Proc. 22nd Int. Conf. on Knowledge-based Systems and Applied Artificial Intelligence ES'02
Cambridge, United Kingdom, December 2002, pp. 33-46.

Simplest Rules Characterizing Classes Generated by δ-Free Sets

Bruno Crémilleux
GREYC CNRS-UMR 6072, University of Caen
F-14032 Caen Cedex, France

Jean-François Boulicaut
Laboratoire d'Ingénierie des Systèmes d'Information
INSA Lyon
F-69621 Villeurbanne Cedex, France

Abstract

We present a new approach that provides the simplest rules characterizing classes with respect to their left-hand sides. This approach is based on a condensed representation (δ-free sets) of data which is efficiently computed. Produced rules have a minimal body (i.e. any subset of the left-hand side of a rule does not enable to conclude on the same class value). We show a sensible sufficient condition that avoids important classification conflicts. Experiments show that the number of rules characterizing classes drastically decreases. The technique is operational for large data sets and can be used even in the difficult context of highly-correlated data where other algorithms fail.

Keywords: characterization of classes, rule conflicts, δ-free sets, association rules, classification rules.

1 Introduction

Context and motivations. Frequent association rules is one popular data mining technique. This kind of process has been studied a lot of times since the definition of the mining task in [1]. Association rules can tell something like "It is frequent that when properties A_1 and A_2 are true within an example, then property A_3 tends to be true". We provide a simple formalization of this task in Section 2.1. Finding rules that characterize classes and classification rules are important research topics as well. Starting from a collection of examples associated with a known class value, classification concerns the design of models that enable to predict accurate class values for unseen examples. The set of examples for which the class value is given is the so-called learning set. Various knowledge representation formalisms have been used for designing classifiers. Classification rules (which are rules that conclude on one class value) are quite popular for that purpose and the literature is abundant (see for example [12]). Mining rules that characterize classes can be viewed as a special form of association rule mining where conclusions of rules are pre-specified.

However, known techniques are not able to handle dense and highly-correlated data (i.e., problems for rule mining) and the large number of produced rules leads to rule conflicts and over-fitting (i.e., rules may be over-specified and miss-classification arise when using them for classifying unseen examples). To cope with these drawbacks, we consider the efficient extraction of the set of the simplest rules characterizing classes w.r.t. their left-hand sides. Furthermore, we found a property that avoids classification conflicts. Such results form a sound basis for the selection of classification rules and the design of a classifier.

Related work. Until recently, extracting rules characterizing classes used to undergo two steps: first, association rules were mined from the learning set and then the identification of the classification rules was performed mainly as a post-processing step ([2, 13]). CBA system (Classification Based on Association) is presented in [13]. The selection of a subset of classification rules is done in two steps: first, all rule concluding on the class are produced, some of them are pruned using the pessimistic error rate [20]. Second, a classifier builder selects the final set of rules according to the numbers of well-classified and miss-classified examples of the training set. Besides a high computational cost, Freitas [9] has shown the limitations of such an approach, emphasizing the differences between classification rules and association rules. Furthermore, the huge number of potential interesting classification rules makes this approach difficult (indeed, how to identify the most relevant rule to classify a new case might be quite difficult). Bayardo [3] suggests to add pruning strategies to control combinatorial explosion in the number of candidates. Target-constraint association rules [4] can be used, but rule sets contain many redundant rules useless for prediction. In [15], Liu et al. add two improvements on their CBA system. On one side, to deal with unbalanced class distribution, they use multiple class minimum frequencies in rule generation. On the other side, to tackle large data sets, they propose a technique to combine CBA with the decision tree method and the Naïve-Bayes method. Recent works revisit these questions and bring improvements. In [11], memory consumption and time complexity have been decreased by features selection and, in a post-processing stage, rules covering most examples are selected. CMAR [14] uses statistical techniques to avoid bias and improve efficiency by relevant data structures. The minimum subset of classification rules having the same prediction power (defined by statistical measures or the confidence) as the complete class rule set is computed in [16].

