ISSN 2250-3153

Architecture of an Automated CBA System Using ERP Model

Shweta Tiwari*, Prof. Sapna Choudhary**

* Computer Sc & Engg ,S.R.I.T
** Computer Sc & Engg,S.G.I

Abstract- In this paper , a model is proposed which integrates the database, customer queries, transactions, and all other specifications used in ERP systems, then use enhanced & latest data mining techniques to integrate decision making and forecast flows. The proposal of the paper is based on the data mining effects using ERP framework. By using the various properties of ERP's and background we collect the data from central database in cluster format which is based on the action taken against the queries generated by the customers. Furthermore, the clustered data used by ARM Algorithm to extract new rules and patterns for the enhancement of an organization. This is a complete architecture of data mining applications on ERP framework to find out the answers of upcoming queries. This will make the best association between the customers and organization. It act as a base for a CRM system as it permits the company itself to recommend other products by e-mail. The model is basically consist of three layers 1) CRM 2) ERP 3) KNOWLEDGE DISCOVERY. Here the third layer is the proposed layer, since the Knowledge discovery can be defined as the extraction of contained, hidden and useful information from the large database. So in this presented model this layer also deal with the central database containing data collected from any department of ERP and CRM layers. Since customer's queries contain unlimited attributes and characteristics of data. By utilizing the benefits of third layer in the proposed model we used enhanced variation of Apriori algorithm i.e distributed cba Algorithm for effective and high-quality results.

Index Terms- Association, Classification, Distributed cba algorithm , Data mining, ERP Model .

I. INTRODUCTION

Over the past few years, the capabilities of data generation & collection have increased rapidly. With the rapid growth of Internet, and the advances in data collection tools have provided us with a vast amount of data. These changes have contributed improvement in quality as the essential component systems have been established, but to shape the quality of infrastructure must provide hardware and software and the infrastructure can be associated with conditions. The problem is that we can increase the concurrency and transaction system for high quality we have. We can do this when we create timing and synchronization between tasks use some relationships to merge exist phases to close them together and monitoring action take places between phases. In order to maximize customer satisfaction levels, synchronization is a new technique that should be used on practical software to create the balance in organizational structure. In order to extend the market competitiveness of enterprises, the structural and functional aspects of ERP system must adapt to the characteristics of the E-commerce times, with the support of internet technology, realize the organic integration of enterprise, partners and customers based on the same Ecommerce platform. Information technology is an attempt to integrate combined methods for providing a suitable solution that increases customer satisfaction level is. In order to present confident and dependable environment we have replaced integrated distributed systems with the integrated central systems. So many comprehensive reviews on the proposed framework and models were used to present new framework. That using these techniques and methods we can solve problems related to repetitive processes using concurrency in the implementation of data mining technique and resolve problems related to it desirable. Within this framework we have proposed distributed cba algorithm

II. EXPERIMENTAL MODEL

The proposed model shown in the Fig.1, described the working of all concerned departments exist in every organization. The traditional structure of an organization is divided into two layers. Each layer has its own departments and officials having specific responsibilities and burden. This is basically the combination of CRM and ERP which are essential part of the enterprise. We extend this layering structure by adding knowledge discovery layer, as a third layer where the association rule generation and classifier generation from the large database will be handled by using data mining technique like distributed cba Algorithm, which is better than Apriori Algorithm.

A. Description of Three Layers:

I. Crm Layer: In our model we presented customer relationship management as an outer view. Whenever a customer contact with the company the customer supporting officer receive customer's request. This layer has much importance because it predicts the customer's behavior having direct communication with the customer. These requests includes queries, complaints, suggestions and orders then forward these requests to the inner view enterprise resource planning. (ERP) through the query generator. After taking action on the perspective request the answer will forward through

the CRM Layer. And result will also be saving in the database for knowledge discovery view.

II. Erp Layer: The important part of the model is ERP Layer. In this view each department have equal access to a single database that holds the customer's data or complaints. In this layer the customer queries ERP's rotating and evaluating by the concern department. For example a customer contact with customer support department (CRM layer) after initial review and statistics this query will throw to the ERP Layer. The department is responsible to find the solution and give proper reply to the customer and forward feedback back to customer through CRM layer as well as in the central database for future assistance[5].

