An Associative Classifier based on Positive and Negative Rules

Maria-Luiza Antonie
Department of Computing Science
University of Alberta,Edmonton AB, Canada
luiza@cs.ualberta.ca

Osmar R. Zaïane
Department of Computing Science
University of Alberta,Edmonton AB, Canada
zaiane@cs.ualberta.ca

ABSTRACT

Associative classifiers use association rules to associate attribute values with observed class labels. This model has been recently introduced in the literature and shows good promise. The proposals so far have only concentrated on, and differ only in the way rules are ranked and selected in the model. We propose a new framework that uses different types of association rules, positive and negative. Negative association rules of interest are rules that either associate negations of attribute values to classes or negatively associate attribute values to classes. In this paper we propose a new algorithm to discover at the same time positive and negative association rules. We introduce a new associative classifier that takes advantage of these two types of rules. Moreover, we present a new way to prune irrelevant classification rules using a correlation coefficient without jeopardizing the accuracy of our associative classifier model. Our preliminary results with UCI datasets are very encouraging.

1. INTRODUCTION

Association rule mining is a data mining task that discovers relationships among items in a transactional database. Association rules have been extensively studied in the literature for their usefulness in many application domains such as recommender systems, diagnosis decisions support, telecommunication, intrusion detection, etc. The efficient discovery of such rules has been a major focus in the data mining research community. From the original *apriori* algorithm [1] there have been a remarkable number of variants and improvements of association rule mining algorithms i.e. [7].

Association rule analysis is the task of discovering association rules that occur frequently in a given data set. A typical example of association rule mining application is the market basket analysis. In this process, the behaviour of the customers is studied when buying different products in a shopping store. The discovery of interesting patterns in this collection of data can lead to important marketing and management strategic decisions. For instance, if a customer buys bread, what is the probability that he/she buys milk as well? Depending on the probability of such an association, marketing personnel can develop better planning of the shelf space in the store or can base their discount strategies on such associations/correlations found in the data.

All the traditional association rule mining algorithms were developed to find positive associations between items. By positive associations we refer to associations between items existing in transactions (i.e. items bought). What about associations of the type: “customers that buy Coke *do not* buy Pepsi” or “customers that buy juice *do not* buy bottled water”? In addition to the positive associations, the negative association can provide valuable information, in devising marketing strategies. Interestingly, very few have focused on negative association rules due to the difficulty in discovering these rules.

Throughout this paper we will refer to positive association rules as being rules of the following type: given two items X and Y, a positive association rule is a rule of the form $X \rightarrow Y$ (X and Y exist together frequently and $X \cap Y = \emptyset$). A negative association rule is one of the following: $\neg X \rightarrow Y$ or $X \rightarrow \neg Y$ (where X means existence and $\neg X$ means absence in enough transactions).

Although some researchers pointed out the importance of negative associations, only one group of researchers [18] proposed an algorithm to mine these types of associations. This not only illustrates the novelty of negative association rules, but also the challenge in discovering them.

Recent studies in the data mining community proposed new methods for classification employing association rule mining [9, 10, 2, 3]. These associative classifiers have proven to be powerful and achieve high accuracy. However, they were only discovering and using positive association rules in the classification process. In this paper we experiment and discuss the potential of negative association rules in the categorization task. To the best of our knowledge, there is no other associative classification system that uses both positive and negative association rules.

1.1 Contributions of this paper

The main contributions of this work are as follows:

1. We devise a new algorithm to generate both positive and negative association rules. There are very few papers to discuss and discover negative association rules. Our algorithm differs from those in the sense that it uses a different interestingness measure and it generates classification rules.

DMKD ’04 June 13, 2004, Paris, France
Copyright 2004 ACM ISBN 1-58113-908-X/04/06 ...$5.00.

2. The potential of strong correlated association rules is demonstrated through the pruning and classification results presented in the experimental section. We demonstrate the ability of the negative association rules by their usage in the classification application.

3. A naïve approach in generating positive and negative association rules would create a very large number of rules. Besides the huge number of rules to deal with, many associations may be of low interest. We propose a new algorithm for classification rules generation that is based on the correlation analysis.

