A Study On Interestingness Measures for Associative Classifiers

Mojdeh Jalali Heravi and Osmar R. Zaïane

Unversity of Alberta, Computing Science Department, Edmonton, Canada
{jalalihe,zaiane}@cs.ualberta.ca

Abstract. Associative classification is a rule-based approach to classify data relying on association rule mining by discovering associations between a set of features and a class label. Support and confidence are the de-facto "interestingness measures" used for discovering relevant association rules. The support-confidence framework has also been used in most if not all associative classifiers. Although support and confidence are appropriate measures for building a strong model in many cases, they are still not the ideal measures because in some cases a huge set of rules is generated which could hinder the effectiveness in some cases for which other measures could be better suited.
There are many other rule interestingness measures already used in machine learning, data mining and statistics. This work focuses on using objective measures for associative classification rules. A wide range of UCI datasets are used to study the impact of different "interestingness measures" on them. The results show that there are interestingness measures that can significantly reduce the number of rules for some datasets while the accuracy of the model is hardly jeopardized or even improved. However, no single measure can be declared an obvious winner.

Key words: associative classifier, interestingness measures

1 Introduction

Associative classification is a rule-based approach recently proposed to classify data by discovering associations between a set of features and a class label [10, 2, 9]. To build an associative classification model, association rules whose consequent is a class label are generated using an association rule mining technique. Research shows promising results for associative classification and its potential for improvement to a more powerful classification paradigm.

Support and confidence are the default "interestingness measures" universally used for discovering relevant association rules. The support-confidence framework is the most common framwork used in most association rule mining methods, and similarly for mining and selecting rules of associative classifiers. Although these two measurs are used widely, they are still not necessarily the ideal measures because in many cases a huge set of rules is generated which could hinder the effectiveness in some cases for which other measures could be better suited. Yet, no sistematic study has been done to identify a better framework or the most appropriate measure.

There are many other rule interestingness measures already used in machine learning, data mining and statistics. Many different measures are introduced in the field of association rule mining as filters to weed-out the least relevant rules. All those measures can directly apply to associative classifiers as well, although never tested or reported in the litterature. This work focuses on probability-based objective rule interestingness measures for associative classification. Our aim is to probe efficient "interestingness measures" in three phases of associative classification: rule generation, pruning and selection, so that the mining algorithm improve both in terms of increasing classification accuracy and decreasing the number of rules. A wide range of UCI datasets are used to study the impact of different "interestingness measures" on them. The results show that there are interestingness measures for each dataset that can significantly reduce the number of rules while the f-measure is hardly jeopardized. In many cases the accuracy is even improved despite the drastic reduction in the number of rules.

The remainder of the paper is organized as follows: Section 2 describes the associative classifiers, and Section 3 introduces the known interestingness measures reportedly used in association rule mining. In Section 4 some related work both in association rule mining and associative classifiers is introduced. The methodology of using interestingness measures in the three different phases of an associative classifier is discussed in Section 5. Experimental results, comparing the impact of interestingness measures on classification accuracy and the number of generated classification rules, is illustrated in Section 6. The paper is concluded in Section 7 with reference to future work.

2 Associative Classifier

The Associative classifier [10] is an interpretable classifier that uses association rule mining in order to generate classification rules. The term interpretable means that the built model is easily human readable and even editable for domain knowledge injection. To use this classifier, datasets have to be transformed in a transactional format. Considering each attribute-value pair in a dataset as an item results in a dataset in which a row of data looks like a transaction of items. Among items of each transaction, one is the class label of the related object. Using an association rule mining technique (e.g., Apriori [19] or Eclat[12]) on the resulting transactional data, frequent itemsets of the form $\{A, c\}$ are extracted where A is a set of feature and c is a class label, i.e., A and c are disjoint subsets of items. Among these frequent itemsets, the confident ones are chosen to build a classification rule of the form $A \Rightarrow c$. Then, these rules are used to predict class labels for objects with an unkown class.

As mentioned above, the support-confidence framework is the standard in association rule mining and inherited by associative classification. For a rule $A \Rightarrow c$, support is the fraction of data samples having A and c together. A rule is frequent if its support is greater than a minimum support threshold. Confidence is the conditional probability that a record is of class c given that it includes A. A rule is confident if its confidence is above a minimum confidence threshold. To build an associative classifier, only strong rules i.e., the rules that are both frequent and confident are used.

The associative classification approach can be summerized in three phases. First, frequent itemsets having a class label are extracted from a training dataset using an association rule mining thechnique and only the strong rulesm that have a class label as a consequent, are selected. Because of the exhaustive search, there might be a huge set of rules, most of which may be redundant or leading to misclassification, hence, in the second phase, a pruning is required in order to keep only the accurate rules by eliminating what might be noise. Lastly, in the third phase, for each object with an unknown class, the best rules is selected for predicting a class label.

2.1 Disadvantages of support and confidence

Although support and confidence are appropriate measures for building a strong model in many cases, they are still not the ideal measures. For example, choosing a large minimum support may lead to having only rules that contain obvious knowledge and missing exceptional cases that are interesting. On the other hand, assigning a low minimum support yields a huge number of rules which could be redundant or noise. Therefore, support is not an appropriate measure and is difficult to tune.

Similarly, confidence is not a perfect measure as it considers nothing beyond the conditional probability of rules. An example clarifies this claim: the confidence of the rule $A \Rightarrow c$ is 90% in the case that $N(A) = 10, N(c) = 9,000, N(Ac) = 9$ and $N = 10,000$ where N is the size of data and $N(X)$ denotes the frequency of X. Although, the confidence is very high, A and c are statistically independent because $P(Ac) = P(A)P(c)$. The rule is not interesting at all because 90% is exactly the support (probability) of c regardless of A. Finally, this confident rule is very rare with the support of 0.0009 indicating that it might be originating from noise. Cases can also abound in which a rule is not rare but still suffers from the afore mentioned drawback.