Contributions. The contribution of this paper is twofold. First, we provide the simplest rules that characterize classes w.r.t. their left-hand sides, i.e., a key point in classification. Given a rule characterizing a class, one wants that any own and proper subset of its left-hand side does not enable to conclude on the same class value. It is possible to work on difficult contexts such as dense and highly-correlated learning sets. Second, we highlight a property to avoid important classification conflicts for unseen examples. This allows to use such rules to design a classifier. Furthermore, we think that this work proposes an original use of frequent sets (frequent δ-free sets are a generalization of frequent sets) and condensed representations [18].

Organization of the paper. The next section defines the concept of δ-strong rule for class characterization. It shows how the simplest rules characterizing classes are derived. We present in Section 3 a property that avoids classification conflicts. Experimental results are given in Section 4.

2 Mining the δ-strong rules characterizing classes

Let us provide a simple formalization of δ-strong rule mining task. We start to recall standard definitions.

2.1 Association Rule Mining

Definition 1 (item, itemset, example) *Assume* $\mathbf{R} = \{A_1, \ldots, A_n\}$*, is a schema of boolean attributes. One attribute from* $\mathbf{R}$ *is called an* item *and a subset of* $\mathbf{R}$ *is called an* itemset*.* $\mathbf{r}$*, an instance of* $\mathbf{R}$*, is a multi-set of* examples*. Thus,* $\mathbf{r}$ *can be considered as a boolean matrix.*

For instance, these attributes can identify molecule properties. In practice, one can have hundreds of thousands of examples and hundreds of attributes. In our experiments (see section 4) data have 6,150 fragments. This is obviously a difficult mining context.

Definition 2 (association rule) *Given* $\mathbf{r}$*, an instance of* $\mathbf{R}$*, an* association rule *on* $\mathbf{r}$ *is an expression* $X \Rightarrow B$*, where the itemset* $X \subseteq \mathbf{R}$ *and* $B \in \mathbf{R} \setminus X$*.*

The intuitive meaning of a potentially interesting association rule $X \Rightarrow B$ is that all the items in $X \cup \{B\}$ are true (value 1) for enough examples and that when an example contains true for each item of X, then this example tends to contain true for item B too. This semantics is captured by the classical measures of *frequency* and *confidence* [1].

Definition 3 (frequency, confidence) *Given* $W \subseteq \mathbf{R}$, $\mathcal{F}(W, \mathbf{r})$ *(or frequency of* W*) is the number of examples in* $\mathbf{r}$ *that contain* 1 *for each item in* W*. The frequency of* $X \Rightarrow B$ *in* $\mathbf{r}$ *is defined as* $\mathcal{F}(X \cup \{B\}, \mathbf{r})$ *and its confidence is* $\mathcal{F}(X \cup \{B\}, \mathbf{r})/\mathcal{F}(X, \mathbf{r})$*. We define an absolute frequency (a number of examples* $\leq |\mathbf{r}|$*). We also use the relative frequency* $\mathcal{F}(X \cup \{B\}, \mathbf{r})/|\mathbf{r}|$*, i.e., a value in* $[0, 1]$*.*

The standard association rule mining task concerns the discovery of *every* rule whose frequency and confidence are greater than user-specified thresholds. In other words, one wants rules that are frequent "enough" and valid. The main algorithmic issue concerns the computation of every frequent set.

Definition 4 (frequent itemset) *Given* γ *a frequency threshold* $\leq |\mathbf{r}|$*. An itemset* X *is said* frequent *or* γ*-frequent if* $\mathcal{F}(X, \mathbf{r}) \geq \gamma$*.*

The complexity of frequent itemset mining is exponential with the number of attributes. Many research works (e.g. [22, 19, 6]) concern the contexts for which such a discovery remains tractable, even though a trade-off is needed with the exact knowledge of the frequencies and/or the completeness of the extractions.

2.2 δ-strong rules

A rule characterizing classes must conclude on class values with a rather high confidence. δ-strong rules introduced in [6] satisfy such a constraint.

