Improvement in Performance of Mobile User Behavior Prediction Using Mining Techniques

E. Rama Kalaivani[1], E. Ramesh Marivendan[2]

[1] Faculty of CSE,Karpagam College of Engineering,Coimbatore,TamilNadu,India.
[2] PG Scholar,Sri Manakula Vinayagar Engineering College,Pondicherry, India.

ABSTRACT

The use of mobile phones and the Location Based Services Technology is growing rapidly. Location Based Services (LBS) use technology to find a location. , to find people, identify devices such as mobile phones, or services such as ATMs. Location-based services can be query-based and can be push-based both providing useful information to the end users.LBS also helps service providers to predict the next behavior of a mobile user. The main objective of the paper is to increase the performance of mobile behavior prediction using optimization technique. In predicting the mobile behavior spatial and temporal factors are considered .Improving the performance of mobile behavior prediction helps the service provider to improve the quality of service. The system includes formation of cluster and time interval table. Framing cluster table involves identifying the similarities of mobile users and then grouping the users. Number of time segmenting points is identified before framing time interval table. Evolutionary concept is used to identify the best time interval. The results obtained by optimizing the segmentation of time intervals achieve a better formation of time interval table. Using these cluster table and time tine interval table, mining can be done efficiently. Thus the accuracy of the prediction process is improved with the mined results.

Index Terms -Data Mining, Mobile Commerce, Location Based Services

I. INTRODUCTION

The tremendous growth of wireless communication technique has changed people to do business in mobile commerce environment**.** Mobile commerce is the buying and selling of goods and services through wireless handheld devices such as cellular telephone and Personal Digital Assistant (PDA). Mobile Commerce provides leading solutions which optimize search monetization and advertising on mobile devices. The products and services of mobile commerce are mobile banking, mobile browsing, location-based services, mobile ATM, content purchase and delivery, mobile ticketing. . One of the major products and new services involved is location based services. Location Based Services enable the mobile users to identify the information about the area around a given physical or geographical point or place. Location Based Services determine the location of the user. There are two types of Location Based Services active service and passive service. User initiating the service is active i.e, where a mobile user wants information to be sent to them on their phone e.g. a request for details of the nearest cash machines, the railway station or bus station, or for a map or directions to a particular address .Located user is the subject of a location based service initiated by someone else is ‘passive’ service. A passive service may help parents know where their children are when they are out and about and have their mobile phone with them. One of the most advantages in using Location Based Services is that mobile users don't have to manually specify ZIP codes or other location. The components of location based services are shown in the Figure 1.

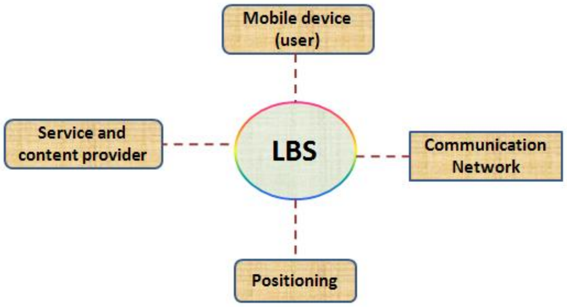


Figure 1: Components of Location Based Services

Mobile device is a tool to request the information which is needed by the user. Communication network is the mobile network that helps in transferring the user data and service request from the mobile terminal to the service provider and then the requested information back to the mobile user. The positioning component helps to identify the physical location of the user in terms of latitude and longitude. The user positions are identified either by using the mobile communication network or by using the Global Positioning System (GPS).

Service and content Provider offers a number of different services to the user. The services includes the calculation of the position, finding a route, searching yellow pages with respect to position or searching specific information on objects of user interest . A geographic base data and location information data will be usually requested from the maintaining authority or business and industry partners.

The mobile user request services through mobile devices and identify the required information using LBS. The major work deals in predicting the service request of a mobile user in Location Based Services environment. The system should identify the next behavior of the mobile user using various techniques like clustering of mobile users, segmentation of time intervals, discovery of CTMPS mining, setting up various prediction strategies .The main purpose of the work is to provide mobile users a precise and efficient mobile behavior prediction system.

