ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 3, May 2012

Optimized Way to Uncover Promising Customer Deeds Using CMAR

G. Srujana Kumari, M.tech (CS), ASCET, Gudur

Prof.C.Rajendra, H.O.D(CS), ASCET,Gudur

V.Sreenatha Sarma, Asst.Prof(cs), ASCET,Gudur

Abstract:-- **By the disappearance of idyllic customer relationship and uninhibited bitterness in the nuptial and worldwide business, the Customer Relationship Management (CRM) has fully fledged to be topics of startled. CRM is a trade scheme engaged to comprehend, foresee and retort to the wants of an endeavour's contemporary and promising customers in order to grow the relationship value.As the promising customers are the grounds for the maturity of the enterprise, most of the datamining techniques propose associative classification to be an optimized way to uncover the promising customers deeds. But this associative classification still endures from the bulky set of mined rules and sometimes prejudiced classification since the classification is pedestald on only solitary high confidence.So in this paper we propose a data mining approach called Classification Basedon Multiple Association Rules(CMAR) by which huge data are mined efficiently,store,retrieve mined association rules efficiently and prunes rules efficiently with high classification accuracy and provide strong flexibility.Thus CMAR provides an optimized way to uncover the promising customers deeds.**

***Keywords:*Customer RelationshipManagement(CRM);optimized way;promising customer deeds;data mining;Classification Basedon Multiple Association Rules(CMAR);**

I. INTRODUCTION

With the desertion of idyllic customer liaison and not reserved sullenness in the nuptial and all-inclusive business, the Customer Relationship Management (CRM) has flattering fledged to be matter of anxious. CRM can be defined as the succession of envisageing customer deeds and plump for actions to sway that activities to achieve the goals of the concern, as in [1], It endow with a work of fiction being to the course of action organization structure and optimizes the trade progression.

By guessing customer requests in advancement, organizations can then flea market the factual wares to the factual segments at the factual moment through the factual rescue means. Customer fulfilment be able to be improved in the course of further efficient marketing. The four proportions of CRM cycle are indispensable efforts to expand customer insight. They are customer identification ,customer attraction customer retention,customer development [2]. Customer identification engrosss aiming the citizens who are on the whole apt to transpire consumers or most lucrative to the company.It involves in analysing customers who are being vanished to the competition and how they can be won back.

.Reminiscent of a mounting topic, data mining is in concert an ever more momentous division in the appraisal seize activity of every pace of life, as in [3]. Data mining uses thorny statistical handing out or artificial intelligence algorithms to conclude accommodating inclinations and patterns from the haul outed data with the objective that it can give up significant nears together with prediction models and associations that can assist companies seize their customer better. Searching and analyzing the data can take turns unprocessed data into precious information concerning customer’s needs. On behalf of replica, countenanced through the increasingly weighty situation in customer agitateing, enterprises require data mining technology to analyze the agitateing in order to procure measures to keep precious customers, and reduce customers agitateing to subsidiary economic defeats.

In view of the fact that a customer classification and prediction is the pedestal of the realize of CRM. It is the prerequisite to analyze and forecast customer’s representation of exploitation, and the assertion of personalized marketing services and management.In this study, we propose a an optimized way to uncover promising customer deeds in commercial bank that uses collected information of customers as inputs to make a prediction for the type of loan taken by the promising customers. In meticulous, we opt

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 3, May 2012

Classification Basedon Multiple Association Rules from the various data mining methods because it gives accurate and efficient results compared to remaining associative classification methodologies.

The paper is organized as follows. Section 2 provides a brief review of previous research and the section 3 describes our proposed model, classification basedon Multiple association and section 4 using a bank example describes the propose model.The final section presents the contributions and future researches of the study.

II. LITERATURE REVIEW

Surrounded by bright lifetime, data mining has enlarge widespread absorption and augment standing in the business world. According to the trained and deal imaginary tale, novel companies are using data mining as the base for strategies that assist them get the better of competitors, recognize novel customers and minor costs, as in [4]. Even the current surveys set up that data mining had grown-up in practice and effectiveness, data mining applications in the commercial world have not been widely.

With the lenient of accusation of business fitness, we necessitate to reinforce the study on data mining applications in the business world. During meticulous, data mining is exte nsively used in marketing, risk management and fraudcontrol as in [5]Datamining techniques can assist to achieve the goals by extracting or detecting hidden customer features and behaviours from huge databases.Appropriate datamining tools are high-quality at extracting and identifying helpful informa tion and knowledge starting massive customer databases.