III. Knowledge Discovery Layer: As per the abstract in the presented model this layer is involved to discover the knowledge from the organization's central database having variety of data deposited from any department of ERP and CRM layers. The datasets collected in database has several attributes and characteristics according to the queried customer. In order to reflect the applicability of this layer we have used data mining techniques for better results. The various techniques are described below-

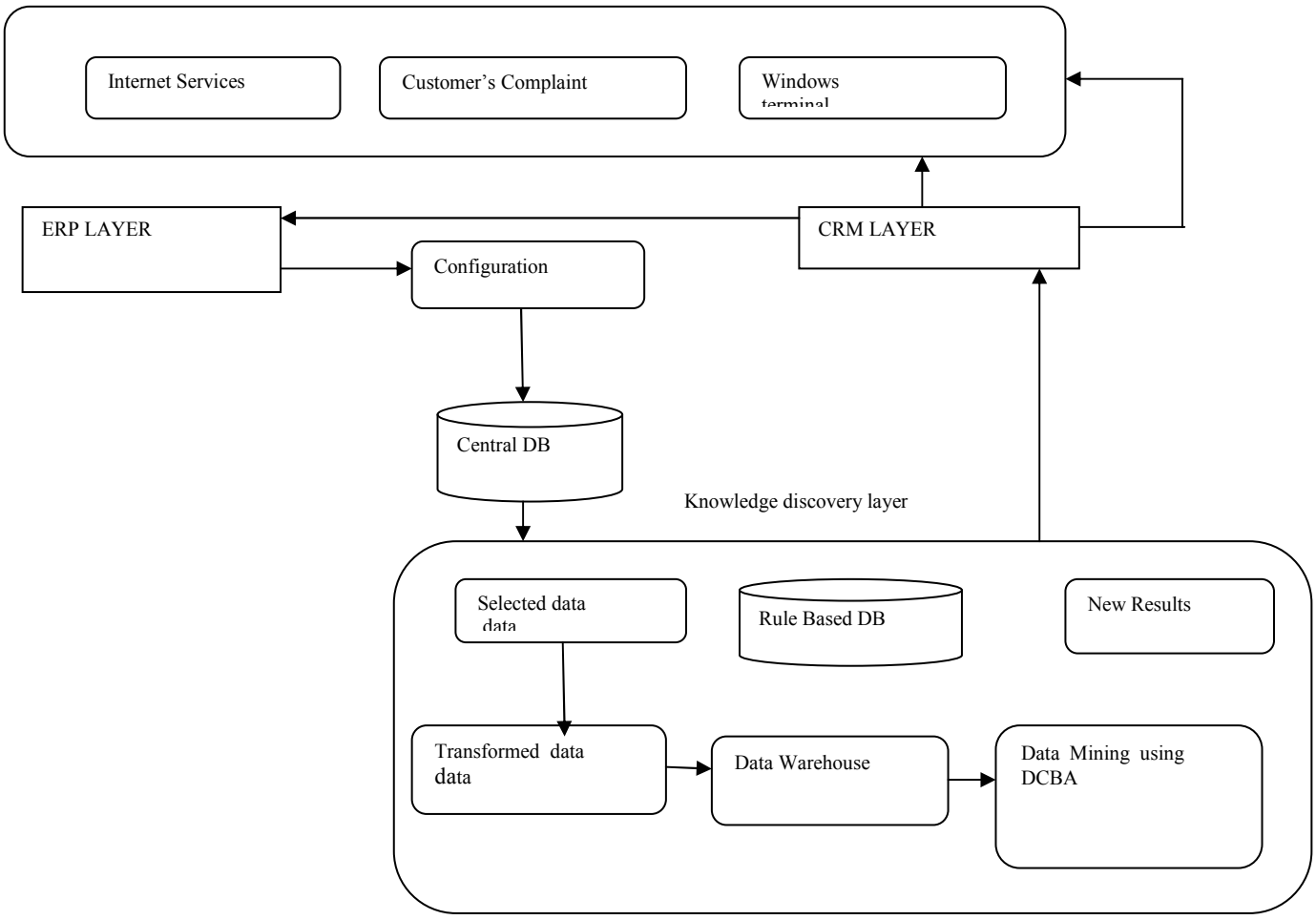

Fig-1 ERP MODEL

b. *Classification*: Classification [Han and Kamber 2000] is to build (automatically)a model that can classify a class of objects so as to predict the classification or missing attribute value of future objects (whose class may not be known). It is a two-step process. In the first process, based on the collection of training data set,a model is constructed to describe the characteristics of a set of data classes or concepts. Since data classes or concepts are predefined, this step is also known as supervised learning (i.e., which class the training sample belongs to is provided). In the second step, the model is used to predict the classes of future objects or data [3].

