Using Significant, Positively Associated and Relatively Class Correlated Rules for Associative Classification of Imbalanced Datasets

Florian Verhein
The School of Information Technologies
University of Sydney
fverhein@it.usyd.edu.au

Sanjay Chawla
The School of Information Technologies
University of Sydney
chawla@it.usyd.edu.au

Abstract

The application of association rule mining to classification has led to a new family of classifiers which are often referred to as “Associative Classifiers (ACs)”.

An advantage of ACs is that they are rule-based and thus lend themselves to an easier interpretation. Rule-based classifiers can play a very important role in applications such as medical diagnosis and fraud detection where “imbalanced data sets” are the norm and not the exception.

The focus of this paper is to extend and modify ACs for classification on imbalanced data sets using only statistical techniques. We combine the use of statistically significant rules with a new measure, the Class Correlation Ratio (CCR), to build an AC which we call SPARCCC. Experiments show that in terms of classification quality, SPARCCC performs comparably on balanced datasets and outperforms other AC techniques on imbalanced data sets. It also has a significantly smaller rule base and is much more computationally efficient.

1 Introduction

Since the introduction of CBA [8] many variations on Associative Classifiers (ACs) have been proposed in the literature [7, 2, 17, 15, 4, 5, 3, 12]. Most of the ACs are based on rules discovered using the support-confidence paradigm and the classifier itself is a collection of rules ranked using confidence or variation thereof. In many application domains, the data sets are imbalanced, i.e., the proportion of samples from one class is much smaller than the other class(es). Additionally, the smaller class is the class of interest. Unfortunately, the support-confidence framework does not perform well in such cases.

Recently Webb [16] has shown the value of using statistically significant rules and has demonstrated that many of the rules mined using support-confidence are spurious and are irregularities in the data rather than properties of the underlying population. We believe that the same holds true from rules used for classification. It is also well known that confidence has non-intuitive properties in imbalanced data sets. For example, high confidence rules can also be negatively correlated. In this paper we combine statistically significant rules with a new measure, the Class Correlation Ratio (CCR), which leads to a better classifier. Furthermore, our method does not use the support-confidence paradigm.

We make the following contributions:

- We propose the Class Correlation Ratio (CCR), which measures the relative class correlation of a rule. A high CCR is desirable because it means the rule is more positively correlated with the class it predicts than the alternative(s). CCR also forms the basis of an effective rule ranking method that can also be employed in other algorithms.
- We prove that confidence and support are biased toward the majority class in imbalanced datasets in the context of CCR.
- We propose an Associative Classifier that is based purely on statistical techniques. We call the method Significant, Positively Associated and Relatively Class Correlated Classification (SPARCCC) because we use only rules that are statistically significant and positively associated, and where the antecedent is more correlated with the class it predicts than with the other class(es). The classifier is parameter-free, in the sense that it does not use thresholds – except standard levels of significance – to prune rules.

The remainder of this paper is organised as follows: We first give a brief summary of ACs. Section 2 describes the class correlation ratio and the significance test we use. Section 3 proves that confidence (and support) is biased against the minority class under CCR. Section 4 describes our technique. Section 5 contains experimental results and we survey related work in Section 6.

In the Associative Classification problem, we assume a discrete dataset \( D \) with attributes \( A = \{a_1, a_2, ..., a_{|A|}\} \), one of which is the class attribute \( a_c \). In every instance
d ∈ D, each attribute a_i ∈ A takes one of a finite number of possible values V_i = {v_{i1}, ..., v_{i|V_i|}}. The Associative Classification Rule Mining task is to find interesting rules X → y where X is a set of legal (an attribute cannot occur more than once) attribute-value pairs and y is one of the class attribute-value pairs. By interesting, we mean rules that, in conjunction with other mined rules, are likely to perform well for classification of unseen data. The support of a set of attribute-value pairs X is \( \text{sup}(X) = |\{d_1 : X \subseteq d_1 \wedge d_1 \in D\}| \). The support of \( X \rightarrow y \) is \( \text{sup}(X \rightarrow y) = \text{sup}(X \cup y) \) and its confidence is \( \text{conf}(X \rightarrow y) = \text{sup}(X \rightarrow y)/\text{sup}(X) \).