3 Interestingness Measures

Generating rules in association rule mining or with associative classifiers can lead to a huge set of rules which make them impossible for users or domain specialists to study. Sifting through thousands or even millions of rules in impractical. Thus, users lose the opportunity to interpret the results, find interesting rules or even modify them for having a more accurate classifier. To solve this problem, interestingness measures can be used for filtering or ranking associations or classification rules. There are many different rule interestingness measures widely used in machine learning, data mining and statistics. However, to the best of

	c	$\neg c$	
A	$N(Ac)$	$N(A\neg c)$	$N(A)$
$\neg A$	$N(\neg Ac)$	$N(\neg A\neg c)$	$N(\neg A)$
	$N(c)$	$N(\neg c)$	N

Table 1. Frequencies shown in a 2×2 contingency table for rule $A \rightarrow c$

our knowledge, there is still no formal definition of "interestingness". In a study, Geng and Hamilton[8] have brouth together 9 diferrent criteria which specify the interestingness of a pattern. These 9 criteria are *conciseness*, *generality*, *reliability*, *peculiarity*, *diversity*, *novelty*, *superisingness*, *utility* and *actionability*. The definition of these criteria may have overlaps or conflicts with others. For example, usually a concise pattern, because of its simplicity, can also be general and generality may also lead to reliability. On the other hand, generality is in conflict with peculiarity and novelty.

In addition to the mentiond criteria that can define the interestingness of a measure, there are 3 main categories that classify interestingness measures: *objective*, *subjective* and *semantics-based* measures[8]. Objective measures are those that are not application or user-specific and depends only on raw data. Subjective measures are those that consider users' background knowledge as well as data. As a special type of subjective measures, semantic-based measures take into account the explanation and the semantic of a pattern which are, like subjective measures, domain specific. The aim of this work is to focus on probability-base objective measures. Some examples of objective rule interestingness measures are shown in Table 2. The measures described in this table are defined based on frequencies of a 2×2 contingency table shown in Table 1, where $P(X) = \frac{N(X)}{N}$.

Table 2: Objective rule interestingness measures for rule in the form of $A \rightarrow c$

No.	Measure	Formula	Ref
1	1-way support	$P(c\|A) \times \log \frac{P(Ac)}{P(A)P(c)}$	[8]
2	2-way support	$P(Ac) \times log \frac{P(Ac)}{P(A)P(c)}$	[8]
3	2-way support variation(modified)[1]	$P(Ac) \times log \frac{P(Ac)}{P(A)P(c)}$ $+P(\neg A\neg c) \times \log \frac{P(\neg A\neg c)}{P(\neg A)P(\neg c)}$ $-P(A\neg c) \times \log \frac{P(A\neg c)}{P(A)P(\neg c)}$ $-P(\neg Ac) \times \log \frac{P(\neg Ac)}{P(\neg A)P(c)}$	[8]
4	Accuracy	$P(Ac) + P(\neg A\neg c)$	[8]
5	Added value	$P(c\|A) - P(c)$	[8]
		Continued on next page	

[1] The modified measures are different from their original version that is cited so that it can satisfy the interestingness for a classification rule

Table 2 – continued from previous page

No.	Measure	Formula	Ref
6	Certainity factor	$\frac{P(A\|c)-P(c)}{1-P(c)}$	[8]
7	Chi-square	$N \times \left(\frac{P(Ac)-P(A)P(c)}{\sqrt{P(A)P(c)P(\neg A)P(\neg c)}}\right)^2$	[17]
8	Class correlation ratio	$\frac{N(Ac)(N(A\neg c)+N(\neg A\neg c))}{N(A\neg c)(N(Ac)+N(\neg Ac))}$	[4]
9	Collective Strength	$\frac{P(Ac)+P(\neg c\|\neg A)}{P(A)P(c)+P(\neg A)P(\neg c)} \times \frac{1-P(A)P(c)-P(\neg A)P(\neg c)}{1-P(Ac)-P(\neg c\|\neg A)}$	[8]
10	Confidence	$P(c\|A)$	[8]
11	Confidence causal	$0.5 \times \left(\frac{P(Ac)}{P(A)} + \frac{P(\neg A\neg c)}{P(\neg c)}\right)$	[7]
12	Confirm causal	$P(Ac) + P(\neg A\neg c) - 2 \times P(A\neg c)$	[7]
13	Confirm descriptive	$P(c\|A)$	[7]
14	Confirm-confidence causal	$0.5 \times (P(c\|A) + P(\neg A\|\neg c)) - P(\neg c\|A)$	[7]
15	Confirm-confidence descriptive	$P(c\|A) - P(\neg c\|A)$	[7]
16	Convictione	$\frac{P(A)P(\neg c)}{P(A\neg c)}$	[8]
17	Correlation coefficient	$\frac{P(Ac)-P(A)P(c)}{\sqrt{P(A)P(c)P(\neg A)P(\neg c)}}$	[8]
18	Cosine/IS	$\frac{P(Ac)}{\sqrt{P(A)P(c)}}$	[8]
19	Dilated chi-square	$(\frac{N}{lmax(\chi^2)})^\alpha \chi^2$ $lmax(\chi^2) = \frac{(n_1 n_2)^2 \times N}{(k_1+k_3)(k_4+k_2)(k_1+k_4)(k_3+k_2)}$ $k_1 = N(Ac), k_2 = N(\neg A\neg c)$ $k_3 = N(A\neg c), k_4 = N(\neg Ac)$ $n_1 = \min(\min(N(Ac) + N(A\neg c)), (N(\neg Ac) + N(\neg A\neg c)),$ $\min(N(Ac) + N(\neg Ac)), (N(A\neg c) + N(\neg A\neg c)))$ $n_2 = \min(\max(N(Ac) + N(A\neg c)), (N(\neg Ac) + N(\neg A\neg c)),$ $\max(N(Ac) + N(\neg Ac)), (N(A\neg c) + N(\neg A\neg c)))$ $\alpha = 0.5$	[22]
20	Example and counterexample rate	$1 - \frac{P(A\neg c)}{P(Ac)}$	[8]
21	F-measure	$\frac{2\times P(c\|A)P(A\|c)}{P(c\|A)+P(A\|c)}$	[15]
22	Gini index	$P(A)\left(P(c\|A)^2 + P(\neg c\|A)^2\right)$ $+P(\neg A)(P(c\|\neg A)^2 + P(\neg c\|\neg A)_2)$ $-P(c)_2 - P(\neg c)^2$	[8]
23	Hyper confidence	$P(C_{Ac} < N(Ac) = \sum_{i=0}^{N(Ac)-1} P(C_{Ac}) = i$	[14]

