Innovative Science and Technology Publications

International Journal of Future Innovative Science and Technology
ISSN: 2454- 194X Volume - 2, Issue - 2

Manuscript Title

ENHANCED TRANSACTION DATA SET FILTERING MODEL USING R-TREE WITH CTMSP ALGORITHM

S.Nanthini[1], R. Sivaraj[2] and R. Devi Priya[3]

[1] PG-Student, Department of Computer science and Engineering
Velalar College of Engineering and Technology
snanthini.s7@gmail.com

[2]Assoc.Professor, Department of Computer science and Engineering
Velalar College of Engineering and Technology
rsivarajcse@gmail.com

[3]Assistant Professor, Department of Information Technology,
Kongu Engineering College
scrpriya@gmail.com

May – 2016
www.istpublications.com

International Journal of Future Innovative Science and Technology, ISSN: 2454- 194X
Volume-2, Issue-2, May - 2016 editor@istpublications.com

ENHANCED TRANSACTION DATA SET FILTERING MODEL USING R-TREE WITH CTMSP ALGORITHM

S.Nanthini[1], R. Sivaraj[2] and R. Devi Priya[3]

[1] PG-Student, Department of Computer science and Engineering
Velalar College of Engineering and Technology
snanthini.s7@gmail.com

[2]Assoc.Professor, Department of Computer science and Engineering
Velalar College of Engineering and Technology
rsivarajcse@gmail.com

[3]Assistant Professor, Department of Information Technology,
Kongu Engineering College
scrpriya@gmail.com

ABSTRACT

Researches on Location-Based Service (LBS) have been emerging in recent years due to a wide range of potential applications. One of the active topics is the mining and prediction of mobile movements and associated transactions. Most of existing studies focus on discovering mobile patterns from the whole logs. However, this kind of patterns may not be precise enough for predictions since the differentiated mobile behaviors among users and temporal periods are not considered. we propose a novel algorithm, namely, Cluster-based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine), to discover the Cluster-based Temporal Mobile Sequential Patterns (CTMSPs). Moreover, a prediction strategy is proposed to predict the subsequent mobile behaviors. In CTMSP-Mine, user clusters are constructed by a novel algorithm named Cluster-Object-based Smart Cluster Affinity Search Technique (CO-Smart-CAST) and similarities between users are evaluated by the proposed measure, Location-Based Service Alignment (LBS-Alignment). Mean while, a time segmentation approach is presented to find segmenting time intervals where similar mobile characteristics exist. To our best knowledge, this is the first work on mining and prediction of mobile behaviors with considerations of user relations and temporal property simultaneously. Through experimental evaluation under various simulated conditions, the proposed methods are shown to deliver excellent performance.

Keywords*: Location-based services, frequent items, Similarity Measures, Cluster-based Temporal Mobile Sequential Pattern Mine*

I.INTRODUCTION

Data mining, or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining tools predict behaviors and future trends, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations (Fig 1). Data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find where the value resides.

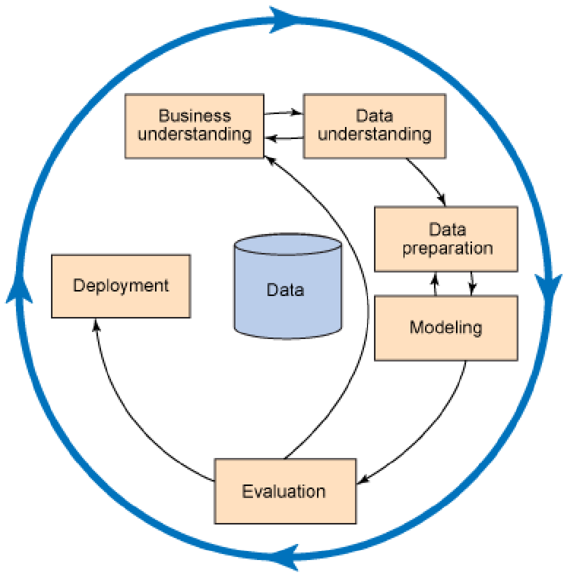

Fig.1 Process of Data Mining

The technique that is used to perform these feats is called modeling. Modeling is simply the act of building a model (a set of examples or a mathematical relationship) based on data from situations where the answer is known and then applying the model to other situations where the answers aren't known. Modeling techniques have been around for centuries, of course, but it is only recently that data storage and communication capabilities required to collect and store huge amounts of data, and the computational power to automate modeling techniques to work directly on the data, have been available.

