Multi-Cluster Based Temporal Mobile Sequential Pattern Mining Using Heuristic Search

M. I. THARIQ HUSSAN
Research Scholar & AP(Sr.G),
Department of Computer Science and Engineering,
Paavai Engineering College, Pachal,
Namakkal, Tamilnadu,
INDIA
thariqhussain@rediffmail.com

Dr. B. KALAAVATHI
Professor & Head
Department of Computer Science and Engineering
K. S. R. Institute for Engineering & Technology
Tiruchengode, Tamilnadu,
INDIA

Abstract: - An enhanced mobile sequential pattern mining using heuristic search technique is explored to predict mobile user's behavior effectively. By analyzing the movement of mobile users with respect to time, location and service request, one can contend that users in different user groups may have different mobile transaction behavior. Similar transaction behavior in a set is grouped by applying heuristic search technique. The heuristic search technique efficiently performs search on mobile transaction database to prune undesired transaction to form cluster and make them refined, meaningful, and relevant to the query. Research on multi-cluster based sequential pattern mining has been emerging in recent years due to a wide range of potential applications. One of the active topics is the facilitation of wireless and web technologies to the mobile user's through the usage of mobile devices at anytime and anywhere. This approach has been evaluated with the transactional dataset and simulation is carried out with data obtained from the real world to generate the required network environment. Compared with Cluster-based Temporal Mobile Sequential Pattern (CTMSP), the evaluation results show that Multi-Cluster based Temporal Mobile Sequential Pattern using Heuristic Search (MCTMSPHS) achieves 30 to 40% more in accuracy, 40 to 50% less in energy usage and 20 to 25% less in execution time.

Key-Words: - Mobile Sequential Pattern Mining, Heuristic Search, Clustering, Mobile Environment

1 Introduction

Mobile Sequential Pattern Mining is a premeditated problem in the context of mobile computing. A special sequential pattern reflects the moving behavior of mobile users which attracts many of the researcher's. Mining sequential pattern has huge implication for effective and efficient location management in wireless communication systems. Moving sequential pattern is a kind of mobile behavior that methodically describes the problem of mining it.

Mobile sequential pattern uses the heuristic approach to forecast the mobile user behavior at minimal time. The heuristic method is planned which combines the most attractive features of the non-hierarchical clustering and course traversal patterns for pulling out mobile web user behavior. However, mobile behaviors disagree among different user clusters or at different time intervals. The calculation of mobile behavior is more specific if it is probable to discover the analogous mobile patterns in every user cluster and time interval.

This work (MCTMSPHS) focuses on mining and calculating the mobile actions connected with user cluster and chronological relation. In this work, heuristic approach is introduced and the multi-cluster temporal position of each agent can consider clustering in the space of events with combined spatial position of the participants and its time.

2 Literature review

A mobile transaction database is very complex because a massive quantity of mobile transaction log is formed with the user's mobile behavior. Data mining is an extensively employed method for determining important information in a multipart data set and a number of improvements have conversed the subject of mobile action mining.

Authors in [1] proposed an algorithm Cluster-based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine) to determine the Cluster-based Temporal Mobile Sequential Patterns. In CTMSP-Mine, the user clusters are created by a technique called Cluster-Object-based Smart Cluster Affinity Search Technique. In this work, the users are evaluated by Location Based Service Alignment. High-dimensional data space is classically very exposed and major clusters are established in inferior dimensional subspaces. Clustering, conversely, build up one-dimensional association and as a result, a more difficult depiction of the data set is generated.

A cluster-adaptive distance [2] introduces the matching cluster based list can present in the one-dimensional table. The competent spatial filtering is performed with a reasonably minute preprocessing storage overhead.

A novel method termed Dark Block Extraction [3] for determining the number of clusters in unlabeled data set is examined which employs frequent widespread image and pointer indulgence subspace techniques. In this work, the energy handling and latency are considered as main concern.

In Upturned list-style indexing technique [5], the upturned lists are positioned and a two-level indexing is considered to reduce the latency.