II. RELATED WORK

Temporal Mobile Sequential Patterns (TMSPs) proposed by Tseng and Tsui[3] considers the factors of moving paths and time intervals. There are two phases involved in it namely data mining phase and prediction phase TMSP-Mine is used to discover the patterns in each time interval. In prediction phase, the most suitable pattern is based on the historical transaction log, moving path, and current time interval . These prediction strategies make the location based services to provide the user’s request and query related service in advance. In time slot calculation genetic algorithm is used to identify the best time interval. The fitness function used in genetic algorithm obtained the results efficiently. The Mobile Sequential Pattern (MSP) was proposed by Yun and Chen [2] to take moving paths into consideration to better reflect the customer usage patterns in the mobile commerce environment. Three algorithms namely algorithm TJLS(Transaction set join with large Transaction set), algorithm TJPT(Transaction set Join With Path Trimming), and algorithm TJPF (Transaction set Join With Pattern Family) was used for determining the frequent sequential patterns, from the mobile transaction sequences. Algorithm TJLS used two level hash based tree in mining large sequential patterns. Algorithm TJPT considers path traversal patterns and improves the performance by path trimming technique. Algorithm TJPF generates the large sequential patterns very efficiently by utilizing the pattern family technique which is developed to exploit the relationship between moving and purchase behaviours. The results identified that by taking both moving patterns and purchase patterns into consideration, a better model for mobile commerce system is developed and able to exploit the intrinsic relationship between two important customer behaviors for the efficient mining of mobile sequential patterns.

Tseng and Lin proposed SMAP-(Sequential Mobile Access Patterns) mine [4] for efficiently mining users sequential mobile access patterns dealing with requested services for mobile users in mobile web systems. Behavior of mobile users became more complex than earlier, so it is necessary to improve the quality of services efficiently. SMAP-mine uses a special data structure named SMAP-tree. SMAP – mine involves in constructing SMAP-tree and extracting sequential access patterns. SMAP tree can be used with one physical database scan for mining the large amount of data. A set of mobility logs consisting of movements and service requests is maintained.SR-tree is constructed by inserting the mobile sequence from root node. When traversing the sequence the label count is increased.SR-node is generated with various links and pointers. In SMAP-mine algorithm, a threshold value is set, the count with the largest one is chosen. To obtain large sequential mobile access patterns various constraints are used to store the patterns. In internet service providers instead of using SMAP-tree, a continuous mobile access patterns are used.

III. SYSTEM FRAMEWORK

The system framework involves

- LBS Alignment process for similarity computation
- Clustering of mobile sequences
- Segmentation of time intervals
- Final mined result dataset and then the prediction is done using these result dataset.

The parameters used for measuring prediction accuracy are precision and recall.

Precision=Number of correct predictions / (Number of correct predictions + Number of incorrect predictions).

Recall= (Number of correct predictions + Number of incorrect predictions) / Total number of service requests

The input to the system consists of user’s time, location, service requests. The final system produces a mined datasets which can be used for predicting the behaviour of other mobile sequences.

Prediction flow is explained in the system framework figure 2.The input to the LBS Alignment algorithm is two mobile sequences s and s’. For example.

S= { (5 A S1),(6 B pi),(7 C S3),(10 D S2),(11 E pi),(13 F {S3 S4}),(18 L pi),(20 K S5),(21 A S1),(25 E pi),(28 F pi),(30 K S2)}

S’= {(5 A S1), (10 B pi), (19 C pi), (20 D S2),(22 B pi),(27 A S1),(28 E pi),(29 F pi),(30 K S2)}.The output is the similarity score between S and S’

Clustering is done using Cluster Affinity Search Technique Co-Smart CAST algorithm. The similarity score is the input to the algorithm. The output is partition of clusters in different user groups .The number of time segmenting points are identified with the help of GETNTSP algorithm using genetic algorithm. The input is mobile transaction database and the time length. The output is number of time segmenting points. Finally the cluster table and time interval table are formed.