Continued on next page

Table 2 – continued from previous page

No.	Measure	Formula	Ref
		$P(C_{Ac}=r)=\frac{\binom{N(c)}{r}\binom{N-N(c)}{N(A)-r}}{\binom{N}{N(A)}}$	
24	Hyper lift	$\frac{N(Ac)}{Q_\delta(C_{Ac})}$ $P\left(C_{Ac} < Q_\delta(C_{Ac})\right)$ $\delta = 0.99$	[14]
25	Information gain	$log(\frac{P(Ac)}{P(A)p(c)})$	[8]
26	Intensity of implication	$1-\Sigma_{k=0}^{N(A\neg c)}\frac{\lambda^k}{k!}e^{-\lambda}$ $\lambda=\frac{N(A)(N-N(c))}{N}$	[22]
27	Interestingness Weighting Dependency	$\left(\left(\frac{P(Ac)}{P(A)P(c)}\right)^k-1\right)\times P(Ac)^m$ $k=m=1$	[8]
28	Jaccard	$\frac{P(Ac)}{P(A)+P(c)-P(Ac)}$	[8]
29	J-measure (modified)	$P(Ac)log\left(\frac{P(c\|A)}{P(c)}\right)-$ $P(A\neg c)log\left(\frac{P(\neg c\|A)}{P(\neg c)}\right)$	[8]
30	Kappa	$\frac{P(Ac)+P(\neg A\neg c)-P(A)P(c)-P(\neg A)P(\neg c)}{1-P(A)P(c)-P(\neg A)P(\neg c)}$	[18]
31	Klosgen	$\sqrt{P(Ac)}(P(c\|A)-P(c))$	[8]
32	K-measure	$P(c\|A)log\left(\frac{P(c\|A)}{P(c)}\right)+$ $P(\neg(c\|A))log\left(\frac{P(\neg(c\|A))}{P(\neg c)}\right)-$ $P(c\|A)log\left(\frac{P(c\|A)}{P(\neg c)}\right)-$ $P(\neg(c\|A))log\left(\frac{P(\neg(c\|A))}{P(c)}\right)$	[15]
33	Laplace correlation	$\frac{N(Ac)+1}{N(A)+2}$	[8]
34	Least contradiction	$\frac{P(Ac)-P(A\neg c)}{P(c)}$	[8]
35	Leverage	$P(c\|A)-P(A)P(c)$	[8]
36	Lift/interest	$\frac{P(Ac)}{P(A)P(c)}$	[8]
37	Loevinger	$1-\frac{P(A)P(\neg c)}{P(A\neg c)}$	[8]
38	Odd multiplier	$\frac{P(Ac)P(\neg c)}{P(A\neg c)P(c)}$	[8]
39	Odds ratio	$\frac{P(Ac)P(\neg A\neg c)}{P(A\neg c)P(\neg Ac)}$	[8]
40	Piatetsky-Shapiro	$P(Ac)-P(A)P(c)$	[8]
41	Recall/local support	$P(A\|c)$	[8]
42	Relative risk	$\frac{P(c\|A)}{P(c\|\neg A)}$	[8]
43	Sebag-Schoenauer	$\frac{P(Ac)}{P(A\neg c)}$	[8]
44	Specificity	$P(\neg c\|\neg A)$	[8]

Continued on next page

Table 2 – continued from previous page

No.	Measure	Formula	Ref
45	Support/global support	$P(Ac)$	[8]
46	Yule's Q	$\frac{P(AB)P(\neg A\neg c)-P(\neg Ac)P(A\neg c)}{P(AB)P(\neg A\neg c)+P(\neg Ac)P(A\neg c)}$	[8]
47	Yule's Y	$\frac{\sqrt{P(AB)P(\neg A\neg c)}-\sqrt{P(\neg Ac)P(A\neg c)}}{\sqrt{P(AB)P(\neg A\neg c)}+\sqrt{P(\neg Ac)P(A\neg c)}}$	[8]
48	Zhang	$\frac{P(Ac)-P(A)P(c)}{max(P(Ac)P(\neg c),P(A\neg c)P(c))}$	[8]

4 Related work

Interestingness measures are used in different aspects of data mining[6]. One of the main purposes of using interestingness measures is to reduce or rank the patterns(e.g., association rules, classification rules, sequential patterns, contingency tables and summaries) in order to find interesting ones. In this paper the focus is on association and classification rules, hence, only the works related to these two areas are introduced in this section.

