Measures of Class Membership in Association Rule based Classification

Viet Phan-Luong
Laboratoire d'Informatique Fondamentale de Marseille
UMR CNRS 6166, CMI de Université Aix-Marseille
39 rue F. Joliot Curie, 13453 Marseille, France
viet.phanluong@lif.univ-mrs.fr

Abstract

In this work, we focus on the measures of class membership defined for classifiers based on associative classification rules. We revisit the χ^2 test for defining an effective measure. For comparison, we adapt the weight of evidence (Wang and Wong TKDE 2003) for a system based on the notions of support and confidence. Some variants of those measures are also defined. The effect of the defined measures is verified through the experimentation on categorical UCI datasets.

1. Introduction

The association rule based classification has received extensive attention in data mining and machine learning research [1], [2], [3], [4], [5], [6], [7]. The approach consists of two main steps: (i) Extraction of a set of class-association rules, called a classifier, from the learning data, and (ii) Selecting significant rules in this set for classifying unseen data. In general, the high quality rules are defined on the notions of support and confidence. The support of a rule shows how frequent the rule classifies correctly the learning data, and the confidence represents the precision of the rule. Further heuristics to enhance the quality of selected rules based on statistical tools, such as the chi-square test [2], the correlation coefficient [6], have been proposed.

Instead of basing on support and confidence, another approach to select interesting rules is to base on residual analysis in statistics [7]. In this approach, a pattern is considered significant if its frequency of occurrence is significantly different from its expected occurrences, using a test based on the notion of standardized residual. For classifying an unseen object, the effect of each class label on the object is considered in a similar manner, using a measure called the weight of evidence. This is a competitive measure in classification.

In this work, we focus in the measures to quantify the effect of a class label on an unseen object, using the rules in the classifier built on small size key itemsets. We revisit the χ^2 test to define an effective measure. We adapt the weight of evidence in the context of classification based on the notions of support and confidence. To improve those measures, some variants are defined. The effect of those measures and of the variants is verified and compared with the classical measure defined on the notion of confidence, through the experimentation on categorical UCI datasets [8].

2. Preliminaries

2.1. Dataset and Association Rules

A dataset is a set of objects (or transactions). Each object is represented by an identifier and a list of valued attributes; each valued attribute is called an item. Let $\mathcal{I}$ be the set of all items of a dataset $\mathcal{D}$. A subset of $\mathcal{I}$ is called an itemset. An itemset consisting of k items is called a *k-itemset*. *Class labels* are special items for defining the class of objects.

The *support* of an itemset I with respect to a dataset $\mathcal{D}$, denoted by $sup(I)$, is the number of the objects in $\mathcal{D}$ that have all the items of I. In practice, one is interested in itemsets with support larger than some threshold $minsup$. These itemsets are called *frequent* itemsets.

An *association rule (AR)* is an expression of the form $X \rightarrow Y$, where X and Y are itemsets and $X \cap Y = \emptyset$. A *class-association rule (CAR)* is an expression of the form $X \rightarrow C$, where X is an itemset and C a class label.

Let r be a CAR $X \rightarrow C$. An object O is *covered by* r (or O *satisfies* r) if O has all the items of X. An object O is *correctly classified by* r if O satisfies r and C is actually the class label of O. The *support* of r with respect to a dataset $\mathcal{D}$, denoted by $sup(r)$, is the number of the objects in $\mathcal{D}$ that are correctly classified by r. The *confidence* of r is defined by $conf(r) = sup(r)/sup(X)$. A rule with confidence 1 is called an *exact rule*.

Example 1: For the training dataset represented in Table 1, $a \rightarrow C$ is a CAR with support 2 and confidence 0.67, because $sup(aC) = 2$ and $sup(a) = 3$.

