

Pattern-Based Decision Tree Construction

Dominique Gay, Nazha Selmaoui
 ERIM - University of New Caledonia
 BP R4 F-98851 Nouméa cedex, France
 {dominique.gay, nazha.selmaoui}@univ-nc.nc

Jean-François Boulicaut
 INSA-Lyon, LIRIS CNRS UMR5205
 F-69621 Villeurbanne cedex, France
 jean-francois.boulicaut@insa-lyon.fr

Abstract

Learning classifiers has been studied extensively the last two decades. Recently, various approaches based on patterns (e.g., association rules) that hold within labeled data have been considered. In this paper, we propose a novel associative classification algorithm that combines rules and a decision tree structure. In a so-called δ -PDT (δ -Pattern Decision Tree), nodes are made of selected disjunctive δ -strong classification rules. Such rules are generated from collections of δ -free patterns that can be computed efficiently. These rules have a minimal body, they are non-redundant and they avoid classification conflicts under a sensible condition on δ . We show that they also capture the discriminative power of emerging patterns. Our approach is empirically evaluated by means of a comparison to state-of-the-art proposals.

1. Introduction

The popular association rule mining task [1] has been applied not only for descriptive tasks but also for class characterization and classifying tasks when considering labeled transactional data [11, 10, 5, 16, 2]. An association rule π is an implication of the form $X \Rightarrow Y$ where X (condition) and Y (consequence) are different sets of features (also called items or Boolean attributes). Such rules provide feedback on attribute value co-occurrences. When Y denotes a single class value, it is possible to look at the predictive power of such association rules and to look at them as candidate classification rules: when conjunctions of value X is observed, is it accurate to predict class value Y . Such a shift between descriptive and predictive tasks has to be done carefully [9], and this is an identified research topic called associative classification. The idea is to post-process the association rules that conclude on class values for using the *fewest* (with least information loss) but the *most interesting* ones for classifying purposes. Interestingness is approached by support and confidence measures which

are used in CBA [11] to rank the rules in a list that will be considered as a classifier. Then, pruning based on database coverage can be applied to reduce the size of such a list. An unseen case t is finally labeled by the first verified classification rule in the list. Other approaches like CMAR [10] or CPAR [16] define class-related scores – respectively *combined effect* of subsets of rules and *average expected accuracy* of the best k rules – then choose the class that maximizes this score. Further properties can be enforced on selected rules like, for instance, a minimal body property [5, 2]. Another important way to support classification is to consider emerging patterns [6], i.e., patterns that are frequent in samples of a given class and barely infrequent for samples of the other classes. Many algorithms have been developed. For instance, SJEPClassifier [7] is based on the aggregation of emerging patterns and the computation of a collective score (see, e.g., [14] for a survey).

Both associative classification and approaches based on emerging patterns follow the same principle : (1) mining set patterns or rules that support class characterization, (2) pruning this set to select best rules, (3) combine extracted rules to classify unseen samples. In this paper, we propose a novel associative classification approach called δ -PDT. It combines decision tree construction and the disjunction of δ -strong rules (i.e., association rules that conclude on a class value and are violated by at most δ samples, δ is assumed to be quite small w.r.t. the size of labeled data). These rules have minimal body (i.e., any subset of the body of a rule cannot conclude on the same class), and [5] points out a condition on δ to avoid classification conflicts. δ -strong rules are based on the so-called δ -free patterns, i.e., a class of pattern which has been introduced for frequency approximation in difficult frequent set mining tasks [4].

We introduce new conditions on the δ parameter such that δ -free sets become emerging patterns that have a discriminating power property. Furthermore, when $\delta > 0$, we can consider a "near equivalence" perspective and thus, roughly speaking, the concept of almost-closed set (maximal one) and the notion of δ -Closure Equivalence Class (δ -CEC). It means that we can consider the disjunction of δ -

free sets of the same δ -CEC. The construction of the δ -PDT tree roughly follows the popular C4.5 decision tree building algorithm. Main difference is that test nodes of a δ -PDT are bodies of disjunctive δ -strong rules selected w.r.t. an extended version of Information Gain Ratio criterion.

In Section 2, we provide the preliminary definitions and the problem setting. Section 3 gives more details on how to use δ -strong rules when considering pattern-based decision tree building. Experiments are reported in Section 4. Section 5 is a brief conclusion.

