

A Novel RSRM Algorithm for Mining services for Better Enhancement in Small HandHeld Devices

Ganesh Raj Kushwaha

M.Tech Scholar, PG Dept. of Computer Science &
Engineering
RKDF, Bhopal, India

Prof. Farha Haneef

PG Dept. of Computer Science & Engineering
RKDF, Bhopal, India

ABSTRACT

This paper presents a case study on data mining services able to support decision makers in strategic planning for the enhancement of small handheld devices. The application provides e-Knowledge services for the analysis of territorial dynamics by processing and modeling huge amount of data, in order to discover rules and patterns in a distributed and heterogeneous content environment. For the analysis of structured data, the application covers the whole Knowledge Discovery process. The purpose of the paper is to show how to implement existing techniques in a flexible architecture for providing new added value services. Finally in our paper, we proposed a RSRM(Read Subset Removal Miner) algorithm for mining on mobile devices.

Keywords

J2ME, DMS, CLDC, MIDP

1. INTRODUCTION

Pattern Analysis is becoming more popular by the help of the recent developments in the computer and communication technologies. Mobility of the users today gives rise to the problem of mobility management. Modeling the behavior patterns of users in the mobile systems benefits not only the users in smart access, but also the mobile service providers in various fields. In the mobile environments and developments, the users may request various kinds of services and applications by cellular phone, PDA, or notebook from arbitrary locations at any time via GSM, GPRS or wireless networks. Obviously, the behavior pattern and behavior, in which the location and the service are inherently coexistent, of mobile users becomes more complex than that of the traditional web systems. To help the user get desired information in a short time is one of the promising applications, especially in the mobile environments because in today's scenario everyone wants the desired result in short duration of time Pattern prediction and recognition can be defined as the prediction of a mobile user's next movement where the mobile user is traveling between the cells of a GSM network. The predicted movement can then be used to increase the efficiency of pattern analysis by using the predicted movement; the system can effectively allocate resources to the most probable-to-move cells instead of blindly allocating excessive resources in the cell neighborhood of a Handheld user. Effective allocation of resources to mobile users would improve resource utilization and reduce the latency in accessing the resources. Accurate prediction of location information is also crucial in processing location-dependent queries of mobile users. When a user submits a location dependent query, the answer to the query will depend on the current location of the user. Many

application areas including health care, bioscience, hotel management, and the military benefit from efficient processing of location-dependent queries. With effective prediction of behavior, it may also be possible to answer the queries that refer to the future positions

of users and like or dislike in terms of frequent pattern mining.

Compared to the amount of work performed on location update, little has been done in the area of mobility prediction.

We also analyze about the Mobile Agent which is an independent computer program which can migrate independently in the heterogeneous network, according to certain regulations, and seek for the appropriate computing resource, the information resource or the software resource. It can process the resources in advantage of these resources are in the same host or network and complete the specific task on behalf of the user. In this paper we analyze some of the good thing available in pattern mining and we concentrate on those rules to improve the mining pattern which is helpful in improving mining pattern in small handheld devices like Mobile devices,PDA,notebook etc. There are lot of patterns and rule discovered until now we focus on some of the rules which is much more useful in terms of small devices and suited the memory requirement of these devices.

The remaining of this paper is organized as follows. We briefly review the Data Mining in J2ME aspects in Section 2. The empirical evaluation on different data mining services in section 3.Review in section 4. Proposed method in section 5.The conclusions and future work are given in Section 6. References are given in Section 6.

2. J2ME IN DATA MINING

The core concept which is used in data mining through mobile is based on j2me.J2me consist of mainly two components first is CLDC and second is MIDP. The j2me architecture enables computing data mining thrives by scaled based on constraints on small computing device. J2ME architecture does not replace the architectural configuration and operating system of small computing device but the J2ME architecture consist of layered application located above the host operating system.

