A PROFICIENT METHOD FOR MINING ECPs FROM TRANSACTIONAL DATASET

[1] Shilpa Sudersahn Kumar, [2] Swaroopa Shastri, [3] Sony
[1]Department of Studies in Computer Applications (MCA) ,Visvesvaraya Technological University
Centre for Post Graduation Studies Kalaburagi, Karnataka, India

[2]Department of Studies in Computer Applications (MCA) ,Visvesvaraya Technological University
Centre for Post Graduation Studies Kalaburagi, Karnataka, India

[3]Department of Studies in Computer Applications (MCA) ,Visvesvaraya Technological University
Centre for Post Graduation Studies Kalaburagi, Karnataka, India

Abstract: In the current internet era the first thing we have to deal with is the huge data. We have many ways to deal with this kind of problem such as pattern mining (PM), data-classification, text-mining and opinion mining. Among all these other, pattern mining is the important issue. While optimizing the product plans of the organization data, we propose the mining erasable patterns as an alternative of the frequent pattern mining, but while working with large amount of data there also product large amount erasable patterns which in turn consumes large amount of memory. Therefore it has become important to produce condensed erasable patterns. Here we propose a system in which we first define the erasable closed patters which again represents the set of erasable patterns without any information loss, after collecting the erasable patterns we will implement method for determining the erasable closed patterns faster based on the product ID set. Next we implement efficient algorithms ECPat and dNC-ECPM for the creation of ECPs. The experimental result says that the ECP suits best for mining of the erasable closed patterns by reducing the process time and memory usage for all datasets.

***Keywords* - Erasable Itemsets (EI), Erasable Closed Pattern Mining (ECPM), Erasable Closed Patterns (ECPs), Data Mining (DM), Pattern Mining (PM), Mining Erasable Itemsets (MEI), Erasable Itemsets for very Dense Datasets (EIFDD).**

I. INTRODUCTION

Mining can be defined as the process of finding the knowledge data or the patterns in the large amount of datasets. As the data in the internet is growing fast as time passes, data mining has become important to deal with this large amount of data. The data mining consists of process like dataset mining, top-rank frequent data mining and sequential pattern mining. In the recent years techniques like EPM and top-rank-k erasable pattern mining has been introduced and used a lot.

Let us consider the example of manufacturing factory to understand these techniques. The products in the factory are manufactured using the number of products. There are some situations where the company cannot be able to buy all the necessary items due to the low budget. Now here is the work for the data mining techniques. The erasable pattern mining finds the items which can be erased in order to reduce the manufacturing factory loss, this erasable pattern mining dataset can be used to generate the new product plan.

There are many erasable pattern mining algorithms such as META, VME, MERIT, and MEI. The experimental study after comparing all these algorithms shows that the MERIT uses an NC Set created from a WPPC-tree, as the MEI uses the dPidset. MEI is currently the most effective algorithm for mining erasable itemsets. However MEI algorithm needs large process time and memory for large amount of dataset. This paper a new and enhanced method that uses the subsume and dPidset to improve the mining process of EI.

II. RELATED WORK

In [1] this paper, while dealing with large amount of data in the internet it quite difficult to determine the associations between the data to generate the pattern. Here we propose two algorithms. The proposed algorithm is not like the other known algorithms. Experiment shows the proposed algorithms works better than the known algorithm by factors ranging from three small problems to more than an order of large problems. These two proposed algorithms works as hybrid algorithm called ApiorHybrid to make use of best features of both algorithms. The scale-up experiment shows that Apriori algorithm remains linear with the increase of the transaction. It has very efficient and excellent scale-up features with respect to the transaction size and the number of dataitems in the database.

In [2] this paper we are dealing with large data transaction of the users. The system has to deal with the large amount data transactions which keep on increasing day by day. Each transaction consists of the purchase details made by the customers. Here in this paper we propose a algorithm that provides association rules to applied between the dataitems in the DB. This algorithm also provides buffer-management and novel estimation and pruning techniques. This paper also describes the results of the association between the data items from the large retailing data set. The experimental result show that the algorithm to be efficient.