The cluster table and time interval table are the input for mining Cluster based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine). The prediction strategies involved are selecting the pattern that matches the recent mobile behaviour, patterns belonging to the same clusters, patterns belonging to same time interval. The prediction process is verified using mined data set and test data sets. The accuracy of the prediction process is performed using precision and recall measures.

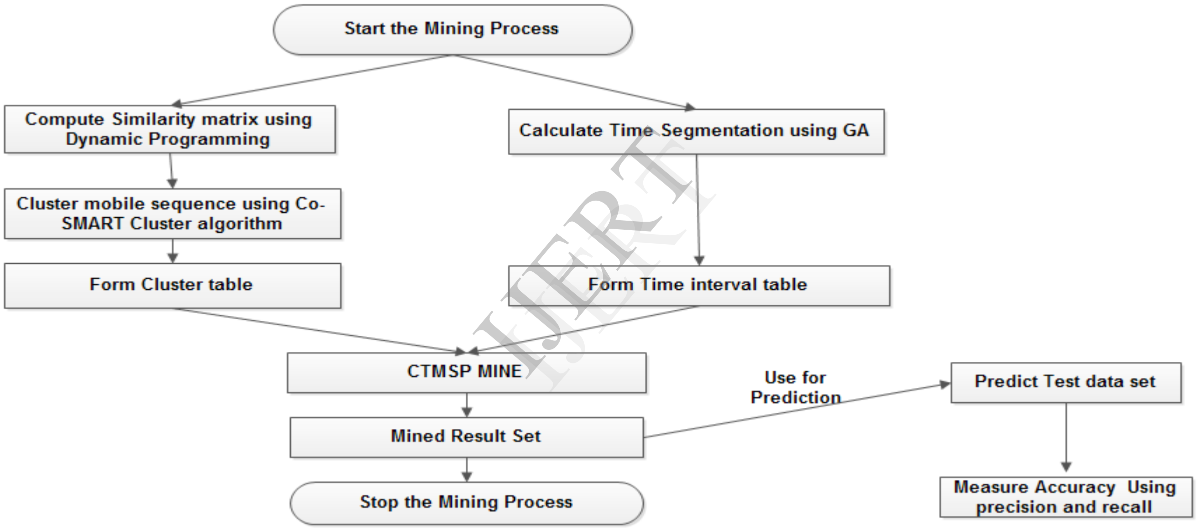

Start the Mining Process
Compute Similarity matrix using Dynamic Programming
Calculate Time Segmentation using GA
Cluster mobile sequence using Co-SMART Cluster algorithm
Form Cluster table
Form Time interval table
CTMSP MINE
Use for Prediction
Predict Test data set
Mined Result Set
Measure Accuracy Using precision and recall
Stop the Mining Process

Figure 2: System framework

IV. SIMILARITY SCORE COMPUTATION ALGORITHM

LBS Alignment algorithm is used to compute the similarity score.The dataset consists of user’s time, location and service[1]. Then the individual time, location and service are allocated in a variable. The individual sequences in the dataset are compared with all other sequences .The algorithm is implemented with the help of dynamic programming concept. In dynamic programming concept the matrix is framed with the maximum number of rows and columns involved in the dataset. Implementation of LBS algorithm for two sample sequences is shown in Figure 3.The following steps are used to compute the similarity score.

