δ-strong classification rules for predicting collagen diseases

Jean-François Boulicaut[1], Bruno Crémilleux[1,2]
[1] Laboratoire d'Ingénierie des Systèmes d'Information
INSA Lyon, Bâtiment Blaise Pascal
F-69621 Villeurbanne Cedex, France
Jean-Francois.Boulicaut@insa-lyon.fr
[2] Université de Caen - GREYC - CNRS UMR 6072
Campus Côte de Nacre
F-14032 Caen Cedex, France
Bruno.Cremilleux@info.unicaen.fr

Abstract

This is a contribution to the PKDD'01 discovery challenge with a classification point of view on Thrombosis Data. The goals of this challenge point out the search for patterns and features to predict thrombosis and classify collagen diseases. Our aim is to show the potential impact of δ-strong classification rules in such a domain. This technique relies on recent results in the association rule mining area in order to mine a condensed representation of potentially interesting rules for the characterization of classes.

1 Introduction

One popular data mining technique concerns knowledge discovery from frequent association rules. 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 "When properties A_1 and A_2 are true within the data, then property A_3 is often true". We provide a simple formalization of this task in Section 2.1.

Classification is another popular data mining technique. 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 building the so-called classifiers. Classification rules are quite popular for that purpose and the literature is abundant (see for example [9, 13]). In that context, a classification rule is a rule that concludes on one class value.

The key point of this research is to use efficient association rule mining techniques in order to identify classification rules and, as a first step, provide a symbolic characterization of the classes.

The starting idea is quite simple. Roughly speaking, we first mine association rules from the learning set and then we select in such a collection the rules that conclude on a class value. In [8], Freitas has shown the limitations of such a naive approach, emphasizing the differences between classification rules and association rules. Other researchers have studied the selection of classification rules from a collection of association rules, e.g., [3, 4, 10]. Interestingly, in these proposals, the identification of the classification rules is performed mainly as a post-processing step on standard association rules.

We use recent results in the association rule mining area in order to mine efficiently a condensed representation of potentially interesting rules, the so-called δ-strong rules, for the characterization of classes. It is possible to work on difficult contexts such as large, dense and highly-correlated learning sets. Let us note that medical data are often highly-correlated. Next, we show that it is rather easy to process the discovered δ-strong rules in order to get a cover of the classification rules that characterize the classes. We see in Section 2.2 that our aim is to produce the simplest rules w.r.t. their left-hand sides and the technique provides *every* simplest classification rule: given a classification rule, one wants that any ownn and proper subset of its left-hand side does not enable to conclude on the same class value.

We tackle this discovery challenge with the classification point of view. Goals of this challenge point out the search for patterns and features to predict thrombosis and classify collagen diseases. Our aim is to show the potential impact of δ-strong classification rules in domains like medical thrombosis data. This is a preliminary work and more investigations have to be done to validate the interest of the extracted rules. Section 2.2 introduces association rule mining and the concept of δ-strong classification rule. Section 3 presents the data preparation stage and Section 4 gives the results on the medical data for collagen diseases patients.

2 δ-strong rules to characterize classes

2.1 Association rule mining

Let us provide a simple formalization of δ-strong rule mining task.

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.*

In the context of this challenge, $\mathbf{R}$ is made up by the values indicated in Table 1. An attribute identifies a patient or disease feature. In practice, one can have hundreds of thousands of examples and hundreds of attributes. We will see in Section 3.2 that the space of data is here rather small, which does not mean that the task is easy from the classification point of view.

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* $B \in \mathbf{R} \setminus X$*.*

Examples of association rules extracted from the challenge data are given in Section 4. The intuitive meaning of a potentially interesting association rule $X \Rightarrow B$ is that when an example contains true, i.e., 1, 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, support) *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* $\mathcal{F}(X \cup \{B\}, \mathbf{r}) / \mathcal{F}(X, \mathbf{r})$*. Note that in this paper, we use an absolute frequency (a number of examples* $\leq |\mathbf{r}|$*) instead of the relative frequency* $\mathcal{F}(X \cup \{B\}, \mathbf{r}) / |\mathbf{r}|$ *in* $[0, 1]$*. Frequency is also called* support*.*

The standard association rule mining task concerns the discovery of *every* rule having its frequency and its confidence higher 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* $\mathcal{F}(X, \mathbf{r}) \geq \gamma$*.*

Algorithmic complexity of frequent itemset discovery is exponential with the number of attributes. Many research works (e.g., [2, 15, 11, 6]) concern the practical contexts for which such a discovery remains tractable, even though a trade-off is needed with the exact knowledge of the frequencies (a fundamental issue for classification, see Section 2.2) and/or the completeness of the extractions.

