ANALYSIS OF EXISTING SYSTEM WITH THEIR PROBLEMS AND WEAKNESS

Gajraj Singh*

Dr. P.K. Yadav**

Abstract:-

To study existing system, association rule mining, classification and class association rule mining, how to train and use classifier, how to incorporate association rules in classification etc needs to be studied.

Keyword: - Existing System, Data Mining

* Research scholar, NIMS University, Sobha Nagar Jaipur (Rajastan)

** (Research Guide), Associate Professor, GC, Kanwali, Rewari (Haryana)

1. INTRODUCTION

The idea of using association rule mining in classification rule mining was first introduced in 1997 and [7, 6] it was named as class association rule mining or associative classification. The use of association rules for classification is restricted to problems where the instances can only belong to a discrete number of classes. The reason is that association rule mining is only possible for nominal attributes. However, association rules in their general form cannot be used directly. We have to restrict their definition. The head Y of an arbitrary association rule X → Y is a disjunction of items. Every item which is not present in the rule body may occur in the head of the rule. When we want to use rules for classification, we are interested in rules that are capable of assigning a class membership. Therefore we restrict the head Y of a class association rule X → Y to one item. The attribute of this attribute-value-pair has to be the class attribute. According to this, a class association rule is of the form X → ai where ai is the class attribute and X ⊂ {a1,a2,a3 . . . , ai−1, ai+1, . . . , an}.

The idea of class association rule mining is as follows. We have given a training database where each transaction contains all features of an object in addition to the class label of that object. We can derive the association rules to always have a class label as consequent i.e. the problem states of finding a subset of an association rule set of the X ⇒ C, where X is association of some or all object features and C is class label of that object. Class association rule mining is a special case of association rule mining. And associative classification finds a subset of class association rule set to predict the class of previously unseen data (test data) as accurate as possible with minimum efforts. This subset of class association rule set is called associative classifier or simply a classifier.

2. Methods

PROCESS IN THE EXISTING SYSTEM

The idea of classification using association rule is as further decomposed in two parts.

1. Find Class Association rule mining based on support threshold.
2. Pruning the week rules based on confidence threshold.

We have taking a training database where each transaction contains all features of an object in addition to the class label of that object. As explained above We can derive the association rules to always have a class label as consequent i.e. the problem states of finding a subset of an

association rule set of the X ⇒ C, where X is association of some or all object features and C is class label of that object.

Let we illustrate the class association rule mining with the training data shown in Table 3.1. It consists three attributes X (X1, X2, X3), Y (Y1, Y2, Y3), Z (Z1, Z2, Z3) and two class labels (C1, C2). We assume the min_sup = 30% and min_conf = 70%.

Table 3.1 Training Database

Training Database				
Ti D	X	Y	Z	Class
1	X2	Y2	Z1	C1
2	X1	Y2	Z2	C2
3	X1	Y3	Z3	C2
4	X3	Y1	Z2	C1
5	X1	Y1	Z3	C2
6	X2	Y3	Z1	C1
7	X3	Y3	Z2	C1
8	X1	Y1	Z1	C1
9	X2	Y3	Z1	C1
10	X1	Y3	Z1	C2

As explained above first step of associative classification finds frequent itemset which satisfies min_support and generate class association rules. Table 3.2 shows the class association rule set.

Table 3.2 Class Association Rule

Class Association Rule		Support	Confidence
Antecedent	Consequent		

X1	C2	4/10	4/5
X2	C1	3/10	3/3
Y3	C1	3/10	3/5
Z1	C1	4/10	4/5
X2Z1	C1	3/10	3/3

Now after that in second step find strong class association rule set by pruning the weak rules which satisfy confidence threshold. Table 3.3 shows the strong class association rules along with their confidence.

Table 3.3 Strong Class Association Rule Set

Strong Class Association Rule		Confidence
Antecedent	Consequent	
X1	C2	4/5
X2	C1	3/3
Z1	C1	4/5
X2Z1	C1	3/3

The rules that shown in Table 3.3 also represent a classifier as the rules are sorted according to confidence they hold.