Table 1. A training dataset

Oid	Itemset	Class label
1	acd	C
2	abe	C
3	abd	C'

Over CARs, the *precedence order* [3] is a partial order defined as follows. Given two CARs r and r', $r \preceq r'$ (r precedes r') if

– $conf(r') < conf(r)$, or

– $conf(r) = conf(r')$ and $sup(r') < sup(r)$, or

– $conf(r) = conf(r')$ and $sup(r) = sup(r')$ and the number of items in $LHS(r)$ (the left-hand side of r) is less than the number of items in $LHS(r')$.

An itemset I is called a *key* itemset (or a *minimal generator*) [9] if $\forall I' \subseteq I, sup(I') = sup(I)$ implies $I = I'$.

An association rule $X \rightarrow Y$, with X and Y being key itemsets, represents an equivalent class of rules, with respect to the confidence and support. Based on this property, classifiers can be built with only key itemsets.

3. Related work

Using an Apriori-like algorithm [10], *CBA* [3] extracts class association rules (CARs) and sorts them in the precedence order. In this order, CARs are selected for building the classifier. A CAR is selected if it classifies correctly at least a training object. Once a CAR is selected, all the objects that are covered by the rule are discarded from the selection process. As a consequence, each training object is covered by at most one CAR of the built classifier.

CMAR [2] is developed on the ideas of CBA, but it uses FP-growth for computing CARs. In contrast to CBA, (i) CMAR selects only positively correlated CARs for the classifier, using the χ^2 test, and (ii) CMAR allows each training object to be covered by several CARs.

HARMONY [5] uses the same strategy as FP-growth for computing CARs. An important difference from CBA and CMAR is that, during the CAR mining process, HARMONY maintains for each training object a top-k list of the highest confidence rules mined so far ($k \geq 1$) that classify correctly the object. At the end of the process, those rules are grouped by class label to form the classifier.

For classifying an object O, CBA searches the classifier, in the precedence order, for the first rule that covers the object to predict the class label. If such a rule does not exist, CBA predicts the majority class label. CMAR and HARMONY follow the vote model. For each class label, HARMONY computes the score of the top-k highest confidence rules that cover O. It predicts the class label with the highest score. The sum of the confidences is a score. CMAR defines a measure called the weighted χ^2 as the score.

In HPWR [7], significant patterns (itemsets) is defined on the deviation between the observed and expected frequency of occurrences. For building classifiers, classification rules are measured in a similar manner. For a rule $r = P \rightarrow C$, using the conditional probability, HPWR defines the weight of evidence of P in favor of C by

$$W(C/\neg C|P) = log\frac{Pr(C|P)}{Pr(C)} - log\frac{Pr(\neg C|P)}{Pr(\neg C)} \quad (1)$$

$$= log\frac{Pr(P|C)}{Pr(P|\neg C)} \quad (2)$$

The weight of evidence is positive if P provides positive evidence supporting C, otherwise, it is negative or zero.

For classifying an object O, the rules that cover O are grouped by class label. The sum of the weights of evidence is computed for each group. The object is classified with the class label that has the largest sum.

4. Class membership Measures

4.1. Confidence

The confidence of rules is a natural measure of class membership. Each rule that covers an object represents a conditional probability space the object can go into. Using the confidence measure, there are two main strategies for classifying an unseen object. CBA looks for the first rule of the classifier, in the precedence order, that covers the object to predict the class of the object. If such a rule does not exist, CBA predicts the majority class. With such a simple strategy of classification, CBA is a very good classifier. HARMONY is also a very competitive classifier with the classification based on the confidence of rules. For classifying an object, the best strategy of HARMONY is to compute, for each class label, the sum of confidences of all the rules of the classifier that cover the object ($\Sigma conf(r_i)$). The class label that has the largest sum is the predicted label.

4.2. C-coefficient

In CMAR, only positively correlated rules are selected for building classifiers. Those rules are determined by the χ^2 test. Given a rule $r : P \rightarrow C$, let

- T be the number of the objects in the dataset.
- p, c and pc be respectively the number of the objects in the dataset that are covered by r, the number of the objects in the dataset that have the class label C, and the number of the objects in the dataset that are correctly classified by r.

