Where we cannot invent, we may at least improve.

– Charles Caleb Colton (1780-1832)

University of Alberta

A STUDY ON INTERESTINGNESS MEASURES FOR ASSOCIATIVE CLASSIFIERS

by

Mojdeh Jalali Heravi

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science

Fall 2009
Edmonton, Alberta

Examining Committee

Osmar R. Zaïane, Computing Science

Lukasz Kurgan, Electrical and Computer Engineering

Davood Rafiei, Computing Science

To mom and dad,

for their endless care and love,

and to my lovely Vahid,

for being in my life.

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 53 different objective measures for associative classification rules. A wide range of UCI datasets are used to study the impact of different "interestingness measures" on different phases of associative classifiers based on the number of rules generated and the accuracy obtained. The results show that there are interestingness measures that can significantly reduce the number of rules for almost all datasets while the accuracy of the model is hardly jeopardized or even improved. However, no single measure can be introduced as an obvious winner.

Acknowledgements

There are many people that I want to thank for all their help and support during my graduate studies in the University of Alberta.

My greatest appreciation goes to my dear parents who have always encouraged me through my whole life. Their love and care through all these years especially in the last two years and from thousands of miles away, cannot be compensated. No success is achievable without their support.

I also thank Shadi, Neshat and Navid, my precious siblings, for being a part of my life. Their love and friendship have made every second of being with them a great memory.

My deepest gratitude also goes to my supervisor, Dr. Osmar Zaïane. His generous guidance and kindness made my graduate studies to be a great experience which I will never forget.

Thanks to all members of the Database Research Group for making a friendly environment for me to feel comfortable during my research. Special thanks to Banafsheh, Atoosa, Farzane, Parisa, Mehdi, Mohsen, Ali Azad, Ali Fakheri, Ali Ketabi and my other friends for making me feel warm in hard winters of Edmonton and for creating me a warm family in Edmonton when I was miles away from my home. All the great moments that we have had together will remain in my heart and my mind for ever.

At last but not least, my warmest appreciations go to my dear Vahid for being next to me in all happy and sad moments. He was the one helping me in the hardest parts of my project and accompanying me in difficult moments. My success during these studies owes him as well. He is the greatest husband and friend ever.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Associative classification [6, 42, 43] is a rule-based approach recently proposed to classify data by discovering associations between a set of features and a class label. 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 framework used in most association rule mining methods, and similarly for mining and selecting rules of associative classifiers. Although these two measures are widely used, they are still not necessarily the ideal measures. This is because in many situations 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 systematic study has been done to identify a better framework or the most appropriate measure.

1.1 Background and Problem Definition

The Associative classifier 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 transactional 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 [4], Eclat [71] or FP-growths [31]) on the resulting transactional data, frequent itemsets are mined and the ones of the form $\{A, c\}$ are extracted where A is a set of features and c is a class label (A and c are disjoint subsets of items). Among these frequent itemsets, the confident ones are chosen to build classification rules of the form $A \rightarrow c$. Then, these rules are used to predict class labels for objects with an unknown class.

As mentioned above, the support-confidence framework is the standard framework 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 (i.e., $P(Ac)$). 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 (i.e., $P(c|A)$). 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. Having these two constraints, still a huge set of rules may be generated. Different approaches are used to prune the rules in the second phase of associative classifiers [17]. Finally, to classify an object, two different approaches are typically used. The first way is to take into account only the best rule with choosing the rule with the "highest rank" based on a defined ordering. The other way is to consider all rules by calculating the "average" value of the measure used in the defined ordering for the matching rules for each class label and choose the label with the highest average as prediction.

The associative classification approach can be summarized in three phases. First, frequent itemsets having a class label are extracted from a training dataset using an association rule mining technique and only the strong rules 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 technique 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, a class label is assigned based on the rules that apply to the object. Support and confidence are obviously used in the first phase, but also in the third phase.