2. Problem setting and definitions

We want to support classification tasks when considering that samples are O/1 data records. We assume a binary database $r = \{\mathcal{T}, \mathcal{I}_C, R\}$ where \mathcal{T} is a set of *objects*, $\mathcal{I}_C = \mathcal{I} \cup \mathcal{C}$ a set of *items* (boolean attributes and classe attributes) and R a subset of $\mathcal{T} \times \mathcal{I}_C$. A *pattern* (or item-set) $I \subseteq \mathcal{I}_C$ is a set of items and it is said to cover a set of objects $T \subseteq \mathcal{T}$ if $I \subseteq t$. T is the *support* of I in r and is noted $supp(I, r)$. $freq_a(I, r) = |supp(I, r)|$ is known as *absolute frequency* of I while *relative frequency* of a pattern I in r , noted $freq_r(I, r)$, is defined as $\frac{|supp(I, r)|}{|r|}$. Given γ a frequency threshold, a pattern I is said to be γ -frequent in r if $freq_a(I, r) \geq \gamma$.

2.1. Emerging patterns

An emerging pattern is a pattern whose relative frequency is significantly high in r_{c_i} (database restricted to objects of class c_i) compared with its relative frequency in the rest of the database ($r \setminus r_{c_i}$). This ratio of frequencies is measured by growth rate. The *growth rate* of a pattern I related to r_{c_i} is formally defined as:

$$GR(I, r_{c_i}) = \begin{cases} 0 & \text{if } freq_r(I, r_{c_i}) = 0 \\ \infty & \text{if } freq_r(I, r_{c_i}) > 0 \\ & \wedge freq_r(I, r_{c_j}) = 0 \\ & (\forall j \neq i) \\ \frac{freq_r(I, r_{c_i})}{freq_r(I, r \setminus r_{c_i})} & \text{otherwise} \end{cases}$$

Given a growth rate threshold $\rho > 1$, r_{c_i} - ρ -emerging patterns (ρ -EPs) related to r_{c_i} are patterns I s.t. $GR(I, r_{c_i}) > \rho$, and *Jumping Emerging Patterns* (JEPs) are the patterns whose growth rate is infinite. Emerging patterns own an intrinsic power of class discrimination whose strength is related to growth rate. Intuitively, higher growth rate implies more power to discriminate classes.

2.2. Closure equivalence classes

A pattern I is said to be *closed* in r if and only if $cl(I, r) = I$, where $cl : \mathcal{P}(\mathcal{I}) \rightarrow \mathcal{P}(\mathcal{I})$ is a *closure*

operator such that $cl(I, r) = \{i \in \mathcal{I} \mid freq_a(I, r) = freq_a(I \cup \{i\}, r)\}$. In other words, there is no superset of I with the same frequency than I in r , i.e. $\nexists I' \supseteq I : freq_a(I', r) = freq_a(I, r)$. Intuitively, a closed pattern is the unique maximal description (w.r.t. \subseteq) commonly shared by a set of objects T_1 . Furthermore, minimal descriptions of T_1 are called free patterns in [4] and key patterns in [3]. Intrinsic properties of free patterns (minimality, concision, non-redundancy) are captured in essential classification rules [2] to manage classification tasks and to improve associative classifiers like, for instance, CBA or CPAR.

If we consider the relation \sim_{cl} , meaning 'has the same closure as', we can group patterns by *closure equivalence classes* (CECs). The unique maximal (w.r.t. set size) element of a CEC C_1 is a closed pattern and its minimal elements are the so-called free patterns, i.e. minimal descriptions for a same set of supporting objects T_1 [3]. Inside CECs, every association rule $\pi : I \Rightarrow i$ (where I is a free pattern and $i \in cl(I, r) \wedge i \notin I$) is an exact rule (confidence 1). We think π is relevant for classification if i is a class label c . Indeed, in this case, π is a rule with confidence 1 and I is a JEP. Unfortunately, in some real (noisy) data sets, such rules tend to be rare. Instead, we propose to consider rules with at most δ exceptions ($\delta \geq 1$).