The χ^2 value of r is defined as follows. Let

$$o_1 = pc, \quad o_2 = p - o_1, \quad o_3 = c - o_1, \quad o_4 = T - c - o_2 \quad (3)$$

$$n_1 = \frac{p \times c}{T}, \quad n_2 = \frac{p \times (T - c)}{T}, \quad (4)$$

$$n_3 = \frac{(T - p) \times c}{T}, \quad n_4 = \frac{(T - p) \times (T - c)}{T} \quad (5)$$

$$\chi^2 = \frac{(o_1 - n_1)^2}{n_1} + \frac{(o_2 - n_2)^2}{n_2} + \frac{(o_3 - n_3)^2}{n_3} + \frac{(o_4 - n_4)^2}{n_4} \quad (6)$$

The authors of [2] remarked that in the vote model, the application of Equation 6 to define the effect of a class label does not yield a good result, because Equation 6 may

be favorable to minority classes. They proposed another sophisticated measure that integrates both information of correlation and popularity, called the *weighted* χ^2. This measure yields the best results in the experimentation of CMAR. However, the authors also remarked that it is hard to verify theoretically the soundness or effect of this measure.

We have another observation. Through the expressions in Equations 3, 4, 5, we can show that

$$(o_1 - n_1)^2 = (o_2 - n_2)^2 = (o_3 - n_3)^2 = (o_4 - n_4)^2 \quad (7)$$

for any rule built from the dataset, and

$$o_1 + o_2 + o_3 + o_4 = n_1 + n_2 + n_3 + n_4 = T \quad (8)$$

That is, for any rule $P \rightarrow C$, the numerators of the fractions in Equation 6 are identical, and the sum of the denominators is constant and equal to T. Hence,

$$\chi^2 = (o_1 - n_1)^2(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}) \quad (9)$$

If $o_1 = n_1$, then from equation 7, $o_1 = n_1$, $o_2 = n_2$, $o_3 = n_3$ and $o_4 = n_4$, and χ^2 is minimal, $\chi^2 = 0$. We have the same result if $o_2 = n_2$ or $o_3 = n_3$ or $o_4 = n_4$. Conversely, if χ^2 is minimal, i.e., $\chi^2 = 0$, then $o_1 = n_1$, $o_2 = n_2$, $o_3 = n_3$ and $o_4 = n_4$. Otherwise, if the classification by $r = P \rightarrow C$ is well distinct from the classification by the equi-probable distribution, either much better or much worse, then χ^2 tends to the maximal value. In the context of this work where the classifier is built with the rules that classify correctly the training objects with the highest confidence, it is the case where the classification by $P \rightarrow C$ is much better than the classification by the equi-probable distribution. Hence, χ^2 has the similar property as the information gain in sense of the weight of evidence of HPWR [7].

However, the experimentation shows that χ^2 is not competitive measure. In addition, by Equation 9, for a non-zero value of $o_1 - n_1$, (i.e., $o_1 \neq n_1$), if one among $n_i, i = 1..4$, tends to zero, then χ^2 tends to the maximal value. This is the case of minority classes.

To improve it, firstly, we delete the last two terms of χ^2, because they concern only the objects that are not covered by r. Let

$$\chi'^2 = \frac{(o_1 - n_1)^2}{n_1} + \frac{(o_2 - n_2)^2}{n_2} \quad (10)$$

χ'^2 preserves the above property of χ^2.

Secondly, in order to deal appropriately with minority classes, we shift Equation 10 to the scale of the dataset to define a measure, called *C-favor*, to estimate the favor of P to a class C. We denote

$$o'_1 = conf(r) \times T, \quad o'_2 = T - o'_1, \quad (11)$$

$$\chi_r = \sqrt{\frac{(o'_1 - c)^2}{c} + \frac{(o'_2 - (T - c))^2}{T - c}} \quad (12)$$

χ_r still has the similar property as the information gain and can be used to measure the class membership. Let O be an object represented by an itemset X. Let the rules with class label C that cover O be $r_i = P_i \rightarrow C$, $i = 1, ..., k$. We define the *C-favor* of X by

$$F_C(X) = \Sigma_{i=1,k}\chi_{r_i} \quad (13)$$

To classify O, we compute $F_C(X)$ for each class label C, and O is predicted in the class that has the largest $F_C(X)$.

