 and line 19) and all of the entities in comparison matrix *S'* are modified to the average similarities between all pairs of parallel cluster results (line 20 to line 24). The total process is frequent until no better *ΓCO* statistic is generated (line 07 to line 26).

C. Segmentation of Mobile Transactions

In a mobile transaction folder, similar mobile behaviors live underneath some certain time segments. Hence, it is main to make suitable settings for time segmentation so as to classify the characteristics of mobile behaviors beneath different time segments. We suggest a genetic algorithm (GA) based method to by design obtain the most suitable time segmentation table with familiar mobile behaviors. Fig. 7 show the method of our proposed the time segmentation method, name Get Number of Time Segmenting Points (Get NTSP) algorithm. The enter data is a mobile transaction database *D* and its time length *T* (line 01). The output data is the number of instant segmenting points (line 02). For every item, we gather the total number of occurrence for each entry in each time point (line 07 to line 11). Then, an item can copy a curve of count distribution, For all curves, we originate the time points with the major change rate (line 13). We defined the change rate as ($c[i+1]$ - $c[i]$) / (1 + $c[i]$), where $c[i]$ correspond to the total number of occurrence for the item at time point i. We add up occurrences of all these time points (line 15), and discover out the happy time points whose count are larger than or like to the average of all occurrence from these ones, and then take these pleased ones as a set of the time point series .Initially, we randomly produce the initial population and classify a suitable fitness function. Through frequent selection, crossover and mutation, we find an optimal solution. There are three operator in Genetic Algorithm: selection, crossover, and mutation. For the collection operator, a proportion of the current population is select to product the next population in every generation. Individual chromosomes which are elected based on their health value. The larger the fitness value of a chromosome, the upper probability the chromosome is selected. For the crossover operator, we relate one point crossover involve a crossover probability to this operator. A crossover point on both parent chromosomes is accidentally selected. All time segmenting points past the crossover point is swap between the two parent chromosomes. The ensuing chromosomes are the children. For the mutation operator, we apply the one-bit mutation to this operator. This operator involve a mutation probability that arbitrary time segmenting point in a chromosome will be altered from its original state.

D. Discovery of CTMSPs

In order to mine the cluster based temporal mobile sequential patterns ably, we future a novel technique named CTMSP Mine to appreciate this mining process. The basic proposal of CTMSP Mine is based on TJPF algorithm proposed in [37]. However, the TJPF algorithm did not think the factors of user cluster and time interval, which are needed in discover the total information relating to personal mobile behaviors. In CTMSP Mine, both factors of user cluster and time interval are taken into account such that the absolute mobile sequential pattern can be exposed. The entire procedures of CTMSP Mine algorithm can be divided into three main steps: Frequent-Transaction Mining, Mobile Transaction Database Transformation and CTMSP Mining. Frequent Transaction mining, in this part, we mine the frequent transactions in every user cluster and time interval by applying a customized Apriori algorithm [2]. Table 1 shows the mobile transaction database. There are two user clusters and two time intervals in the database, i.e., $C1$ = {1,2, 4, 7}, $C2$ = {3, 5, 6}, $T1$ = {1-20}, and $T2$ = {21-32}. At first, the carry of every cell and service is counted in every user cluster and time interval according to the user cluster.

Mobile Transaction Database Transformation. We use Frequent Transactions to convert every mobile transaction sequence S into a frequent mobile transaction sequence S'. If a transaction T in S is frequent, T would be distorted into the parallel Frequent Transaction. Otherwise, the cell of T would be altered into a part of path. The result of repeated mobile transaction database altered. Take the limited sequence (C1, T1, A, LS1) BC (C1, T1, D, LS2) in Uid 2 as an example, (C1, T1, A, LS1) and (C1, T1, D, LS2) are inelegantly misshapen from the contact (5, A, S1) and (20, D, S2), since they are recurrent transactions. The path BC among them is generate from (10, B, ∅) and (19, C, ∅). The major objectives and advantages are: 1) service set can be represent by symbols for capably processing; 2) transactions whose hold is less than the least support threshold can be eliminate to ease the size of database.