4.1 Interestingness measures in association rule mining

Bayadro and Agrawel [20] have proposed an algorithm to mine optimized rules under partial ordering of the rules (instead of the typical total ordering on rules) according to different interestingness measures such as support, confidence, conviction, lift, gain, entropy gain, gini, chi-square, etc .

Tan et al.[18] have introduced 21 different rule objective measures which can be used to evaluate association rules. Having a specific set of patterns, these measures can be used to rank the patterns and then using a similarity measure, like Pearson correlation, the similar rankings and consequently the similar measures can be found. They also proposed an appraoch to find the best interestingness measure for patters of a specific domain. For this purpose, first a domain specialist should rank a set of patterns in that domain manually. Then the most similar ranking using different measures shows the best measure that can be used for that specific application domain. In cases where there is a huge number of patterns, only those having high standard deviation on the evaluation of different measures are chosen as samples to be presented to domain experts.

In another work, Lenca et al.[16] have ranked the measures based on their properties rather than using a set of patterns. For each application domain, a specialist assigns weights to each property of measures (e.g., symmetric property). Each weight shows the importance of that property in the given domain. Then using all properties and also the weights assigned to each of them, measures are ranked by applying a multicriteria decision process.

Vaillant et al.[21] have compared 20 different interestingness measures based on their properties and also based on experimental results from 10 different datasets. In the first comparison, they cluster the measures based on 6 different properties. Then they use a preorder agreement coefficient to find the similarity between two different rankings done by two different measures. These similarities are used to cluster the measures. They have found 4 main clusters using 10 different datasets. Finally, they compare these two clusterings to show that most of the measures are in the same group.

Also, there are some application specific research done in this area. Merceron and Yacef [13] have tried 3 different interestingness measure including cosine, lift and added value on association rules and found the impact of these measures on educational data. In another work, Ohsaki et al.[15] have applied different interestingness measures on mining association rules to examine the usefulness of these measures for finding interesting rules extracted from clinical data.

In addition to these related works, there are two surveys [8] and [6] having many useful information about interestingness measures in general.

4.2 Interestingness measures for associative classifiers

In an effort to present better alternatives to confidence in associative classifiers, Lan et al.[22] have proposed two novel interesting measures, intensity of implication and dilated chi-square. These measures, which are used to sort generated rules, statistically reveal the interdependence between the antecedent of a rule and its class. Their experiments on 16 UCI datasets show the impact of these measures on having a more accurate and more compact set of rules.

After showing that even confident rules can have negative correlations, Arunasalam and Chawla [3] propose a new measure called Complement Class Support (CCS) which guarantees rules to be positively correlated. Based on anti-monotonic property of CCS and the fact that "good" rules have low CCS values, they discover strong rules by a row enumeration algorithm [5].

SPACCC [4], an associative classifer, was introduced by Verhein and Chawla. This classifier utilizes the Fisher Exact Test's ρ-value to extract only statistically significant rules. They also use a new measure called Class Correlation Ratio(CCR) to select only the rules that are more positively correlated to the class they predict rather than the other classes. For classifying, they use a strength score to rank the rules. This score is a combination of ρ-value, confidence and CCR. They show that they can outperform other algorithms when an imbalanced dataset is used for training.

These related works show that different interestingness measures are also being used in associative classifers in order to improve their performanc. However, to the best of our knowledge there has not been any previous study on comparing different interestingness measures specifically for associative classifiers.

5 Using interestingness measures for associative classifiers

As described in section 2.1 the support and confidence are not necessarily the ideal measures for associative classifiers. The aim of this work is to study the impact of different interestingness measures on three phases of this classifers: rule generating, rule pruning and rule selection. For this purpose, a simple associative classifier is chosen which is as follows: first, using an association rule mining technique, rules with a classlabel as a consequent are generated. For generating rules, an antiminotonic measure, am, and a threshold, t_a is required to prune the search space and make the searching algorithm efficient. Then, all the conflicting rules are eliminated. Second, for filtering redundant rules, a measure, fm, and a threshold, t_f is needed . Hence, all rules having values less than the t_f for their fm value are eliminated. Finally, for classifying a new object, a selecting measure, sm, is used. Among all rules that apply to that object, the class label of those with the highest average on their sm value is selected to be assigned to that unlabeled object. If no rule can match the object, the dominant class is assigned to it.

This study is based on many different interestingness measures already used in association rule mining and a few that have been used for associative classifiers. The list of all measures used can be found in Table 2. For the rule generation phase, only the measures with antimonotonic property can be used. This property for measures listed in Table 2 is not known, however, in most of the cases, it is obvious that the measure is not anti-monotone. Hence, for this phase only support(global support) and local support are used. For removing the conflicting rules, confidence is the simplest measure that can be used. Here we use confidence with giving it the lowest possible threshold(%51) only to eliminate conflicting rules. In the pruning phase, all the measures except global and local support are used in this study to see the impact of rule reduction while still trying to keep the f-measure for accuracy stable or with little change. In the last phase, all the 48 measures are used to study the impact of different measures on increasing the accuracy when selecting the rules to apply for labelling.

6 Experimarntal Results

To compare the impact of measures listed in Table 2, 20 different datasets have been chosen from the UCI repository [1]. The datasets and their properties are shown in Table 3. To be able to use these datasets, all numeric attributes are discretized. The same entropy-based discretization method [11] used in CBA [10] is also used here to categorize the continues attributes.