As the confidence is an important measure, we think that the combination of confidence and F_C can be a better measure. The combined measure, denoted by $cF_C(X)$, is defined as follows. For each class label C,

$$cF_C(X) = F_C(X) \times \Sigma_{i=1,k} conf(r_i) \quad (14)$$

The class label that corresponds to the largest value $cF_C(X)$ is used to predict O. To verify the effect of F_C and cF_C, we shall compare them with a measure that we get by adapting the weight of evidence (HPWR [7]) in the context of the systems based on the notions of support and confidence.

4.3. Weight of evidence

For a rule $r = P \rightarrow C$, HPWR [7] defines a measure of evidence provided by P in favor of C as the difference in gain of information when predicting C on P instead of predicting some other class label. Remind that, in Section 3, this measures is defined by Equation 2. HPWR applies this measure in the context of significant association patterns defined by standardized residual. In the context of frequent itemsets defined on the notions of support and confidence, the confidence of r, $conf(r) = sup(r)/sup(P)$, corresponds to the conditional probability $Pr(C|P)$. Moreover, we have

$$Pr(P|C) = \frac{Pr(C|P) \times Pr(P)}{Pr(C)} \quad (15)$$

Hence,

$$\frac{Pr(P|C)}{P(P|\neg C)} = \frac{conf(r) \times (1 - rsup(C))}{(1 - conf(r)) \times rsup(C)} \quad (16)$$

where $rsup(C) = sup(C)/T$, the relative support of C with respect to the number of the objects of the dataset.

Let $r_i = X_i \rightarrow C$, $i = 1, ..., k$ be the rules that cover an object O represented by an itemset X. Then the adapted weight of evidence of X in favor of C is

$$E_C(X) = log\frac{conf(r_1) \times (1 - rsup(C))}{(1 - conf(r_1)) \times rsup(C)} + \\ ... + log\frac{conf(r_k) \times (1 - rsup(C))}{(1 - conf(r_k)) \times rsup(C)} \quad (17)$$

In this work, we omit the condition $X_i \cap X_j = \emptyset, \forall i \neq j, 1 \leq i, j \leq k$, required in [7], because checking this

condition is a cost operation and also because when the omission is applied to all class labels, each one can share almost the same effect.

For classifying O, the adapted weight of evidence of X is computed in favor of each class label. The object O is predicted in the class that corresponds to the highest weight of evidence.

4.4. Variants

It is easy to see that when $conf(r) = 1$ Equation 16 becomes infinite, as well as $W(C/\neg C|P)$. In such a case, from Equation 17, the class label of any exact rule can be selected for the prediction without consideration of the information gain. Under this observation, we define a variant of the weight of evidence as follows. The expression

$$log\frac{conf(r) \times (1 - rsup(C))}{(1 - conf(r)) \times rsup(C)} \quad (18)$$

is modified into

$$log(1 + \frac{T.conf(r)}{c}) - log(1 + \frac{T.(1 - conf(r))}{1 - c}) \quad (19)$$

Let $vE_C(X)$ denote the variant of the weight of evidence as defined in Equation 17, but use Equation 19 in the place of Equation 18. This variant preserves the property of the weight of evidence and allows to compare the gain of information of exact rules.

Finally, we consider the combination of the measures $E_C(X)$ and $vE_C(X)$ with the measure by the sum of confidences to try to improve them. These combined measures are denoted respectively $cE_C(X)$ and $cvE_C(X)$.

$$cE_C(X) = E_C(X) \times \Sigma_{i=1,k} conf(r_i) \quad (20)$$

$$cvE_C(X) = vE_C(X) \times \Sigma_{i=1,k} conf(r_i) \quad (21)$$

5. Implementation

We apply those measures of class membership to a method for building classifiers [11] to classify unseen objects. This method builds classifiers on small key itemsets, using a prefix tree structure for extracting class association rules. It has the following particularities.