4. To avoid adding new parameters that would make the classifier difficult to tune and thus impractical, we introduce an automatic thresholding on the correlation coefficient. We automatically and progressively slide the threshold to find strong correlations.

The remainder of the paper is organized as follows: Section 2 gives an overview of related work in association rule mining and associative classifiers. In Section 3 we introduce our approach for positive and negative rule generation based on correlation measure. Experimental results are described in Section 4 along with the performance of our system compared to known systems. We summarize our research and discuss some future work directions in Section 5.

2. RELATED WORK AND TERMINOLOGY

This section introduces association rules terminology and some related work on negative association rules and associative classification systems.

2.1 Association Rules

Formally, association rules are defined as follows: Let $\mathcal{I} = \{i_1, i_2, ...i_n\}$ be a set of items. Let $\mathcal{D}$ be a set of transactions, where each transaction T is a set of items such that $T \subseteq \mathcal{I}$. Each transaction is associated with a unique identifier TID. A transaction T is said to contain X, a set of items in $\mathcal{I}$, if $X \subseteq T$. An *association rule* is an implication of the form "$X \Rightarrow Y$", where $X \subseteq \mathcal{I}, Y \subseteq \mathcal{I}$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ has a *support* s in the transaction set $\mathcal{D}$ if s% of the transactions in $\mathcal{D}$ contain $X \cup Y$. In other words, the support of the rule is the probability that X and Y hold together among all the possible presented cases. It is said that the rule $X \Rightarrow Y$ holds in the transaction set $\mathcal{D}$ with *confidence* c if c% of transactions in $\mathcal{D}$ that contain X also contain Y. In other words, the confidence of the rule is the conditional probability that the consequent Y is true under the condition of the antecedent X. The problem of discovering all association rules from a set of transactions $\mathcal{D}$ consists of generating the rules that have a *support* and *confidence* greater than given thresholds. These rules are called *strong rules*.

2.1.1 Negative Association Rules

Example 1. Suppose we have an example from the market basket data. In this example we want to study the purchase of organic versus non-organic vegetables in a grocery store. Table 1 gives us the data collected from 100 baskets in the store. In Table 1 "organic" means the basket contains organic vegetables and "¬ organic" means the basket does not contain organic vegetables. The same applies for non-organic. On this data, let us find the positive association

Table 1: Example 1 Data

	organic	¬organic	$\sum_{row}$
non-organic	20	60	80
¬non-organic	20	0	20
$\sum_{col}$	40	60	100

rules in the "support-confidence" framework. The association rule "non-organic → organic" has 20% support and 25% confidence (supp(non-organic ∧ organic)/supp(non-organic)). The association rule "organic → non-organic" has 20% support and 50% confidence The support is considered fairly high for both rules. Although we may reject the first rule on the confidence basis, the second rule seems a valid rule and may be considered in the data analysis. Now, let us compute the statistical correlation between the *non-organic* and *organic* items. A more elaborated discussion on the correlation measure is given in Section 3.1. The correlation coefficient between these two items is -0.61. This means that the two items are negatively correlated. This measure sheds a new light on the data analysis on these specific items. The rule "organic → non-organic" is misleading. The correlation brings new information that can help in devising better marketing strategies.

The example above illustrates some weaknesses in the "support-confidence" framework and the need for the discovery of more interesting rules. The interestingness of an association rule can be defined in terms of the measure associated with it, as well as in the form an association can be found.

Brin *et al.* [5] mentioned for the first time the notion of negative relationships in the literature. Their model is chi-square based. They use the statistical test to verify the independence between two variables. To determine the nature (positive or negative) of the relationship, a correlation metric was used. In [13] the authors present a new idea to mine strong negative rules. They combine positive frequent itemsets with domain knowledge in the form of a taxonomy to mine negative associations. However, their algorithm is hard to generalize since it is domain dependant and requires a predefined taxonomy. A similar approach is described in [19]. Wu *et al.* [18] derived a new algorithm for generating both positive and negative association rules. They add on top of the support-confidence framework another measure called *mininterest* (the argument is that a rule $A \rightarrow B$ is of interest only if $supp(A \cup B) - supp(A)supp(B) \geq mininterest$). Although they introduce the "mininterest" parameter, the authors do not discuss how to set it and what would be the impact on the results when changing this parameter. In [15] the authors use only negative associations of the type $X \rightarrow \neg Y$ to substitute items in market basket analysis.

