ISSN: 2250–3676

[IJESAT] INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE & ADVANCED TECHNOLOGY Volume-2, Issue-6, 1582 – 1586

EXPLOITING PERVASIVE DATA MINING OF DATABASES FROM MOBILE ENVIRONMENT

Ms. S. A. Ghotkar [1], Prof. R. R Shelke [2]

[1,2]H. V. P. M. College of Engineering and Technology, Amravati, India ,
ghotkarsonal@gmail.com

Abstract

Data mining services play an important role in the field of Communication industry. Data mining is also called knowledge discovery in several database including mobile databases. In this paper, the consumptive behavior based on data mining technology will be discussed and analyzed. Mobile search is becoming increasingly important for mobile users as mobile devices are more widely used. Mobile search is quite different from standard PC-based web search in a number of ways: (a) the user interfaces and I/O are limited by screen real estate, (b) key pads are tiny and inconvenient for use, (c) limited bandwidth and (d) costly connection fees. These limitations result in more navigational queries in the mobile search. Furthermore, user location, activities, preferences, and interaction history can also improve accuracy in determining relevance for mobile search. In the past, most personalized search algorithms are studied in the context of PC-based web search. Personalized mobile search should however play a bigger role at improving the user experiences. This paper focuses on the personalization strategies which explicitly and implicitly infer user search context at individual user level. We show that personalized mobile search perform well for ambiguous queries and localized searches. In this paper, a data mining method for generating mobile clients' location-aware service rules are used and we use a data mining algorithm which involves mining for location-based services.

Keywords **:** *Mobile Search, Personalized Search, location aware services.*

---*******---

1. INTRODUCTION

Mobile search is the second most used application only after social networking in wireless internet [1].Search engine like Google appear in top three of the most visited web sites in terms of wireless internet usage. Most mobile search queries are short due to the hardware limitations such as tiny keypads and small screen. Early studies [2] attempted to provide solutions to mitigate the hardware limitations of the wireless devices. The top 100 mobile queries at AT&T [3] reveal that a great number of search queries are navigational [4] in nature. The navigational searches, for example "Google", usually steer mobile users to specific web sites conveniently. Unlike navigational queries, words like "images" and "free" which are informational and transactional are ambiguous to search engine. A housewife and an iPhone user interpret "apple" differently in search context. A housewife is likely to know the apple variety and prices at the local grocery stores. While an iPhone user is interested in service or products related to iPhone. Researchers studied methods and models to determine the query ambiguity. Clarity score [5] was proposed to evaluate the relative entropy between the query language model and the collection language model. A large click entropy indicates that user clicks more web pages to solve the query, thus the query is ambiguous. A small click entropy means mobile users have common understanding for a search query. Song [6] developed classifier to automatically identify three types of queries, ambiguous, broad, or clear query. We believe these methods and algorithms work equally well to identify the query ambiguity in the mobile search. Researchers have explored personalized search to improve topical relevance of result documents in PCbased web search. Shen et al. [7] studied user's immediate and short-term search context to expand the current query. Qiu and Cho [8] learned user interest from the click history and developed ranking mechanism based on the user interest. Chirita et al. [9] proposed personalized search and summarization algorithms which assist search keywords expansion based on extracted information from local desktop. Duo et al. [10] and Teevan et al. [11] investigated the personalized search strategies and stated that personalization improves the search accuracy on ambiguous queries. So far the personalization is studied only for PC based web search. Most of such personalization strategies are limited to the user search history, returned search results, and documents stored in PC.

2. MOBILE DATA MINING

The goal of mobile data mining is to provide advanced techniques for the analysis and monitoring of critical data from mobile devices. Mobile data mining has to face with the typical issues of a distributed data mining environment, with in addition technological constraints such as low bandwidth networks, reduced storage space, limited battery power, slower processors, and small screens to visualize the results [1]. The mobile data mining field may include several application scenarios in which a mobile device can play the role of data producer, data analyzer, client of remote data miners, or a combination of them. More specifically, we can envision three basic scenarios for mobile data mining:

1. The mobile device is used as terminal for ubiquitous access to a remote server that provides some data mining services. In this scenario, the server analyzes data stored in a local or distributed database, and sends the results of the data mining task to the mobile device for its visualization. The system we describe in this chapter is based on this approach.
2. Data generated in a mobile context are gathered through a mobile device and sent in a stream to a remote server to be stored into a local database. Data can be periodically analyzed by using specific data mining algorithms and the results used for making decisions about a given purpose.
3. Mobile devices are used to perform data mining analysis. Due to the limited computing power and storage space of today's mobile devices, currently it is not realistic to perform the whole data mining task on a small device. However, some steps of a data mining task (i.e, data selection and preprocessing) could be run on small devices.