Step1: Initialize the parameters. Location penalty P=0.5/|s.length+s’.length| (length of each sequences).M[0][0]=0.5,$M_{i,0}$=$M_{i-1,0}$-P,$M_{0,j}$ = $M_{0,j-1}$ -P,foralli={1,2,3..........s.length},j={1,2,3......s’.length}

Step2: If the location involved in the sequence is same perform step3, 4,5else 6

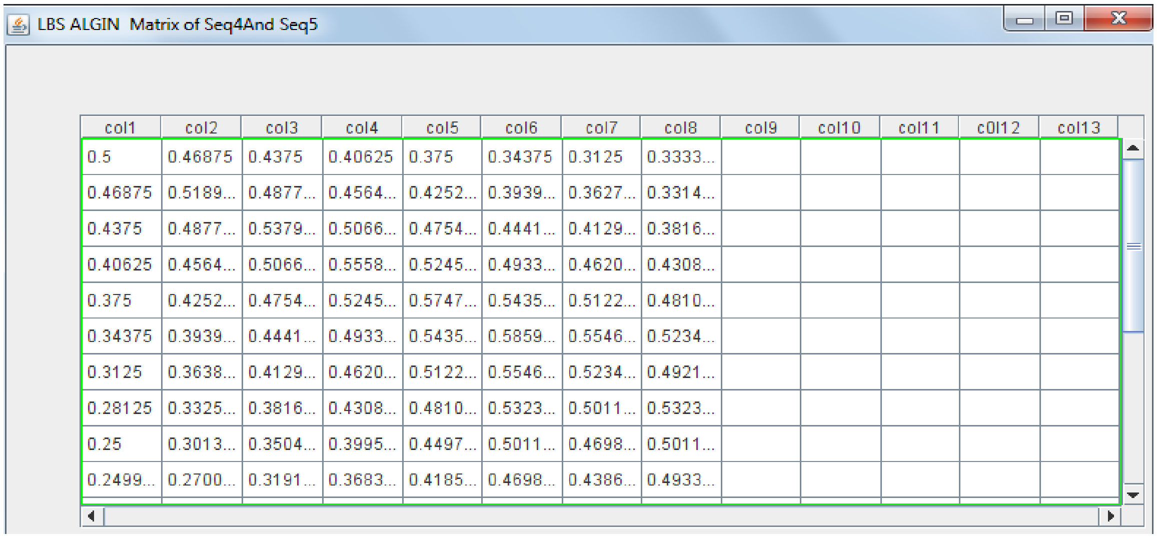
Step3: Calculate Time penalty.TP=0.5/|s.time-s;.time| for all the sequences.

Step4: Calculate Service reward SR= P*S_i Service intersection S_j.Service / S_i .Service Union S_j.Service

Step5: Compute similarity score M_{ij} = Max ($M_{(i-1),(j-1)}$ – TP+SR,$M_{(i-1),j}$ –P,$M_{i,(j-1)}$-P)

Step6: Compute the similarity M_{ij} = Max ($M_{(i-1),j}$ – P,$M_{i,(j-1)}$-P).

Step7: Store the N X N position value of Mij in the final similarity matrix in i X j and j X i position.



LBS ALGIN Matrix of Seq4And Seq5

col1	col2	col3	col4	col5	col6	col7	col8	col9	col10	col11	c0l12	col13
0.5	0.46875	0.4375	0.40625	0.375	0.34375	0.3125	0.3333...					
0.46875	0.5189...	0.4877...	0.4564...	0.4252...	0.3939...	0.3627...	0.3314...					
0.4375	0.4877...	0.5379...	0.5066...	0.4754...	0.4441...	0.4129...	0.3816...					
0.40625	0.4564...	0.5066...	0.5558...	0.5245...	0.4933...	0.4620...	0.4308...					
0.375	0.4252...	0.4754...	0.5245...	0.5747...	0.5435...	0.5122...	0.4810...					
0.34375	0.3939...	0.4441...	0.4933...	0.5435...	0.5859...	0.5546...	0.5234...					
0.3125	0.3638...	0.4129...	0.4620...	0.5122...	0.5546...	0.5234...	0.4921...					
0.28125	0.3325...	0.3816...	0.4308...	0.4810...	0.5323...	0.5011...	0.5323...					
0.25	0.3013...	0.3504...	0.3995...	0.4497...	0.5011...	0.4698...	0.5011...					
0.2499...	0.2700...	0.3191...	0.3683...	0.4185...	0.4698...	0.4386...	0.4933...					