The J2ME host architectures system comprises mainly three software layers. The first layer is the configuration part which includes the java virtual machine which is the remedy key for byte codes generated with WTK (wireless tool kit) used for J2ME devices. It directly interacts with the native operating system. The second layer is the profile part which consists of application programming interface (API) for the mobile devices.

The Connected, Limited Device Configuration (CLDC) is one of the special component is J2ME which is installed top on the operating system. The CLDC is implements a stripped down J2SE because of the limited resource in small computing devices. A small computing device has two components which are based on the applications which is fruitful for data mining that is classes and application which is supplied by original equipment manufacturer (OEM). The package, javax.microedition.io, implements the Connection of generic Framework, a generalized API for making network connections.

It is the result of a Java Community Process [1] expert group JSR 30 [2] consisting of a number of industrial partners. The main goal of the CLDC is to standardize a highly portable Java application development platform for resource-constrained, connected devices.

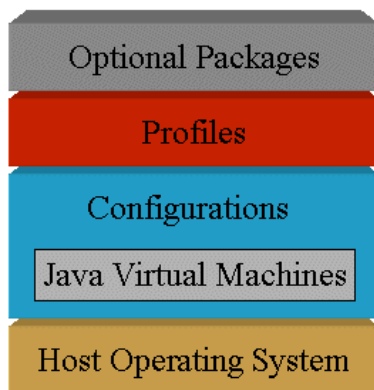


Fig 1.1 CLDC Position in J2ME Architecture.

Fig 1.1 and Fig 1.2 shows that CLDC is core technology designed to be the basis for one or more profiles. CLDC defines a minimal subset of functionality from the J2SE platform. Hence, the CLDC does not define the functionality which is operating system dependent, but instead defines the basic Java libraries and functionality available from the Kilo Virtual Machine (KVM). The KVM got its name because it includes such a small subset of the J2SE .CLDC position shows that it is the core part for connections and networking use in handheld devices. Host Operating system is the part of PDA and mobile devices.

The Mobile Information Device Profile (MIDP) is a part for a J2ME profile. It is layered on top of CLDC and adds APIs for application life cycle, user interface, networking, and persistent storage which are useful in data mining application. An application written for MIDP is called a MIDlet. MIDlet applications are subclasses of the javax.microedition.mid let. MIDlet class that is defined by MIDP.

2.1 MIDlet Suites

MiDlets are packaged in J2ME suite and distributed as MIDlet suites. A MIDlet suite can contain one or more MIDlets. The MIDlet suite consists of two files:

- Java Application Descriptor (.jad) file

The Java Application Descriptor file lists the archive file name, the names and class names for each MIDlet in the suite, and other information. This file is used by the mobile device to ensure that device has the minimum requirements to

run the application which is used in Data Mining Services (DMS).

- A Java Archive file (.jar) file.

The archive file contains the MIDlet classes and resource files which is ported on J2ME. These archival values are contained in DMs which are used in mobile devices through J2ME and CLDC.

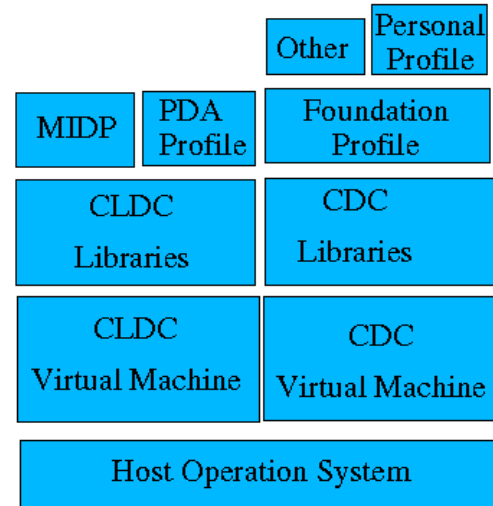


Fig 1.2 MIDP Position in J2ME Architecture

The MIDP is a set of APIs that resides on top of the CLDC as shown in Fig 1.2, providing features such as user interface, networking support and persistent storage. Specifications, APIs and other MIDP-related information are available at Sun Microsystems' MIDP website [2].