Definition 1 (δ -strong rule, δ -free pattern, almost closure). Let δ be an integer. A δ -strong rule is an association rule $I \Rightarrow^\delta i$ which is violated in at most δ transactions, and where $I \subseteq \mathcal{I}$ and $i \in \mathcal{I} \setminus I$. Thus, $I \Rightarrow^\delta i$ is a valid δ -strong rule in r if $freq_a(I, r) - freq_a(I \cup \{i\}) \leq \delta$. A pattern $I_1 \subseteq \mathcal{I}$ is a δ -free pattern iff there is no valid δ -strong rule $I \Rightarrow^\delta i$ s.t. $I \cup \{i\} \subseteq I_1$. If $\delta = 0$, δ is omitted, and then we talk about strong rules and free patterns. When $\delta > 0$, the almost-closure of a δ -free pattern I is $cl_\delta(I, r) = \{i \in \mathcal{I} \mid freq(I, r) - freq(I \cup \{i\}) \leq \delta\}$.

One can consider relation \sim_{cl_δ} denoting 'has the same almost-closure as' to group patterns by *almost-closure equivalence classes* (δ -CECs); δ -free patterns being minimal elements of these equivalence classes. A δ -strong rule $I \Rightarrow^\delta i$ can be built from a γ -frequent δ -free pattern I and an item $i \in cl_\delta(I, r)$ ($i \notin I$) [4]. It is known to have a minimal body and a confidence greater or equal than $1 - \delta/\gamma$. To characterize classes, [5] used δ -strong classification rules (δ -strong rules concluding on a class). To build δ -strong classification rules, we need to compute γ -frequent patterns not containing class but whose almost-closure contains a class. This could be managed efficiently with the AC_like implementation of the algorithm from [4] (see <http://liris.cnrs.fr/jeremy.besson/>). These rules are known to avoid common classification conflicts under a simple constraint. Indeed, δ -strong rule extraction can not produce $\pi_1 : I \Rightarrow^\delta c_1$ and $\pi_2 : J \Rightarrow^\delta c_2$ with $I \subseteq J$ (bodies conflicts) if we state $\delta \in [0; \lfloor \gamma/2 \rfloor]$.

Moreover, let us see how δ -strong classification rules may catch the discriminating power of emerging patterns. Consider a p -class learning problem. Let $\pi : I \Rightarrow^\delta c \ \{\delta_1\}$ be a δ -strong classification rule built from a γ -frequent δ -free pattern I s.t. $\gamma_1 \geq \gamma$ is the exact frequency of I and $\delta_1 \leq \delta$ the exact number of violations of π in r . $GR(I, r_c) = \frac{freq_r(I, r_c)}{freq_r(I, r \setminus r_c)}$ and relevant growth rate is greater than 1, i.e., when:

$$\begin{aligned} \text{i.e.,} \quad & \frac{(\gamma_1 - \delta_1)/|r_c|}{\delta_1/|r \setminus r_c|} \geq \frac{(\gamma - \delta)}{\delta} \cdot \frac{|r \setminus r_c|}{|r_c|} > 1 \\ \text{i.e.,} \quad & \frac{\gamma}{\delta} > \frac{|r_c|}{|r \setminus r_c|} + 1 \\ \text{i.e.,} \quad & \frac{\gamma \cdot |r \setminus r_c|}{|r_c| + |r \setminus r_c|} > \delta \\ \text{i.e.,} \quad & \gamma \cdot \frac{|r \setminus r_c|}{|r|} > \delta \end{aligned}$$

Thus, $\delta < \gamma \cdot \frac{|r \setminus r_c|}{|r|}$ is a sufficient condition s.t. δ -free patterns (bodies of δ -strong classification rules) are ρ -emerging patterns ($\rho > 1$).

Proposition 1. *A γ -frequent δ -free pattern I s.t. a class $c \in cl_\delta(I, r)$ is an emerging pattern if $\delta \in [0; \gamma \cdot (1 - freq_r(c_i, r))]$ [where $|r_{c_i}| \geq |r_{c_j}| \ \forall j \neq i$].*

These classification rules have minimal bodies, they avoid common classification conflicts, and they have class discriminating power of EPs. In the following, when talking about δ -CECs or δ -strong classification rules, we only consider couples (γ, δ) that produce no body conflict and which satisfy the constraint induced by Proposition 1.

3. Classification using δ -CECs

Our idea is to build a classifier combining δ -CECs with a decision tree. We consider that a δ -CEC C_1 is *relevant* for our purpose if its maximal element (w.r.t. \subseteq) contains a class and at least one of its minimal elements (that are EPs) does not contain a class attribute. Indeed, potentially interesting classification rules related to a δ -CEC C_1 are those δ -strong rules $\pi : I \rightarrow c$ where I is a minimal element of C_1 (i.e., a γ -frequent δ -free pattern) and c the class label which C_1 is related to.