We define as *generalized negative association rule*, a rule that contains a negation of an item (i.e a rule for which its antecedent or its consequent can be formed by a conjunction of presence or absence of terms). An example for such association would be as follows: $A \wedge \neg B \wedge \neg C \wedge D \rightarrow E \wedge \neg F$. To the best of our knowledge there is no algorithm that can determine such type of associations. Deriving such an algorithm is not an easy problem, since it is well known that the itemset generation in the association rule mining process is an expensive one. It would be necessary not only to consider all items in a transaction, but also all possible items absent from the transaction. There could be a considerable exponential growth in the candidate generation phase. This is especially true in datasets with highly correlated attributes. That is why it is not feasible to extend the attribute space by adding the negated attributes and use the existing association rule algorithms. Although we are currently investigating this problem, in this paper we generate and use in the classification process a subset of the generalized negative association rules. We refer to them as *confined negative association rules*. A confined negative association rule is one of the follows: $\neg X \rightarrow Y$ or $X \rightarrow \neg Y$, where the entire antecedent or consequent must be a conjunction of negated attributes or a conjunction of non-negated attributes.

2.2 Associative Classifiers

The main idea of the classification task is to discover interesting patterns in training sets of data to build a classifier model that is used later in the classification process. Classification has multiple applications and has already been applied in many areas such as text categorization, medical analysis, space exploration, etc. Although classification has a long history and there exist many popular techniques for classification, there is still room for improvement. Besides decision trees [11], Bayesian classifier [17], neural networks [12], support vector machines [16], the classification based on association rule mining started attracting attention in the past few years [10, 9, 2, 3].

The main steps in building an associative classifier when a data set is given are the following:

1. *Generating the set of association rules from the training set.* In this phase association rules of the form $set\ of features \Rightarrow class\_label$ are discovered by using a mining algorithm.

2. *Pruning the set of discovered rules.* In the previous phase a large set of association rules can be generated especially when low support is given. That is why pruning techniques are a challenging task to discover the best set of rules that can cover the training set. This phase is employed to weed out those rules that may introduce errors or are overfitting in the classification stage.

3. *Classification phase.* At this level a system that can make a prediction for a new object is built. The task here is how to rank and make use of the set of rules from the previous phase to give a good prediction.

The existing associative classifiers mine the training set in an apriori-like fashion or some use FP-tree as association rule discovery algorithm. Essentially, although the mining methods are slightly different, all approaches generate the same kind of association rules for classification since they are discovered in the "support-confidence" framework. Some of the algorithms base their classification decisions on the first matching rule or on a set of matching rules. There are different ways to combine the rules that could classify a new object. Some algorithms average the confidences for each category, while others compute a weighted chi-square for each category.

Example 2. Let us consider the following example: we have two classes ($C_1 \quad and \quad C_2$) and three attribute values (X, Y and Z) that are found strongly associated with these classes. In the support-confidence framework there are found positive associations between the attribute values and classes. Let us consider that the following classification rules are discovered in this context: $X \rightarrow C_1$, $Y \rightarrow C_1$, $X \rightarrow C_2$, $Y \rightarrow C_2$ and $Z \rightarrow C_2$. Without considering the confidences and any scoring scheme for now, let us assume that a new tuple, with attribute values X and Y, is presented for classification. Since both X and Y are associated with both classes a decision has to be made between these two.

Let us consider that positive and negative association rules are discovered. In this case, the following classification rules are generated: $X \rightarrow C_1$, $Y \rightarrow C_1$, $\neg Z \rightarrow C_1$, $X \rightarrow C_2$, $Y \rightarrow C_2$ and $Z \rightarrow C_2$. In this situation when a new tuple, with attribute values X and Y, is presented for classification, an easier decision can be made since we have an extra rule ($\neg Z \rightarrow C_1$) to reinforce the decision in favour of C_1.

The above example points out the potential and the usefulness of negative association rules in the classification process.