Delineation mobile behavior among the user's and chronological periods are not considered [5], so a calculation strategy is proposed [6] to determine the behavior pattern of successive mobile user's over a period of time, but the indexing scheme is not efficiently handled.

Subspace grouping is an emergent task [7] which aims to differentiate clusters inbuilt in subspaces. Oriented Synchronized Clusters, a new effective and experienced technique were introduced in [8] for subspace clustering stimulated by the organization.

Energy saving in WSNs is extended in the Information Controlled Transmission Power adjustment [9], where nodes with additional sequence employ superior transmit power than those that are less enlightening to split its objective state information with the adjoining nodes. Narrative subspace clustering model [10] determines the clusters based on the comparative section compression in the subspaces. The clusters are scrutinized as region whose thickness is reasonably high in a subspace.

A chronological data clustering structure through a weighted clustering groups the numerous partitions created by initial clustering examination on different temporal

data representation. The work presented in [11] proposed a narrative weighted method by clustering legalization measure to come together primary partition varied viewpoints.

The CTMSP [1] for prediction uses the moving path of user's and considers only the individual user not the user group for mining. CTMSP uses the Location-Based Service which does not provide efficient prediction of mobile user behavior. It consumes more time to predict mobile user behavior and also it lacks in accuracy. A major distinction among these literatures is the concerned information of proposed system.

3 Proposed Heuristic Search Technique for Mobile Sequential Pattern Mining

A heuristic search mechanism is presented for the clustering of a mobile transaction database with different user groups. The users in the different user groups may have different mobile transaction behavior. The main goal of MCTMSPHS is to present a cluster in a list form where tightly related transactions are near to each other. MCTMSPHS work is not concerned with the effectiveness of the search engine, but the way the outcome is presented.

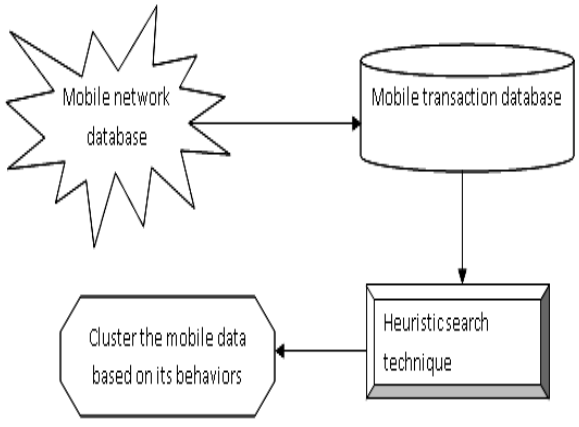

Fig 1. Architecture diagram of the proposed MCTMSPHS

The architecture diagram of MCTMSPHS is shown in the fig 1. The architecture diagram includes four blocks.

- Mobile Network Database
- Mobile Transaction Database
- Heuristic Search Technique
- Clustering the Mobile Data

3.1 Mobile Network Database

The mobile network database in MCTMSPHS comprises of the database connected to by a mobile computing device, including smart phones, PDAs, etc.

3.2 Mobile Transaction Database

The second block of MCTMSPHS holds the mobile transaction database. The operation of clustering is performed in this block. A heuristic search technique is presented for clustering the mobile transaction behaviors stored in the mobile transaction database. The database involved in mobile transaction is highly complex due to the massive amount of mobile transaction logs formed.

3.3 Heuristic Search Technique

The third block comprises of Heuristic Search, an Artificial Intelligence searching method that concerns heuristics. It makes use of supplementary knowledge concerning the crisis that assist the straight search to more capable path. Heuristics assist to diminish the number of substitutes from an exponential number to a polynomial number. Heuristics take part in positioning and searching the most favorable path in the web search result graph. It relates beam exploration in the search outcome graph that is analogous to conventional clustering technique on the clusters.

3.4 Clustering the Mobile Data

Finally, the last block concentrates on clustering the mobile data which comprises of agent, seeker node, associated node, singleton cluster and negotiator. Singleton cluster hold the query result for the specific group of user query. The idea of heuristic search used only in clipping the undesired communication of mobile chronological pattern that is conserved

in the chronological database to figure out clusters.