1.1.1 Why Associative Classifiers?

There are many different classification methods vastly used in different problem domains, each being useful and accurate for some specific domains and not useful or not accurate for other domains. Rule-based classifiers are preferred by most of domain experts because of their interpretability. The classifiers built based on rules are highly expressive and easy to understand. Their interpretability also makes it possible for the domain specialists to modify the rules based on their previous knowledge in order to have a more accurate classifier.

Besides from associative classifiers, decision trees are also well-known rule-based classifiers. Decision trees [13, 53] are simple and easy to understand and they can also be built relatively fast. For constructing a model base on decision trees, a greedy search is used to heuristically select the most promising features. However, this greedy (local) search may prune the important rules. Veloso et al. [64] have shown that the rules derived from decision trees are a subset of rules generated from associative classifiers based on information gain assuming a reasonable low minimum support threshold, and associative classifiers make better or at least equal predictions compared to decision trees according to the information gain principle. This is because associative classifiers exhaustively search for rules that satisfy some constraints. Although this method searches for rules globally, the search is done in a controlled manner using computationally efficient algorithms such as Apriori, Eclat or FP-growth. Hence, the rules are still generated fast enough to make it a better choice compared to decision trees.

1.1.2 Disadvantages of Support and Confidence

The antimonotonicity of support makes the support-confidence framework an appropriate approach for building a strong model in many cases, however, they are still not the ideal measures. This framework has been criticised by many authors

[14, 15, 2, 1]. 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 noisy. 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.

Brin et al. [14] show that with high support and confidence, a rule can even have negative correlation between its antecedent and consequent. Table 1.1 shows an example for this problem. In this example, rule $t \rightarrow c$ have 20% support and 80% confidence which make it a strong rule. However, if we calculate the lift for this rule it is equal to $\frac{P(tc)}{P(t)P(c)} = 0.89$. The fact that this value is less than 1 indicates that this rule is negatively correlated; however, using only support and confidence, this characterization of the rule will remain hidden.

	c	$\neg c$	row total
t	20	5	25
$\neg t$	70	5	75
col total	90	10	100

Table 1.1: An example to show negative correlation for a strong rule, $t \rightarrow c$ [14]

1.2 Approach

There are many 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 or rankers to weed-out the least relevant rules. All those measures can be directly applied to associative classifiers as well, although never tested or reported in the literature. This work focuses on probability-based objective rule interestingness measures for associative classification. Using these interestingness measures, there are two questions that should be answered:

First, can "interestingness measures" have any effect on the associative classifiers on its three different phases: rule generation, pruning and selection, so that the mining algorithm improves both in terms of increasing classification accuracy and decreasing the number of rules?

Second, if there are any improvements, is it possible to probe the best measure or measures which can beat the other measures for improving the results base on either the accuracy or the number of rules in all cases? There is a possibility that no one measure can be found to be effective in all circumstances. In this case, are there any relevant dataset characteristics or measure properties that can help build a classifier in order to predict an efficient measure for a dataset?

To answer the above questions 20 different UCI datasets are used with 53 different measures to study the impact of "interestingness measures" on associative classifiers.

1.3 Thesis Statements

Associative classifiers form an important new paradigm competing with other classification paradigms. They still use a support-confidence framework. Yet, the existence of many different interestingness measures used in association rule mining and other data mining tasks, leads us to postulate the following three hypothesis:

Hypothesis 1 Support-confidence is not the ideal framework for associative classifiers. There are other interestingness measures that can have impact on associative classifiers, both in the term of number of rules and classification accuracy.

Hypothesis 2 There is no specific measure that can be the best measure for all different datasets. Different measures are appropriate in different contexts.

Hypothesis 3 There should be learning method that can find a relationship between some dataset characteristics and efficient interestingness measures.

1.4 Dissertation Organization

The remainder of the dissertation is organized as follows: Chapter 2 describes interestingness measures and their properties and introduces 53 different probability-based objective measures reportedly used in association rule mining. In Chapter 3 some related works using different interestingness measures both in association rule mining and associative classifiers are studied. At the end of this chapter, different properties for each measure and a study on clustering measures based on their properties are explained. The methodology of using interestingness measures in the three different phases of an associative classifier is discussed in Chapter 4. Experimental results, comparing the impact of interestingness measures on classification accuracy and the number of generated classification rules, are illustrated in Chapter 5. The paper is concluded in Chapter 6 with reference to future work.