To generate the rules, Eclat[12] is used with local and global support wih a threshold of 1%. To keep only the confident rules, we use a confidence threshold set at 51%. Rule sets generated only using local/global support and confidence are called “original rule sets”. The results of original rulesets are shown in Table 4 in terms of number of rules, f-measure and accuracy. All results in this section are base on 10 fold cross validations. The results of original datasets show that using local support yields a very large number of generated rules, specialy when the class labels are imbalanced, but it also creates more accurate models for this kind of datasets as it also finds frequent patterns in small classes.

Datasets	# of att-val pairs	Avg. # of items per transaction	# of transactions	# of classes	Stdev of class destributions
Anneal	70	13.31	898	5	0.32
Breast	25	8.98	699	2	0.22
Census	149	12.87	32561	2	0.37
Colic	61	14.52	368	2	0.18
Credit	54	14.90	690	2	0.08
Diabetes	17	7.00	768	2	0.21
German	57	15.00	1000	2	0.28
Glass	20	7.00	214	6	0.14
Heart	28	10.98	303	2	0.06
Hepatitis	35	16.14	155	2	0.42
Iris	13	4.00	150	3	0.00
Labor	33	8.40	57	2	0.21
Led7	14	7.00	3200	10	0.00
Pima	17	7.00	768	2	0.21
Tictactoe	27	9.00	958	2	0.22
Vote	32	15.10	435	2	0.16
Vowel	62	13.00	990	11	0.00
Waveform	107	19.00	5000	3	0.00
Wine	36	13.00	178	3	0.06
Zoo	132	17.00	101	7	0.13

Table 3. 20 different datasets from UCI repository [1]

We conducted two different experiments for finding the impact of using 46 different filtering measures(all the measures listed in Table 2 except local and global support) on reducing the number of rules and improving the f-measure. Table 5 shows the results for decreasing the number of rules. Each row of this table shows the maximum percentage of rule reduction for a dataset using different measures while keeping the f-measure above 98% of its original f-measure. As it is shown, in all datasets the number of rules has decreased significantly and even in some datasets the f-measure also improved. The second experiment shows the impact of using filtering measures on improving the f-measure. Each row in Table 6 shows the maximum percentage of f-measure improvements using different measures. The results show that filtering the rules does not have the same impact on improving the f-measure as it does on rule reduction, however, there are still improvements in all datasets and some of them are significant.

Moreover, the effect of using different selecting measures (all 48 measures) in the third phase of the associative classifier, are also studied. This effect is only on the improvement of the f-measure. There is no change in the number of rules per-se. This effect is shown on both the original and filtered datasets. Table 7 shows the results on original rulesets and Tables 8 and 9 shows the results on filtered rulesets with maximum rule reduction and maximum f-measure improvement respectively. From the results, it can be inferred that in almost half of the cases

	Local support			Global support		
Datasets	**# of rules**	**Fmeasure%**	**Acc%**	**# of rules**	**Fmeasure%**	**Acc%**
Anneal	791,998	88.59	93.33	309.828	66.64	89.76
Breast	11,338	95.74	96.13	6,936	96.08	96.42
Census	139,508	75.44	84.74	63,226	73.50	83.89
Colic	794,762	79.46	80.98	188,278	80.36	82.57
Credit	868,705	86.92	87.25	299,311	87,64	87.97
Diabetes	1,113	68.26	73.70	922	68.54	74.09
German	512,204	61.89	73.00	223,508	43.62	70.10
Glass	2,779	50.32	63.19	1,599	54.85	66.97
Heart	62,833	79.40	79.87	41,095	80.37	80.85
Hepatitis	1,150,690	67.17	83.65	1,150,690	67.17	83.65
Iris	131	91.73	92.00	108	91.06	91.33
Labor	44,203	82.02	82.33	44,203	82.02	82.33
Led7	988	71.49	72.44	473	70.97	72.00
Pima	1,200	68.50	74.22	988	68.64	74.48
Tictactoe	19,534	97.21	97.49	7,397	88.06	90.19
Vote	1,539,000	95.69	95.87	955,659	95.69	95.87
Vowel	1,176,990	72.28	72.32	18,500	56.24	58.38
Waveform	489,169	79.49	80.14	35,625	75.57	76.54
Wine	479,564	92.51	92.60	185,942	95.60	95.48
Zoo	971,581	91.26	94.99	971,581	91.26	94.99

Table 4. Results on 20 datasets using local and global support with threshold of 1%

non of the measures could beat confidence to improve the f-measure however in the other half the f-measures have been improved by a small percentage.

To summarize the results, there are interestingness measures that can be used as filtering measures and be able to reduce the number of rules significantly in all datasets without jeopardizing the accuracy of the model. In other words, the filters are capable of identifying unnecessary rules from the model. However, this drastic improvement in the number of rules is not observed in terms of accuracy. The change in accuracy remains stable, however, not negligeable since some positive improvements in the accuracy (f-measure) were noted. Another observation is that, no single measure can declared as a winner for all types of datasets. There are some measures that have more impact than others.