In the perspective of CRM, the datamining can be seen as a trade determined process intended at the detection and reliable use of gainful knowledge from organizational data.CRM rudiments can be supported by diverse datamining models which usually comprise association ,classification,clustering,forecasting ,regression,sequence discovery and visualization[6].

Previous studies have developed heuristic/greedy search techniques for building classifiers, such as decision trees [7], rule learning [8], naıve-Bayes classification [9],and statistical approaches [10]. These techniques induces a represen tative sub set of rules (e.g., a decision tree or a set of rules) from training data sets for quality prediction.Recent studies propose the extraction of a set of high quality assoc iation rules from the training data set which satisfy certain user-spec ified frequency and confidence thresholds.

Effective and efficient classifiers have been built by careful selection of rules, Such a method takes most effective rule(s) from among all the rules mined for classification.Since association rules explore highly confident associations among multiple variables, it may overcome some constraints intro duced by a decision-tree induction method which examines one variable at a time.Extensive performance studies [11] show that associationbased classification may have better accuracy in general.However, this approach may also suffer some weakness.

On one hand, it is not easy to identify the most effective rule at classifying a new case. Some method, such as [12], simply selects a rule with a maximal user-defined measure, such as confidence. As we will see later, such a selection may not always be the right choice in many cases.Such a simple pick may affect the classification accuracy.

On the other hand, a training data set often generates a huge set of rules. It is challenging to store, retrieve, prune,and sort a large number of rules efficiently for classification.Many studies have indicated the inherent nature of a combina torial explosive number of frequent patterns and hence associa tion rules that could be generated when the support threshold is small (i.e., when rare cases are also be included in the consi deration). To achieve high accuracy,a classifier may have to handle a large set of rules,including storing those generated by association mining methods, retrieving the related rules, and pruning and sorting a large number of rules

At times it is rigid for banks to market their services and locate people that might be in necessitate of what they can tender .Loan officers and managers now and then have to come across leads on people probing for loans.A lead is information that is obtained so as to undeviating you to potential customers.

.

III.CLASSIFICATION BASED ON MULTIPLE ASSOCIATION RULES:

The method extends an efficient frequent pattern mining method, FP-growth, constructs a class distribution-associated FP-tree, and mines large database efficiently. Moreover, it applies a CR-tree structure to store and retrieve mined association rules efficiently,and prunes rules effectively based on confidence,correlation and database coverage.

To improve both accuracy and efficiency,CMAR employs a novel data structure, CR-tree, to compactly store and efficiently retrieve a large number of rules for classification. CR-tree is a prefix tree structure to explore the sharing among rules, which achieves substantial compactness. CR-tree itself is also an index structure for rules and serves rule retrieval efficiently.

To speed up the mining of complete set of rules, CMAR adopts a variant of recently developed FP-growth method. FP-growth is much faster than apriori like methods
used in previous association-based classification, such as especially when there exist a huge number of rules, large training data sets, and long pattern rules

CMAR consists of two phases: rule generation and classification.In the first phase, *rule generation*, *CMAR* computes the*complete* set of rules in the form of R:P→c where P is a pattern in the training data set, and is a class label such that sup(R) and conf(R)pass the given support and confidence thresholds, respectively. Furthermore, *CMAR* prunes some rules and only selects a subset of high quality rules for classification.In the second phase, *classification*, for a given data object *CMAR* extracts a subset of rules matching the object and predicts the class label of the object by analyzing this subset of rules

IV.AN EXAMPLE

Suppose commercial bank wants to find out the promising customers in the bank and wants to find out in which loans they are preferring the most,so that according to this information the bank management wants to provide some extra features for that particular loans in order to attract the customers so that the customers will be benifited by which the organization will be benifited.

For this there is a large number of valuable customer information in huge amounts of data accumulated by commercial banks,which is used to uncover promising customers in an efficient,flexible and accurate way and provide decision support.

We wish to uncover the class label of an unknown sample using classification basedon multiple association rules,given the training data as Table1.The data samples are described by the attributes:gender,age,qualification,occupation.The class label attribute ,type of loan has three distinct values(namely,{car,business,housing}).