In this paper we explore methods for mining class association rules. The first classifier based on association rules was Classification based Association (CBA) [5] given by Liu et al in 1998. Later, some improved classifiers were given by Li et al. Classification based on Multiple Association Rules (CMAR) [4] in 2001, Yin et al. Classification based on Predictive Association Rules **(**CPAR) [3] in 2003, and Fadi et al. MCAR in 2005. More research is going on to design even improved classifiers. There are some good numbers of associative classification algorithms available now. All claim to offer some benefits, either in accuracy or in

reduction of computation time. Here is a brief description of the major association classification algorithms:

3. RESULT

Classification Based on Association (CBA)

B. Liu, W. Hsu, and Y. Ma proposed a framework, named associative classification, to integrate association rule mining and classification. The integration is done by focusing on mining a special subset of association rules whose consequent parts are restricted to the classification class labels, called "Class Association Rules" (CARs). This algorithm first generates all the association rules and then selects a small set of rules to form the classifiers. When predicting the class label for a coming sample, the best rule is chosen. It consists of two parts, a *rule generator* (called CBA-RG), which is based on algorithm Apriori for finding association rules and a *classifier builder* (called CBA-CB). The key operation of CBA-RG is to find all *ruleitems* that have support above *minsup*. A *ruleitem* is of the form : <*condset*, y> where *condset* is a set of items, *y is* a class label. *Ruleitems* that satisfy *minsup* are called *frequentRuleitems*. Again the confidance of frequent *ruleitem* is greater than *minconf*, we say the rule is *accurate*. The set of *class association rules* (CARs) thus consists of all the rules that are both frequent and accurate. In classifying an unseen case, the first rule that satisfies the case will classify it. If there is no rule that applies to the case, it takes on the default class.

Classification based on Multiple Association Rules (CMAR)

W. Li, J. Han, and J. Pei proposed an algorithm "Classification based on Multiple Association Rules" (CMAR), which utilizes multiple class-association rules for accurate and efficient classification. This method extends an efficient mining algorithm, FP-growth [1], constructs a class distribution- associated FP-trees, and predicts the unseen sample within multiple rules, using weighted χ 2. Liu and Li's approaches generate the complete set of association rules as the first step, and then select a small set of high quality rules for prediction. These two approaches achieve higher accuracy than traditional classification approaches such as C4.5. However, they often generate a very large number of rules in association rule mining, and take efforts to select high quality rules from among them.

Classification based on Predictive Association Rules (CPAR)

Yin et al proposed "Classification based on Predictive Association Rules" (CPAR), which combines the advantages of both associative classification and traditional rule based classification. CPAR adopts a greedy algorithm to generate rules directly from training data, and hence generates and tests more rules than traditional rule-based classifiers to avoid missing important rules, and uses expected accuracy to evaluate each rule and uses the best k rules in prediction to avoid over fitting.

Association rules and classification rules are represented as *if-then* type rules. However, there are some differences between them. Association rules are generally used as descriptive tools, which give the association relationships to the specific application experts, while classification rules are used for predicting the unseen testing data. Therefore, the evaluations of the two types of rules are different. Association rules are typically evaluated by the application experts, while Classification rules are evaluated by the classification accuracy of testing data. In classification rule mining, the most important point to evaluate the quality of rules is the classification accuracy. Therefore, there usually is not expert which could provide the expected results. For classification rule mining, given a specific application, the interestingness measure which can provide the highest classification accuracy would be the appropriate measure.

4. Discussion

ASSOCIATIVE CLASIFICATION AT A GLANCE

Various methods are common to accomplish the class association rule mining process. The algorithmic approach for classification using association rules can be divided into three fundamental parts: association rule mining, pruning and classification. Figure 3.2 provides a graphical overview of the entire process. Mining of association rules is a typical data mining task that works in an unsupervised manner. A major advantage of association rules is that they are theoretically capable of revealing all interesting relationships in a database. But for practical applications the number of mined rules is usually too large to be exploited entirely. This is why the pruning phase is stringent in order to build accurate and compact classifiers.