MobiMine [2] is an example of data mining environment designed for intelligent monitoring of stock market from mobile devices. MobiMine is based on a clientserver architecture. The clients, running on mobile devices such as PDAs, monitor a stream of financial data coming through a server. The server collects the stock market data from different Web sources in a database and processes it on a regular basis using several data mining techniques. The clients query the database for the latest information about quotes and other information. A proxy is used for communication among clients and the database. Thus, when a user has to query the database, she/he sends the query to the proxy which connects to the database, retrieves the results and sends them to the client. To efficiently communicate data mining models over wireless links with limited bandwidth, MobiMine uses a Fourierbased approach to represent the decision trees,which saves both memory on mobile device and network bandwidth. Another example of mobile data mining system is proposed in [3]. Such system considers a single logical database that is split into a number of fragments. Each fragment is stored on one or more computers connected by a communication network, either wiredly or wirelessly. Each site is capable of processing user requests that require access to local or remote data. Users can access corporate data from their mobile devices. Depending on the particular requirements of mobile applications, in some cases the user of a mobile device may log on to a corporate database server and work with data there. In other cases the user may download data and work with it on a mobile device or upload data captured at the remote site to the corporate database. The system defines a distributed algorithm for global association rule mining, which does not need to ship all of local data to one site, thereby not causing excessive network communication cost. Another promising application of mobile data mining is the analysis of streams of data generated from mobile devices. Some possible scenarios are patient health monitoring, environment surveillance, and sensor networks. The VEhicle DAta Stream mining (VEDAS) system [4] is an example of mobile environment for monitoring and mining vehicle data streams in real time. The system is designed to monitor vehicles using onboard PDAbased systems connected through wireless networks. VEDAS continuously analyzes the data generated by the sensors located on most modern vehicles, identifies the emerging patterns, and reports them to a remote control center over a lowbandwidthwireless network connection. The overall objective of VEDAS is supporting drivers by characterizing their status, and helping the fleet managers by quickly detecting security threats and vehicle problems.

Location-aware service mining

To understand the behaviours of the clients, mining data log will play an important role. The issue of multilevel association mining was first proposed by Han (Han, & Fu, 1999). It is incorporated the taxonomy to demonstration the data hierarchy relations. Different minimum supports may be assigned for each level. Tseng (Tseng, & Tsui, 2004) exploited the multilevel association rule mining to discover the location-aware service rules. The service and location concept hierarchy trees were constructed to mining frequent patterns. Let HL and HS be two different hierarchy trees and representing the location and the service hierarchical concept respectively. Let P={P1,P2,…PK} be a set of items of pairs. Each pair PI consists of two elements, location_a and service_b, where location_a is a location item from HL and service_b is a service item from HS. In firstly, the location and service hierarchies are represented by using an encoding method. Then, the mining process is performed from root level to the leaf level, and in each level have different minimum support value to generate large itemsets. Afterward, they iterative find out all large itemsets in the combinatory pairs of levels in HL and HS. According the large itemsets generate the location-aware service rules which satisfied the minimum confidence. Those rules represent the strong association between the services and locations.

Periodic sequential mining

In past studies, many researches in association rules assume the items of transactions are unordered. This assumption seems reasonable, but Agrawal (Agrawal, & Srikant, 1995) claimed that both the transaction items and transaction times are the keys to find customer behaviours. Nowadays, there are many problems to deal with the ordered data for solving the practical applications, such as travelling sequence, shopping path design and plan failure predictions. In general, mining sequential data can be classified into three main categories: similar pattern mining, periodic pattern mining, and frequent pattern mining (Chen, Chen, & Hsu, 2002; Masseglia, Poncelet, & Teisseire, 2004). For periodic sequence mining, Han et al. (Han, Dong, & Yin, 1999) proposed a finding periodic patterns methodology which based on association mining from time series database. A pattern s = {s1 s2…sp }as a non-empty sequence which is a set of itemsets ordered according to their time-stamp. A k-pattern is a sequence of k-items (also called length k). Moreover, a sub-pattern of a pattern s is a pattern s' = s'1 s'2 … s'p such that s and s' have the same length, and s'i s⊆i for every position i. In this paper, we focus on periodic pattern mining for finding the cyclic patterns in a time-stamped database.