Definition 5 (δ-strong rules) *Given* $\mathbf{R}$*, a matrix* $\mathbf{r}$*, a frequency threshold* γ*, and an integer* δ*, a* δ-strong rule *on* $\mathbf{r}$ *is an association rule* $X \Rightarrow B$*, where* $\mathcal{F}(X \cup \{B\}, \mathbf{r}) \geq \gamma$*,* $\mathcal{F}(X, \mathbf{r}) - \mathcal{F}(X \cup \{B\}, \mathbf{r}) \leq \delta$*,* $X \subseteq \mathbf{R}$*, and* $B \in \mathbf{R} \setminus X$*.*

A δ-strong rule is violated by at most δ examples. In other words, its confidence is at least equal to $1 - (\delta/\gamma)$. From a technical perspective, δ-strong rules can be built from δ-free sets that constitute their left-hand sides [6]. Let us provide the key intuition for the concept of δ-free set. An itemset X is called δ-free if there is no δ-strong rule that holds between two of its own and proper subsets. We illustrate this notion with Table 1 (this table consists of 8 examples, each one identified by its Id, and there are 4 items denoted $A_1 \ldots A_4$).

Id	Items			
	A_1	A_2	A_3	A_4
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1		1	
5		1	1	
6			1	
7	1		1	1
8	1	1		1

Table 1: A set of 8 examples described by 4 items

The case $\delta = 0$ (corresponding to 0-free-sets) is important: no rule with confidence equal to 1 holds between proper subsets of X. For instance, $A_1 A_2$ is a 0-free set because all rules constructed from proper subsets of $A_1 A_2$ have at least one exception: examples numbers 4 and 7 for the rule $A_1 \Rightarrow A_2$ and example number 5 for the rule $A_2 \Rightarrow A_1$. If $\delta = 1$, $A_1 A_2$ is not a 1-free set owing to the rule $A_2 \Rightarrow A_1$ which has only one exception. On the contrary, A_1 and A_2, for instance, are 1-free sets.

δ-free sets are related to the concepts of closed itemsets in [19] and almost-closures in [5]. In the case $\delta = 0$, let us provide the relationship with the *closure* operator.

Definition 6 (closure) *Given an itemset X, the* closure *of X is the maximal superset (w.r.t. set inclusion) of X that has the same frequency than X.*

For instance in Table 1, with $X = A_1A_2$, A_4 belongs to the closure of X (i.e. A_4 is always true when A_1 and A_2 are true). In that case, frequencies of A_1A_2 and $A_1A_2A_4$ are the same. The closure of X can be computed efficiently during the computation of the frequency of X and, in our example, one avoids to count the frequencies of $A_1A_2A_4$. In highly-correlated data, it reduces drastically the extraction time [19, 5]. Frequent closed itemsets are the closures of 0-free sets.

When $\delta > 0$, we are interested in the *almost-closures* of a frequent δ-free set X: B belongs to the almost-closure of X if $\mathcal{F}(X, \mathbf{r}) - \mathcal{F}(X \cup \{B\}, \mathbf{r}) \leq \delta$. It is easy to provide δ-strong rules from the γ-frequent δ-free sets and their almost-closures: in this case, we have the rule $X \Rightarrow B$ with at most δ exceptions. Following our example given in Table 1, A_4 belongs to the almost-closure of A_1 with $\delta = 1$ (there is only one exception, example number 4).

A collection of frequent δ-free sets is a condensed representation of the collection of frequent itemsets. If $\delta = 0$, one can compute the frequency of every frequent itemset. If $\delta > 0$, one can approximate the frequency of every frequent itemset X with a bounded error: when an itemset X is not δ-free, its frequency is approximated from the frequency of the largest δ-free set included in X [6]. In [6], it is shown that the error is very low in practice.

One interesting property of *freeness* is its anti-monotonicity w.r.t. itemset inclusion (a property ρ is *anti-monotone* iff for all itemsets X and Y, $\rho(X)$ and $Y \subseteq X$ implies $\rho(Y)$). It gives a safe pruning criterion for level-wise search in the itemset lattice [17, 6]. Accordingly, it is possible to design efficient algorithms for frequent δ-free set discovery and have tractable extractions for practical mining tasks that are intractable with `apriori`-like algorithms (see Section 4).

2.3 δ-strong rules and minimal body

We have seen that δ-strong rules have a number of exceptions bounded by δ and can be efficiently extracted from large sets of highly-correlated data. Furthermore, this formalism offers a property of minimal body which is a key point for a classification purpose.