Table 2: 2x2 Contingency table for binary variables

	Y	$\neg Y$	$\sum_{row}$
X	f_{11}	f_{10}	f_{1+}
$\neg X$	f_{01}	f_{00}	f_{0+}
$\sum_{col}$	f_{+1}	f_{+0}	N

3. DISCOVERING POSITIVE AND NEGATIVE ASSOCIATION RULES

The most common framework in the association rules generation is the "support-confidence" one. Although these two parameters allow the pruning of many associations that are discovered in data, there are cases when many uninteresting rules may be produced. In this paper we consider another framework that adds to the support-confidence some measures based on correlation analysis. Next section introduces the correlation coefficient, which is the measure added to the support-confidence framework.

3.1 Correlation coefficient

Correlation coefficient measures the strength of the linear relationship between a pair of two variables. It is discussed in the context of association patterns in [14]. For two variables X and Y, the correlation coefficient is given by the following formula: $\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$. In this equation, $Cov(X, Y)$ represents the covariance of the two variables and σ_X stands for the standard deviation. The range of values for ρ is between -1 and +1. If the two variables are independent then ρ equals 0. When $\rho = +1$ the variables considered are perfectly positive correlated. Similarly, When $\rho = -1$ the variables considered are perfectly negative correlated. A positive correlation is evidence of a general tendency that when the value of X increases/decreases so does the value of Y. A negative correlation occurs when for the increase/decrease of X value we discover a decrease/increase in the value of Y.

Let X and Y be two binary variables. Table 2 summarizes the information about X and Y variables in a dataset in a 2x2 contingency table. The cells of this table represent the possible combinations of X and Y and give the frequency associated with each combination. N is the size of the dataset considered.

Given the values in the contingency table for binary variables, Pearson introduced the ϕ *correlation coefficient* which is given in the equation 1:

$$\phi = \frac{f_{11}f_{00} - f_{10}f_{01}}{\sqrt{f_{+0}f_{+1}f_{1+}f_{0+}}} \tag{1}$$

We can transform this equation by replacing f_{00}, f_{01}, f_{10}, f_{0+} and f_{+0} as follows:

$$\phi = \frac{Nf11 - f_{1+} * f_{f+1}}{\sqrt{f_{1+}(N - f_{1+})f_{+1}(N - f_{+1})}} \tag{2}$$

The measure given in Equation 2 is the measure that we use in the association rule generation.

Cohen [6] discusses about the correlation coefficient and its strength. In his book he considers that a correlation of 0.5 is large, 0.3 is moderate, and 0.1 is small. The interpretation of this statement is that anything greater than 0.5 is large, 0.5-0.3 is moderate, 0.3-0.1 is small, and anything smaller than 0.1 is insubstantial, trivial, or otherwise not worth worrying about as described in [8].

We use these arguments to introduce an automatic progressive thresholding process. We start by setting our correlation threshold to 0.5. If no strong correlated rules are found the threshold slides progressively to 0.4 and 0.3 until some rules are found with moderate correlations. This progressive process eliminates the need for manually adjusted thresholds. It is well known that the more parameters a user is given, the more difficult it becomes to tune the system. Association rule mining is certainly not immune to this phenomenon.

3.2 Our Algorithm

In this section we introduce the algorithm used to mine positive and negative association rules (**Algorithm** Classification Rule Generation - All Itemsets). The algorithm generates a set of rules which is the union of PCR (Positive Classification Rules) and NCR (Negative Classification Rules). This set of rules is later used in the classification stage. Since in this approach we are interested in generating only classification rules, the algorithm creates only rules of the form *set_of_features* $\rightarrow$ *class_label*. We denote by C a set that keeps all class labels existing in the data set. The algorithm is an apriori-like process. It generates first the set of frequent 1-itemsets. Once the 1-frequent itemsets is generated the candidate sets C_2 to C_n are found as a join between F_{k-1} and F_1. Those candidates that exceed minimum support threshold are added to the corresponding frequent set. For each candidate the PONERG function is called to generate the positive and negative association rules which is described below.