3. EVALUATION

In 1999 Dan Lim proposed mobile agent systems [3] and ubiquitous mobile computing which offer various kinds of dynamic services to the user with versatile devices at anytime and anywhere. Dam Lim proposed ubiquitous environment intelligent mobile agents are mandated to communicate with users and it is enabled by capturing interesting user's behavior patterns. Further, how services and access methods can be provided to user in ubiquitous mobile computing environment is critical problems recently. This environment provides the flexibility on different operating system .Therefore, effective modeling the behavior patterns of user in mobile agent systems is becoming very important. Effective modeling the behavior pattern of user benefits not only the users in smart access by caching or prefetching, in 2002 C.Y Chang et al. [4] and J.L. Chen et al. [5] but also the mobile service providers in financial profit like advertising [6]. Temporal association rules can be used to decide the next likely user's request services based on significant dynamic correlations. In the past, sequential association rule [7][8] have been used to capture the co-occurrence of user's sequential movement pattern in mobile web systems domains.

Service-oriented architecture (SOA) provides a flexible and dynamic paradigm to compose business processes from individual services [9, 14]. Using SOA, enterprise systems can define and execute transactions across multiple server domains at distributed locations.

Enterprises can use SOA's plug-and-play interoperability to integrate different services on the fly to enable cooperation among business partners.

Mobile path prediction is an important issue in both mobile calculating and wireless network, and it has great effect on all the protocols and QoS of wireless network. Shadow Cluster [15], LZ-tree [16], PPM [17], and order-k Markov predictor [18], all proposed prediction methods respectively. Thesis [19] pointed that order-k Markov performed the same or even better than the others, and among order-k Markov predictors, the order-2 was the best in predicting effect, however, order-k Markov predictors have the disadvantage of space inflation and this is a great challenge when it is put into application in this area.

With the advancements of the wireless communication techniques and popularity of mobile devices, i.e., mobile phone, PDA, contribute to a new business model. In any time and any place, user can get the service through mobile devices from Information Service and Application Provider (ISAP). In [20] and [21], this business model is called Mobile Commerce. Mobile phones of this new model even provide digital wallet, some banks also provide digital payment through mobile phones and with the approach of 3G, multimedia information will be transferred by mobile devices. In this way, diverse services and applications will be moved on. In the future, Mobile commerce is expected to be as popular as e-commerce [22].

A number of studies have discussed the issue of analysis of mobile behavior mining. Yun et al. [23] proposed a method of mining mobile sequential patterns. To increase the accuracy of prediction, the factor of moving path was taken into consideration by the authors. However, users' mobile behaviors might be different among different user groups or different time intervals. Higher precision of mobile behavior prediction can be achieved if we can discover the representative patterns in each pair of user group and time interval. In order to provide appropriate services for users with different behaviors, we need an effective personalized mobile behavior pattern mining approach.

In mobile environment, moving log size is very large. It will increase the overhead to integrate all the moving logs into one database server [24]. The centralized mobility pattern algorithm was discussed in [25] which cannot be efficient for large data size. Many parallel and distributed variants of sequential apriori algorithm [26, 29] have been discussed in other resources. In [30,34], grid implementation of frequent item sets in a grid environment deals with sales transaction of a company, these algorithms cannot be used directly in our soft computing environment, because this algorithm does not take into account the network topology while generating the candidate patterns. The generation of candidate pattern in is not same as the candidate pattern in mobile environment. In PCS, only the sequence of neighboring location of the network can be considered as the mobility pattern.

We propose incremental parallel and distributed knowledge grid mining approach based on apriori algorithm for mining mobility patterns in mobile environment.