Definition 7 (rule with a minimal body) *Given a frequency threshold γ and an integer δ, a rule $X \Rightarrow B$ has a* minimal body *if there is no frequent rule $Y \Rightarrow B$ with $Y \subset X$ and a confidence greater or equal to $1 - (\delta/\gamma)$.*

This definition means that we consider only minimum sets of items to end up on B, the uncertainty being controlled by δ. In practical applications, more specified rules concluding on B can exist. Nevertheless, Section 3 gives a property showing that, under a sensible assumption, any specified rule R characterizes the same class as the rule with a minimal body which is included in R.

Property 1 (minimal body) *Given a frequency threshold γ, an integer δ and $X \Rightarrow B$ a rule with a minimal body, then X is a δ-free set.*

This result comes from properties of δ-free sets [6]. The key intuition of this result is that a δ-free set X is a minimal conjunction of items to know the frequencies of a set of items $\mathcal{C}$ ($\mathcal{C}$ is bounded by the closed itemset defined by X and its closure). As all itemsets of $\mathcal{C}$ have the same frequency, X is a minimal conjunction of items which has to be used to produce all rules having their left-hand sides stemming from $\mathcal{C}$. The whole collection of δ-free sets ensures to produce all rules with minimal body.

Nevertheless, let us note that it can happen that a δ-strong rule has not a minimal body. For instance, in Table 1, $A_1A_2A_3$ is a 0-free set having A_4 in its closure. It means that the rule $A_1A_2A_3 \Rightarrow A_4$ exists (with a frequency value of 3). We have seen in the previous section that there is also the rule $A_1A_2 \Rightarrow A_4$ (with a frequency value of 4) which is the simplest rule to conclude on A_4 with $\delta = 0$. Our prototype can be seen as an instance of the level-wise search algorithm presented in [17] which allows to recognize a minimal conjunction of items concluding on B. Let us consider X a δ-free set with k items (i.e. a k-δ-free set produced at level k). When the almost-closure of X is computed, we start to merge the almost-closures of all $(k-1)$-δ-free sets included in X. δ-strong rules with a minimal body will be infer only from new items added in the almost-closure of X (i.e. items which do not belonging to the set of the almost-closures of all $(k-1)$-δ-free sets included in X).

We argue that this property of minimal body is a fundamental issue for class characterization. Not only it prevents from over-fitting [21] (i.e. over-specified rules acquired on the learning set and leading to miss-classified unseen examples) but also it makes the characterization of an example easier to explain. It provides a feedback on the application domain expertise that can be reused for further analysis.

2.4 δ-strong rules characterizing classes

Let us consider a classification task with k class values. Assuming $C_1, \ldots, C_k$ are the k items that denote class values.

Definition 8 (δ-strong rule characterizing classes) *A* δ-strong rule characterizing classes *is a δ-strong rule with a minimal body that concludes on one class value (i.e., C_i).*

Hereafter, we consider the typical case where each example is associated to a unique class value. Thus, we have the following equality (see equation 1). Let us notice that when it is clear from the context, $\mathbf{r}$ is omitted:

$$\sum_{i=1}^{k} \mathcal{F}(C_i) = |\mathbf{r}| \tag{1}$$

The value of δ is fundamental to obtain relevant rules. Let us recall that when $\delta = 0$, every extracted rule has a confidence value of 1. This is a problem for the practical impact. Indeed, it is rare that rules characterizing classes with confidence 1 hold in real data. In many application domains, such as biology or medicine, where the same cause does not always produce the same effect and/or where parameter values that are used for decision making are unknown, we must accept exceptions for rules. Experts are trained to cope with some bounded uncertainty.

The more δ raises, the more the confidence decreases. Intuitively, we feel that when the confidence decreases (i.e. the uncertainty increases), it can give rise to more and more classification conflicts. In order to characterize classes, it is useful to verify whether this formalism enables to point out situations (and under which assumptions) where rule conflicts can be avoided. The next section studies the relationship between the values for δ and γ and classification conflicts. We show that with respect to a simple property, we can avoid some types of classification conflicts.

3 Avoid rule conflicts

Let us consider now a systematic study of three pairs of rules that lead to classification conflicts.