Table 1 Training data from the customer database

S .no	Gender	Age	Student	Type of loan
1	male	16-25	yes	edu
2	male	30-40	no	car
3	male	16-25	yes	edu
4	male	30-40	no	car
5	male	30-40	no	car
6	female	16-25	no	house

Let the support threshold is 2 and confidence threshold is 30%.CMAR mines class association rules as followes.First, CMAR scans the training data set T once, find the set of attribute values happening at least twice in D . The set is F={Male,no,30-40,16-25,car,yes,edu}and is called frequent item set.All other attribute values, which fail the support threshold,cannot play any role in the class-association rules, and thus can be pruned.

Then, CMAR sorts attribute values in F in support descending order, i.e., F-list=Male,no,30-40,16-25,car,yes,edu Then,CMAR scans the training data set again to construct an FP-tree, as shown in Figure 1.FP-tree is a prefix tree w.r.t. F-list. For each tuple in the training data set, attributes values appearing in F-list are extracted and sorted according to F-list. For example, for the first tuple,(male,16-25,yes) are extracted and inserted in the tree as the left-most branch in the tree. the class label is attached to the last node in the path

Tuples in the training data set share prefixes. For example, the second tuple carries attribute values(male,no,30-40) in F-list and shares a common prefix malewith the first tuple. So, it also shares the male sub-path with the left-most branch. All nodes with same attribute value are linked together as a queue started from the header table

Third, based on F-list, the set of class-association rules

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 3, May 2012

can be divided into subsets without overlap (1)The ones having yes: (2)The ones having 30-40 :(3)The ones having 16-25 but no degree nor govt:(4)The ones having no CMAR finds these subsets one by one.

Fourth, to find the subset of rules having govt. CMAR traverses nodes having attribute value yes and look "upward" to collect a yes-projected database,which containes one tuples:(male,16-25,yes):edu It contains all the tuples having yes.Recursively, in yes-projected database,male,16-25,yes are the frequent attribute values i.e they pass support threshold.We can mine the projected database recursively by constructing FP-trees and projected databases

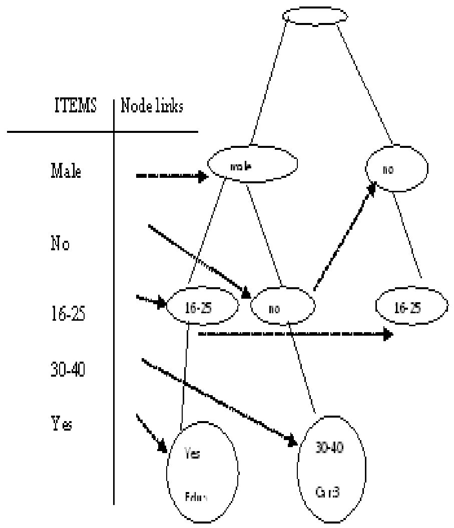

Figure 1. *FP-tree* **in Example 1.**

There are two major differences in the rule mining in CMAR and the standard FP-growth algorithm.On one hand, CMAR finds frequent patterns and generates rules in one step.The difference of CMAR from other associative classification methods is that for every pattern, CMAR maintains the distri bution of various class labels among data objects matching the pattern.This is done without any overhead in the proce dure of counting (conditional) databases.Thus, once a freq uent pattern (i.e., pattern passing support threshold) is found, rules about the pattern can be generated immediately. Therefore, *CMAR* has no separated rule generation step.

Storing rules:

Once a rule is generated, it is stored in a *CR-tree*, which is a prefix tree structure. We demonstrate the general idea of *CR-tree* in the following example. After mining the training data set we are considering the following two association rules

s.no	Rules	support	confidence
1	Male,no,30-40->car	50	100%
2	Male,16-25,yes->edu	33	100%

Table 2. Rules found in a training data set.

A *CR-tree* is built for the set of rules, as shown in Figure 2, while the construction process is explained as follows

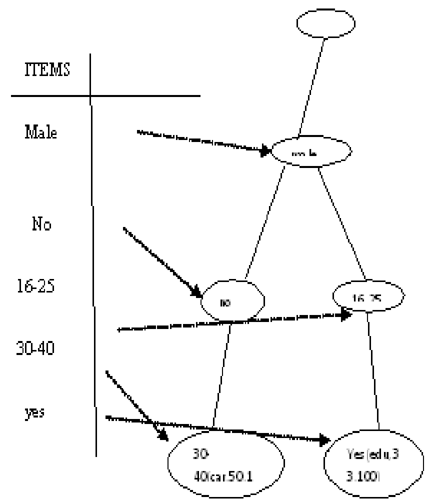

Figure 2. *CR-tree* **in Example 1.**

A *CR-tree* has a *root node*. All the attribute values appearing at the left hand side of rules are sorted according to their frequency, i.e., the most frequently appearing attribute value goes first.