Multidimensional data

In a multidimensional data model, user requests are modeled as facts, and the values that characterize the user requests are organized into dimensions. For our scenario, we will have three dimensions. The TIME dimension captures the time of the user requests and has categories (levels) such as Second, Minute, Hour, etc. The USER dimension captures aspects of the users issuing the request with categories such as Spoken Language, Personal Interest, Actual Age, Main Occupation, etc. The LOCATION dimension captures the, possibly changing, locations of the users when the requests were issued. Entity types in the Location ER diagram are then represented as categories in the hierarchy of categories that makes up the LOCATION.

Data Discovery Using Location-based Cyclic Sequence (LBCS)

Due to recent advances in computer hardware technology, a vast number of mobile users (we also called it mobile clients) are accessing various information systems by notebooks or personal digital assistant (PDAs) via wireless communication from anywhere at any time. At present, the wireless services do not support the personalization and localization for mobile clients. If wireless internet service provider (WISP) has the ability to explore the user behaviours, and support Location-based Services (LBS), it will increase the client's loyalties and satisfaction. In this paper, we use the data mining technologies to trace out the mobile clients' behaviours and the sequences of service requests. An intelligent broadcast disk is organized to decrease the access latency by scheduling the subsequently requested service items closed to each other. We propose a cyclic sequential mining to discover a set of location-based cyclic sequential patterns (LBCS) which are frequently requested by a number of mobile client over time. In this section, a set of location-based cyclic sequential rules are obtained predicatively prefetch sequential service patterns to clients' caches in the mobile computing environments.

Since some service items have been requested frequently, we could infer that service requests have cyclic occurrence characteristic. Therefore, we exploit the characteristic to propose a location-based cyclic sequential mining method in a specific location. The aim of this mining approach is to discovery the clients' cyclic service patterns for reorganizing the broadcast service items in an intelligent way and reduce the access latency for the clients. For mining location-based cyclic sequences (abbreviated as LBCS), we select data fragments of each client requested records from the service history in a specific location.

LBCS Algorithm

Input

1. Given a fragment Fi of mobile client Ui service request Tj; Each Tj contains sequence ID and a service sequence.
2. min_sup

Output: a set of large LBCS patterns
Method
Step 1. scan Fi and generate a set of L1
Step 2. C2 = apriori-gen(L1)
Step 3. L2 = large_sequences_gen(C2)
Step 4. for (k=2; Lk-1 ≠∅ ; k++)
for (i, j=1; i, j ≦number of Lk; i, j ++)
if (Lki.equiv_class ∩ Lkj.equiv_class)
if (suffix(Lki, k-1) equal to prefix(Lkj, k-1))
Ck+1 = Lki + suffix(Lkj, k-1)
end
Lk = large_sequences_gen(Ck+1)
end
Procedure: large_sequences_gen(Cm)
for (r.sequences ∈Fi)
if (subsequence (Cm, r.sequences))
Cm.equiv_class = Cm.equiv_class ∪{r.ID}

ISSN: 2250–3676
[IJESAT] INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE & ADVANCED TECHNOLOGY Volume-2, Issue-6, 1582 – 1586

Cm.count ++
end
Lm = Lm ∪{ Cm│Cm.count ≧min_sup}
return (Lm, Lm.equiv_class)
LBCS algorithm

In the previous work, there are services which can access by users are not relevant because lots of irrelevant data. The queries from users are not mined so the speed was slow of small devices, waste of memory and low performance screen. So we would implementing data mining techniques on small devices such as mobile phone, PDA so that only required data will be getting due to filtration. User can access web services in short period of time and with good performance. The small size of the screen is one of the main limitations of mobile device applications. In data mining tasks, in particular, a limited screen size can affect the appropriate visualization of complex results representing the discovered model. The mobile phone has becoming an important device for providing information anytime anywhere. However, due to the limitation of its hardware such as small display screen and input capacities, it is not as easy as using a personal computer. Using information services and mobile internet on mobile devices has several difficulties such as information overload, small screen and input capacities limitation. There is a huge amount of information available online today. It is sometimes common that the information accessed by the desktop PC users could also be accessed by the mobile internet users. One possible way to make the accessing of online information easier with the help of data mining techniques. In our system we overcome the limitation by splitting the result in different parts and allowing a user to select which part to visualize at one time. Moreover, users can choose to visualize the mining model either in textual form or as an image. In both cases, if the information does not fit the screen size, the user can scroll it by using the normal navigation facilities of the mobile device. We would implementing data mining techniques on small devices such as mobile phone, PDA so that only required data will be getting due to filtration. User can access web services in short period of time and with good performance. Mobile users retrieve information and services efficiently i.e. speed increases and saves memory.