3.1 Identical body conflicts

An identical body conflict is characterized by a pair of rules that have the same body but conclude on different class values. It means that a same δ-free set gives rise to at least two δ-strong rules characterizing classes. For example :

$$R_1 : X \Rightarrow C_1 \qquad R_2 : X \Rightarrow C_2$$

The pair $\{R_1, R_2\}$ does not characterize properly classes and leads to a conflict when an example to be classified contains X. Obviously, the more δ is large w.r.t. γ, the more that identical body conflicts can appear.

Let us now consider a sufficient condition on δ and γ values such that identical body conflicts are impossible. Assume that we have the δ-strong rule characterizing classes $R_1 : X \Rightarrow C_1\ \{\delta_1\}$ with frequency γ_1 and δ_1 exceptions (figure between braces indicated the exact number of exceptions). Note that the choice of C_1 for the class value is arbitrary. Given that R_1 is a δ-strong rule, we get:

$$\gamma_1 = \mathcal{F}(X \cup \{C_1\}) \geq \gamma$$
$$\delta_1 = \mathcal{F}(X) - \mathcal{F}(X \cup \{C_1\}) \leq \delta$$

Let us look now for conditions on γ and δ values that prevent the existence of a δ-strong rule characterizing classes $R_2 : X \Rightarrow C_2\ \{\delta_2\}$ with frequency γ_2 and δ_2 exceptions such that C_2 is a class value different from C_1. As each example is associated to a unique class value, we have from the Equation 1 (the inequality becomes an equality if the class has just the two values C_1 and C_2):

$$\mathcal{F}(X \cup \{C_1\}) + \mathcal{F}(X \cup \{C_2\}) \leq \mathcal{F}(X)$$

It points out a lower bound for δ w.r.t. γ_2:

$$\gamma_2 = \mathcal{F}(X \cup \{C_2\}) \leq (\mathcal{F}(X) - \mathcal{F}(X \cup \{C_1\}) \leq \delta \tag{2}$$

If R_2 is a δ-strong rule, we have $\gamma_2 \geq \gamma$. This inequality combined with inequality 2 enables to compare γ and δ:

$$\gamma \leq \gamma_2 \leq \delta \tag{3}$$

The inequality 3 shows that it is sufficient to take $\delta < \gamma$ to enforce that R_2 is not a δ-strong rule characterizing classes.

Property 2 *If $\delta < \gamma$, there is no identical body conflict.*

3.2 Included bodies conflicts

Given the pair $\{R_1, R_2\}$, there is an included bodies conflict if the left-hand side of R_1 is included in the left-hand side of R_2 (or vice-versa) and if R_1 and R_2 have different right-hand sides. For instance:

$$R_1 : X \Rightarrow C_1 \qquad R_2 : X \cup Y \Rightarrow C_2$$

It means that the δ-free set that produces one of the rules is included in the δ-free set from which the other rule is derived. The pair $\{R_1, R_2\}$ is not a reliable characterization of classes since it leads to a conflict as soon as an example supports X (however, this pair suggests that when Y is added to X there may be an interesting exceptional situation if the frequency of the rule $X \cup Y \Rightarrow C_2$ is small w.r.t. the frequency of the rule $X \Rightarrow C_1$).

Let us now discuss about a sufficient condition on δ and γ such that there is no included bodies conflict. Assume that there is a δ-strong rule characterizing classes $X \Rightarrow C_1$ $\{\delta_1\}$ with frequency γ_1 and δ_1 exceptions. Let us look now for conditions on γ and δ values that prevent the existence of a δ-strong rule characterizing classes R_2 $X \cup Y \Rightarrow C_2$ $\{\delta_2\}$ with frequency γ_2 and δ_2 exceptions with $Y \subseteq \mathbf{R} \setminus X$. As previously, the following inequality holds:

$$\mathcal{F}(X \cup \{C_1\}) + \mathcal{F}(X \cup \{C_2\}) \leq \mathcal{F}(X)$$

and thus: $\mathcal{F}(X \cup \{C_2\}) \leq \mathcal{F}(X) - \mathcal{F}(X \cup \{C_1\}) \leq \delta$
and we have: $\gamma_2 = \mathcal{F}(X \cup Y \cup \{C_2\}) \leq \mathcal{F}(X \cup \{C_2\})$

As for the first kind of conflict, we can get a lower bound for δ w.r.t. γ_2:

$$\gamma_2 = \mathcal{F}(X \cup Y \cup \{C_2\}) \leq \mathcal{F}(X \cup \{C_2\} \leq \delta \tag{4}$$

To enforce that R_2 is δ-strong, we must have $\gamma_2 \geq \gamma$. This bound, combined with the one from inequality 4 enables to compare γ and δ:

$$\gamma \leq \gamma_2 \leq \delta \tag{5}$$

Inequality 5 shows that it is a sufficient condition to take $\delta < \gamma$ to enforce that R_2 cannot be a δ-strong rule characterizing classes.