CTMSP Mining. In this phase, we excavate all the CTMSPs from the recurrent mobile transaction database. Frequent 1 CTMSPs are obtained in the recurrent transaction mining phase. In the mining algorithm, make use of a two level tree named Cluster-based Temporal Mobile Sequential Pattern Tree. The internal nodes in the tree store the common mobile transactions, and the leaf nodes store up the matching paths.

IV CONCLUSIONS AND FUTURE WORK

In this paper, we have future a new technique, named Cluster based Temporal Mobile Sequential Pattern Mine for discovering Cluster based Temporal Mobile Sequential Patterns in Location-Based Service environments. besides, we have planned new prediction strategies to expect the successive user mobile behaviors using the exposed CTMSPs. In CTMSP-Mine, we first advise a transaction clustering algorithm named Cluster Object based Smart Cluster Affinity Search Technique to form user clusters based on the mobile transactions using the proposed Location-Based Service Alignment comparison measurement. Then, we utilize the genetic algorithm to produce the most apposite time intervals. To our finest knowledge, this is the first work on mining and prediction of mobile behaviors connected with user clusters and temporal relations. A series of experiment were conduct for evaluate the performance of the proposed methods. The new results show that CAST methods get high quality clustering results and the proposed CBSS strategy obtain highly exact results for user classification. Meanwhile, our GA based technique obtain the most appropriate and correct time intervals. For behavior calculation, CTMSP is exposed to outperform other prediction technique in terms of precision and F-measure.

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rule between Sets of Items in Large Databases," Proc. ACM SIGMOD Conf. on Management of Data, pp. 207-216, May 1993.
[2] R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules," Proc. 20th Int. Conf. on Very Large Databases, pp. 478- 499, Sept. 1994.

[3] R. Agrawal and R. Srikant, "Mining Sequential Patterns," Proc. Int. Conf. on Data Engineering, pp. 3-14, Mar. 1995.
[4] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman,"Basic Local Alignment Search Tool," J. of Molecular Biology,vol. 215, no. 3, pp. 403-410, Oct. 1990.
[5] A. Ben-Dor and Z. Yakhini, "Clustering Gene Expression Patterns," J. of Computational Biology, vol. 6, no. 3, pp. 281-297, Jul. 1999.
[6] L. Chen and R. Ng, "On the marriage of Lp-norms and edit distance," Proc. 30th Int. Conf. on Very Large Databases, pp. 792- 803, Aug. 2004.
[7] L. Chen, M. Tamer Özsu, and V. Oria, "Robust and Fast Similarity Search for Moving Object Trajectories," Proc. ACM SIGMOD Conf. on Management of Data, pp. 491-502, Jun. 2005.
[8] M.-S. Chen, J.-S. Park, and P. S. Yu, "Efficient Data Mining for Path Traversal Patterns," IEEE Trans. on Knowledge and Data Engineering, vol. 10, no. 2, pp. 209-221, Apr. 1998.
[9] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, pp. 226-231, Aug. 1996.
[10] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," Proc. 13th Int. Conf. on Knowledge Discovery and Data Mining, pp. 330-339, Aug. 2007.
[11] M. Halvey, T. Keane, and B. Smyth, "Time-Based Segmentation of Log Data for User Navigation Prediction in Personalization," Proc. IEEE Int. Conf. on Web Intelligence, pp. 636-640, Sept. 2005.
[12] J. Han and M. Kamber, "Data Mining: Concepts and Techniques, 2nd Edition," Morgan Kaufmann, Sept. 2000.
[13] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," Proc. ACM SIGMOD Conf. on Management of Data, pp. 1-12, May 2000.
[14] R. J. Hathaway and J. C. Bezdek, "Visual Cluster Validity for Prototype Generator Clustering Models," Pattern Recognition Letters, vol. 24, no. 9-10, pp. 1563-1569, Jun. 2003.
[15] J. Holland, "Adaptation in Natural and artificial system," University of Michigan Press, Ann Arbor, 1975.