– It can adapt the level-wise computation of Apriori, but it is different from Apriori: by enumerating the itemsets and counting their occurrences in the prefix tree, it combines the two phases, generating candidates and computing their support, into one phase.

– It builds the classifiers with the highest confidence rules that classify correctly each training object, as in HARMONY. In contrast to HARMONY, it postpones the search for those rules until the end of the class association rule extraction process.

– Another contrast to HARMONY is that we do not consider the itemsets of all sizes, but we limit the maximal size of key itemsets to 5. Under this limit, we can explore key itemsets with low supports. In fact, we apply the support constraint, using $minsup$, only to i-itemsets, with $i \leq 2$. This constraint is not applied to itemsets with size > 2, excepting those with support 1. Though itemsets with very small supports can be considered, only rules with maximal confidences and supports that classify correctly each training object, are added to the classifier.

6. Experimental Results

Table 2. The 23 UCI datasets.

13 small datasets			
Dataset	**#obj**	**#item**	**#class**
anneal	798	106	5
auto	205	142	7
breast	699	48	2
glass	214	52	7
heart	303	53	5
hepatitis	155	58	2
horseColic	368	94	2
ionosphere	351	104	2
iris	150	23	3
pimaIndians	768	42	2
ticTacToe	958	29	2
wine	178	68	3
zoo	101	43	7
10 large datasets			
adult	48482	131	2
chess	28056	66	18
connect	67557	66	3
led7	3200	24	10
letRecog	20000	106	26
mushroom	8124	127	2
nursery	12960	32	5
pageBlocks	5473	55	5
penDigits	10992	90	10
waveform	5000	100	3

The program for verifying the effect of the class membership measures is implemented in C and experimented on a laptop with a Pentium 4, 1.7 GHz mobile processor and 768 MB memory, running Linux 9.1. Let us call SIM the implemented program. For comparison, we evaluate the measures on the same 23 UCI categorical datasets (13 small and 10 large datasets) obtained from the author of [8], using the 10-fold cross validation. Table 2 represents the characteristics of these 23 UCI datasets.

To have an idea on the effect of the measures, we use the experimental results of HARMONY as the reference. The reasons are:

(i) HARMONY is a good classification system, its performance is well compared [6] with the important systems such as FOIL [12], CPAR [13], and SVM [14].

Table 3. Experimentation on 13 small datasets

	HAR.	SIM							
Dataset	Σconf	Σconf	F_C	cF_C	E_C	cE_C	vE_C	cvE_C	**Sz**
anneal	94.04	94.27	91.35	92.58	**95.17**	**94.94**	92.02	93.37	4
auto	72.00	76.50	**78.50**	78.00	77.00	77.00	**78.50**	78.00	3
breast	91.18	90.14	92.03	92.03	**93.19**	91.88	92.17	**93.19**	3
glass	68.58	**71.90**	70.48	**71.90**	71.43	**71.90**	71.43	71.43	5
heart	56.33	**59.00**	56.67	**58.67**	**58.67**	**58.67**	58.00	**58.67**	5
hepatitis	82.01	82.67	81.33	82.00	**84.67**	**85.33**	82.67	80.67	3
horseColic	81.68	**83.89**	83.61	**83.89**	83.61	**83.89**	**83.89**	**83.89**	4
ionosphere	89.15	90.00	**90.86**	90.29	90.00	90.00	90.29	90.29	4
iris	93.98	94.00	94.00	94.00	94.00	94.00	94.00	94.00	3
pimaIndians	69.22	69.08	73.42	**73.95**	73.29	**73.82**	73.42	73.42	3
ticTacToe	96.42	97.89	98.11	98.21	97.89	97.89	98.21	**98.63**	4
wine	90.57	**91.76**	91.18	91.18	91.76	**91.76**	91.18	91.18	4
zoo	89.00	92.00	**93.00**	**93.00**	92.00	92.00	**93.00**	92.00	3
Total/Avg.	82.63	84.08	84.20	**84.59**	**84.82**	**84.85**	84.52	84.52	