In [3] this paper, there are different techniques in the patter mining methodology among them top-k frequent pattern is a popular technique. This technique is mainly used in transactional business in which needs to find out the patterns in the transaction database. However several processes have been proposed to perform this kind of task but the computational cost was expensive. As a solution to these issues we propose efficient algorithm named BTK. This algorithm works on the tree-structure name TB-tree to store all the important information about the patterns generated using frequent pattern mining technique. Along with TB-tree structure the BTK algorithm also implement the B-List structure for the information storage. This structure depends on the subsume indexes to decrease the search space and to increase the speed in the discovery of the frequent pattern mining. It's also pruning strategy and a mechanism to increase the system execution speed and reduce the execution time. Additional to all these techniques BTK algorithm implements two efficient mechanisms to generate subsume indexed list and intersecting B-lists. An experimental result which was conducted using different kinds of dataset shows that BTK algorithm is efficient and competitive.

In [4] this paper, whenever dealing with the large amount of data we need to implement data mining techniques. Databases like transaction database, time series databases and other databases have been using frequent pattern mining techniques. This frequent pattern mining technique has been gaining lot of popularity among these kind of databases. Most of the previous papers have implemented the Apriori algorithm which works on the generate and test methodology however these kind of algorithms were still costly and required large space and time while dealing with large number of datasets. We propose a NFP tree-structure which is the extension of the prefix tree. This prefix tree structure stores the important pattern details generated by the algorithm and compression details. This method will help the system to generate the frequent pattern and for the mining of the frequent patterns. To develop the efficient frequent pattern we need to follow the below steps (1) we need to compress the large data into the small structure.(2) the FP uses the PF growth method to avoid the large number of data which makes the system costly. (3) We use a divide and conquer method to divide the task into smaller parts to make the mining process faster in conditional database, which in turn reduces the execution time and the search space. The experimental result shows that the FP-growth structure id efficient even while dealing with large amount of data and works faster when compared to previous algorithms.

In **[5]** this paper, Cloud computing is considered as a basic infrastructure in a growing service industry. This has the advantage of the reduced cost by allowing the users to share the resources and merging with the on demand and pay as you go mechanism. This mechanism allows the users to dynamically allocate the resources and based on the payment as they pay. It will also affect the traditional and trust management. We have lot of advantages while dealing with the cloud computing environment such as its scalability, accessibility, storage of data at remote place and sharing services dynamically.

In **[6]** this paper, maintaining the trust between the users and the cloud is a very challenging and important fact in the cloud environment. The current environment is greatly based on the trust, reputation of the cloud providers in the cloud. Here we will show the existing mechanism to build the trust between the users and the CSP and we will also list the challenges and the limitation in the traditional system of trust based management.

In **[7]** Cloud computing provides various opportunities for the company owners to increase their economy by providing different services to users such as resource sharing , providing remote access to data, data storage, etc. Generally, cloud providers will provide the trust based policies on the service level Agreements for the services they are providing. The trust policies are different among different CSP even though they provide same services to the user. The proposed system will provide a way for the users to decide the cloud service provider who are trust worthy based on different features and services.

In **[8]** There are many trust management system which are develop in the cloud computing industry which are based on the fact that each entity will use the attribute the scoring function in the system. The previous developed system which provides trust based services also focuses on the increase in the reputation in the system. In proposed work we will focus on the providing the trust based management framework which are supported and analysed the trust related feedback given from the user and from different entities. The system also allows using different scoring functions to analyse the same feedback data for the evaluation of the trust.

In **[9]** this paper, Customer's feedback is the most reliable and a good source to analyse the trustworthiness of the cloud. These feedbacks are collected from the users. However, it is common that the user show the malicious behaviour by the users in the cloud. In proposed work we proposed a method to detect the attacks the form the users which will help the users to decide the trust worthy cloud service providers out of many. This system not only provides detect the malicious behaviour but also detect the misleading feedback from the users. This identification is done using the collision attack and also identifies the Sybase attack.