CONCLUSION

This paper discussed the technique of pervasive data mining of databases from mobile Environment (devices) through the use of Web Services. By implementing mobile Web Services we allow remote users to execute data mining tasks from a mobile phone or a PDA and receive on those devices the results of a data analysis task. In our system we overcome the limitation by splitting the result in different parts and allowing a user to select which part to visualize at one time. Moreover, users can choose to visualize the mining model either in textual form or as an image. In both cases, if the information does not fit the screen size, the user can scroll it by using the normal navigation facilities of the mobile device. We would implementing data mining techniques on small devices such as mobile phone, PDA so that only required data will be getting due to filtration. User can access web services in short period of time and with good performance. Mobile users retrieve information and services efficiently i.e. speed increases and saves memory.

REFERENCES

[1]. S. Pittie, H. Kargupta, B. Park. Dependency detection in MobiMine: a systems perspective. Information Sciences, 155(34): 227243 (2003).
[2]. H. Kargupta, B. Park, S. Pitties, L. Liu, D. Kushraj, K. Sarkar. Mobimine: monitoring the stock marked from a PDA. ACM SIGKDD Explorations, 3(2): 3746 (2002).
[3].F. Wang, N. Helian, Y. Guo, H. Jin. A Distributed and Mobile Data Mining System. Proc. Int. Conf. on Parallel and Distributed Computing, Applications and Technologies (2003).
[4]. H. Kargupta, R. Bhargava, K Liu, M. Powers, P. Blair. S. Bushra, J. Dull. VEDAS: A Mobile and Distributed Data Stream Mining System for RealTime Vehicle Monitoring. Proc. SIAM Data Mining Conference (2003).
[5]. Domenico Talia† and Paolo Trunfio, “Mobile Data Mining on Small Devices Through Web Sevices”(2010).
[6].Pittie, H. Kargupta, B. Park, ”Dependency detection in MobiMine:a systems perspective” Information Sciences, 155(34):227243 (2003).
[7].H.Kargupta,B.Park,S. Pitties, L.Liu, D. Kushraj,K.Sarkar, ”Mobimine: monitoring the stock marked from a PDA” ACM SIGKDD Explorations, 3(2):3746 (2002).
[8].M.Adaçal,A.B.Bener,”MobileWebServices:ANewAgentBasedFramework”IEEE Internet Computing, 10(3): 5865 (2006).
[9].M.Tian,T.Voigt,T.Naumowicz,H.Ritter,J.Schiller, ”Performance Considerations for MobileWeb Services” Computer Communications, 27(11): 10971105(2004).
[10].H.Chu,C.You,C.Teng,”Challenges:WirelessWeb Services” Proc.Int.Conf.Parallel and Distributed Systems (ICPADS 04), IEEE CS Press (2004).
[11].W.Zahreddine,Q.H.Mahmoud, ”An Agentbased Approach to Composite Mobile Web Services”, Proc.Int. Conf.on Advanced Information Networking and Applications (AINA’05), IEEE CS Press (2005).
[12]. Khaled M. Hammouda and Mohamed S . Kamel, “Hierarchically Distributed Peer-to-Peer Document Clustering and Cluster Summarization”, IEEE Transactions on

Knowledge and Data Engineering, Vol. 21(5), pp.681-698 (2009).
[13].Fenggui, Malek, Adjouadi, Naphtali Rishe "Personalized Approach for Mobile Search", IEEE DOI 10.1109/CSIE.2009.718(2008).
[14].Hyun-suk,Hwang,Seon-hyunShin,Ki-ukKim,Seok-CheolLee,Chang-sookim,"A Context-aware System Architecture using Personal Informatiom based on Ontology",5th IEEE Int'(2007).
[15]. Chih-Hao Liu Jason Jen-Yen Chen,Jhong-Li, Taiwan, "Mobile User Agent with User Ontology for Personalized Web Service Access" IEEE (2010).