Association rule mining is the most important, frequently used and fundamental technique in data mining which discovers the association and correlation among the item sets in large database. Now a day's data collection is huge and omnipresent in social and business areas. Many organizations are interested in association rule

mining in collaborations to achieve joint benefits. But the drawback is it leaks some sensitive information. This should not happen and for the security of such information, application of privacy preserving association rules mining is compulsory [35],[36]. In Association rule mining databases are transactional database. In transactional database, each transaction represents set of items and these set of items are those items which customers purchase together. By this transactional database, we are interested to find out those set of items, which are purchased by customers frequently. This helps the marketing department to design strategy, mainly to design of display layout with associated items placed together. The format of an association rule is " $X \Rightarrow Y$ ", where X & Y are the item sets. " $X \Rightarrow Y$ " can be interpreted as presence of X item set in a particular transaction implies the presence of Y item set. The extraction of such association rules from transactional database is referred to as association rule mining. Association rule mining is a two step process, in the first step all set of frequent item sets are identified, while in the second step association rules from the set of frequent item sets can be extracted. First part is very expensive and has exponential time complexity. Apriori algorithm is developed by Agrawal and Shrikant [37].

4. RECENT SCENARIO

The concept of mobile agent was put forward by the General Magic Company in 1990s when he introduced the Telescript business system [38]. It is an executable program and can move between all of computers of network along with its code and state.

The new method is proposed which stream is mining, with the constraints of storage; it is desirable to identify the outcome in an on-the-fly manner. The modified method was proposed to incorporate the two major mobile constraints: distributed server environments and large data stream [39]. The 4+1 bit method shows a promising result.

In 2009, Toshihiko Yamakami [40] proposed that a user with multiple visits in one day with a long interval has a larger revisiting possibility in the following month than the others. He investigates the minimum number of bits to incorporate this empirical law in order to cope with the two major mobile restrictions: distributed server environments and large data stream. He shows that the method with 2+1 bits can provide usable results to classify regular users in the case study. It gives the lower-bound of memory needed to identify revisiting users under mobile-specific constraints.

In 2009, S. Krishnaswamy proposed a hybrid architectural model for Distributed Data Mining (DDM), which is tailored to meet the needs of e-businesses where application service providers sell DDM services to e-commerce users and systems. The hybrid architecture integrates the client-server and the mobile agent chronologies. This model focuses on the optimization and costing issues of DDM, which are particularly relevant in the context of billing user for data mining services.

5. PROPOSED METHOD

We proposed a RSRM (Read Subset Removal Miner) for mobile devices. In this algorithm we first take the input from res file which is the part of wireless tool kit(WTK), we add all the data in the list variable. Then we check for the tokens and check the patterns in the tokens. Find the count of each subset and increment the count variable every time when reoccurrence is done. We only add those values which count value is greater than the minimum support value.

If the list is successfully searched then apply pruning on the list. Finally apply the subset removal algorithm and consider only the superset which is useful.

Assumptions:

File- res file to read the data
read()- predefined method of java.io
list-for storing the value
count-stores the occurrences
list1-stores the final result.

Algorithm:

RSRM (file, count)

```
Step1: [Input from res]
While(object.read() != -1)
{
Append(list);
}
```

```
Step2:[Check for Tokens]
If(object.hasMoreTokens())
{
Count++;
}
Else
{
Break;
}
```

```
Step 3:while(list != -1)
3a. Computing the occurrences
3b.Store in any integer variable let x
3c. count=count+x
3d.if successfully searched
Prune(list)
```

```
Step 4:[Subset Removal]
4a. if (count >= min-support)
List1=list
4b. if(list1)
Remove subset from list
```

```
4c. add relevant superset
Else remove the set from the list
```

Step 5. Finish

6. CONCLUSIONS AND FUTURE WORK

In this paper we survey on different mining techniques along with their efficient way sequentially. By analyzing the above paper we conclude some of good things that we apply for the handheld devices like sequential pattern and other approaches which we apply with CLDC and MIDP components. In future different techniques and algorithm will be our interest to discuss. Similar idea will be discussed for some other major technique of data mining like clustering, etc. How authentication will be achieved between database owner and third party will also be a work of future interest.