4 Transaction Process of MCTMSPHS

Initially every mobile node of the database locates transaction obtained from the second block, which is then allocated with a negotiator. The agent node has a list which is distributed to all its associated nodes that are present in the same cluster.

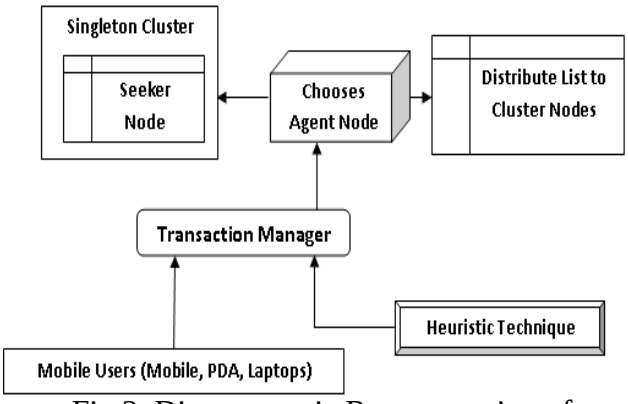

Fig 2. Diagrammatic Representation of overall transaction process

Fig 2 describes the process flow of seeker node and transaction manager activities. Every agent chooses a seeker node from the catalog based on transportation of nodes mobile transaction. Initially, the transportable source node is the seeker node, which is a node to be included in a particular cluster. During each process, the transactional manager handles the entire queries from mobile users by choosing a seeker node to discover its associated nodes and insert them to the list. This variety of seeker node is completed employing heuristic technique to discover a node with eminent number of statement. Clusters in a path can be recognized by elevated figure of edges within and fewer numbers of edges among them. Afterward the lists attained by agent are traversed. If a frequent node is established in two lists, the groups are combined that their source nodes fits. In this method, the communication which can be obtained is topologically comparable.

The query is also symbolized as a vector in a k-dimensional space. The query vector can then be evaluated to all offered text feature vectors in a cluster and the documents are graded by their resemblance (nearness) to the query. One widespread evaluation of similarity is the cosine measure which computes similarity between the query and mobile operation vector. Given two vectors, of attributes X and Y, the cosine measure cos(θ) is given as

$$\text{Cosine Measure} = \frac{X \cdot Y}{\|X\| \|Y\|}$$

$$= \frac{\sum X_i * Y_i}{\sqrt{\sum (X_i)^2} * \sqrt{\sum (Y_i)^2}} \quad (1)$$

where i = 1,2,…n. This assessment of query with all the mobile operation nodes in all clusters will be completed similar. The ‘X’ are PDA users and ‘Y’ are mobile users where 1, 2…..n are the ‘n’ number of queries by different users. The mobile operation which has the maximum comparison with the query is situated initially in the cluster list. The other results stored in mobile database are located stable with their comparison with the query in the list in lessening order. The cluster with the uppermost number of associates is presented on the primary phase of the clustering system. The algorithmic representation is as given below:

Step 1	Parse the exploration outcome and figure a network.
Step 2	Allocate an agent to every mobile node of the network.
Step 2.1	Every agent holds a list of associated nodes.
Step 2.2	Initially, the agent holds the source mobile node.
Step 2.3	Every agent chooses capabile mobile node "seeker node".
Step 2.4	During each iterations, more seeker nodes are explored.

Step 3 Each associated node expands list of nodes.
Step 3.1 Filters seeker node using heuristic.
Step 3.2 Update seeker node after search.
Step 4 Every mobile seeker node is allocated to singleton cluster.
Step 4.1 During every iteration, the agent updates the list by choosing a seeker node applying heuristic.
Step 5 If widespread node established for two source nodes, cluster are merged.
Step 5.1 Process followed for predefined number of iterations.
Step 5.2 Finally, the clusters are obtained.
Step 6 Haul out the contents of the mobile nodes of every cluster and preprocess them.

The above algorithm describes the process of clustering the mobile data based on the behavior of user transactions and processed in terms of heuristic search technique. At first, it selects the mobile node which should act as an agent for identifying the clusters in the form of transactional behavior. With the mobile node behavior, the transactional items are clustered and formed a group. Through the group, the transactional behavior of the user are identified and formed in a respective manner.