Property 3 *If $\delta < \gamma$, there is no included bodies conflict.*

3.3 Distinct bodies conflicts

A distinct bodies conflict can not be foreseen: it occurs only when an unseen example is to be classified. Assume an example is supported by a pair of rules (that have distinct but not included bodies) concluding on two different class values. For instance, the pair:

$$R_1 : X \Rightarrow C_1 \qquad R_2 : Y \Rightarrow C_2$$

leads to a conflict when the unseen example satisfies $X \cup Y$. Note that we can have $X \cap Y \neq \emptyset$.

It is not possible to avoid such a conflict when producing the rules. On the other hand, interestingly, one can see that if $\delta < \gamma$, there is no distinct bodies conflict within the pair $\{R_1, R_2\}$ if there is a δ-strong rule characterizing classes $X \cup Y \Rightarrow C_i$. Indeed, we saw that if $\delta < \gamma$, the existence of a δ-strong rule $X \cup Y \Rightarrow C_i$ avoids the existence of simpler rules (with bodies that are subsets of $X \cup Y$, like R_1 and R_2) that conclude on classes.

The property $\delta < \gamma$ enforces no identical body conflict and no included bodies conflict. It also enforces no distinct bodies conflict for pairs of rules that are such that the union of their left-hand sides gives rise to a δ-strong rule characterizing classes. If there is not such a δ-strong rule, that means that the learning set does not contain enough examples, according to the thresholds, in order to solve the ambiguity on the relationship between the union of the left-hand sides of the rules and classes.

4 Experiments

The purpose of the first experiment is to compare the number of all classification rules versus the number of δ-strong rules characterizing classes. Data come from the discovery challenge on thrombosis data (see `http://lisp.vse.cz/challenge/pkdd2001/`), classes are collagen diseases. After a data preparation work [8], the resulting file gathers 721 tuples described by 56 binary items. As this training set is small, we were able to run an `apriori`-based algorithm even with a low frequency threshold ($\gamma = 3$, i.e., 0.5%).

We performed three experiments (see Table 2) with different values of γ and δ. On the one hand, we mined δ-strong rules characterizing classes. On the other hand, we extracted with the same value of γ with an `apriori`-based algorithm all classification rules with a single item in their right-hand side and with a confidence threshold equals the lowest confidence value possible for a δ-strong rule characterizing classes (this value, denoted $minconf.$, depends on δ and γ: see Table 2). In this table, each percentage is the number of δ-strong rules characterizing classes divided by the number of classification rules. We see that mining the δ-strong rules characterizing classes strongly reduces the number of rules.

	No. of rules	
	δ-strong rules characterizing classes	classification rules
$\gamma = 3$ $\delta = 2$ $(minconf. = 33\%)$	1,342 (1.77%)	76,004
$\gamma = 6$ $\delta = 3$ $(minconf. = 50\%)$	526 (4.25%)	12,366
$\gamma = 6$ $\delta = 5$ $(minconf. = 16\%)$	801 (2.65%)	30,260

Table 2: Comparison of numbers of rules

The goal of the second experiment is to show that δ-strong rules characterizing classes can be extracted efficiently from large data sets and we give a first approach to design a classifier stemming from these rules. Data are those used in the Predictive Toxicology Challenge (PTC) for 2000-2001 (see `http://www.informatik.uni-freiburg.de/~ml/ptc/`). The aim is to predict chemical carcinogens (i.e. the class) on several rodents populations (each file contains about 350 examples). Each molecule is represented by 6,150 binary attributes (i.e. chemical fragment). It is a hard context for association rule mining since we have a huge number of attributes. On these data, with most values of γ indicated in Table 3, `apriori`-like algorithms fail due to an excessive memory requirement (we used a PC with 768 MB of memory and a 500 MHz Pentium III processor with Linux operating system). For different values of δ and γ (but always with $\delta < \gamma$), Table 3 gives the extraction time, the number of δ-free sets and almost-closures that contain a class value on male rats. This last number can be seen as the number of potential δ-strong rule characterizing classes (i.e., with any frequency and confidence values). Results on other populations are given in [7].