c. Association: Association analysis is the discovery of what are commonly called association rules. It studies the frequency of items occurring together in transactional databases, and based on a threshold called support, identifies the frequent item sets. Data can be use to find association between several attributes, generate rules from data sets, this task is known as association rule mining [12]. Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction. The goal of association rule mining is to find all rules having support ≥ minsup (minimum support) threshold and confidence ≥ minconf (minimum confidence) threshold [3].Moreover, association rule mining can be viewed as a two-

step process, first, find all frequent itemsets: items satisfying minimum support. Second, generate strong association rules from the frequent itemsets: these rules must satisfy minimum support and minimum confidence Furthermore, data mining can be apply in ERP data also where association between the several attributes of customer queries provides the result of future prediction of expected solution of the customer queries and department activities.

D. *Distributed Classification based on Association rule (DCBA) Algorithm:*

Association-based classification attracts special interests in the past several years. Essentially, association based classification integrates the classification technique with the association technique by building a classification model based on the association rules mined. It differs from other approaches in that it examines different variables in the classification space simultaneously instead of examining only one variable at one time. Association-based classification has been proven effective for very sparse and high-dimensional data.

Reference [7] Bing Liu et al proposed Classification Based on Association rules (CBA) algorithm as an integration of classification rule mining and association rule mining (Bing Liu et al, 1998). The Integration was done by focusing on mining a special subset of association rules called class association rules (CARs).

The Problem for CBA algorithm can be stated as follows:
Assume a relational table D with n attributes. An attribute can be discrete or continuous. (CBA also works with transactional data)
There is a discrete class attribute (for classification).For a continuous attribute, it is discretized into intervals.

Item: (attribute, value) Let I be the set of all items in D, and Y be the set of class labels.A class association rule (CAR) is an implication of the form: $X \rightarrow y$, where $X \subseteq I$, and $y \in Y$.

A rule $X \rightarrow y$ holds in D with confidence and support (as in normal association rule mining).

CBA algorithm was carried out in three stages:

(1) Find the CARs set using CBA-RG algorithm, which is based on Apriori algorithm
(2) Build Classifier based on CARs set using training data
(3) Apply the Classifier for data mining (predict which class a new item belongs to)

Therefore, the main focus to improve CBA algorithm is to How can we better choose the CARs set using association rule mining how can we generate more accurate classifier.Bing Liu suggested using multiple s_{min} to solve the above problem using the following algorithm: Bing Liu et al has attacked on both frontiers. As we know, the key parameter in association rule mining is the s_{min}. It controls how many rules and what kinds of rules are generated. Earlier CBA system follows the original association rule model and uses a single s_{min} in its rule generation. However, this is inadequate for mining of CARs since many practical classification datasets have uneven class frequency distributions. Using a single s_{min} will result in one of the following two problems:

If we set the s_{min} value too high, we may not find sufficient rules of infrequent classes.

If we set the s_{min} value too low, we will find many useless and overfitting rules for frequent classes. $s_{min,i}$: For each class ci, a different minimum class support is assigned. The user only gives a total s_{min}, denoted by $t\_ s_{min,}$, which is distributed to each class according to their class distributions as follows:

$s_{min,i} = t\_ s_{min} \times freqDistr(ci)$

The formula gives frequent classes higher s_{min} and infrequent classes lower s_{min}. This ensures that sufficient rules for infrequent classes will be generated and will not produce too many overfitting rules for frequent classes. Regarding c_{min}, it has less impact on the classifier quality as long as it is not set too high since we always choose the most confident rules. With a sound set of CARs, accurate classifier can be obtained using the high precedent rules on a training data set. Various algorithms are available for fulfilling this job proposed a simplest algorithm [6].