The first rule, Male,no,30-40->car is inserted into the tree as a path from root node. The class label as well as the support and confidence of the rule, denoted as (car,50,100%)

Are registered at the last node in the path, i.e., node for this rule.

The second rule, Male,16-25,yes->edu , shares a *prefix* Male with the first rule. Thus, it is inserted into the tree byextending a new node govt to the path formed by the first rule.Again,the class label, support and confidence of the rule are registered at the last node.

CR-tree is a compact structure. It explores potential sharing among rules and thus can save a lot of space on storing rules. Most experimental results show that, in many cases, about 50-60% of space can be saved using CR-tree. *CR-tree* itself is an index for rules.

Once a *CR-tree* is built, rule retrieval becomes efficient. That facilitates the pruning of rules and using rules for classification dramatically.

Pruning rules:

To make the classification effective and also efficient, we need to prune rules to delete redundant and noisy information According to the facility of rules on classification, a global order of rules is composed. Given two rules R_1 and R_2 . R_1 is said **having higher rank** than R_2, denoted as $R_1 > R_2$, if and only if (1)conf(R_1)>conf(R_2);(2) conf(R_1)=conf(R_2) but sup(R_1)>sup(R_2);or(3)conf(R_1)=conf(R_2), sup(R_1)=sup(R_2) but R_1 should have fewer attributes in its left hand side than R_2 does.

CMAR employs the following methods for rule pruning.

Using general and high-confidence rule to prune more specific and lower confidence ones. This pruning is pursued when the rule is inserted into the *CR-tree*. When a rule is inserted into the tree, retrieval over the tree is triggered to check if the rule can be pruned or it can prune other rules that are already inserted. Most experimental results show that this pruning is effective.

Thus the most promising customers in the above example are the customers who are male^age of 30-40^who is not a student is the first and foremost promising customer and the type of loan taken by him is car loan.And the second promising customer was male^age of 16-25^who is a student and the type of loan taken by him is the education loan.Therefore by using CMAR we uncover the promising customers for the given commercial bank in an efficiently and accurately.

V. CONCLUSION

Data mining meet the expense of the proficiency to analyze accretion competence of reality and/or distinguish obscured patterns in data to revamp unprepared data into luxurious information. This document fundamentally unrelenting on the survey of the optimized way for finding promising customers deeds in Customer Relation Management apprehensive with data mining based on Classification based on Multiple Association Rule algorithm, which have a seek to the optimization of the dealing practice. The cram will aid the production to analyze and foretell promising customer's pattern of consumption, and the foundation of tailored marketing forces and supervision. While the document focuses chiefly on the banking sector, the concerns and functions thrash outed are pertinent to other sectors.

REFERENCES

[1] Jenkins D(1999). "Customer relationship management and the data warehouse" [J]. Call center Solutions, 1892):88-92

[2] Rosset,S,Neumann,E,Eick,U,& Vatnic,N,(2006),Customer Lifetime value models for decision support.Data Mining and knowledge Discovery,7,321-339.

[3]Chung HM and P Gray(1999). "Data mining" [J]. Journal of MIS,16(1):11-13

[4]Davis B(1999). "Data mining transformed" [J]. Informationweek, 751:86-88

[5]Kuykendall L(2000)."The data-mining toolbox"[J]. Credit Card Management,12(6):30-40 .

[6] Kincaid,j,w(2003),Customer Relationship Management:Getting it right,upper saddle river,NJ:Pentice Hall

[7] J. R. Quinlan. C4.5: Programsfor Machine Learning. Morgan Kaufmann, 1993

[8] P. Clark and T. Niblett. The CN2 induction algorithm. Machine Learning, 3:261–283, 1989.

[9] R. Duda and P. Hart. Pattern Classification and Scene Analysis. JohnWiley & Sons, 1973.

[10] T.-S. Lim, W.-Y. Loh, and Y.-S. Shih. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine Learning, 39, 2000.

[12] B. Lent, A. Swami, and J. Widom. Clustering association rules. In ICDE'97, England, April 1997.

[3] G. Dong, X. Zhang, L. Wong, and J. Li. Caep: Classification by aggregating emerging patterns. In DS'99 (LNCS 1721), Japan, Dec. 1999