5 Experimental Evaluation

In this section, a series of experiments are performed to appraise the performance of the proposed MCTMSPHS underneath diverse system condition. N*N sample are taken to define and all users are split up into N user groups consistently. When a user activity inside mobile nodes for challenging in the network is analyzed, the mobile operation progression is produced by this user. The technique for creating mobile transaction series is referred**.** In mobile system, a node is formed with a base station in every part. The mobile user's progress in arbitrary from one position to another and admit more services from the web during ISAP. The user's association from one part to another part, their demanded services and the time are supervised and logged into massive database. Users in the diverse user groups might contain diverse mobile transaction behavior. LBS-Alignment method proceeds with the similarity amongst numerous users and gives it to MCTMSPHS. The performance of the proposed MCTMSPHS is measured in terms of accuracy, execution time and energy usage.

Accuracy: It defines the rate of accuracy measured using heuristic technique for deriving the mobile user's behavior. The classification accuracy (CA) is evaluated using the formula as given below:

$$CA = \sum \max \left\{ \frac{|P_i * Q_{mj}|}{|P_i| + |Q_{mj}|} \right\} \Big/ N \quad (2)$$

where $P = (P_i,\ldots P_k)$ refers to the transaction dataset and $Q_m = (Q_{m1},\ldots Q_{mk})$ refers to the partition obtained by the algorithm given above and N is the total no. of samples (i.e) mobile devices taken.

Execution Time: The execution time is denoted as ET, where ET measures the time taken to execute and classify the transaction behavior for constructing the user group transactional dataset.

$$ET = \frac{\text{Instruction count (P)} * \text{CPI}}{\text{Clock rate}} \quad (3)$$

where CPI represents the Cycle Per Instruction and P denotes the transactional dataset.

Energy Usage: It measures the energy requirement for the user query and denoted as EU. The mathematical formula to evaluate energy usage for processing the query from transactional dataset P is given as follows:

$$EU = W_{cpu} * N_{tuples}\,[P] \quad (4)$$

where W_{cpu} represents the energy utilization weight to process one tuple in CPU and N_{tuples} denote the number of fetched tuples from transactional dataset P.

Experimental Setup: The experiment was run on an Intel Core-2 Duo with 6 GB

memory and implemented in Java. The size of the database consists of large number of mobile transactions. For experimental purposes, the 6-7 user groups are taken for evaluation with varying size of database. The accuracy, execution time and energy usage for MCTMSPHS is analyzed and the results are discussed below with tabulation and graph represented respectively.

6 Results and Discussion

The mobile transaction behavior is monitored and clustered using heuristic search. User's different behavior such as the user request service (eg., (ri,si) requesting service si at location ri) are monitored and processed based on the algorithm explained under chapter 4. Experimental evaluation has been conducted to evaluate the performance of the multi-cluster based temporal mobile sequential pattern mining using heuristic search technique.

The accuracy of user classification based on the number of user groups formed is illustrated in table 1 for both MCTMSPHS and CTMSP.

User Groups	Accuracy (%)	
	Existing CTMSP	Proposed MCTMSPHS
5	0.96	0.36
10	0.93	0.32
15	0.89	0.33
20	0.87	0.30
25	0.85	0.28
30	0.84	0.26
35	0.82	0.24