Figure 3. LBS -Algorithm between two Sequences

V. CLUSTREING OF MOBILE USERS

With the similarity matrix obtained in LBS Alignment algorithm, the mobile users are clustered.CO-SMART CAST algorithm is used to cluster the mobile user. Mobile transaction database consists of mobile user's time, location and services [1]. The objective of CO-SMART CAST algorithm is to cluster the mobile users according to the similarity matrix obtained. Clustering result with the highest quality is computed. Following are the steps used to obtain the user clusters.

Introduce new parameters T_{co} (to return the best quality of clusters), CR_i (to identify the users in C_{open} and C_{closed}).S and S' are used to represent the similarity matrix.S-original object similarity matrix and S last cluster similarity matrix (Initially both are same).

Compute the following.

1)CR_i← CAST (Clustering result,S',P_i) CAST(to find the best clustering result with highest T_{co})

2)T_{co}←2 X T_{clu} X T_{obj} / (T_{clu} + T_{obj}) .

T_{obj} ← Hubert's Γ statistic (CR_i,S') ,T_{clu}← Hubert's Γ statistic (CR_i,S).F1 score is used to compute the harmonic mean between T_{obj} and T_{clu} as T_{co}. CAST requires N X N similarity matrix and affinity threshold t as input values. R value is initialized for setting t from 0 to 1.R is equally divided in to five points P_0,P_1,P_2…..P_5,where $P_0<P_1<P_2<P_3<P_4$.The value of P_i is considered as affinity threshold which is the node membership to the cluster. The cluster under construction is C_{open}. Affinity of a user is defined as sum of similarity value between a user and all users in C_{open}.Initially when a new cluster C_{open} is started the initial affinity of all users are zero. Since C_{open} is empty. While computing the clustering result each time higher affinity values of a user are included C_{open} and lower are moved to C_{closed}. A user is said to have higher affinity if its affinity value is greater than or equal to t X $|C_{open}|$.Clusters are formed repeatedly by adding or removing users from current cluster until such time that changes no longer occur or a maximum of iterations have been executed. Hubert's Γ statistics which represents the point serial correlation is computed for measuring the quality of produced clustering.T_{obj} ← Hubert's Γ statistic(CR_i,S') ,T_{clu}← Hubert's Γ statistic (CR_i,S).When the execution of clustering is computed ,the clustering at point P_b is considered the highest T_{co} value(better clustering quality).The

testing range R is limited with in a new range (Pb-1,Pb+1)b←$\arg_{j\varepsilon(0,1\ldots.4)}$max($Tco_j$).The process is repeated until R value is smaller than $10^{-5.}$Both Pb value and Tco values are compared .Higher value is recorded and entries of similarity matrix S' is updated to the average similarities between all pairs of corresponding cluster results. Finally the clustering result with highest quality is obtained. In implementing CO-SMART CAST algorithm it was identified that users 3, 5 and 6 belong to one group and users 1, 2, 4 and 7 belong to another group.

VI. SEGMENTATION OF MOBILE TRANSACTIONDATABASE

In a mobile transaction database, similar mobile behavior exists under certain time segments. It is necessary to segment the time intervals.GETNTSP algorithm is used to identify the segmenting points. Input to the algorithm is mobile transaction data and time length .Output is number of time segmenting points. Genetic Algorithm is used to obtain the most suitable time segments. The following are the steps used in GETNTSP algorithm

1) Identify the total occurrences at each time point.
2) Compute the accumulative count $C_{l,s}$ [t]← $C_{l,s}$ [t] +1
3) Draw the accumulative distribution as shown [1]
4) The change rate is defined as (C[i+1]-C[i] / (1+C[i])),where C[i] represents total number of occurrences for the item at time point i
5) The occurrences of all the time points are counted and find out the satisfied time points whose counts are larger than or equal to the average of all the occurrences from these ones and then take these satisfied ones as a time point sequence
6) Compute the average time distance (a) between two neighboring points
7) Identify the number of neighboring time point pairs in which the time distance is higher than a, the results represents the time segmentation count. In implementing the above steps number of time segmenting points identified was found to be one.