2.2 δ-strong rules

For class characterization, interesting association rules 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 $1 - (\delta/\gamma)$. When using δ-strong rules, we assume that δ is rather small.

From a technical perspective, δ-strong rules can be built from δ-free itemsets that will constitute their left-hand sides [6]. Due to the space limitation, we do not explain in details what is a δ-free itemset and how they are extracted (see [6, 7] for details). We just provide an intuition for the concept of δ-free itemset. It is clearly related to the concepts of closed itemsets in [11] and almost-closures in [5]. An itemset X is called δ-free if there is no a δ-strong rule that holds between two of its own and proper subsets. The case $\delta = 0$ is important: no rule with confidence equal to 1 hold between proper subsets of X. In fact, frequent closed itemsets are the closures of 0-free sets: the closure of an itemset X is the largest superset of X (w.r.t. set inclusion) that has the same frequency as X. 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 a γ-frequent δ-free set and its almost-closure. We use the research prototype `ac-miner-12` [6] for that purpose[1]. Given thresholds γ and δ, it provides the collection of frequent δ-free itemsets, their frequencies and the attributes in their almost-closures. It has been shown efficient even in the case of dense and higly-correlated data, i.e., in practical applications where `apriori`-like algorithms clearly fail.

Let us now indicate the property of minimal body which allows us to build δ-strong rules with a minimal left-hand side.

Property 1 (minimal body) *If X is a δ-free itemset and $X \Rightarrow B$ $[\delta]$ is a δ-strong rule with exactly δ exceptions, then X is the minimal set of items from which we can conclude on B with at most δ exceptions.*

It means that if $X \Rightarrow B$ $[\delta_1]$ is a δ-strong rule with δ_1 exceptions, there is no itemset Y, $Y \subset X$, such that $Y \Rightarrow B$ $[\delta_2]$ is a δ-strong rule with $\delta_2 \leq \delta_1$. In other words, it is possible to get the simplest rules, i.e., a cover of δ-strong rules. We argue that it is a fundamental issue for classification. Not only it prevents from over-fitting [14] but also it makes the classification of an example easier to explain. Experts are generally interested in an explicit characterization of the concepts that support classification. It provides a feedback on the application domain expertise that can be reused for further analysis.

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

Definition 6 (δ-strong classification rule) *A* δ-strong classification rule *is a δ-strong rule that concludes on one class value (i.e., C_i).*

It is shown in [7] that if $\delta < \gamma$, then some rule conflicts are avoided. For instance, if there is the δ-strong classification rule $R_1 : X \Rightarrow C_i$, then a δ-strong classification rule $R_2 : X \Rightarrow C_j$ with $i \neq j$ cannot appear. Furthermore, if $\delta < \gamma$,

[1] `ac-miner-12` has been implemented by A. Bykowski at INSA Lyon

there is no δ-strong classification rule $R_3 : X \cup Y \Rightarrow C_j$ with $i \neq j$. As this sufficient condition on γ and δ is quite reasonable in practice, our experiments have been done under this assumption.

3 Data preparation

The seven tables available on the web (`http://lisp.vse.cz/challenge/pkdd2001/`) have been loaded using the relational database management system (`PostgreSQL 7.0.3`). We noticed that some attributes were missing for table `LAB_EXAM` (i.e., delimiters between attributes are missing, which is different from missing values). For example, the patient of `ID`#2110 and `ed` 860609 has 34 attributes whereas 44 are expected. So, we decided not to use this table in this work. It is unfortunate because table `LAB_EXAM` embeds 57,452 tuples described by 44 attributes. It is the largest table in this dataset.

One advantage of using a relational database is to provide an overview of the data (see Section 3.1). Such a preliminar inspection helps for the understanding of the required data transformations, i.e., to prepare the data for the data mining tool. Let us explain some transformations that we performed on the three main tables (`PATIENT_INFO`, `DIAGNOSIS` and `ANTIBODY_EXAM`).

3.1 Tables `PATIENT_INFO`, `DIAGNOSIS` and `ANTIBODY_EXAM`

Table `PATIENT_INFO` contains 1239 tuples. 217 patients have a missing `description date` attribute, 248 a missing `first date` attribute and 365 either one or other missing date. The birth day of patient having `ID`#4500676 is unknown. There is a temporal inconsistency: the patient having `ID`#5713181 is born in 2007, May 28. At first, we included the age of a patient when he/she was recorded (it is computed by the difference between `description date` and `birthday`). But, it leads to define an attribute with a lot of missing values. Finally, we decided to drop that information.