2 Significance and Class Correlation Ratio

There are strong arguments for mining statistically significant rules [16]. These also hold true when the rules are used for classification, as we would like to make a decision based on significant evidence. We are interested in rules \( X \rightarrow y \) that are statistically significant in the positively associated direction. Toward that end, we use Fisher's Exact Test (FET) on contingency tables of the form of Figure 1. FET is an exact test (permutation test). Given a table \([a, b, c, d]\), FET will find the probability (p-value) of obtaining the given table or a table where \( X \) and \( y \) are more positively associated under the null hypothesis that \( \langle X, \neg X \rangle \) and \( \langle \neg y, y \rangle \) are independent, and that the margin sums are fixed. The p-value is given by:

\[
p([a, b; c, d]) = \sum_{i=0}^{\min(b,c)} \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!(a+i)!(b-i)!(c-i)!(d+i)!}
\]

We only use rules whose p-values are below the level of significance desired as they are statistically significant in the positively associated direction. We also prune the search space using significance as outlined in Section 4.2.

Correlation also forms a very important component of our technique. We are interested in rules \( X \rightarrow y \) where \( X \) is more positively correlated with \( y \) than it is with \( \neg y \). In this paper we use the following definition of correlation:\(^1\)

\[
\text{corr}(X \rightarrow y) = \frac{\text{sup}(X \cup y) \cdot |D|}{\text{sup}(X) \cdot \text{sup}(y)} = \frac{a \cdot n}{(a+c) \cdot (a+b)}
\]

\( X \) and \( y \) are positively (negatively) correlated if \( \text{corr}(X \rightarrow y) > 1 \) (1), and independent otherwise. Note that \( \text{corr}(X \rightarrow y) = I(X, y) \), where \( I(X, y) \) is the Interest Factor [11]. This measure has downsides when used by itself. It is clear to see that increasing the size of the dataset by increasing \( d \) (refer to Figure 1) will increase the correlation between \( X \) and \( y \) even though it is actually increasing the association between \( \neg X \) and \( \neg y \). The reverse holds for decreasing \( d \). For example, consider the table \( T_1 = [100, 20; 20, 10] \) where \( X \) and \( y \) have a strong association but \( \text{corr}(X \rightarrow y) = 1.04 \) (almost independent!). If we increase \( d \) to get \( T_2 = [100, 20; 20, 200] \) then clearly \( \neg X \) and \( \neg y \) are strongly associated, but \( \text{corr}(\neg X \rightarrow \neg y) = 1.4 \) while now \( \text{corr}(X \rightarrow y) = 2.36! \) This is clearly undesirable. This problem arises only in imbalanced datasets however - note how changing \( d \) changes the class distribution. We therefore do not search for positively correlated rules using it. When we speak of a rule being positively associated or correlated, we mean by using the one sided test of significance described above. The FET does not have this downside because of the constant margin sum restriction. Indeed, \( p(T_1) = 0.041 \) (significant at the 0.05 level) and \( p(T_2) = 1.07 \cdot 10^{-44} \) (highly significant).

We measure how correlated \( X \) is with \( y \) compared to \( \neg y \) using what we call the Class Correlation Ratio (CCR):

\[
\text{CCR}(X \rightarrow y) = \frac{\text{corr}(X \rightarrow y)}{\text{corr}(X \rightarrow \neg y)} = \frac{a \cdot (b+d)}{b \cdot (a+c)}
\]

This measures how much more positively the antecedent is correlated with the class it predicts, relative to the alternative class(es). This avoids the downsides of using an absolute correlation measure – indeed, terms cancel out. It also makes a lot of sense intuitively – you would not want to use a rule that is more correlated with classes other than that it predicts! Returning to the example, \( \text{CCR}(\cdot) = 1.25 \) for \( T_1 \) and \( \text{CCR}(\cdot) = 9.17 \) for \( T_2 \). This also says that \( X \rightarrow y \) is a better rule under \( T_2 \) than under \( T_1 \). This is true – it is much more discriminative because under \( T_1, y \) is already the majority class and therefore the rule does not provide much additional information. In fact, the Information Gain of using \( X \rightarrow y \) over \( \emptyset \rightarrow y \) is only 0.072 bits under \( T_1 \) but is 0.215 bits under \( T_2 \). Recall also that the rule was much more significant under \( T_2 \). We only use rules with \( \text{CCR} > 1 \), so that no rules are used that are more positively associated with the classes they do not predict. Furthermore, we use CCR in our Strength Score.