VII. GENETIC ALGORITHM

A genetic algorithm is a class of adaptive stochastic optimization algorithms involving search and optimization. In a genetic algorithm genes represent individual components of a solution Individual genes are not modified as the organisms evolve. It is the chromosomes that evolve by changing the order and makeup of their genes. Genetic algorithm begins by creating an initial population. Fix up the maximum number of iterations based on the application.

This population consists of chromosomes that are given a random collection of genes. Initial population consists of Time (used in mobile transaction database) at which various users access the service. Hence time of various users generated as individuals in a solution space. Evaluate the fitness or "suitability" of each chromosomes(time) that makes up the population In general, a fitness function F(i)is first derived from the objective function and used in successive genetic operations. Fitness in biological sense is a quality value which is a measure of the reproductive efficiency of chromosomes. Fitness function used is

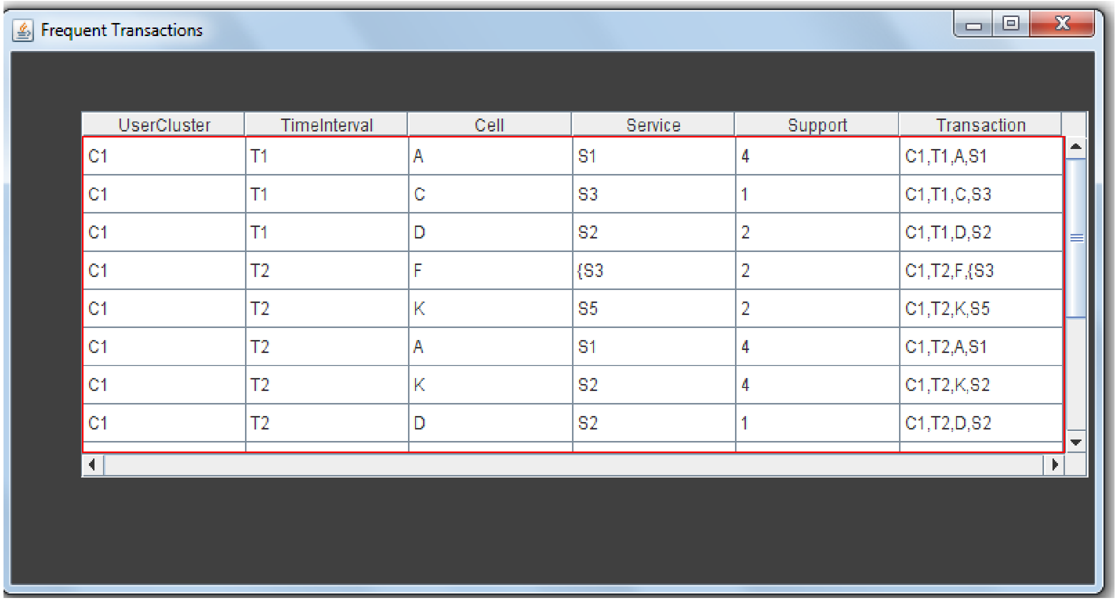
$$\sum_{i=1}^{Len(x)+1} \cdot \sqrt{\frac{1}{NC\cdot Ns}\left(\sum_{c=1}^{NC} \cdot \sum_{s=1}^{Ns} (Ti[c,s] - Ti)\hat{\ }2)\right)}$$

Individuals with higher fitness value will have higher probability of being selected as candidates for further examination. The population is then operated by three main operators; reproduction, crossover and mutation to create a new population of points. GAs can be viewed as trying to maximize the fitness function, by evaluating several solution vectors. . Based on this fitness, select the chromosomes that will mate or those that have the "privilege" to mate. Reproduction (or selection) is an operator that makes more copies of better strings in a new population. Reproduction is usually the first operator applied on a population. Reproduction selects good strings in a population and forms a mating pool. Roulette wheel selection is applied. Next step is to Cross over or mate the selected chromosomes and produce offspring.