III. METHODOLOGY USED

MEI uses the dPidset structure to quickly determine the information of erasable itemsets. Although mining time and memory usages are better than other algorithms. MEI's performance can be enhanced by the EIFDD algorithm.

The EIFDD algorithm reduces the search space for erasable itemset mining using the subsume concept. The EIFDD algorithm is used for mining the itemsets. Firstly, the algorithm scans the dataset to determine E1 with their pidsets, and then sorts E1 in descending order of pidset length. Secondly, the algorithm calls Algorithm 1 to generate the subsume index associated with E1. Thirdly, the algorithm puts all EIs in E1 to the results. Finally, the algorithm calls the Expand E procedure, which uses the divide-and-conquer strategy and the subume index associated with E1 to mine all erasable itemsets.
The following steps are used:

Step 1: consider e as variable consists of remaining erasable 1- itemsets. Because Subsume (e) = {c}, ec is added to the results without determining its information. ed, ef, and eh are disqualified because their revenues exceed the threshold

Step 2: d which consists of the result generated at first step is considered with the remaining erasable 1- itemsets (h, f , and c). h and f belong to Subsume(d); therefore, the algorithm adds df, dh, and dfh to the results without determining their information. d is combined with c to create dc with g(dc) = 750 dollars. dc and the erasable itemsets {dch, dcf , dchf } created from Subsume(d) and dc are added to the results.

Step 3: h is considered with the remaining erasable 1- itemsets (f and c) to create hf and hc with g(hf) = 500 dollars and g(hc) = 700 dollars. hf and hc are added to the results. They are used to create hfc with g(hf c) = 750 dollars, which is added to the results.

IV. RESULT AND DISCUSSION

As the registered users purchase goods their transaction data is stored. At this point the EIFDD algorithm is used and the information of same items being purchased frequently by the users is collected. Now data obtained is used by this system to produce the patterns for the erasable itemsets. All these phases mentioned require more time and memory. But, this system uses EIFDD algorithm that is much more efficient that uses less memory space and less time to produce the desired results.

V. CONCLUSION

In this paper we have successfully achieved the results to show that the system can efficiently generate the list of erasable patterns in an efficient manner with fastest execution time and also uses less memory compared to other algorithms. This system uses the divide and conquers method along with the MEI to generate the erasable item list. This system is proven to be work efficiently even with the large amount of data. To prove the efficiency of the proposed method we compare the VME, MERIT with the MEI method and the result shows that our algorithm is efficient in mining with less execution time and it has also shown less memory usage.

REFERENCES

[1] An efficient algorithm for mining top-rank-*k* frequent patterns

[2] Fast Algorithms for Mining Association Rules

[3] Mining Association Rules between Sets of Items in Large Databases (1993)

[4] Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach

[5] S. Pearson, "Privacy, Security and Trust in Cloud Computing," in Privacy and Security for Cloud Computing, ser. Computer Communications and Networks, 2013, pp. 3–42.

[6] J. Huang and D. M. Nicol, "Trust Mechanisms for Cloud Computing," Journal of Cloud Computing, vol. 2, no. 1, pp. 1–14, 2013.

[7] S. Habib, S. Ries, and M. Muhlhauser, "Towards a Trust Management System for Cloud Computing," in Proc. of TrustCom'11, 2011

[8] W. Conner, A. Iyengar, T. Mikalsen, I. Rouvellou, and K. Nahrstedt, "A Trust Management Framework for

Service-Oriented Environments," in Proc. of WWW'09, 2009.

[9] T. H. Noor, Q. Z. Sheng, and A. Alfazi, "Reputation Attacks Detection for Effective Trust Assessment of Cloud Services," in Proc. of TrustCom'13, 2013.