7 Conclusion and future work

Associative classification is a new paradigm for classification relying on association rule mining and naturally inherits the most commonly used interestingness measures, support and confidence. These are not necessarily the best choice and no systematic study was undertaken to identify the most appropriate measures from the myriad measures already used as filters for relevant rules in different fields. This study is to answer the question whether other measures are more suited for the different phases of the associative classifier, and an attempt to

	Local support				Global support			
Datasets	**RC %**	**FC %**	**AC %**	**Filtering measure**	**RC %**	**FC %**	**AC %**	**Filtering measure**
Anneal	-58.85	-0.57	0.00	chi2	-93.04	+1.34	+0.02	klosgen
Breast	-99.90	-1.75	-1.63	klosgen	-98.73	-1.82	-1.62	collstrength
Census	-99.90	-0.66	-2.49	klosgen	-99.95	-1.90	-4.17	weightdep
Colic	-100.00	-1.30	+1.91	piatetsky	-100.00	+4.70	3.98	giniindex
Credit	-100.00	-1.66	-2.01	ccr	-99.97	-1.87	-2.14	confirmcau
Diabetes	-99.43	0.38	-5.31	giniindex	-99.30	-0.02	-5.81	giniindex
German	-98.17	-0.52	-2.19	chi2	-100.00	-0.86	-4.25	hyperlift
Glass	-89.59	-1.85	-5.04	kappa	-80.22	-1.01	-3.02	leastcont
Heart	-99.99	-0.29	+0.85	confirmcau	-99.99	-1.49	-0.37	confirmcau
Hepatitis	-99.87	+5.87	-8.86	klosgen	-99.87	+5.87	-8.86	klosgen
Iris	-94.00	+3.06	+2.90	specificity	-92.90	+2.97	+2.17	accuracy
Labor	-99.99	-0.49	+4.86	weightdep	-99.99	-0.49	+4.86	weightdep
Led7	-79.68	-1.95	-1.86	cf	-57.56	-1.24	-1.26	cf
Pima	-99.48	+0.95	-5.46	giniindex	-99.37	+0.73	-5.79	giniindex
Tictactoe	-99.74	+0.93	+0.86	relativerisk	-99.57	+3.01	+1.01	klosgen
Vote	-100.00	-0.24	-0.24	accuracy	-100.00	-0.24	-0.24	accuracy
Vowel	-72.23	-1.09	-0.84	klosgen	-61.57	-0.46	-23.60	sebag
Waveform	-99.53	-1.88	-2.52	hyperlift	-99.00	-1.01	-6.94	infogain
Wine	-99.33	-1.37	-1.72	jaccard	-96.41	-1.16	+2.03	piatetsky
Zoo	-62.41	+0.42	-0.29	ccr	-62.41	+0.42	-0.29	ccr

Table 5. Maximum rule reduction while the f-measure does not decrease more than 2% of its original f-measure using filtering measures on original datasets.(RC: # of rule change, FC: f-measure change, AC: accuracy change)

identify the best measure for each phase. The results clearly indicate that many interestingness measures can indeed provide a better set of classification rules (i.e. a dastic reduction in the number of rules). However there was no single measure that was consistently impacting the rule set for all datasets tested, even though for eatch dataset, some interestingness measure was successful in reducing the rule set without jeopardizing the effectiveness of the classifier. An interesting future study would be to identify the relevant features of a dataset that would help indicate the appropriate interestingness measure to use, and in this way exploit these features to build a predictor for best measure to use in the associative classifier given a specific training set. Our study also shows that confidence is a good choice in the third phase of the associative classification. Selecting the label of the rules with the highest confidence average among applicable rules, is a good and safe bet. Using other measures did not provide significant improvement in classification effectiveness on most tested datasets.

References

1. D.J. Newman A. Asuncion. UCI machine learning repository, 2007.

	Local support				Global support			
Datasets	**RC %**	**FC %**	**AC %**	**Filtering measure**	**RC %**	**FC %**	**AC %**	**Filtering measure**
Anneal	-2.19	+1.08	+0.83	dilatedchi2	-16.51	+10.78	+4.34	kmeasure
Breast	-4.50	+0.18	+0.15	hyperconf	0.00	0.00	0.00	N/A
Census	-52.39	+2.56	-0.63	oddmulti	-61.49	+4.03	+0.12	oddmulti
Colic	-100.00	+5.99	+6.03	2waysupport	-100.00	+4.81	+3.98	2waysupport
Credit	-90.38	+0.79	+0.65	intensofimpl	-0.88	+0.01	0.00	intensofimpl
Diabetes	-36.82	+5.94	+1.59	addedvalue	-40.43	+5.34	+0.87	addedvalue
German	-30.07	+8.78	-3.70	exncounterex	-99.56	+45.78	-5.71	1waysupport
Glass	-43.83	+13.04	+11.31	intensofimpl	-7.30	+10.62	+7.15	ccd
Heart	-87.63	+4.15	+4.08	fmeasure	-81.09	+2.90	+2.82	fmeasure
Hepatitis	-83.87	+9.76	-2.20	chi2	-83.87	+9.76	-2.20	chi2
Iris	-65.96	+4.62	+4.35	klosgen	-58.67	+5.38	+5.11	klosgen
Labor	-99.40	+13.16	+13.36	jaccard	-99.40	+13.16	+13.36	jaccard
Led7	-35.66	+2.02	+1.61	lift	-33.69	+3.11	+2.52	lift
Pima	-56.36	+5.49	+0.53	1waysupport	-60.72	+5.09	+0.01	1waysupport
Tictactoe	-99.59	+1.90	+1.71	correlation	-98.91	+12.49	9.96	correlation
Vote	-1.03	+0.97	+0.97	infogain	-4.46	+0.43	+0.49	loevinger
Vowel	-41.30	+4.48	+4.61	intensofimpl	-14.48	+4.56	+3.98	loevinger
Waveform	-68.84	+2.12	+1.55	intensofimpl	-97.58	+3.43	+2.14	conviction
Wine	-97.63	+4.64	+4.43	2waysupport	-93.89	+1.27	+1.27	2waysupport
Zoo	-43.67	+1.88	+0.88	cf	-43.67	+1.88	+0.88	cf