Crossover is simply the chance that two chromosomes will swap their bits. In crossover operation, recombination process creates different individuals in the successive generations by combining material from two individuals of the previous generation. It is important to note that no new strings are formed in the reproduction phase. If the length of the chromosome is one cross over cannot be applied. Length of the chromosome is equal to number of time segmenting points..Mutation adds new information in a random way to the genetic search process and ultimately helps to avoid getting trapped at local optima. One bit mutation is applied Repeat above steps until a maximum iteration is reached. The algorithm ends when the best solution has not changed for a present number of generations. The best time interval identified using GA was {1-19} and {19 -32}

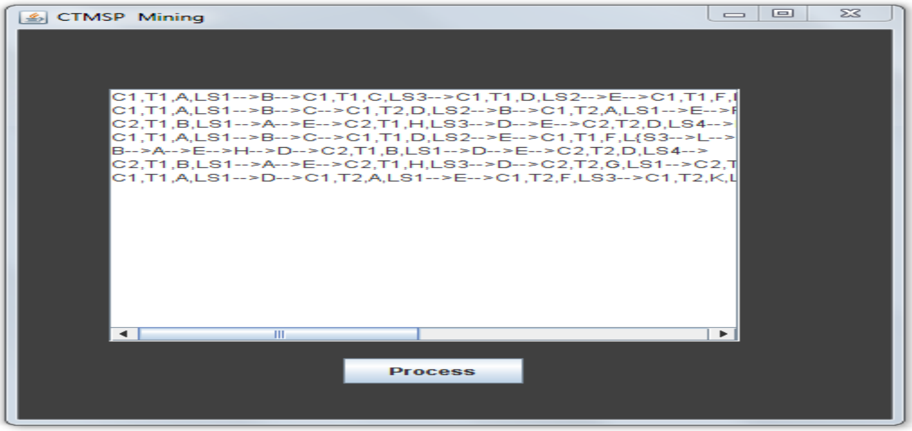
VIII. CLUSTER BASED TEMPORAL MOBILE SEQUENTIAL PATTERN MINING

The procedure of Cluster Based Temporal Mobile Sequential Pattern mining involves frequent transaction mining, Mobile Transformation Database and CTMSP mining. The cluster table contains two groups namely C1={3,5,6} and C2={1,2,4,7}.Time interval table consists of T1= {1-19} and T2={19 -32}.Frequent transactions in each user cluster and time interval are mined using apriori algorithm[7] is used. Frequent 1- transactions are obtained by setting the minimal support threshold to two. Figure 4. Shows Frequent Transactions. A candidate 2-transaction is generated by joining two frequent 1-transactions if their user cluster, time intervals and cells are same. In mobile transaction database transformation F –transactions are used to transform each mobile transaction sequence S into frequent mobile transaction sequence S'. If a transaction T in S is frequent,T would be transformed into the corresponding F- Transaction. CTMSPs are mined from frequent mobile transaction database. In CTMSP mining phase Cluster Based Temporal Mobile sequential Pattern Tree [1] is used. Finally the cluster based temporal mobile sequential pattern mined results are obtained. Figure 5. Shows the CTMSP mined sets. These sets are used for prediction with the test datasets. The prediction process is done as follows. The patterns are selected only from the corresponding cluster a user belongs to. The patterns are selected only from the time interval corresponding to current time. The patterns are selected only from the ones that match the user's recent mobile behaviors. Finally the prediction process is evaluated and accuracy is measured using precision value. Precision=Number of correct predictions / (Number of correct predictions + Number of incorrect predictions).