Table 1. User Groups Vs Accuracy

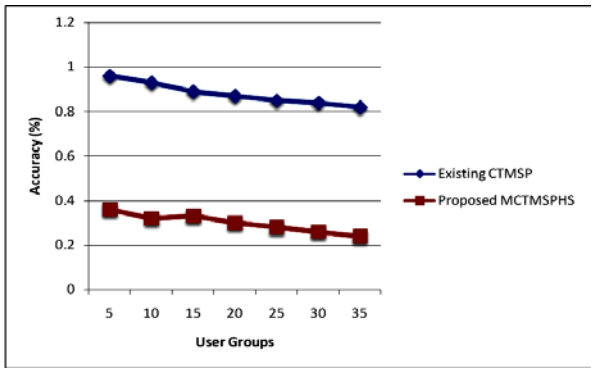

Fig 3. User Groups Vs Accuracy

Fig 3 shows that the accuracy decreases in the existing CTMSP method with increase in the number of predefined user groups. This is because as the number of predefined user groups increases, users are not easily classified in the correct cluster using classification methods. But in the proposed MCTMSPHS, the classification accuracy increases dramatically. The main reason is that MCTMSPHS scheme employed heuristic search which taken into account the time interval at a particular period of time, string and categorical data at the same time, whereas the existing CTMSP method consider only the similarity of time series string. Contrast to existing CTMSP, the proposed MCTMSPHS provides 30 to 40% more classification accuracy.

The execution time needed to analyze the transaction behavior of the user for constructing the user groups in mobile transaction database is illustrated in table 2.

Size of Database (KB)	Execution Time (seconds)	
	Existing CTMSP	Proposed MCTMSPHS
25	8	3
50	14	5
75	20	8
100	32	13
125	45	17
150	50	18
175	54	20

Table 2. Size of Database Vs Execution Time

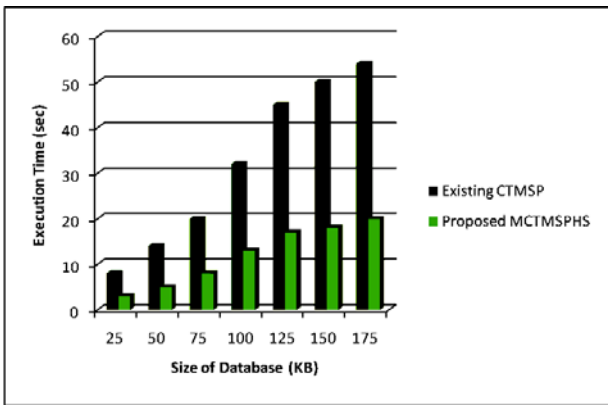

Fig 4. Size of Database Vs Execution Time

Fig 4 shows the measurement of execution time of proposed MCTMSPHS and CTMSP. Execution time is measured by the time taken to retrieve the information from real time experiment consisting of 7 records with varying database size. To check the performance of information retrieval from the database, setup a test which considers the execution time of MCTMSPHS mine on database with payload ranging in size from 25 Kilo bytes to 175 Kilo bytes. The execution time needed for classifying the behavior of the user transaction is measured in terms of seconds. From the fig 4, it can be seen that the time taken to execute using MCTMSPHS is faster in execution when compared to the CTMSP. This is because the expressive sequences generated by the proposed heuristic search mechanism are based on the common similar user groups. This mechanism is capable of symbolizing the features of every cluster better than the expressive sequences produced by the existing CTMSP. Consequently, MCTMSPHS consumes 20 to 25% less execution time than CTMSP in user classification.

The consumption of energy required to process the user's query is depicted in the table 3.

No. of Queries	Energy Usage (J)	
	Existing CTMSP	Proposed MCTMSPHS
5	1.2	0.7
10	1.7	0.9
15	1.9	1.1
20	2.3	1.3
25	2.5	1.5
30	2.8	1.6
35	2.9	1.7

Table 3. No. of Queries Vs Energy Usage

Fig 5 illustrates the energy usage corresponding to number of queries. Energy is measured in terms of Joules. In MCTMSPHS, energy is measured while predicting mobile user's behavior and makes them cluster. In the MCTMSPHS, less energy usage in prediction may lead to accuracy in the clustered transaction behavior. The MCTMSPHS determine more specific mobile sequential pattern using the chronological order and predict the successive mobile behavior efficiently by consuming approximately 40 to 50% lesser energy.