When C_1 leads to several δ -strong rules ($\pi_1 : I_1 \Rightarrow c \ \{\delta_1\}, \dots, \pi_k : I_k \Rightarrow c \ \{\delta_k\}$), we gather the k rules within a unique classification rule π .

Definition 2 (disjunctive δ -strong classification rule). *The unique disjunctive δ -strong classification rule of a δ -CEC C_1 is a classification rule of the form $\pi : I_1 \vee \dots \vee I_k \rightarrow c$ where I_1, \dots, I_k are δ -free patterns of C_1 that do not contain a class attribute.*

Intuitively, using a disjunctive δ -strong classification rule $\pi : I_1 \vee \dots \vee I_k \rightarrow c$ related to C_1 , a new object t matches C_1 descriptions and is similar to the group of objects described by C_1 if t respects at least one minimal element of C_1 (i.e., t could be labeled with c). Disjunction of

(δ) -free patterns is a way to choose a unique representant of C_1 for description. Disjunctive δ -strong classification rules differ from extracted rules of associative classifiers (CBA and CPAR) since rules concerning a same amount of objects are packed into a unique concise rule.

3.1. Towards a pattern-based decision tree

The construction of a Pattern Decision Tree (Algorithm 1) is based on the recursive procedure of C4.5 [13] algorithm to build decision trees – the main difference is that we are not dealing with attributes but with disjunctive δ -strong classification rules. At each step, the procedure determines which rule is the *most* discriminant (call it *BestRule*), then it splits training data set r into two groups – the one ($r_{BestRule}$) covered by *BestRule* and the other ($r_{\neg BestRule}$). Next, another rule is chosen for each of the groups for further splitting until all of groups are considered to be *terminal*. In such a tree, each test node is a disjunction of δ -free patterns (bodies of δ -strong classification rules used for splitting) and each branch leads to a subset of the training data.

Following the C4.5 principle, the most discriminant disjunctive δ -strong classification rule (*BestRule*) is chosen w.r.t. an entropy-based measure (*InformationGainRatio IGR*). We extend *IGR* definition – based on *Entropy (E)*, *Information Gain (IG)* and *Split Information (SI)* – for a classification rule $\pi : I \rightarrow c$ where I could be a disjunction of δ -free patterns.

Definition 3 (Information gain ratio).

$$E(r) = - \sum_{i=1}^{i=|C|} freq_r(c_i, r) \cdot \log_2(freq_r(c_i, r))$$

$$\begin{aligned} IG(\pi, r) &= E(r) - \left[freq_r(I, r) \cdot E(r_\pi) \right. \\ &\quad \left. + (1 - freq_r(I, r)) \cdot E(r_{\neg \pi}) \right] \\ SI(\pi, r) &= -freq_r(I, r) \cdot \log_2(freq_r(I, r)) \\ &\quad - (1 - freq_r(I, r)) \cdot \log_2(1 - freq_r(I, r)) \\ IGR(\pi, r) &= \frac{IG(\pi, r)}{SI(\pi, r)} \end{aligned}$$

Note that deciding if a position is terminal and pruning is also performed “à la C4.5”.

Remark 1. *PDTBuild is launched with following parameters: r is the involved binary database, $S_r^{\gamma, \delta}$ is the set of disjunctive δ -strong classification rule extracted from r according to user-defined values for γ and δ , and PDT is an empty tree.*

Algorithm 1 $PDTBuild(r, S_r^{\gamma, \delta}, PDT)$

```
if  $IsTerminal(r, S_r^{\gamma, \delta})$  then
   $PDT.class \leftarrow \arg \max_{c_i \in \mathcal{C}} |r_{c_i}|$ 
else
   $BestRule \leftarrow \arg \max_{\pi \in S_{\gamma, \delta}} GainRatio(\pi, r)$ 
   $PDT.left \leftarrow PDTBuild(r_{BestRule}, S_{r_{BestRule}}^{\gamma, \delta}, \phi)$ 
   $PDT.right \leftarrow PDTBuild(r_{\neg BestRule}, S_{r_{\neg BestRule}}^{\gamma, \delta}, \phi)$ 
end if
return  $PDT$ 
```

4. Empirical validation

We performed many tests to evaluate the accuracy of a δ -PDT classifier. We processed 8 data sets from the UCI Machine Learning Repository [12] and another real-world data set meningitis¹. When necessary, all data sets have been discretized with entropy-based method [8], then binarized to enable 0/1 data analysis by means of AC_like, i.e., the implementation we use for computing δ -strong rules. We compared δ -PDT with other well-known state-of-the-art techniques (i.e., using C4.5, CBA, CPAR, SJEP-classifier).