3 Relative Correlation Bias of Confidence (and Support) on Imbalanced Datasets

Confidence is widely used as a measure of strength of a classification rule \( X \rightarrow y \) because it is an estimate (the dataset is a sample) of the probability that, given the attribute-value pairs in \( X \) appear in an instance \( d \) generated

<table>
<thead>
<tr>
<th>y</th>
<th>a</th>
<th>b</th>
<th>a+b</th>
</tr>
</thead>
<tbody>
<tr>
<td>¬y</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>∑cols</td>
<td>a+c</td>
<td>b+d</td>
<td>n = a+b+c+d</td>
</tr>
</tbody>
</table>

Figure 1. 2 × 2 Contingency Table for \( X \rightarrow y \). We will often use the notation \([a, b; c, d]\).
by the underlying process, the instance will have the class label $y$. That is, $conf(X \rightarrow y) \sim P(y \in d | X \in d)$. The confidence of a significant rule (it does not make sense to use insignificant rules, and their confidences are unlikely to mean anything) is therefore a useful measure of the rule strength in classification—

\textit{but only in balanced datasets.} We show that confidence (and support, while we’re at it) are biased toward the majority class under the $CCR$. In our previous example, note that $conf(X \rightarrow y) = 0.83$ in both $T_1$ and $T_2$, despite the rule being clearly better in $T_2$.

\textbf{Lemma 1} Confidence (and support) are biased toward the majority class under the Class Correlation Ratio. Specifically (statements in parentheses are defined in Figure 2):

1. If $X \rightarrow y$ is more positively correlated than $X \rightarrow \neg y$ but has a lower confidence (support), then $y$ must be the majority class: $(B \land C' \implies A)$.  
2. If $X \rightarrow y$ is more positively correlated and more confident (frequent) than $X \rightarrow \neg y$, we cannot say anything about whether $y$ is the minority or majority class: $(B \land C \iff A$ and $B \land C \iff \neg A)$.  
3. If $y$ is the minority class and $X \rightarrow y$ is more confident (frequent) than $X \rightarrow \neg y$, then it is also more positively correlated: $(A \land C \implies B)$.  
4. If $y$ is the minority class and $X \rightarrow y$ is less confident (frequent) than $X \rightarrow \neg y$, there is no relationship between the correlation of the rules: $(A \land C' \iff B$ and $A \land C' \iff \neg B)$.  
5. If $y$ is the minority class and $X \rightarrow y$ is less positively correlated than $X \rightarrow \neg y$, it is also less confident (frequent): $(A \land B' \implies C')$.  
6. If $y$ is the minority class and $X \rightarrow y$ is more positively correlated than $X \rightarrow \neg y$, then we cannot say anything about their confidences (supports): $(A \land B \iff C'$ and $A \land B \iff \neg C'$).

\textbf{Proof:} Please see [14].

Suppose we have a two class problem and $y$ describes the minority class. 3) tells us that if $X \rightarrow y$ is more confident than $X \rightarrow \neg y$, then it is also more positively correlated. However, the reverse does not hold as described by 4). That is, if $X \rightarrow \neg y$ is more confident than $X \rightarrow y$, then it may or may not be more positively correlated. This means we may have a highly confident rule for the majority class, $X \rightarrow \neg y$ (that is more confident than $X \rightarrow y$), but is actually less positively correlated than $X \rightarrow y$—very undesirable! In the opposite case, 5) tells us that a rule in the minority class, $X \rightarrow y$, with lower relative correlation will also have lower confidence than $X \rightarrow \neg y$. Again, this does not hold for the majority class. Since higher confidence (support) for a rule in the minority class implies higher relative correlation ($CCR > 1$), and lower relative correlation ($CCR < 1$) in the minority class implies lower confidence, but neither of these are true for the majority class, we say that confidence (support) tends to bias the majority class—because confidence (support) and $CCR$ can only ‘contra-
dict’ each other in the majority class. In a related manner, 1) tells us that if $X \rightarrow y$ is more positively correlated than $X \rightarrow \neg y$ but is less confident, then $y$ must be the minority class. Again, the reverse does not hold in general. Hence, if we choose high confidence (support) rules we are more likely to miss rules that have $CCR > 1$ applying to the minority class than in the majority class. Furthermore, when ranking by confidence (support) we are likely to use rules that with $CCR < 1$ predicting the majority class over rules with $CCR > 1$ predicting the minority class.