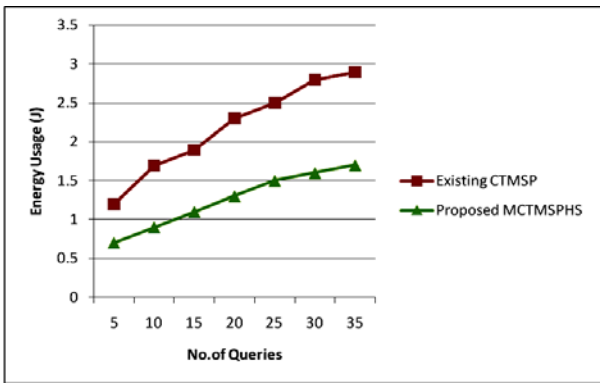

Fig 5. No. of Queries Vs Energy Usage

This section discussed the various methods of temporal mobile pattern based on sequential pattern mining factors and cluster based metrics. The MCTMSPHS and existing CTMSP schemes are examined and their performance is evaluated on three criteria: Execution Time, Energy usage and Accuracy. From the experimental results the MCTMSPHS performs well compared with the CTMSP method.

6 Conclusion

A mobile transaction database is highly significant in varied applications and at the same time very complicated as large amount of transaction logs involved in mobile. It is produced on the basis of the behavior pattern of the user. Even though analogous documents connect with each other, the results obtained do not provide a precise result. One of the solutions to the above mentioned problem is clustering. MCTMSPHS framework is presented for providing heuristic search for combining the new technique for the domain of web documents since there are numerous links or paths between transactions and it is used in pruning of undesired links. Systematic analysis is performed on real system to estimate the performance of MCTMSPHS scheme and the experimental results have

shown that the heuristic search technique efficiently grouped the set of similar behavior of user transactions. Also in real datasets, MCTMSPHS achieves 30 to 40% more accuracy, 20 to 25% minimization in execution time and an improvement of energy level with a variance of 40 to 50%.

References:

[1] Eric Hsueh-Chan Lu, Vincent S. Tseng, Philip S. Yu, "Mining Cluster-Based Temporal Mobile Sequential Patterns in Location-Based Service Environments", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, No. 6, June 2011.

[2] Elhamifar E,Vidal R, "Sparsity in unions of subspaces for classification and clustering of high-dimensional data", *Communication, Control And Computing, Annual Allerton Conference* on 2011.

[3] Sharadh Ramaswamy, Kenneth Rose, "Adaptive Cluster Distance Bounding for High-Dimensional Indexing", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, No. 6, June 2011.

[4] Liang Wang, Christopher Leckie, Kotagiri Ramamohanarao, and James Bezdek, "Automatically Determining the Number of Clusters in Unlabeled Data Sets", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, No. 3, March 2009.

[5] V. C. Belokar, P.S. Kulkarni, "Mobile Sequential Pattern Mining In Location Based Service Environment", *International Journal of Electronics and Computer Science Engineering*.

[6] Yon Dohn Chung, Sanghyun Yoo, and Myoung Ho Kim, "Energy and Latency Efficient Processing of Full-Text Searches on a Wireless Broadcast Stream", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22, No. 2, February 2010.

[7] Mansaf Alam et, "Web Search Result Clustering using Heuristic Search and Latent Semantic Indexing", *International Journal of Computer Applications* (0975 – 8887), Vol. 44, No. 15, April 2012.

[8] Vidal R, "Subspace Clustering", *IEEE Signal Processing Magazine*, Vol. 28, Issue: 2, 2011.

[9] Junming Shao, Plant C, Qinli Yang, Bohm C, "Detection of Arbitrarily Oriented Synchronized Clusters in High-Dimensional Data", *IEEE International Conference on Data Mining*, 2011.

[10] Tolga Onel, Cem Ersoy, Hakan Delic, "Information Content-Based Sensor Selection and Transmission Power Adjustment for Collaborative Target Tracking", *IEEE Transactions on Mobile Computing*, Vol. 8, No. 8, August 2009.

[11] Yi-Hong Chu, Jen-Wei Huang, Kun-Ta Chuang, De-Nian Yang, Ming-Syan Chen, "Density Conscious Subspace Clustering for High-Dimensional Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22, No. 1, January 2010.

[12] Yun Yang and Ke Chen, "Temporal Data Clustering via Weighted Clustering Ensemble with Different Representations", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, No. 2, February 2011.