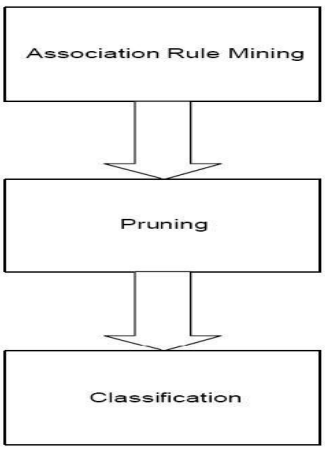

Figure 3.2 The algorithmic steps in classification using association rules

The smaller the number of rules a classifier needs to approximate the target concept satisfactorily. It is able to reveal all interesting relationships, called associations, in a potentially large database. However, how interesting a rule is depends on the problem a user wants to solve. Existing approaches employ different parameters to guide the search for interesting rules. Classification is one of the several tasks of data mining and it is a very important area of learning in the database field of the real world. One of the comprehensible models is the association rule-based classification that combines the advantages of traditional classification and association rule discovery [2]. A class association rule is generally expressed as IF-THEN rule, *i.e.*, IF [*term*1 AND *term*2 AND ...] THEN [class]. Each *term* of the antecedent is a pair of [attribute, value]. The consequent is the result of classification, that is, the class value of the attribute.

Classification using association rules combines association rule mining and classification, and is therefore concerned with finding rules that accurately predict a single target (class) variable. The key strength of association rule mining is that all interesting rules are found. The number of associations present in even moderate sized databases can be, however, very large – usually too large to be applied directly for classification purposes. Therefore, any classification learner using association rules has to perform three **major steps**: Mining a set of potentially accurate rules, evaluating and pruning rules, and classifying future instances using the found rule set.

5. Conclusion

PROBLEM AND WEAKNESS OF EXISTING SYSTEM

Recently, extensive research has been carried out to develop enhanced methods for classification and higher classification accuracy is obtained than traditional classifiers. However, recent studies

show that the associative classifiers suffer from some problems inherited from association rule mining such as the limited support-confidence framework. To address this weakness, several measures have been proposed to evaluate the significance of the rules and to focus only on those that are significant accurately and statistically. On the other hand, a correlation measure can be used to enhance the support-confidence framework for association rules, that is, *A -> B*[*support, confidence, correlation*] is used, where the rule is measured not only by its support and confidence but also by the correlation between *A* and *B*.

6. Reference

[1] Chen Peiyou and Han Jingyao, "Application of association rule in product sale" IEEE, 2010 pp554-557.

[2] Wang Yaqin and Song Yuming, "Classification Model Based on Association Rules in Customs Risk Management Applic ation", International Conference on Intelligent System Design and Engineering Application, 2010, pp 436-439.

[3] X Yin and J. Han, "CPAR: Classification based on Predictive Association Rules", Proceedings of the Third SIAM International Conference on Data Mining (SDM-03), SIAM, San Francisco, CA, United States, 2003, pp. 331-335.

[4] W. Li, J. Han, and J. Pei, "CMAR: Accurate and Efficient Classification based on Multiple Class association Rules", In Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM-01), IEEE Computer Society, San Jose, CA, United States, 2001, pp. 369-376.

[5] B. Liu, W. Hsu, and Y. Ma, "Integrating Classification and Association Rule Mining", Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD-98), AAAI Press, New York City, NY, United States, 1998, pp. 80-86.

[6] K. Ali, S. Manganaris, and R. Srikant, "Partial Classification using Association Rules", Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97), AAAI Press, Newport Beach, CA, United States, August 1997, pp. 115-118.

[7] R. Bayardo, "Brute-force mining of high-confidence classification rules", Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97), AAAI Press, Newport Beach, CA, United States, August 1997, pp. 123-126.