Table 4. Experimentation 10 large datasets

	HAR.	SIM							
Dataset	Σconf	Σconf	F_C	cF_C	E_C	cE_C	vE_C	cvE_C	**Sz**
adult	83.40	**84.23**	82.02	83.73	83.41	**84.29**	82.61	83.78	3
chess	44.93	**60.57**	61.58	**61.79**	60.80	60.56	**61.70**	**61.76**	5
connect	77.30	**77.67**	78.24	78.88	**79.72**	79.21	78.88	**79.45**	4
led7	74.35	**74.37**	74.19	74.22	74.22	74.31	**74.42**	**74.44**	5
letRecog	70.82	71.29	71.54	71.47	70.82	71.25	69.74	**72.34**	4
mushroom	100	100	100	100	100	100	100	100	3
nursery	92.94	**98.33**	96.71	97.85	**98.77**	**98.66**	97.21	97.85	5
pageBlocks	91.17	90.93	91.52	**92.08**	91.74	**92.08**	91.52	**91.90**	4
penDigits	96.03	**97.05**	**97.02**	**97.02**	**97.02**	**97.04**	**97.02**	**97.02**	4
waveform	77.92	**79.86**	**79.78**	**79.82**	75.72	76.24	**79.78**	**79.78**	3
Total/Avg.	80.89	**83.45**	83.26	**83.69**	83.22	83.36	83.27	**83.83**	

(ii) SIM shares with HARMONY on many points in the method for building classifiers and is also different from HARMONY on many other points.

(iii) Thanks to the authors of HARMONY, we get the runnable program of HARMONY and can use it to experiment HARMONY and SIM on the same UCI datasets and on the same computer.

For the experiments conducted in this work, the parameter setting for HARMONY is done as described in [6]: for the 13 small datasets, $minsup = 10$, and for the 10 large datasets, $minsup = 50$; items are sorted in the correlation coefficient ascending order (the order with which HARMONY gets the best results in general). In particular, for the dense datasets *connect* and *ionosphere*, only the items with supports no greater than $20,000$ and 190 respectively are considered to generate class association rules. The same consideration is applied to SIM.

Tables 3 and 4 represent the experimental results of the classification accuracy of HARMONY (abbreviated by HAR.) and SIM using the measures $\Sigma conf$, F_C, cF_C, the adapted weight of evidences (E_C) and its variants (cE_C, vE_C, cvE_C). In these tables, the column *Sz* represents the maximal size of itemsets extracted by SIM.

7. Discussions and conclusion

Notice that the results represented in column HAR. (HARMONY) of Tables 3 and 4 are different and better than what was reported in [6]. The differences can be due to the method for dividing data randomly in the 10-fold cross validation.

On the results of the experiments conducted in this work, Tables 3 and 4 show that:

– For the same measure $\Sigma conf$, on average the classification by SIM is about 1.45% and 2.56% more accurate than that by HARMONY for the 13 small and 10 large datasets, respectively. The good accuracy of SIM can be explained by the selection of CARs having the maximal confidence and support among those having small supports. These rules may be very specific. However, as they are built on the small size key itemsets, they are in general not specific.

– F_C (the measure based on χ^2 revisited) and E_C (the adapted weight of evidence) are comparable to $\Sigma conf$. Moreover, we can say that F_C is comparable to E_C and the modified version of E_C (i.e. vE_C) is correct and effective.

– The combinations of F_C, E_C, and vE_C with $\Sigma conf$ improve slightly F_C, E_C, vE_C, and $\Sigma conf$. We can say that the sum of confidences is an important measure and these combined measures are useful, in particular in the context where the accuracies of the classification by the sum of confidence, by F_C, and by E_C are almost very good.