As a simple example of building a model, consider the director of marketing for a telecommunications company. He would like to focus his marketing and sales efforts on segments of the population most likely to become big users of long distance services. He knows a lot about his customers, but it is impossible to discern the common characteristics of his best customers because there are so many variables. From his existing database of customers, which contains information such as age, sex, credit history, income, zip code, occupation, etc., he can use data mining tools, such as neural networks, to identify the characteristics of those customers who make lots of long distance calls. For instance, he might learn that his best customers are unmarried females between the age of 34 and 42 who make in excess of $60,000 per year. This, then, is his model for high value customers, and he would budget his marketing efforts to accordingly.

II.CLUSTERING OF MOBILE TRANSACTION SEQUENCES

In a mobile transaction database, users in the different user groups may have different mobile transaction behaviors. The first task to tackle is to cluster mobile transaction sequences. In this module, a parameter-less clustering algorithm called CAST is proposed. Before performing the CAST, a similarity matrix S is to be generated, based on the mobile transaction database. The entry Sij in matrix S represents the similarity of the mobile transaction sequences i and j in the database, with the degrees in the range of [0, 1].

A mobile transaction sequence can be viewed as a sequence string, where each element in the string indicates a mobile transaction. The major challenge to tackle is to measure the content similarity between mobile transactions. The LBS-Alignment algorithm is proposed, which can obtain the similarity. LBS-Alignment is based on the consideration that two mobile transaction sequences are more similar, when the orders and timestamps of their mobile transactions are more similar. CAST algorithm is used to cluster the users.

1) Similarity Inference Model (SIM) for measuring the similarities among stores and items, which are two basic mobile commerce entities considered in this paper;

2) Personal Mobile Commerce Pattern Mine (PMCP-Mine) algorithm for efficient discovery of mobile users‟ Personal Mobile Commerce Patterns (PMCPs); and

3) Mobile Commerce Behavior Predictor (MCBP) for prediction of possible mobile user behaviors.

III.RELATED WORKS

Zhou.L Location-based services (LBS) have become more and more ubiquitous recently. Existing methods focus on finding relevant points-of-interest (POIs) based on users' locations and query keywords. Nowadays, modern LBS applications generate a new kind of spatio-textual data, regions-of-interest (ROIs), containing region-based spatial information and textual description, e.g., mobile user profiles with active regions and interest tags. To satisfy search requirements on ROIs, we study a new research problem, called spatio-textual similarity search: Given a set of ROIs and a query ROI, we find the similar ROIs by considering spatial overlap and textual similarity.

Spatio-textual similarity search has many important applications, e.g., social marketing in location-aware social networks. It calls for an efficient search method to support large scales of spatio-textual data in LBS systems. To this end, we introduce a filter-and-verification framework to compute the answers. In the filter step, we generate signatures for the ROIs and the query, and utilize the signatures to generate candidates whose signatures are similar to that of the query. In the verification step, we verify the candidates and identify the final answers. To achieve high performance, we generate effective high-quality signatures, and devise efficient filtering algorithms as well as pruning techniques. Experimental results on real and synthetic datasets show that our method achieves high performance.In this paper, we formalize the problem of spatio-textual similarity search, and study the research challenges that naturally arise in this problem. A challenge is how to evaluate the similarity between two ROIs.

Another challenge is how to achieve high search efficiency as LBS systems are required to support millions of users and respond to queries in milliseconds. Given a query ROI, there may be a huge amount of ROIs having significant overlaps with the query, thus it is rather expensive to find similar answers. Take the real dataset Twitter in our experiments as an example.To provide high performance, we introduce a filter-and-verification framework to compute the answers. In the filter step, our method generates signatures for spatio-textual objects and queries, and utilizes the signatures to generate candidates whose signatures are similar to those of the queries. In the verification step, it verifies the candidates and identifies the final answers. We develop effective techniques to generate signatures and devise efficient filtering algorithms to prune dissimilar objects.