\textbf{Example 1} Consider an imbalanced dataset with $sup(y) = 15$ and $sup(\neg y) = 100$. A possible contingency table is $[5, 10; 10, 90]$. Despite $conf(X \rightarrow y) = \frac{1}{3} < conf(X \rightarrow \neg y) = \frac{2}{3}$, $X$ has a significant positive association with $y$ ($p_{value} = 0.02$). Also, $corr(X \rightarrow y) = 2.56$ and $corr(X \rightarrow \neg y) = 0.77$ so this rule has a high $CCR (CCR = 3.32 >> 1)$ and is thus a very good rule at distinguishing between classes.

\section{SPARCCC}

There are four components to SPARCCC:

1. The Interestingness and Rule Ranking technique (Section 4.1) determines which potentially interesting rules are interesting and assigns them a Strength Score.
2. The Search and Pruning Strategy (Section 4.2) determines how the space of all possible rules is examined and pruned. This determines the candidate rules—the potentially interesting rules. The choice of strategy determines the computational performance.
3. The Rule Selection Method determines which of the interesting rules are to be used for classification. It outputs selected rules.
4. The Classification Method determines how we classify an unseen instance by using the selected rules.

Components 3. and 4. are based on using groups of equally highly ranked rules. Due to limited space, we refer the reader to [14] for details.

\subsection{Interestingness and Rule Ranking}

We perform the following tests to determine whether a \textit{potentially interesting rule} is \textit{interesting}:

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
A & $sup(y) < sup(\neg y)$ \\
\hline
B & $CCR(X \rightarrow y) > 1$ \\
\hline
B' & $CCR(X \rightarrow y) < 1$ \\
\hline
C & $conf(X \rightarrow y) > conf(X \rightarrow \neg y)$ \hspace{1cm} $\equiv sup(X \rightarrow y) > sup(X \rightarrow \neg y)$ \\
\hline
C' & $conf(X \rightarrow y) < conf(X \rightarrow \neg y)$ \hspace{1cm} $\equiv sup(X \rightarrow y) < sup(X \rightarrow \neg y)$ \\
\hline
\end{tabular}
\caption{Figure 2. Statements for Lemma 1. $\neg y$ means all class attribute-values other than $y$.}
\end{table}
• We check the significance of a rule \( X \rightarrow y \) by performing Fisher’s Exact Test on the contingency table of Figure 1, as earlier described. We record the \( p_{value} \).

• We check whether \( CCR(X \rightarrow y) > 1 \). If this is not the case, the rule is not interesting because it is more correlated with the alternative class(es) than it is with the class it predicts.

The interesting rules – those that pass the above two tests – are candidates for the classification task.

In order to use the rules to make a classification, we need a ranking (ordering) of the rules that captures the ability of the rule to make a correct classification. This ordering is defined by the Strength Score of the rule. Based on the discussions in Sections 2 we may use:

\[
SS_{p,CCR}(X \rightarrow y) = (1 - p_{value}) \cdot CCR(X \rightarrow y)
\]

Confidence is an estimate of the probability that, given \( X \) occurs, \( y \) will occur. Therefore in balanced datasets, choosing the rule with the highest confidence gives the highest expected probability of making a correct classification. Therefore, for comparison, we also evaluate:

\[
SS_{p,conf}(X \rightarrow y) = (1 - p_{value}) \cdot conf(X \rightarrow y)
\]

But as Lemma 1 shows, confidence has a bias toward the majority class. While \( SS_{p,conf} \) performs well on balanced datasets, it performs very poorly on imbalanced datasets. Recall that a) a highly confident rule predicting the majority class may in fact be more negatively correlated than the same rule predicting the other class(es), and b) a rule that is more positively correlated but predicts the minority class may have much lower confidence than the same rule predicting the other class(es). Now, our interestingness criteria above excludes case a), but it does not correct for the bias in confidence for less extreme cases and it does nothing to fix case b). We propose to correct this using \( CCR \):

\[
SS_{p,conf,CCR}(r) = (1 - p_{value}) \cdot conf(r) \cdot CCR(r)
\]

For the rule \( r = X \rightarrow y \). This works by giving poor rules a lower score (in comparison to better rules) and scaling up cases of b): \( CCR(X \rightarrow y) \geq 1 \).

In terms of a suitable classification performance \( P(\cdot) \), experiments show that on relatively balanced datasets:

\[
P(SS_{p,CCR}) \approx P(SS_{p,conf,CCR}) \approx P(SS_{p,conf})
\]

While on imbalanced datasets (as would be expected):

\[
P(SS_{p,CCR}) \gg P(SS_{p,conf,CCR}) \gg P(SS_{p,conf})
\]

That is, the use of \( CCR \) achieves the highest performance on imbalanced datasets while performing comparably on balanced datasets. Note that in a completely balanced dataset, \( CCR(X \rightarrow y) \) reduces to \( \frac{\text{conf}(X \rightarrow y)}{\text{conf}(X \rightarrow \neg y)} \) which we call the Class Support Ratio and the Class Confidence Ratio respectively.