(1) Sort the set of generated rules R according to the relation "> ", which is defined as Given two rules, ri and rj, ri > rj (also called ri precedes rj or ri has a higher precedence than rj) if

i. the confidence of r i is greater than that of r j , or
ii. their confidences are equal, but the support of r i is greater than that of r j , or
iii. both the confidences and supports of r i and r j are the same, but r i is generated earlier than r j;

(2) Using the rules in R to cover the training data (in sorted sequence). After each rule, the covered cases by the rule are removed. A set of rules C is selected from R that covers all training data, where, R = <r1, r2, …, rn, default_class>.

(3) Discard those rules in C that do not improve the accuracy.

III. FUTURE WORK

The proposed layered architecture will be enhanced by using more enhanced data mining techniques and new rules will be generated in future for the enhancement of an organization. Advanced ERP tools will be used to extend the existing work and create convenient method for the customers to avail the organization's facilities with ease. Since CBA is useful to integrate classification and association mining so it is beneficial where there is a need to-

1. mine highly distributed data
2. produce an accurate classifier. (compared with C4.5)
3. produce better quality data necessary for business enterprises.
4. Solve a problems in classification systems.
5. Able to create classfier with data on disk rather than in memory.

IV. CONCLUSION

The layered model in this paper will be able to solve the problems regarding all the customers queries i.e according to the customer need They can easily contact the organization and can purchase the organization products very easily. The knowledge discovery layer utilizing CBA algorithm generates

new rules and classifiers for the betterment of an organization for future correspondence to improve the growth of the customers for an organization. This algorithm performs better than the Apriori algorithm since the strength of CBA is its ability to use the most accurate rules for classification. However, the existing techniques based on exhaustive search face a challenge in the case of huge amount data due to its computation complexity. In today's technological environment, the databases may be scattered over different locations and heterogeneous. We will combine CBA and distributed techniques to develop a distributed CBA algorithm to mine distributed and heterogeneous databases. CBA running decently on top of a meta database layer, which hide the distributed nature of the underlying data sources.

ACKNOWLEDGMENT

My special thanks goes to Prof. Sapna Chaudhary H.O.D ,M.tech in S.G.I for supporting and providing her best suggestions in my research work.

REFERENCES

[1] Abdullah S. Al-Mudimigh,Zahid Ullah, Farrukh Saleem , A Framework of an Automated Data Mining Systems Using ERP. Model International Journal of Computer and Electrical Engineering, Vol. 1, No. 5 December, 2009 1793-8163.

[2] Apriori Algorithm, http://en.wikipedia.Org /wiki/Apriori algorithm #Algorithm,Accessed Date: 12th April, 2009.

[3] Abdullah Saad Almalaise Alghamdi Chairman of CIT, faculty of Computing and Information Technology, King Abdulaziz rw University, Jeddah Rules Generation from ERP Database: A Successful Implementation of Data Mining.

[4] Arash Ghorbannia Delavar, Behrouz Noori Lahrood, Mohsen Nejadkheirallah and Mehdi Zekriyapanah Gashti ,ERDQ: A Real-time Framework for Increasing Quality of Customer Whit Data Mining Mechanisms International Journal of Information and Education Technology, Vol. 1, No. 1, April 2011 ISSN: 2010-3689.

[5] Y.K.Alp , O.Arikan , "Erp Source Reconstruction By Using Particle Swarm Optimization ", Ieee Icassp 2009.

[6] Davy Janssens , Geert Wets ,Tom Brijs, Integrating Classification and Association Rules by proposing adaptations to the CBA Algorithm.

[7] Bing Liu, Wynne Hsu, Yiming Ma, "Integrating Classification and Association Rule Mining." Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98, full paper), New York, USA, 1998.

[8] Hong Seng Woo , " Implementing ERP in China: Lessons from Family-Owned Chinese Enterprises " PICMET 2009 Proceedings, August 2-6, Portland, Oregon USA ,IEEE PICMET 2009 .

AUTHORS

First Author – Shweta Tiwari, B.E, S.R..I.T., Email: Shweta0944@gmail.com
Second Author – Prof. Sapna Chaudhary, M TECH, S.G.I, Email:Chaudharysapnajain@gmail.com