As expected, the more we increase the value of δ, the more we can have tractable extractions for lower frequency thresholds. Note that with $\delta = 0$, there is no rule characterizing classes for the frequency threshold we can use. It illustrates the added-value of the relaxed constraint on δ.

It is not sensible to use the whole collection of δ-strong rules characterizing classes to classify unseen examples because some rules can have low frequency (and/or confidence if δ is large w.r.t. γ). These rules with a poor quality may introduce errors and have to be deleted. We are giving now a first approach to design a classifier stemming from the δ-strong rules characterizing classes. This approach has techniques in common with the method used by the classifier

$\gamma/\|\mathbf{r}\|$	δ	Time (sec.)	No. of δ-free sets	No. of almost-closures with a class value
0.15	15	intractable	-	-
0.15	17	3814	24671	2835
0.15	20	1563	17173	4529
0.20	10	3300	26377	0
0.20	15	850	12071	8
0.20	20	323	7109	305
0.30	10	69	3473	0
0.40	0	intractable	-	-
0.40	10	36	922	0
0.50	0	201	56775	0

Table 3: Time, δ-free sets and almost-closures w.r.t. δ and γ

builder of CBA [13] that we have mentioned in Section 1, except that we use a test file to select rules (and we think that this point is important). Data are split into a training file (4/5 of data) and a test file (1/5 of data). Class has the same frequency distribution in each file and in the whole data. For each rule, we compute a score (denoted Δ) which is the difference between the well-classified and the miss-classified examples of the test file. Then rules are sorted w.r.t. Δ and, by varying Δ, we define a family of nested sets of rules by the following way: for a value Δ_1 of Δ, we kept rules having Δ higher or equal to Δ_1. All rules belonging to the set defined by Δ_1 belong to sets defined by Δ_2 with $\Delta_2 < \Delta_1$. Then, we use a set of rules with a high value of Δ to classify unseen examples.

We tested this approach on a validation file (185 new chemicals not used to build and select the δ-strong rules characterizing classes) provided by the organizers of the ECML-PKDD 2001 Predictive Toxicology Challenge. When there is a distinct bodies conflict, the class value which maximizes the amount of products of the frequency by the confidence of triggered rules is predicted. According to the population of rodents, the used sets of rules to predict have between 9 and 76 rules. The rate of well-classified chemicals varies between 61.1% and 76.8% with an average of 70.55% Note that predict chemical carcinogens has been identified as a difficult classification task (the correct classification score for experts in the domain ranges from 28% to 78% [10]). We also used the learning classification rules software `C4.5` [20]. Decision trees are built from the same training sets and also tested with the validation file. With 5 as the minimum value of examples for subtrees, trees have around 50 nodes and rates of well-classified 60%. This approach achieved good results against decision trees.

5 Conclusion

We have developed an original technique based on δ-free sets to mine the simplest rules characterizing classes w.r.t. their left-hand sides. We claim that extracting such a set of rules brings several advantages. As the number of rules decreases and as rules are simpler, sets of rules characterizing classes are more understandable. Secondly, it prevents from over-fitting which is fundamental in real-world domains for classes characterization. Experiments show a significant reduction of the number of rules. This approach is effective even in the case of huge, dense and/or highly correlated learning data sets.

A straightforward use of such rules is the characterization of classes. Then, we have shown a sensible sufficient condition that avoids important classification conflicts. Such results form a sound basis for the selection of classification rules and the design of a classifier.

Acknowledgements

The authors thank Christophe Rigotti for his contribution to this research and his constructive comments on the paper.

References

1. Agrawal, R. and Imielinski, T. and Swami, A. Mining association rules between sets of items in large databases, In *Proceedings SIGMOD'93*, ACM Press, pp. 207-216, 1993.