Finally, we note that the \( p_{value} \) has little impact in the final score, because it varies at most by the significance level. It’s inclusion therefore favors more significant rules only if the other components of \( SS \) are similar.

### 4.2 Search and Pruning Strategies

The overall strategy is a bottom up item enumeration technique, as all the rules \( X' \rightarrow y : X' \subseteq X \) will be examined before \( X \rightarrow y \) and the search is over the item space (attribute-value space). The underlying algorithm used to do this is a yet-to-be-published variation of GLIMIT [13]. It performs this task in a depth first fashion. It uses linear space in the number of instances, linear time in the number of itemsets (classification rules and their antecedents) that need to be considered, and one pass over the dataset. While this is faster than alternatives such as Apriori [1] or FP-Growth [6], either of these could potentially be used.

The idea of a rule being statistically significant is not anti-monotonic. To avoid examining all rules, we use search strategies that ensure the concept of being potentially interesting is anti-monotonic – i.e. \( X \rightarrow y \) might be considered as potentially interesting if and only if all \( \{X' \rightarrow y | X' \subset X \} \) have been found to be potentially interesting:

• Select a new attribute-value in such a way that it makes a significant positive contribution to the rule, when compared to all immediate generalizations. Specifically, Figure 3 describes how we test for the significance of the rule \( X \rightarrow y \) in comparison to one of its generalizations \( X - \{z\} \rightarrow y \). The rule \( X \rightarrow y \) is potentially interesting only if the test passes for all immediate generalizations \( \{X - \{z\} \rightarrow y : z \in X \} \). This technique prunes the search space most aggressively, as it performs \(|X| \) tests per rule. However, this also means that it greatly favors shorter rules, as they have fewer tests to pass. This approach is borrowed from Webb [16]. We call it Aggressive-S.

• Use FET as described in Section 2 and force it to be anti-monotonic\(^2\). We call this Simple-S. It performs one test per rule and examines more of the search space.

• For comparison, we also use a minimum support threshold. All rules with \( \text{sup}r(X \rightarrow y) \geq \text{minSup} \) are potentially interesting. We call this Support. For Aggressive-S and simple-S, we define \( \text{sup}(\emptyset) = |D| \) so that we can evaluate a \( p_{value} \) (usually high) for so-called “default rules” – rules with no antecedent.

\(^2\)That is, if and only if all rules \( \{X - \{z\} \rightarrow y : z \in X \} \) are potentially interesting, then we use the contingency table of Figure 1 to determine whether \( X \rightarrow y \) is potentially interesting. Note that this is recursive.
5 Experiments

We performed experiments on relatively balanced, well-known UCI datasets [9] (Australia, breast, Cleve, Diabetes, Heart, Horse) as well as imbalanced variations of them. In the tables, our methods are denoted by “SPARC-CCC” with the search strategy in parentheses. For comparison we also use a purely support and confidence based technique denoted by “Support-Confidence”. It finds all rules satisfying the support and confidence thresholds and uses confidence as the strength score. Due to limited space we show only average results. For more details and further discussion of our experiments please refer to [14].

Original (Balanced) Datasets: Figure 4(a) shows that (on average) SPARC-CCC performs comparably to CBA, CMAR and C4.5, and is insensitive to the choice of SS. However, there are large differences in the search space examined and hence the run times, as shown in Figures 4(c). Also, much fewer rules are found as can be seen in Figure 4(d). So picking the best accuracy (83.6%, “Aggressive-S” using significance of 0.001 and \(SS_{p,conf,CCR}\)) we can obtain comparable accuracy while searching only 1.3% of the space, using 0.08% of the time and finding 0.03% of the rules, when compared to support based methods – for example; CBA and CMAR.

Highly imbalanced versions of the datasets were obtained by keeping the majority class and randomly selecting a subset of the minority class so that the ratio was 1 : 9. Figure 4(b) shows the True Positive Rate (TPR) of the minority class. The effect of using CCR in the Strength Score is dramatic. Clearly:

\[
TPR(SS_{p,CCR}) \gg TPR(SS_{p,conf,CCR}) \gg TPR(SS_{p,conf})
\]

For example, when using “Aggressive-S”, \(SS_{p,conf,CCR}\) is on average (over datasets and significance levels) 2.87 times better than \(SS_{p,conf}\) and \(SS_{p,CCR}\) is 1.58 times better than \(SS_{p,conf,CCR}\) and 4.44 times better than \(SS_{p,conf}\).