Table `DIAGNOSIS` contains 1956 tuples. It includes the attribute `diagnosis` (e.g., collagen diseases) which corresponds to one of the goals of this challenge. There are 160 distinct diagnosis, but as indicated in [16] some of them occur under different names. Even if same diagnosis are grouped, there are still more than 100 and most of them have just few occurrences. It is not a sensible classification task to try to predict more than one hundred diagnosis from 1239 patients. So, among diagnosis belonging to collagen diseases (selected by a join with table `DISEASE`), we grouped the 7 most frequent diagnosis[2]. Our aim is then to predict these 7 collagen diseases (see Table 2 for their frequencies) from 1539 patients affected by them. We deleted the attribute `fromtable` of `DIAGNOSIS` which indicates from which table the data was taken in the raw data.

[2] APS ; BEHCET groups BEHCET, BEHCET(ENTERO), BEHCETINEUROJ, BEHCET(NEURO), BEHCET(SS), BEHCET (VASCULO), BEHCET-VASCULO ; MCTD ; PSS groups PSS, PSS(CREST), PSS(SCLERODERMATOMYOSITIS) ; RA groups RA, RA (SERONEGATIVE), RA(SERONEGATIVE) ; SJS groups SJS, SJS(CNS), SJSLUPOID HEPATITIS, SJSMCTD, SJS RA ; SLE groups SLE, SLE(DIFFUSE LE), SLE PREG

Table ANTIBODY_EXAM (801 tuples) reports on special laboratory examinations performed on some patients and the degree of thrombosis when it happened. 31 patients have the value UNKNOWN in ID. Attributes dealing with the concentration of three anti-cardiolipin antibodies (IGG, IGM, IGA) are continuous and have to be discretized to produce δ-strong classification rules. Each of them is split in order to obtain a number of new values between 4 and 6. At the moment, we did a rather pragmatic discretization which is a trade-off between the two following constraints. On the one hand, for an attribute, we tried to balance the number of tuples among each of its new values. On the other hand, we avoided to choose a split value in an area where there are a large number of tuples. Table 1 depicts the used thresholds. It is likely that these thresholds can be improved by knowledge on angiology. Moreover, if we are interested in binary attributes, discretization can be done according the frequencies of classes [12]. But, in this context, we can think that binary splits would lead to much loss of information.

Frequency of the attribute ANA (Anti-Nucleus Antibody concentration) shows that there is a single tuple with the value 4094 (patient ID#3184022). It might be a mistake (the proper value should be 4096?). As there are just 45 tuples with the value 4096, we merged tuples with values 1024, 4094 and 4096. The new value (≥ 1024) groups 104 tuples. There are a lot of missing values for attributes kct (656), rvvt (656, same tuples that contain missing values for kct) and lac (580). When kct (or rvvt) value is present, lac is always known. Finally, we deleted the attribute exam date as we do not search for temporal patterns in this preliminary experiment.

3.2 Resulting file

We joined the three recoded tables as explained in Section 3.1. We obtain then a file with 721 tuples described by 12 qualitative attributes (see Table 1). The table we got after this preparation stage is called COLLAGEN_DISEASE. Table 2 shows frequencies of the class. The loss of tuples is uneven according to collagen diseases.

Let us give some comments on the remaining tables. 657 ANA patterns are stored in the table ANA_PATTERN which has just two attributes: the ID of patients and the pattern which is an ordinal data with 4 values. Two of them have few tuples (12 in total) and pattern P (peripheral) and S (speckle) gather most of the tuples. Table THROMBOSIS contains 198 tuples about symptoms and attacks per patient. This table has only 76 distinct patients and among them, 66 have had one attack or more. If we join data coming from THROMBOSIS with COLLAGEN_DISEASE , we obtain just 76 patients with the whole information, that means few examples. The last table DISEASE lists all the diseases (65 tuples). We used it to select the collagen diseases from the diagnosis in table DIAGNOSIS.

attribute	values
diagnosis	APS, BEHCET, MCTD, PSS, RA, SJS, SLE
thrombosis	0, 1, 2, 3
sex	F, M
admission	+, -
status	~, +
acl_igg	0, $]0, 0.8]$, $]0.8, 1.2]$, $]1.2, 2.4]$, > 2.4
acl_igm	0, $]0, 1.5]$, $]1.5, 1.9]$, $]1.9, 2.7]$, $]2.7, 5]$, > 5
acl_iga	0, $]0, 3.4]$, $]3.4, 7.5]$, > 7.5
ana	0, 4, 16, 64, 256, $\geq$ 1024
kct	-, +, unknown
rvvt	-, +, unknown
lac	-, +, unknown