For conclusions, firstly, the adapted weight of evidence is a good class membership measure built on the gain of information. F_C, a measure built on the revisited χ^2 test, provides another view of information gain. It is comparable to the adapted weight of evidence. The sum of confidence is a simple and natural measure with the good performance. The combinations of the sum of confidence with the previous measures are interesting and useful to improve their performance.

Next, based on the average accuracy values of different measures, we recommend to use the combined measures, because the average accuracy values of the combined measures are in general better than that of the non-combined measures. Finally, through the results on each dataset, between the combined measures based on χ^2 and the weight of evidence, we suggest the following propositions.

- For the 13 small datasets, though the average accuracy value of cE_C is slightly better than that of cF_C, we can observe that cF_C is much often wins cE_C. Indeed, cE_C wins cF_C on only 3 datasets while cF_C wins cE_C on 6 datasets. Hence, we can recommend cF_C for the small datasets.
- For the 10 large datasets, the average accuracy value of cvE_C is slightly better than that of cF_C, and cvE_C wins cF_C on 4 datasets and cF_C wins cvE_C 3 datasets. Hence, we can recommend cvE_C for large datasets.

References

[1] B. Lent, A. Swami, and J. Widom, "Clustering association rules," Proc. Intl. Conf. on Data Engineering (ICDE'97), IEEE Computer Society, 1997, pp. 220-231.

[2] W. Li, J. Han, and J. Pei, "CMAR: Accurate and Efficient Classification based on multiple class-association rules," Proc. IEEE Intl. Conf. on Data Mining (ICDM'01), San Jose, CA, IEEE Computer Society, 2001, pp. 369-376.

[3] B. Liu, W. Hsu, and Y. Ma, "Integrating Classification and Association Rule Mining," Proc. 4th Intl. Conf. on Knowledge Discovery and Data Mining (KDD'98), AAAI Press, 1998, pp. 80-86.

[4] Y. Sun, Y. Wang, and A.K.C. Wong, "Boosting an Association Classifier," in IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 7, IEEE Computer Society, 2006, pp. 988-992.

[5] J. Wang and G. Karypis, "HARMONY: Efficiently Mining the Best Rules for Classification," Proc. SIAM Intl. Conf. on Data Mining (SDM'05), 2005, pp. 205-216.

[6] J. Wang and G. Karypis, "On Mining Instance-Centric Classification Rules," in IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 11, 2006, pp. 1497-1511.

[7] Y. Wang and A.K.C. Wong, "From Association to Classification: Inference using Weight of Evidence," in IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 3, 2003, pp. 764-767.

[8] F. Coenen, "The LUCS-KDD Implementations of the FOIL, PRM, and CPAR algorithms," http://www.csc.liv.ac.uk/˜ frans/KDD/Software/FOIL_ PRM_CPAR/foilPrmCpar.html, Computer Science Department, University of Liverpool, UK, 2004.

[9] Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, and L. Lakhal, "Mining Frequent Patterns with Counting Inferences," in ACM SIGMOD Explorations, vol. 2, no. 2, 2000, pp. 66-75.

[10] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," Proc. 20th Intl. Conf. on Very Large Databases, VLDB'94, Santiago, Chile, 1994, pp. 487-499.

[11] V. Phan-Luong and R. Messouci, "Building Classifiers with Association Rules based on Small Key Itemsets," Proc. 2nd IEEE International Conf. on Digital Information Management (ICDIM'07), France, 2007, pp. 200-205.

[12] J. Quinlan and R. Cameron-Jones, "FOIL: A Midterm Report," Proc. European Conf. on Machine Learning (ECML'93), 1993, pp. 3-20.

[13] X. Yin and J. Han, "CPAR: Classification based on Predictive Association Rules," Proc. 3rd SIAM Intl. Conf. on Data Mining (SDM'03), San Francisco, CA, SIAM, 2003, pp. 369-376.

[14] C. Cortes and V. Vapnik, "Support-Vector Networks," in Machine Learning, vol. 20, no. 3, 1995, pp. 273-297.