Our methods also score much higher than other rule based techniques such as CBA and CCCS. The highest average TPR overall is for “Aggressive-S” with a significance level of 0.05. This was 45.8% better than CBA and 26.1% better than CCCS. Finally, the computational performance favors our techniques even more on imbalanced datasets.

Figure 4(c) for example shows that “Aggressive-S”, at a significance level of 0.05, explores only 0.29% of the space considered by a support based method with \(minSup = 1\%\) – and the training time is even less.

6 Related Work

CBA [8] was the first Associative Classifier (AC) proposed and almost all other ACs are variations on the original CBA design. For rule mining, CBA mines all rules passing support and confidence thresholds (\(minSup\) and \(minConf\)). Additionally, it ignores rules based on a “pessimistic error based pruning method” borrowed from C4.5 [10]. Unfortunately this still generates thousands of rules – most of which perform poorly. Therefore, a rule selection process is needed to select a small subset likely to perform well. New instances are classified according to the highest ranked rule that is applicable. Rules are ranked according to confidence, support, and size.

CMAR [7] is a variation of CBA which uses the chi-square measure for rule selection. A more detailed analysis of CMAR and related associate classifiers can be found in the longer version of the paper [14]. CCCS [3] is the first associative classifier to propose a way to handle imbalanced classes. However, it provides no guarantees about the statistical significance of the rules mined. We borrow the idea from Webb [16] in our Aggressive-S pruning approach.

Acknowledgments: We are grateful to Bavani Arunasalam for providing the preprocessed data and the CCCS results. This research was partially funded by the Australian Research Council (ARC) Discovery Grant, Project ID: DP055900.

References


<table>
<thead>
<tr>
<th>(t : X \subseteq t)</th>
<th>(t : X - {z} \subseteq t \land z \not\in t)</th>
<th>(t : X - {z} \subseteq t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t : y \in t)</td>
<td>(a = sup(X \rightarrow y))</td>
<td>(b = sup(X - {z} \rightarrow y) - sup(X \rightarrow y))</td>
</tr>
<tr>
<td>(t : \neg y \in t)</td>
<td>(c = sup(X \rightarrow \neg y))</td>
<td>(d = sup(X - {z} \rightarrow \neg y) - sup(X \rightarrow \neg y))</td>
</tr>
<tr>
<td>(a + c = sup(X))</td>
<td>(b + d = sup(X - {z}) - sup(X))</td>
<td>(a + b + c + d = sup(X - {z}))</td>
</tr>
</tbody>
</table>

Figure 3. The contingency table \([a, b; c, d]\) used to test for the significance of the rule \(X \rightarrow y\) in comparison to one of its generalizations \(X - \{z\} \rightarrow y\) for the Aggressive-S search strategy.
Figure 4. Classification and computational performance on original and imbalanced datasets.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>minSup</th>
<th>minConf</th>
<th>significance</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPARCC</td>
<td>na</td>
<td>na</td>
<td>0.01</td>
<td>81.2</td>
</tr>
<tr>
<td>(Aggressive-S)</td>
<td>na</td>
<td>na</td>
<td>0.01</td>
<td>82.8</td>
</tr>
<tr>
<td>SPARCC</td>
<td>na</td>
<td>na</td>
<td>0.01</td>
<td>83.4</td>
</tr>
<tr>
<td>(Simple-S)</td>
<td>na</td>
<td>na</td>
<td>0.01</td>
<td>82.8</td>
</tr>
</tbody>
</table>

Support Confidence
- conf: 1% 0.5 0.6 0.7
- CBA: 1% 0.5 0.6 0.7
- CHAR: 1% 0.5 0.6 0.7
- C4.5: 1% 0.5 0.6 0.7

(a) Accuracy on Original Datasets. Average over datasets and folds.

(b) True Positive Rate (Recall, Sensitivity) of the Minority Class on Imbalanced Versions of the Datasets. Average over datasets and folds.

(c) Search space size on original datasets, training time on original datasets and search space size on imbalanced versions of the datasets. Average over all datasets and folds.

(d) Number of rules found (prior to rule selection) on the original datasets. Average over all datasets and folds.