Table 1: Used data (collagen diseases)

collagen disease	No. of tuples in `DIAGNOSIS`	No. of tuples in `COLLAGEN_DISEASE`	Loss of tuples (%)
APS	73	61	16.4
BEHCET	108	16	85.2
MCTD	75	41	45.3
PSS	104	27	74.0
RA	275	73	73.4
SJS	450	214	52.4
SLE	454	289	36.3

Table 2: Number of tuples of each used collagen diseases

4 Results and discussion

Data have to be translated in a binary format to be performed by `ac-miner-12`. This process is automated by producing a binary item for each pair of attribute/value and missing values. We got 56 binary items. To better evaluate results in classification, the file coming from `COLLAGEN_DISEASE` was split into a training file (580 examples, i.e., 4/5 of data) and a test file (141 examples, i.e., 1/5 of data). Class has the same frequency distribution in each file and in the whole data.

The prototype `ac-miner-12` is implemented in C++. We used a PC with 768 MB of memory and a 500 MHz Pentium III processor under Linux operating system. The training set is small and `ac-miner-12` has been designed for large databases. As a result, extraction time is very fast even with low frequency thresholds (experiments with $\gamma = 3$, i.e., 0.5%). The longest extraction time has been 81 seconds. In the following, we do not report the extraction time.

For each pair of γ and δ, Table 3 shows the number of δ-free and almost-closures that contain an item belonging to the class. This last number can be seen as the number of potential δ-strong classification rules (i.e., with any support and confidence values).

γ	δ	No. of δ-free	No. of almost-closure with class
3	0	21404	373
3	1	13884	1054
3	2	9263	1342
6	0	14067	48
6	1	9604	154
6	2	6792	317
6	3	5396	526
6	4	4448	715
6	5	3792	801

Table 3: δ-free sets and almost-closures

γ	δ	well-classified
3	0	34.04
3	1	41.84
3	2	46.81
6	0	15.60
6	1	29.79
6	2	41.84
6	3	48.94
6	4	46.81
6	5	46.81

Table 4: Classification results (test file)

In this context, we are here able to extract classification rules with a confidence value of 100% (i.e., $\delta = 0$). With $\gamma = 6$, there are 48 rules, 39 concluding on `SLE` and 9 on `SJS`. With $\gamma = 3$, there are 373 rules, 21 concluding on `APS`, 3 on `BEHCET`, 6 on `MCTD`, 2 on `RA`, 71 on `SJS` and 270 on `SLE`. For instance, we have the following rules ($\mathcal{F}(n)$ denotes the frequency of the rule):

acl_igg > 2.4 and acl_igm $\in]1.9, 2.7]$ and kct = - $\Rightarrow$ SLE $\mathcal{F}(10)$
thrombosis $= 2$ and acl_igg > 2.4 and kct = - $\Rightarrow$ SLE $\mathcal{F}(9)$
thrombosis $= 0$ and sex = F and acl_igm $\in]0, 1.5]$ and ana $= 4$ $\Rightarrow$ SJS $\mathcal{F}(8)$
acl_igg $\in]0, 0.8]$ and acl_iga $\in]0, 3.4]$ $\Rightarrow$ SJS $\mathcal{F}(7)$

Here are some rules extracted with $\delta > 0$:

ana = 256 and rvvt = - $\Rightarrow$ SLE	$\mathcal{F}(24)$	confidence = 86%
acl_igm $\in]2.7, 5]$ and lac = - $\Rightarrow$ SLE	$\mathcal{F}(20)$	confidence = 87%
acl_igm $\in]0, 1.5]$ and ana = 4 $\Rightarrow$ SJS	$\mathcal{F}(8)$	confidence = 73%
sex = M and ana = 0 $\Rightarrow$ BEHCET	$\mathcal{F}(7)$	confidence = 58%
thrombosis = 1 and admission = + and acl_iga $\in]3.4, 7.5]$ $\Rightarrow$ APS	$\mathcal{F}(6)$	confidence = 86%

Sets of δ-strong classification rules can be used to predict collagen diseases from patient test file. Results are given in Table 4. When there is a conflict (several rules with different conclusions are triggered from a same patient), a score incorporating the support and the confidence of each rule is computed and the class value having the best score is predicted.