Table 6. Maximum f-measure improvement using filtering measures on original rule-sets. (RC: # of rule change, FC: f-measure change, AC: accuracy change)

2. Maria-Luiza Antonie and Osmar R. Zaïane. Text document categorization by term association. In *In Proc. of the IEEE 2002 International Conference on Data Min ing*, pages 19–26, Maebashi City, Japan, 2002.
3. Arunasalam Bavani and Chawla Sanjay. Cccs: a top-down associative classifier for imbalanced class distribution. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 517–522, New York, NY, USA, 2006. ACM.
4. Verhein F. and Chawla S. Using significant positively associated and relatively class correlated rules for associative classification of imbalanced datasets. In *Proceedings of the Seventh IEEE International Conference on Data Mining (ICDM '07)*, pages 679–684, Los Alamitos, 2007. IEEE Computer Society Press.
5. Cong Gao, Tung Anthony K. H., Xu Xin, Pan Feng, and Yang Jiong. Farmer: finding interesting rule groups in microarray datasets. In *SIGMOD '04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 143–154, New York, NY, USA, 2004. ACM.
6. McGarry Ken. A survey of interestingness measures for knowledge discovery. *Knowl. Eng. Rev.*, 20(1):39–61, 2005.
7. Yves Kodratoff. Comparing machine learning and knowledge discovery in databases: An application to knowledge discovery in texts. In *In: ECCAI summer*, pages 1–21. Springer, 2000.

	Local support			Global support		
Datasets	**Fmeasure change %**	**Acc change %**	**Selecting measure**	**Fmeasure change %**	**Acc change %**	**Selecting measure**
Anneal	0.00	0.00	confidence	11.45	2.85	confidencecau
Breast	0.70	0.60	confidencecau	0.68	0.59	confidencecau
Census	4.44	-0.26	confidencecau	5.21	-0.08	confidencecau
Colic	1.79	1.70	exncounterex	0.89	0.35	confidencecau
Credit	0.01	0.00	ccc	0.00	0.00	confidence
Diabetes	5.99	0.87	confidencecau	5.58	0.33	confidencecau
German	8.53	-0.55	confidencecau	32.87	3.71	conviction
Glass	8.10	4.25	confidencecau	8.25	4.01	confidencecau
Heart	2.02	2.07	laplace	0.83	0.85	exncounterex
Hepatitis	6.95	-1.45	confidencecau	6.95	-1.45	confidencecau
Iris	3.88	3.62	conviction	4.54	4.38	leverage
Labor	1.48	5.26	laplace	1.48	5.26	laplace
Led7	0.67	0.90	oddmulti	1.73	1.60	sebag
Pima	7.53	1.40	confidencecau	7.09	1.22	leverage
Tictactoe	0.50	0.43	ccc	8.93	6.83	confidencecau
Vote	0.00	0.00	confidence	0.00	0.00	confidence
Vowel	1.90	2.23	confirmcau	5.48	3.63	hyperconf
Waveform	0.31	0.27	confidencecau	2.57	2.14	sebag
Wine	0.00	0.00	confidence	0.04	0.07	exncounterex
Zoo	0.14	0.00	exncounterex	0.14	0.00	exncounterex

Table 7. Maximum f-measure improvement using selecting measure on original rule sets.

8. Geng L. and Hamilton H. J. Interestingness measures for data mining: A survey. *ACM Comput. Surv.*, 38(3):9, 2006.
9. W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on mul tiple class-association rules. In *IEEE International Conference on Data Mining (ICDM'01)*, San Jose, California, November 29-December 2 2001.
10. Bing Liu, Wynne Hsu, and Yiming Ma. Integrating classification and association rule mining. In *KDD*, pages 80–86, 1998.
11. Fayyad U. M. and K. B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *In Proceedings of the Thirteenth International Joint Conference on AI*, pages 1022–1027, 1993.
12. Zaki M., Parthasarathy S., Ogihara M., and Li W. New algorithms for fast discovery of association rules. In *Proc. 3rd Int. Conf. on Knowledge Discovery and Data Mining (KDD97)*, pages 283–296, Menlo Park, CA, USA, 1997. AAAI Press.
13. Agathe Merceron and Kalina Yacef. Interestingness measures for association rules in educational data. In *EDM '08: Proceedings of Educational Data Mining Conference*, pages 57–66, 2008.
14. Hahsler Michael and Hornik Kurt. New probabilistic interest measures for association rules. *Intell. Data Anal.*, 11(5):437–455, 2007.
15. Ohsaki Miho, Abe Hidenao, Tsumoto Shusaku, Yokoi Hideto, and Yamaguchi Takahira. Proposal of medical kdd support user interface utilizing rule interestingness measures. In *ICDMW '06: Proceedings of the Sixth IEEE International*