7. REFERENCES

- [1] "Java 2 Platform Micro Edition (J2ME) Technologies for Creating Mobile Devices", White Paper, Sun Microsystems, Inc., 2000,
- [2] "Java 2 Platform, Micro Edition Datasheet", Sun Microsystems, Inc., 2002,
- [3] Dan Lim, "Ubiquitous mobile Computing: UMC's model and success", Educational Technology & society 2(4) 1999 ISSN 1436-4522.
- [4] C. Y Chang, M.S Chen, "Integrating web caching and Web prefetching in client-side proxies", in: Proceedings of the ACM 11th International conference Information and Knowledge Management, 2002.
- [5] J. L. Chen, "Resource allocation for cellular data services using multivalent schemes", IEEE Trans. Syst. Man Cybern. 31 (6) (2001) 864-869.
- [6] Jiawei Han and Micheline Kamber, *Data Mining Concepts and techniques*, www.mkp.com or www.books.elsevier.com.
- [7] R. Agrawal and R. Srikant, "Mining sequential Patterns," 11th International Conference on Data Engineering, Taipei, Taiwan, Mar. 1995.
- [8] Vincent S. Tseng, Kawuu W. Lin, "Efficient Mining and prediction of user behavior patterns in mobile web systems", Information and software technology, 2005.
- [9] M. Bichler and K.-J. Lin. Service-oriented computing. IEEE Computer, 39(3):99–101, March 2006
- [10] J. S. Ide and F. G. Cozman. Generating random bayesian networks. Proceedings on 16th Brazilian Symposium on Artificial Intelligence (SBIA 2002) , November 2002.
- [11] D. S. Johnson. Approximation algorithms for combinatorial problems. Journal of Computer and System Sciences, 9:256– 278, 1974.
- [12] K. B. Korb and A. E. Nicholson. Bayesian Artificial Intelligence. Chapman & Hall/CRC, London, UK, 2004.
- [13] K.-J. Lin, J. Y. Hsu, Y. Zhang, and T. Yu. A distributed reputation broker framework for web service applications. Journal of E-Commerce Research, 7(3), 2006.
- [14] M. P. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann. Service-oriented computing: State of the art and Research challenges. IEEE Computer, 40:38–45, November 2007.
- [15] David A. Levine. The shadow cluster concept for resource allocation and call admission in ATM-based wireless networks[J]. Mobile Computing and Networking, pages 142–150, 1995.
- [16] Amiya Bhattacharya: An information-theoretic approach to track mobile users in PCS networks[J]. *ACM/Kluwer*