The PONERG (POsitive and NEgative Rule Generation) function generates the positive and negative rules based on the item correlation with a class label. This function takes as input an itemset and the set of class labels. The correlation coefficient between the item and the class label is computed as discussed in Section 3.1. If the correlation in absolute value is greater than the correlation threshold given, than the classification rule is of interest. If the correlation is positive, a positive association rule is discovered. When the correlation is negative, negative rules are generated. Given two items X and Y, a positive association rule is a rule of the form $X \rightarrow Y$. A negative association rule is one of the follows: $\neg X \rightarrow Y$ or $X \rightarrow \neg Y$. Once the rules are generated, they are added to PCR or NCR if their confidence exceeds the minimum confidence threshold.

The values for the correlation coefficient are chosen based on the values discussed in the previous section. First, we consider as correlation threshold the value 0.5, since we want to discover strong correlations. However, there were two datasets where no strong correlations were discovered between attribute values and class labels. For these cases, the threshold was lowered to the 0.3 value to discover moderate correlations.

Algorithm Classification Rule Generation - All Itemsets
Input Transactional Database(TD); Support(minsupp); Correlation(corr); Minimum Confidence(minconf)
Output Positive Classification Rules (PCR) and Negative Classification Rules (NCR)
Method:

```
PCR ← ∅ /*PCR stands for Positive Classification Rules*/
NCR ← ∅ /*NCR stands for Negative Classification Rules*/
C ← {all classes in TD} /*keeps all the class labels*/
scan the database and find 1-frequent itemset (F1)
foreach i ∈ F1 {
    {PCR, NCR} ← PONERG(i,C)
}
for (k = 2, F_{k-1} ≠ ∅, k++){
    C_k= F_{k-1} ⋈ F1
    foreach i ∈ C_k {
        s=support(TD,i) /*support of item i is computed*/
        if s≥minsupp
            F_k ← F_k ∪ {i} /*item i is added to F_k*/
        {PCR, NCR} ← PONERG(i,C)
    }
}
return PCR and NCR
```

Algorithm Positive and Negative Classification Rule Generation (PONERG)
Input Itemset(i); Set of Classes(C); Correlation(corr); Minimum Confidence(minconf)
Output PCR and NCR
Method:

```
foreach c ∈ C { /*for each class in C*/
r=corr(i,c) /*compute correlation btw i and c*/
if r>corr{
    pr = i → c // generate positive rule pr
    if conf(i → c) ≥ minconf
        PCR ← PCR ∪ {i → c}
}
else
    if r<-corr{
        nr1 = ¬i → c // generate negative rule 1
        nr2 = i → ¬c // generate negative rule 2
        // add negative rules to NCR
        if conf(¬i → c) ≥ minconf
            NCR ← NCR ∪ {¬i → c}
        if conf(i → ¬c) ≥ minconf
            NCR ← NCR ∪ {i → ¬c}
    }
}
```

(19) return PCR and NCR

3.3 Associative Classification

The set of rules that were generated as discussed in the previous section represent the actual classifier. This categorizer is used to predict to which classes new objects are attached. Given a new object, the classification process searches in this set of rules for those classes that are relevant to the object presented for classification. The set of positive and negative rules discovered as explained in the previous section are ordered by confidence and support. This sorted set of rules represents the associative classifier ARC-PAN(Association Rule Classification with Positive And Negative). This subsection discusses the approach for labelling new objects based on the set of association rules that forms the classifier.

Algorithm Classification of a new object

Input A new object to be classified o; The associative classifier (ARC-PAN);The confidence margin τ;

Output Category attached to the new object

Method:

(1) $S \leftarrow \emptyset$ /*set of rules that match o*/
(2) **foreach** r in ARC-PAN /*the sorted set of rules*/
(3)   **if** $(r \subset o)$ { count++ }
(4) $S \leftarrow S \cup r$
(4)   **if** (count == 1)
(5)   fr.conf ← r.conf /*keep the first rule confidence*/
(6) $S \leftarrow S \cup r$
(7)   **else if** (r.conf > fr.conf-τ)
(8) $S \leftarrow S \cup r$
(9)   **else** break
(10) divide S in subsets by category: $S_1, S_2...S_n$
(11) **foreach** subset $S_1, S_2...S_n$
(12)   sum/substract the confidences of rules and divide by the number of rules in S_k
(12) $score_i = \frac{\sum r.conf}{\#rules}$
(13) put the new object in the class that has the highest confidence score
(13) $o \leftarrow C_i$, with $score_i = max\{score_1..score_n\}$

In the above algorithm (Classification of a new object), a set of applicable rules is selected in the lines 1-8. The set of applicable rules is selected within a confidence margin. The interval of selected rules is between the confidence of the first rule and this confidence minus the confidence margin as checked in line 7. The prediction process is starting at line 10. The applicable set of rules is divided according to the classes in line 10. In lines 11-12 the groups are ordered according to the average confidence per class. In line 13 the classification is made by assigning to the new object the class that has the highest score.