UserCluster	TimeInterval	Cell	Service	Support	Transaction
C1	T1	A	S1	4	C1,T1,A,S1
C1	T1	C	S3	1	C1,T1,C,S3
C1	T1	D	S2	2	C1,T1,D,S2
C1	T2	F	{S3	2	C1,T2,F,{S3
C1	T2	K	S5	2	C1,T2,K,S5
C1	T2	A	S1	4	C1,T2,A,S1
C1	T2	K	S2	4	C1,T2,K,S2
C1	T2	D	S2	1	C1,T2,D,S2

Figure 4. Frequent Transactions



```
C1,T1,A,LS1-->B-->C1,T1,C,LS3-->C1,T1,D,LS2-->E-->C1,T1,F,
C1,T1,A,LS1-->B-->C-->C1,T2,D,LS2-->B-->C1,T2,A,LS1-->E-->
C2,T1,B,LS1-->A-->E-->C2,T1,H,LS3-->D-->E-->C2,T2,D,LS4-->
C1,T1,A,LS1-->B-->C-->C1,T1,D,LS2-->E-->C1,T1,F,L{S3-->L-->
B-->A-->E-->H-->D-->C2,T1,B,LS1-->D-->E-->C2,T2,D,LS4-->
C2,T1,B,LS1-->A-->E-->C2,T1,H,LS3-->D-->C2,T2,G,LS1-->C2,T
C1,T1,A,LS1-->D-->C1,T2,A,LS1-->E-->C1,T2,F,LS3-->C1,T2,K,
```

Figure 5. CTMSP Mined sets

IX. CONCLUSION

The similarity value obtained was in the range of 0 to 1 using LBS algorithm. The computed similarity score was the input to the clustering process. The clustering process and time interval segmentation process forms the cluster and time interval table. Genetic Algorithm identified the best time interval. Thus the improvement in accuracy of predicting the next mobile behavior is identified using CTMSP mining. Quality of service was improved to mobile service providers.

REFERENCES

[1] Eric Hsueh-Chan Lu, Vincent S. Tseng and 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] C.H. Yun and M.S. Chen, "Mining Mobile Sequential Patterns in a Mobile Commerce Environment," IEEE Trans. Systems, Man, and Cybernetics, Part C, vol. 37, no. 2, pp. 278-295, Mar. 2007.

[3] V.S. Tseng, H.C. Lu, and C.H. Huang, "Mining Temporal Mobile Sequential Patterns in Location Based Service Environments," Proc. 13th IEEE Int'l Conf. Parallel and Distributed Systems, pp.18, Dec.2007.

[4] V.S. Tseng and W.C. Lin, "Mining Sequential Mobile Access Patterns Efficiently in Mobile Web Systems," Proc. 19th Int'l Conf.Advanced Information Networking and Applications, pp. 867-871, Mar. 2005.

[5] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman,"Basic Local Alignment Search Tool," J. Molecular Biology, vol. 215,no. 3, pp. 403-410, Oct. 1990.

[6] H. Jeung, Q. Liu, H.T. Shen, and X. Zhou, "A Hybrid PredictionModel for Moving Objects," Proc. 24th Int'l Conf. Data Eng., pp. 70-79, Apr. 2008.

[7] J. Han and M. Kamber, Data Mining: Concepts and Techniques,second ed., Morgan Kaufmann, Sept. 2000.

[8] S.C. Lee, J. Paik, J. Ok, I. Song, and U.M. Kim, "Efficient Mining of User Behaviors by Temporal Mobile Access Patterns," Int'lJ. Computer Science Security, vol. 7, no. 2, pp. 285-291, Feb. 2007.

[9] J. Pei, J. Han, B. Mortazavi-Asl, and H. Zhu, "Mining Access Patterns Efficiently from Web Logs," Proc. Fourth Pacific Asia Conf.Knowledge Discovery and Data Mining, pp. 396-407, Apr. 2000.

[10] J.-S. Park, M.-S. Chan, and P.S. Yu, "An Effective Hash BasedAlgorithm for Mining Association Rules," Proc. ACM SIGMODConf. Management of Data, pp. 175-186, May 1995.