Cong.G Geographic objects associated with descriptive texts are becoming prevalent. This gives prominence to spatial keyword queries that take into account both the locations and textual descriptions of content. Specifically, the relevance of an object to a query is measured by spatial-textual similarity that is based on both spatial proximity and textual similarity. In this paper, we define Reverse Spatial Textual k Nearest Neighbor (RSTkNN) query, i.e., finding objects that take the query object as one of their k most spatial-textual similar objects. Existing works on reverse kNN queries focus solely on spatial locations but ignore text relevance. To answer RSTkNN queries efficiently, we propose a hybrid index tree called IUR-tree (Intersection-Union R-Tree) that effectively combines location proximity with textual similarity. Based on the IUR-tree, we design a branch-and-bound search algorithm. To further accelerate the query processing, we propose an enhanced variant of the IUR-tree called clustered IUR-tree and two corresponding optimization algorithms.

To process the RSTkNN queries efficiently, we propose a hybrid indexing structures and an efficient approach that take into account the fusion of location proximity and document similarity. The contributions and the organization of this paper are summarized as follows. 1. We propose and analyze the problem of Reverse Spatial and Textual k Nearest Neighbor (RSTkNN) search. 2. We propose an efficient algorithm to process RSTkNN queries. The algorithm is based on an effective hybrid indexing structure called Intersection-Union-R tree (IUR-tree) that stores spatial and textual information for the database objects (Section 4). By effectively computing spatial-textual similarities between index nodes, we exploit a branch-and-bound algorithm to prune the irrelevant subtrees. We also theoretically analyze the performance of the algorithm based on IUR-tree. 3. IUR-trees organize the data points by considering only spatial distance, which may damp the pruning power due to the diversity of textual contents in one node. we propose an enhanced hybrid index, called Clustered IUR-tree (CIUR-tree) incorporating textual clusters and two optimization algorithms based on CIUR-tree.This paper addressed the new problem RSTkNN query, which is the extension of RkNN query with the fusion of spatial information and textual description, making it much richer and more complex for the construction and traverse of the index.

IV.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.

V.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.

VI.CTMSP MINING ALGORITHM:

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.

VII.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 . For every item, we gather the total number of occurrence for each entry in each time point . 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.

VIII.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.

IX.CONCLUSION

In this thesis, a novel method named CTMSP-Mine is proposed, for discovering CTMSPs in LBS environments. Furthermore, novel prediction strategies are proposed to predict the subsequent user mobile behaviors using the discovered CTMSPs. In CTMSP-Mine, first a transaction clustering algorithm is proposed named CO-Smart-CAST to form user clusters based on the mobile transactions using the proposed LBS-Alignment similarity measurement. Then, the time segmentation algorithm is utilized to generate the most suitable time intervals. To our best of mobile behaviors associated with user clusters and temporal relations. A series of experiments were conducted for evaluating the performance of the proposed methods. The experimental results show that CO-Smart-CAST method achieves high-quality clustering results and the proposed CBSS strategy obtains highly precise results for user classification. Meanwhile, the algorithms obtain the most proper and correct time intervals. For behavior prediction, CTMSP is shown to outperform other prediction methods in terms of precision and F-measure.

REFERENCES

[1] Cao.X et al.,(2010) „Retrieving top-k prestige-based relevant spatial web objects‟ Proc. VLDB Endowment, vol. 3, No. 1, pp. 373–384.
[2] Cong.G,(2011) „Reverse spatial and textual k nearest neighbor search,‟ in Proc. ACM SIGMOD Int. Conf. Manage. Data, pp. 349–360.
[3] Hadjieleftheriou.M, (2010) ‟Approximate string search in spatial databases,‟ in Proc. Int. Conf. Data Eng., pp. 545–556.
[4] Mrs. K.Maheswari et al.,(2015) "Traffic Incident Detection Using Weighted Durable Quries" International Journal On Engineering Technology and Sciences, Volume 2 Issue 4,pp55-59.
[5] Miss. S.Nanthini et al.,(2016) "An Effective Filtering Dynamic Transaction Datasets Using Serieswith Location Base Service Model" Vol. 2, Special Issue 10, March 2016.
[6] Ooi.B.C,(2011) „Collective spatial keyword querying‟, in Proc. ACM SIGMOD Int. Conf. Manage.Data, pp. 373–384.
[7] Ooi.B.C(2010),"Locating mapped resources in web 2.0,‟ in Proc. Int. Conf. Data Eng., pp. 521–532.