	Local support				Global support			
Datasets	**FC %**	**AC %**	**Filtering measure**	**Selecting measure**	**FC %**	**AC %**	**Filtering measure**	**Selecting measure**
Anneal	2.76	3.10	chi2	conviction	2.79	0.86	klosgen	conviction
Breast	0.00	0.00	klosgen	confidence	0.62	0.46	collstrength	ccc
Census	0.01	0.00	klosgen	oddsratio	0.00	0.01	weightdep	collstrength
Colic	0.00	0.00	piatetsky	confidence	0.00	0.00	giniindex	confidence
Credit	0.00	0.00	ccr	confidence	0.33	0.33	confirmcau	2waysupport
Diabetes	0.00	0.00	giniindex	confidence	0.00	0.00	giniindex	confidence
German	11.38	0.14	chi2	confidencecau	0.00	0.00	hyperlift	confidence
Glass	10.05	0.89	kappa	hyperconf	6.20	2.78	leastcont	hyperconf
Heart	0.41	0.43	confirmcau	intensofimpl	0.41	0.43	confirmcau	intensofimpl
Hepatitis	1.16	2.31	klosgen	cf	1.16	2.31	klosgen	cf
Iris	1.47	1.41	specificity	leverage	0.82	0.71	accuracy	correlation
Labor	0.00	0.00	weightdep	confidence	0.00	0.00	weightdep	confidence
Led7	0.95	1.15	cf	oddmulti	0.95	1.15	cf	oddmulti
Pima	0.06	0.00	giniindex	1waysupport	0.06	0.00	giniindex	1waysupport
Tictactoe	0.00	0.00	relativerisk	confidence	0.00	0.00	klosgen	confidence
Vote	0.00	0.00	accuracy	confidence	0.00	0.00	accuracy	confidence
Vowel	2.46	2.68	klosgen	ccr	0.46	-0.37	sebag	infogain
Waveform	1.67	1.82	hyperlift	sebag	0.31	0.29	infogain	oddsratio
Wine	0.00	0.00	jaccard	confidence	0.00	0.00	piatetsky	confidence
Zoo	0.00	0.00	ccr	confidence	0.00	0.00	ccr	confidence

Table 8. Maximum f-measure improvement using selecting measures on filtered rulesets that have the maximum rule reduction. (FC:f-measure change, AC: accuracy change)

Conference on Data Mining - Workshops, pages 759–764, Washington, DC, USA, 2006. IEEE Computer Society.

16. Lenca P., Meyer P., Vaillant B., and L allich S. A multicriteria decision aid for interestingness measure selection. Technical Report LUSSI-TR-2004-01-EN, LUSSI Department, GET/ENST, Bretagne, France, 2004.
17. Tan P. and Kumar V. Interestingness measures for association patterns: A perspective. Technical Report 00-036, Department of Computer Sciences, University of Minnesota, 2000.
18. Tan Pang-Ning, Kumar Vipin, and Srivastava Jaideep. Selecting the right interestingness measure for association patterns. In *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 32–41, New York, NY, USA, 2002. ACM.
19. Agrawal R. and Srikant R. Fast algorithms for mining association rules. In *The International Conference on Very Large Databases*, pages 487–499, 1994.
20. Bayardo Roberto and Agrawal Rakesh. Mining the most interesting rules. In *KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 145–154. ACM, 1999.
21. Benoit Vaillant, Philippe Lenca, and Stephane Lallich. A clustering of interestingness measures. In *DS '04: Proceedings of the seventh Internation Conference on Discovery Science*, pages 290–297, 2004.

	Local support				Global support			
Datasets	**FC %**	**AC %**	**Filtering measure**	**Selecting measure**	**FC %**	**AC %**	**Filtering measure**	**Selecting measure**
Anneal	0.00	0.00	dilatedchi2	confidence	2.97	0.96	kmeasure	confirmdes
Breast	0.52	0.45	hyperconf	confidencecau	0.68	0.59	N/A	confidencecau
Census	1.99	0.42	oddmulti	oddmulti	1.49	0.33	oddmulti	yulesy
Colic	0.00	0.00	2waysupport	confidence	0.00	0.00	2waysupport	confidence
Credit	0.00	0.00	intensofimpl	confidence	0.00	0.00	intensofimpl	confidence
Diabetes	1.25	0.52	addedvalue	hyperconf	1.53	0.69	addedvalue	hyperconf
German	0.00	0.00	exncounterex	confidence	0.00	0.00	1waysupport	confidence
Glass	7.89	1.98	intensofimpl	confidencecau	0.00	0.00	ccd	confidence
Heart	1.03	1.16	fmeasure	exncounterex	1.03	1.16	fmeasure	exncounterex
Hepatitis	2.71	0.10	chi2	ccc	2.71	0.10	chi2	ccc
Iris	0.70	0.69	klosgen	oddsratio	0.70	0.69	klosgen	yulesy
Labor	0.00	0.00	jaccard	confidence	0.00	0.00	jaccard	confidence
Led7	0.97	0.85	lift	confirmcau	0.16	0.17	lift	laplace
Pima	0.48	0.35	1waysupport	hyperconf	0.58	0.17	1waysupport	confirmdes
Tictactoe	0.00	0.00	correlation	confidencecau	0.00	0.00	correlation	confidence
Vote	0.00	0.00	infogain	confidencecau	0.00	0.00	loevinger	confidence
Vowel	1.04	1.34	intensofimpl	ccr	2.27	1.001	infogain	confidencecau
Waveform	0.30	0.29	intensofimpl	leverage	0.68	0.69	conviction	conviction
Wine	0.00	0.00	2waysupport	confidence	0.00	0.00	2waysupport	confidence
Zoo	0.00	0.00	cf	confidence	0.00	0.00	cf	confidence

Table 9. Maximum f-measure improvement using selecting measures on filtered rulesets that have the maximum f-measure improvment. (FC:f-measure change, AC: accuracy change)

22. Lan Yu, Janssens Davy, Wets Geer, and Chen Guoqing. Improving associative classification by incorporating novel interestingness measures. In *ICEBE '05: Proceedings of the IEEE International Conference on e-Business Engineering*, pages 282–288, Washington, DC, USA, 2005. IEEE Computer Society.