2. Ali, K. and Manganaris, S. and Srikant, R. Partial classification using association rules, In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, KDD'97*, AAAI Press, pp. 115-118, 1997.

3. Bayardo, R. J., Brute-force mining of high-confidence classification rules, In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, KDD'97*, AAAI Press, pp. 123-126, 1997.

4. Bayardo, R. J. and Agrawal, R. and Gunopulos, D. Constraint-based rule mining in large, dense database, In *Proceedings ICDE'99*, pp. 188-197, 1999.

5. Boulicaut, J. F. and Bykowski, A. Frequent closures as a concise representation for binary data mining, In *Proceedings of the Fourth Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD'00*, LNAI 1805, Springer-Verlag, pp. 62-73, Kyoto, Japan, 2000.

6. Boulicaut, J. F. and Bykowski, A. and Rigotti, C. Approximation of frequency queries by means of free-sets, In *Proceedings of the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD'00*, LNAI 1910, Springer-Verlag, pp. 75-85, Lyon, France, 2000.

7. Boulicaut, J. F. and Crémilleux, B. Delta-strong classification rules for characterizing chemical carcinogens, In *Proceedings of the Predictive Toxicology Challenge for 2000-2001 co-located with PKDD'01*, Freiburg, Germany, 2001.

8. Boulicaut, J. F. and Crémilleux, B. Delta-strong classification rules for predicting collagen diseases, In *Discovery Challenge on Thrombosis Data for 2000-2001 co-located with PKDD'01*, pp. 29-38, Freiburg, Germany, 2001.

9. Freitas, A. A. Understanding the crucial differences between classification and discovery of association rules - a position paper, In *SIGKDD Explorations*, Vol. 2(1), pp. 65-69, 2000.

10. Helma, C. and Gottmann, E. and Kramer, S. Knowledge Discovery and data mining in toxicology Technical Report, University of Freiburg, 2000.

11. Jovanoski, V. and Lavrac, N. Classification Rule with Apriori-C, In *Proceedings of the Integrating Aspects of Data Mining, Decision Support and Meta Learning workshop, co-located with PKDD'01*, 81-92, Freiburg, Germany, 2001.

12. King, R.D. and Feng, C. and Sutherland, A. Statlog: Comparison of classification algorithms on large real-world problems, In *Applied Artificial Intelligence*, 1995.

13. Liu, B. and Hsu, W. and Ma, Y. Integrating classification and association rules mining, In *Proceedings of the Fourth International Conference on Knowledge Discovery & Data Mining, KKDD'98*, AAAI Press, pp. 80-86, 1998.

14. Li, W. and Han, J. and Pei, J. CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules, In *Proceedings of the IEEE International Conference on Data Mining, ICDM'01*, San Jose, California, 2001.

15. Liu, B. and Ma, Y. and Wong, C. K. Classification using association rules: weaknesses and enhancements, In *Data mining for scientific applications*, Kumar, V. et al (eds), pp. 1-11, 2001.

16. Li, J. and Shen, H. and Topor, R. Mining the Smallest Association Rule Set for Predictions, In *Proceedings of the IEEE International Conference on Data Mining, ICDM'01*, San Jose, California, 2001.

17. Mannila, H. and Toivonen, H. Levelwise search and borders of theories in knowledge discovery In *Data Mining and Knowledge Discovery*, vol. 3(1), pp. 241-258, 1997.

18. Mannila, H. and Toivonen, H. Multiple uses of frequent sets and condensed representations, In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96*, pp. 189-194, Portland, Oregon, 1996.

19. Pasquier, N. and Bastide, Y. and Taouil, R and Lakhal, L. Efficient mining of association rules using closed itemset lattices. In *Information Systems* 24(1), pp. 25-46. 1999.

20. Quinlan, J. R. C4.5 Programs for machine learning Morgan Kaufmann, San Mateo, Californie, 1993.

21. Schaffer, C. Overfitting avoidance as bias, In *Machine Learning*, vol. 10, pp. 153-178, 1993.

22. Toivonen, H. Sampling large databases for association rules, In *Proceedings of the 22nd International Conference on Very Large Databases, VLDB'96*, Morgan Kaufmann, pp. 134-145, 1996.