Chapter 2

Interestingness Measures

Generating rules in association rule mining or with associative classifiers can lead to a huge set of rules which make it impossible for users or domain specialists to study. Sifting through thousands or even millions of rules is impractical. Thus, users lose the opportunity to interpret the results, find interesting rules or even modify them for having a more accurate model. To solve this problem, interestingness measures can be used for filtering or ranking association or classification rules.

There are many different rule interestingness measures widely used in machine learning, data mining and statistics. However, to the best of our knowledge, there is still no formal definition of "interestingness". In a study of 38 different measures, Geng and Hamilton [25] have brought together 9 different criteria which specify the interestingness of a pattern. These 9 criteria are as follows:

1. *Conciseness.* To be concise, a pattern should be easy to understand, hence, it should contain a small set of items (attribute-value pairs).

2. *Generality.* To be general, a pattern should be able to cover a large subset of instances. For example, support is a measure for evaluating generality.

3. *Reliability.* To be reliable, a pattern should show an association that appears in a large subset of related instances. Confidence is a measure for evaluating reliability.

4. *Peculiarity.* To be peculiar, a pattern should be extracted from a part of data that have a large difference from the rest of the data (i.e., outlier instances).

5. *Diversity.* To be diverse, elements of a pattern should differ significantly.

6. *Novelty.* To be novel, a pattern should give new information that was not known to the user before and it is not deducible from other patterns.

7. *Surprisingness.* To be surprising, a pattern should describe a relationship which contradicts a user's existing knowledge or another more general pattern.

8. *Utility.* To be of utility, a pattern should contribute to reach a specific goal.

9. *Actionability.* To be actionable in a domain, a pattern should make future decision makings possible in that domain.

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 mentioned criteria that can define the interestingness of a measure, there are 3 main categories that classify interestingness measures: *objective*, *subjective* and *semantics-based* measures [25]. Objective measures are those that are not application-specific or user-specific and depend 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. In practice, the combination of both objective and subjective measures should be used [22]. First, an objective measure can be use to only keep the rules that are potentially interesting. Then, depending on the domain, a subjective measure can be used as a final filter to retain only the truly interesting rules. However, for simplicity, our work only focuses on objective measures.

2.1 Objective Interestingness Measures

There is a large number of objective interestingness measures available in the literature. 53 different examples of probability-based objective rule interestingness measures are shown in Table 2.2. The formula of each measure can be find in Table 2.3. The measures described in this table are defined based on frequencies of a 2×2 contingency table shown in Table 2.1. Descriptions of some of these measures for a rule in the form of $A \rightarrow c$ are as follows: *Support* or *global support* is the fraction of all records containing A and c together. *Local support* measures the support of the rule only in that part of data that has the same class label as the rule's label. Local support is favorable when finding frequent rules from the minority class in an imbalanced data is required. *Accuracy* measures the support of the rule plus the support of its contraposition (i.e., $\neg A \rightarrow \neg c$). *Confirm-descriptive* measures the difference between the support of $A \rightarrow c$ and $A \rightarrow \neg c$. *Complement class support* measures the support of the rule $A \rightarrow \neg c$. However here the reverse of it is used. Hence, the lower the support of the complement class, the higher the value of its reverse and the more interesting the rule.

Confidence is the conditional probability of having A and c together given A. *Confidence-causal* adds the confidence of rule's contraposition to the confidence of the rule. *Confirmed-confidence-descriptive* measures the difference between the confidence of rule $A \rightarrow c$ and rule $A \rightarrow \neg c$. *Confirmed-confidence-causal* measures the same thing for confidence causal. *Laplace* is a variation of confidence that estimates the confidence and becomes more pessimistic as the support of the antecedent decreases. *Ganascia* is another variation of confidence.