The association rules of the type $X \rightarrow C$ and $\neg X \rightarrow C$ can be treated in the same way. Both of them have a confidence attached and they have an association with the class label. These types of rules can be considered together and their confidence can be added to the C class total. However, the rules of the type $X \rightarrow \neg C$ have to be treated differently. Currently, we chose to substract their confidences from the total confidence of their corresponding class.We are currently investigating other methods to score these kind of rules.

Table 3: Classification Results (Error rates)

Datasets	c4.5	CBA		ARC-PAN		
	rules	w/o prun	prun	rules+	rules+-	rules_all
breast	3.9	4.2	4.2	5.5	4.8	**3.8**
diabetes	27.6	24.7	25.3	**23.3**	25.4	25.1
heart	18.9	18.5	18.5	16.3	17.0	**16.2**
iris	**5.5**	7.1	7.1	6.6	6.6	6.0
led7	**26.5**	27.8	27.8	28.7	28.7	28.9
pima	27.5	27.4	27.6	27.4	27.1	**26.9**

4. EXPERIMENTAL RESULTS

We tested our algorithm on some datasets from UCI ML Repository [4]. For lack of space we report results for only some of the UCI datasets. On each dataset we performed C4.5's shuffle utility [11] for shuffling the datasets. The shuffle ensures a more accurate classification as observations become randomized. A 10-fold cross validation was performed on each dataset and the results are given as average of the errors obtained for each fold. In addition, to have a fair comparison with the other algorithms that we wanted to compare, we used the same discretization method for continuous attributes as in [10]. The parameters for C4.5 were set to their default values. For all three association rule based methods the minimum support was set to 1% and the minimum confidence to 50%. In our approach the confidence margin was set at 10%.

The experimental results are shown in Table 3. The average error results for CBA and C4.5 are taken from [10]. The last three columns give the results for our classification method. **Column 5** presents the results when only positive rules are considered for classification. **Column 6** shows the results when positive rules and negative rules of the form $\neg X \rightarrow C$. **Column 7** lists the results when all rules as described in the previous sections are considered for categorization.

As presented in Table 3 the results for the classification with classifier based on the positive and negative rules (ARC-PAN) seem encouraging. When all types of rules are used the classification accuracy increases on three datasets when compared with the state-of-the-art classifier C4.5 and with the CBA [10]. The first column under ARC-PAN shows that the classification accuracy can be improved as well with only the generation of positive association rules that are strongly correlated. We ran classification with the negative rules only as well, but the results decreased in this case. The results are not presented in the table.

Table 4 shows the drastic reduction in rule number when the correlation measure is used to derive interesting rules. Under the second column in the table, the approximate number of rules derived in the "support-confidence" framework are given. As it can be seen from the table, when compared to the forth column (where the rules discovered in the correlation framework), there is a large decrease in the rule number from one framework to the other. Moreover, as observed from the error results presented in the third and fifth columns, the error rate remains in the same range, or even decreases in some cases.

A small number of classification rules is very desirable. When a small set of strong classification rules is presented, the classification phase is faster, which can be important for some applications. Another advantage is that a small set becomes human readable. It is realistically feasible to read, edit and augment hundreds of rules, but thousands of rules

Table 4: Comparison in number of rules

Datasets	strong rules		correlated rules	
	#rules	error	#rules	error
breast	17000	5.0	1000	5.5
diabetes	4000	21.8	40	23.3
heart	200000	24.7	80	16.3
iris	140	7.3	60	6.6
led7	4000	34.3	500	28.7
pima	4000	22.0	50	27.4

is impractical. Because of the transparency of the associative classifier, manually updating some rules is favorable and practical in many applications.

5. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this paper we introduced a new algorithm to generate positive and negative associations discovered in transactional data. The interestingness measure that our algorithm relies on is the correlation coefficient. We demonstrated the potential of strong positive and negative correlated rules in the classification context. The results of the classification show that a much smaller set of positive and negative association rules can perform similar or outperform existing categorization systems.

Since both the negative association rule mining and associative classification are two relatively new domains of research, there are still unsolved problems. One of them would be to derive an efficient and effective algorithm for generating *generalized negative association rules*. It is well known that the itemset generation in the association rule mining process is an expensive one, since there may be an exponential growth in the candidate generation phase when absent items are considered. It is simply not feasible to extend the attribute space by adding the negated attributes and use the existing association rule algorithms. The search space would be excessively large.

Another interesting problem to investigate is the scoring scheme in the classification phase of the associative classifier. Although the confidence averaging seems to work well in many cases, there are some examples where the classification is inaccurate while good rules exist. Improving could be possible by selecting the right rules.

Acknowledgements: This work was partially supported by Alberta Ingenuity Fund, iCORE and NSERC Canada.

6. REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data*, pages 207–216, May 1993.

[2] M.-L. Antonie and O. R. Zaïane. Text document categorization by term association. In *In Proc. of the IEEE 2002 International Conference on Data Mining*, pages 19–26, Maebashi City, Japan, 2002.

[3] E. Baralis and P. Garza. A lazy approach to pruning classification rules. In *In Proc. of the IEEE 2002 International Conference on Data Mining*, pages 35–42, Maebashi City, Japan, 2002.

[4] C. Blake and C. Merz. Uci repository of machine learning databases. http://www.ics.uci.edu/∼mlearn/MLRepository.html.

[5] S. Brin, R. Motwani, and C. Silverstein. Beyond market basket: Generalizing association rules to correlations. In *Proc. 1997 ACM-SIGMOD Int. Conf. Management of Data*, pages 265–276, May 1997.

[6] J. Cohen. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum, New Jersey, 1988.

[7] B. Goethals and M. J. Zaki, editors. *FIMI'03: Workshop on Frequent Itemset Mining Implementations*, volume 90 of *CEUR Workshop Proceedings series*, November 2003. http://CEUR-WS.org/Vol-90/.

[8] W. Hopkins. A new view of statistics. http://www.sportsci.org/resource/stats/, 2002.

[9] W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on multiple class-association rules. In *IEEE International Conference on Data Mining (ICDM'01)*, San Jose, California, Novermber 29-December 2 2001.

[10] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In *ACM Int. Conf. on Knowledge Discovery and Data Mining (SIGKDD'98)*, pages 80–86, New York City, NY, August 1998.

[11] J. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

[12] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, et al., editors, *Parallel Distributed Processing: Volume 1: Foundations*, pages 318–362. MIT Press, Cambridge, 1987.

[13] A. Savasere, E. Omiecinski, and S. Navathe. Mining for strong negative associations in a large database of customer transactions. In *In Proceedings of the ICDE 1998 International Conference on Data Engineering*, pages 494–502, 1998.

[14] P.-N. Tan and V. Kumar. Interestingness measures for association patterns: A perspective. In *In Proc. of Workshop on Postprocessing in Machine Learning and Data Mining in conjunction with Seventh ACM SIGKDD*, 2000.

[15] W.-G. Teng, M.-J. Hsieh, and M.-S. Chen. On the mining of substitution rules for statistically dependent items. In *In Proc. of the IEEE 2002 International Conference on Data Mining*, pages 442–449, Maebashi City, Japan, 2002.

[16] V. N. Vapnik. *The nature of statistical learning theory*. Springer Verlag, Heidelberg, DE, 1995.

[17] S. Weiss and C. Kulikowski. *Computer Systems That Learn. Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*. Morgan Kaufmann, San Francisco, 1991.

[18] X. Wu, C. Zhang, and S. Zhang. Mining both positive and negative association rules. In *In Proceedings of 19th International Conference on Machine Learning*, pages 658–665, Sydney, Australia, 2002.

[19] X. Yuan, B. P. Buckles, Z. Yuan, and J. Zhang. Mining negative association rules. In *In Proceedings of the Seventh International Symposium on Computers and Communications*, pages 623–629, Italy, 2002.