Best results are around 48%. According to us, such results are explained by the following reasons:

- The prediction is here based on a cover of the classification rules that characterizes the classes. This cover includes rules with low support and/or confidence. Such rules with a poor quality may introduce errors. The design of a classifier stemming from the classification rules cover still needs some research. As a preliminary result, for each cover of rules, we computed for each rule a score (denoted Δ) which is the difference between the well-classified and the miss-classified examples of the test file. For each cover, we sorted its rules w.r.t. Δ and we defined a family of nested sets of rules: 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 $\Delta_2 \leq \Delta_1$). Then, we used the sets of rules obtained with the higher Δ values to classify again examples of test files. When δ tends to have the same value than γ, the well-classified rate increases of 10% (for instance, with $\gamma = 3$ and $\delta = 2$: 58%, with $\gamma = 6$ and $\delta = 4$: 56%). More investigations are required to confirm such results.

- We have to cope with a classification task with seven classes, which is harder than a binary decision. Moreover, number of examples w.r.t. the number of classes is rather low and the class frequency is uneven.

- Medical domain intrinsically embeds uncertainty. In other words, the same cause does not always produce the same effect and/or a number of parameter values which could explain the phenomenon are unknown. That is why such tasks are generally hard.

Table 3 and 4 highlight the role of δ. For instance, with $\gamma = 6$ and $\delta = 0$, we have seen above that only classification rules concluding on `SLE` and `SJS` are extracted. With $\delta = 2$ (and same γ), classification rules on all classes (except `PSS`) are found and with $\delta = 4$, there are rules concluding on all classes. That is typical for uncertain domains like medicine. Finally, let us note that prediction performances are better with $\delta > 0$. This is due to the lack of rules on some classes with $\delta = 0$ but it highlights also over-fitting: rules without exceptions (or too few exceptions) w.r.t. γ may be over-specified and do not reflect a sound knowledge about the domain.

5 Conclusion and further work

We showed the potential impact of δ-strong classification rules in discovery challenge data for predicting collagen dieases.

A straightforward idea is to perform new experiments when keeping only examples for which the diagnosis is confirmed (attribute `status` has the value `+`). The file contains then 662 tuples (instead of 771). The loss of data is quite low and we can think that information will be more reliable, then better for prediction.

A similar work can be done to predict the attribute `thrombosis` coming from `ANTIBODY_EXAM` . In this case, we might use table `THROMBOSIS`. Nevertheless, as seen in Section 3.2, the join of `THROMBOSIS` with `COLLAGEN_DISEASE` leads to get the whole information just on few examples.

Acknowledgments. The authors thank Christophe Rigotti for stimulating discussions.

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, pages 207–216, 1993.

[2] Agrawal, R. and Mannila, H. and Srikant, R. and Toivonen, H. and Verkano, I. Fast discovery of association rules, In *Advances in Knowledge Discovery and Data Mining*, AAAI Press, pages 307–328, 1996.

[3] Ali, K. and Manganaris, S. and Srikant, R. Partial classification using association rules, In *Proceedings KDD'97*, AAAI Press, pages 115-118, 1997.

[4] Bayardo, R. J. Brute-force mining of high-confidence classification rules, In *Proceedings KDD'97*, AAAI Press, pages 123-126, 1997.

[5] Boulicaut, J. F. and Bykowski, A. Frequent closures as a concise representation for binary data mining, In *Proceedings PAKDD'00*, Springer-Verlag LNAI 1805, pages 62-73, 2000.

[6] Boulicaut, J. F. and Bykowski, A. and Rigotti, C. Approximation of frequency queries by means of free-sets, In *Proceedings PKDD'00*, Springer-Verlag LNAI 1910, pages 75-85, 2000.

[7] Boulicaut, J. F. and Crémilleux B. δ-strong rules to characterize classes, Research Report INSA Lyon-LISI, July 2001, 12 p. Submitted.

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

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

[10] Liu, B. and Hsu, W. and Ma, Y. Integrating classification and association rules mining, In *Proceedings KDD'98*, AAAI Press, pages 80-86, 1998.

[11] 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), pages 25-46. 1999.

[12] Rabaséda, S. and Sebban, M. and Rakotomalala, R. Discretisation of continous attributes: a survey of methods. In *Proccedings JCIS'95*, pages 164-166, 1995.

[13] Salzberg, S. On comparing classifiers: pitfalls to avoid and a recommended approach, In *Data Mining and Knowledge Discovery*, Vol. 3(1), pages 317-327, 1997.

[14] Schaffer, C. Overfitting avoidance as bias, In *Machine Learning*, Vol. 10, pages 153-178, 1993.

[15] Toivonen, H. Sampling large databases for association rules, In *Proceedings VLDB'96*, Morgan Kaufmann, pages 134-145, 1996.

[16] Zytkow, J. M. and Tsumoto, S. and Takabayashi, K Medical (Thrombosis) data description In *Proceedings Discovery Challenges, PKDD'00*, 2000.