- Wireless Networks*, 8(2-3):121–135, March–May 2002.
- [17] John G. Cleary and W. J. Teahan. Unbounded length contexts for PPM[J]. *The Computer Journal*, 40(2/3), 1997.
 - [18] Christine Cheng, Ravi Jain. Location prediction algorithms for mobile wireless systems[M]. In . Illyas and B. Furht, editors, *Handbook of Wireless Internet*. CRC Press, 2003.
 - [19] Libo Song, David Kotz, Ravi Jain, and Xiaoning He, Evaluating location predictors with extensive Wi-Fi mobility data [A], in Proc. of INFOCOM[C], March 2004, pp. 1414 – 1424.
 - [20] U. Varshney, R. J. Vetter, and R. Kalakota, “Mobile Commerce: A New Frontier”, *IEEE Computer*, Vol. 33, No. 10, pp. 32-38, October 2000.
 - [21] J. Veijalainen, “Transaction in Mobile Electronic Commerce”, in Proceedings of the International Workshop on Foundations of Models and Languages for Data and Objects, Dagstuhl Castle, Germany, pp. 203-227, September 1999.
 - [22] J. B. Schafer, J. Konstan, and J. Riedl, “Recommender systems in Ecommerce”, Denver, Co, pp. 158-166, November 1999.
 - [23] C. H. Yun and M. S. Chen, “Mining Mobile Sequential Patterns in a Mobile Commerce Environment”, *IEEE Transaction on Systems, Man, and Cybernetics, Part C*, Vol. 37, No. 2, pp. 278-295, March 2007.
 - [25] R. Agrawal, T. Imielinski and A. Swami, “Mining Association Rules between Sets of Items in Large Databases”, *Proceedings of ACM SIGMOD International Conference on the Management of Data*, 1993, pp. 207-216.
 - [26] Gokhan Yavas, Dimitrios Katsaros, Ozgur Ulssoy, and Yannis Manolopoulos, “A Data Mining Approach for Location Prediction in Mobile Web Environments”, *IEEE Transaction on Data and Knowledge Engineering*, 2005.
 - [27] R. Agrawal and J. Shafer, “Parallel Mining of Association Rules”, *IEEE Transaction on Knowledge and Data Engineering*, 1996, pp. 962-969.
 - [28] R. Agrawal, T. Imielinski and A. Swami, “Mining Association Rules between Sets of Items in Large Databases”, *Proceedings of ACM SIGMOD International Conference on the Management of Data*, 1993, pp. 207-216.
 - [29] Gokhan Yavas, Dimitrios Katsaros, Ozgur Ulssoy, and Yannis Manolopoulos, “A Data Mining Approach for Location Prediction in Mobile Web Environments”, *IEEE Transaction on Data and Knowledge Engineering*, 2005, pp.121-146.
 - [30] Assaf Schuster, Ran Wolff, and Dan Trock, “A High-Performance Distributed Algorithm for Mining Association Rules”, *Knowledge and Information Systems (KAIS) Journal*,
 - [31] Luo, Anil L. Pereira, and Soon M. Chung, “Distributed Mining of Maximal Frequent Item sets on a Data Grid System”, *The Journal of Super Computing*, 2006, pp.71-90.
 - [32] Mario Cannataro, Antonio Congiusta, Andrea Pugliese, Talia, and Paolo Trunfio, ”Distributed Data Mining on Grids: Services, Tools and Applications Dec. 2004, pp.2451-2465.
 - [33] Ian Foster, “Globus Toolkit Version 4: Software for Service-Oriented System”, *Journal of Computer Science and Technology*, Jul. 2006, pp. 513-520.
 - [34] Cristian Aflori and Mitica Craus, “Grid Implementation of Apriori Algorithm”, *Advances in Engineering Software*, pp. 295-300.
 - [35] C. Tang, Su. Chunhua and Sa. Kouichi. A Distributed Privacy-Preserving Association Rules Mining Scheme Using Frequent-Pattern Tree, In ADMA, 2008.
 - [36] W. Jing, L. Huang, Y. Luo, W. Xu and Y. Yao. An Algorithm for Privacy-Preserving Quantitative Association Rules Mining. In Proceedings of the 2nd IEEE International Symposium on Dependable, Autonomic and Secure Computing, 2006.
 - [37] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In VLDB, 1994.
 - [38] H.M. Wang, Y.J. Wang, M. Xian, “Data Collection Technique of Computer Network Attack Effect Evaluation Based on Mobile Agent”, *Computer Engineering*, vol. 33, 2007, pp. 160-162.
 - [39] “A 4+1 bit month-scale regularity mining algorithm With one-path and distributed server constraints for mobile internet,” in *NBiS 2008*, vol. LNCS 5186. Springer Verlag, September 2008, pp. 232–241.
 - [40] Toshihiko Yamasaki, A Space-optimal Month-scale Regularity Mining Method with One-path and Distributed Server Constraints for Mobile Internet, 2009.