C4.5 experiments were performed with WEKA platform [15] and 10-folds stratified cross validation. We used CBA (2.0 demo version available online); frequency threshold and minimum confidence were respectively set to 1% and 50%. We also used CPAR online release. Reported results for SJEP-classifier come from [7]. For δ -PDT, relative frequency of δ -free patterns varies within [0%;10%] (except for car data set where relative frequency varies within [0%;0.8%] since higher thresholds lead to irrelevancy) and δ is constrained w.r.t. Property 1.

We report in Table 1 accuracy results for δ -PDT. The (average column indicates accuracy average on γ and δ values (w.r.t. guided choice, see end of Section 4) and the best column stands for best accuracy obtained for a certain combination of γ and δ : relative frequency and δ are reported) for comparison with other classifiers. Bold faced results are the maximum obtained for a data set and blank cells are due to absence of results in the published papers.

It appears that δ -PDT is generally comparable to classifiers obtained by means of the state-of-the-art methods when considering weighted average overall γ and δ tested values. It often achieves the highest accuracies for some values of γ and δ (see last column).

Thus, δ -PDT has two parameters: $minfreq_r$ (relative frequency threshold) and δ . On one hand, frequency threshold dilemma is well-known: lower thresholds imply huge amount of rules among which some are useless since

¹meningitis concerns children hospitalized for acute bacterial or viral meningitis.

they only concern a few data; but higher thresholds generate too few rules to reasonably cover training data. On the other hand, δ is constrained by frequency threshold ($\delta \in [0; \gamma \cdot freq_r(c_i, r)[$ where c_i is the major class in database): when $\delta = 0$, free patterns (bodies of strong classification rules) become JEPs with high discriminating power and cover a low proportion of training data, but high values of δ generally lead to bodies of δ -strong rules with low growth rate which is less discriminant. Note that the higher the coverage proportion is, the more our rules set will be representative of the training data. Graph in Fig. 1 shows the evolution of coverage proportion for heart training data according to frequency and δ values. For other data sets, graphs are similar: we observe that coverage proportion grows with δ , then stabilizes. We think δ values around the stabilization point is a good choice, since lower values imply a loss of coverage proportion and higher values generate patterns with lower growth rate. In Table 1, average column reports average of accuracy results obtained with these δ values.

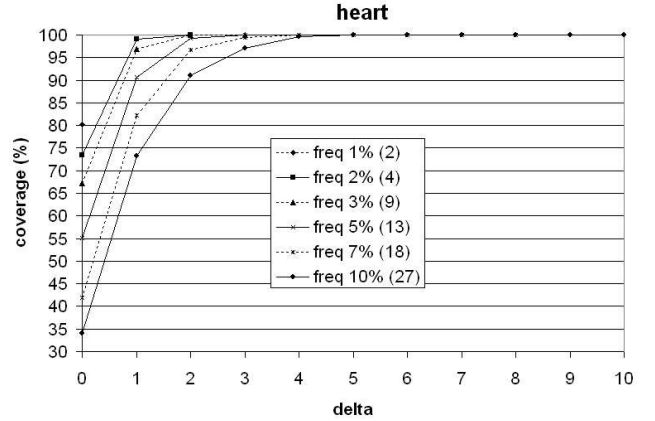


Figure 1. Effect of frequency and δ on coverage of training data.

5. Conclusion and future work

We proposed a novel classification approach based on both association rules and decision trees. More precisely, we integrated the discriminating power of disjunctive δ -strong classification rules into a decision tree structure. Disjunctive δ -strong classification rules are built from γ -frequent δ -free patterns having a same δ -closure equivalence class. These rules are concise and they have a minimal body property. Under a condition on δ , they are also known to avoid some classification conflicts and we demonstrated that they capture the discriminating power of EPs. The collection of extracted rules can be post-processed thanks to an

Data sets	C4.5	CBA	CPAR	SJEPs	δ -PDT		
					average	best	($freq_r; \delta$)
car	92.36	88.90	92.65	-	93.98	94.10	{(0;2;0)}
cleve	78.88	83.83	83.61	82.41	82.07	83.83	{(10;{3,4})}
heart	83.70	81.87	83.70	82.96	83.49	85.93	{(5;3)}
hepatic	82.58	81.82	84	83.33	82.74	85.16	{(10;2),(5;0)}
horse-colic	85.05	81.02	84.14	84.17	81.93	84.51	{(10;{12,13})}
iris	93.33	95.33	94.67	-	95.11	95.33	{(7;{2,3,4}),(5;{2,3,4}),(3;2),(2;2)}
labor	82.46	86.33	91.17	82.00	85.61	87.72	{(10;0),(7;0),(5;0)}
lymph	77.03	84.50	80.28	-	81.98	86.49	{(10;3)}
meningitis	94.83	91.79	91.52	-	92.25	95.13	{(3;2)}
sonar	79.81	79.81	84.07	85.10	79.33	81.25	{(10;5)}
vehicle	69.98	67.99	73.3	71.36	70.28	71.04	{(1;5)}
wine	96.07	94.96	97.54	95.63	96.63	97.19	{(10;6)}
Average	84.67	84.85	86.72	83.37	85.45	87.31	-

Table 1. Accuracy comparison for δ -PDT

extended version of *Information Gain Ratio* while building the final decision tree. Thus, our method needs for only two parameters (frequency threshold γ and δ). We also gave indications to guide user choice of parameters w.r.t. training data coverage. Experimental comparisons with state-of-the-art methods are quite encouraging. Since δ tunes the number of errors allowed for a classification rule, we now plan to investigate noise-tolerant opportunities of our classifying method.

Acknowledgments. The authors wish to thank P. François and B. Crémilleux who provided meningitis data. This work is partly funded by EU contract IST-FET IQ FP6-516169.

References

- [1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings 20th International Conference on Very Large Data Bases VLDB'04*, pages 487–499. Morgan Kaufmann, 1994.
- [2] E. Baralis and S. Chiusano. Essential classification rule sets. *ACM Transactions on Database Systems*, 29(4):635–674, 2004.
- [3] Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, and L. Lakhal. Mining frequent patterns with counting inference. *SIGKDD Explorations*, 2(2):66–75, 2000.
- [4] J.-F. Boulicaut, A. Bykowski, and C. Rigotti. Approximation of frequency queries by means of free-sets. In *Proceedings 4th European Conf. on Principles of Data Mining and Knowledge Discovery PKDD'00*, volume 1910 of *LNCS*, pages 75–85. Springer-Verlag, 2000.
- [5] J.-F. Boulicaut and B. Crémilleux. Simplest rules characterizing classes generated by delta-free sets. In *Proceedings 22nd Int. Conf. on Knowledge Based Systems and Applied Artificial Intelligence ES'02*, pages 33–46.
- [6] G. Dong and J. Li. Efficient mining of emerging patterns: discovering trends and differences. In *Proceedings 5th ACM Int. Conf. on Knowledge Discovery and Data Mining KDD'99*, pages 43–52, New York, USA, 1999. ACM Press.
- [7] H. Fan and K. Ramamohanarao. Fast discovery and the generalization of strong jumping emerging patterns for building compact and accurate classifiers. *IEEE Transactions on Knowledge and Data Engineering*, 18(6):721–737, 2006.
- [8] U. M. Fayyad and K. B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pages 1022–1027, 1993.
- [9] A. A. Freitas. Understanding the crucial differences between classification and discovery of association rules - a position paper. *SIGKDD Explorations*, 2(1):65–69, 2000.
- [10] W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on multiple class-association rules. In *Proceedings IEEE Int. Conf. on Data Mining ICDM'01*, pages 369–376, San Jose, California, 2001. IEEE Computer Society.
- [11] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In *Proceedings 4th Int. Conf. on Knowledge Discovery and Data Mining KDD'98*, pages 80–86. AAAI Press, 1998.
- [12] D. Newman, S. Hettich, C. Blake, and C. Merz. UCI repository of machine learning databases, 1998.
- [13] J. R. Quinlan. *C4.5 : programs for machine learning*. Morgan Kaufmann, San Francisco, USA, 1993.
- [14] K. Ramamohanarao and H. Fan. Patterns based classifiers. *World Wide Web*, 10(1):71–83, 2007.
- [15] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques (2nd edition)*. Morgan Kaufmann Publishers Inc., San Francisco, USA, 2005.
- [16] X. Yin and J. Han. Cpar : Classification based on predictive association rules. In *Proceedings 3rd SIAM Int. Conf. on Data Mining SDM'03*. SIAM, 2003.