Correlation coeffitient(ϕ) indicates the strength and the direction of a linear relationship between the antecedent and consequent of a rule. This measure is closely related to *chi-square* ($\chi^2 = \phi^2 \times N$). However, chi square is often used for goodness of fit testing rather than being a measure for association because it depends on the size of dataset. Lan et al. [38] believe that chi-square is suitable only when the distributions of row total and column total are close. *Dilated chi-square* is introduced to overcome this drawback by adjusting chi-square to a more uniform and fare situation.

Lift measures the dependency between A and c. The value of 1 means that A and c are independent. Values above 1 shows the positive correlation between A and c. *Class correlation ratio* uses lift to measure how much more positively A

	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 2.1: Frequencies shown in a 2×2 contingency table for rule $A \rightarrow c$

is correlated to c relative to $\neg c$. *Conviction* is somehow similar to lift. However, it measures the dependency between A and $\neg c$. Values above 1 shows negative correlation between A and $\neg c$ which leads to positive correlation between A and c. *Leverage* calculates the deviation of A and c from independence. *Cosine* calculate the geometric mean between lift and support to measure the correlation between the antecedent and consequent of the rule. Other variations of lift are *added value*, *certainty factor*, *collective strength* and *Piatetsky-Shapiro*.

If two variables are highly dependent and the value of one of them is known, then the error in predicting the other variable would be small. *Goodman-Kruskal* measures the amount of reduction in the prediction error.

Odds ratio can be used to determine the degree to which antecedent and consequent of the rule are associated with each other. *Yule's Q* and *Yule's Y* are normalized variants of odds ratio.

Mutual information is an entropy-based measure for calculating the dependencies between variables. Entropy is large for a uniform distribution and is small for a skewed distributions. Mutual information calculates the amount of reduction in entropy. If the variables are strongly associated this amount is high. *J-measure* and *Gini index* are other measures based on the probability distribution of variables.

Kappa measures the agreement between a pair of variables. The more the variables agree, the higher the values for $P(Ac)$, $P(\neg A\neg c)$, and consequently kappa.

Jaccard is used to measure the overlap that A and c share in records. And finally, *relative risk* is a ratio of risk in exposed and unexposed groups.

Table 2.2: A list of objective rule interestingness measures, their abbreviations and references.

No.	Measure	Abbreviation	Ref
1	1-way support	1waySup	[67, 25]
2	2-way support	2waySup	[67, 25]
3	2-way support variation	2waySupVar	[25]
4	Accuracy	Acc	[25]
5	Added value	AddVal	[56, 25]
6	Certainty factor	CerFac	[58, 25]
7	Chi-square	Chi2	[60]
8	Class correlation ratio	CCR	[65]
9	Collective Strength	CollStr	[2, 25]
10	Complement class support	CCS	[7]
11	Confidence	Conf	[25]
12	Confidence causal	ConfC	[35]
13	Confirm causal	CnfrmC	[35]
14	Confirm descriptive	CnfrmD	[35]
15	Confirmed-confidence causal	CCC	[35]
16	Confirmed-confidence descriptive	CCD	[35]
17	Conviction	Conv	[14, 25]
18	Correlation coefficient	Corr	[5, 25]
19	Cosine/IS	Cos	[60, 25]
20	Dilated chi-square	D-Chi2	[38]
21	Example and counterexample rate	Ex&Cex	[25]
22	F-measure	FM	[51]
23	Ganascia	Gan	[23, 37]
24	Gini index	Gini	[13, 25]
25	Goodman-Kruskal	GK	[27, 25]
26	Hyper confidence	HConf	[29]
27	Hyper lift	HLift	[29]
28	Implication index	ImpInd	[41, 40]
29	Information gain	InfoGain	[25]
30	Intensity of implication	IntImp	[38]
31	Interestingness Weighting Dependency	IWD	[28, 25]
32	Jaccard	Jacc	[54, 25]
33	J-measure	JM	[59, 25]
34	Kappa	Kappa	[19, 61]
35	Klosgen	Klos	[34, 25]
36	K-measure	KM	[51]
37	Laplace correlation	Lap	[18, 25]
38	Least contradiction	LC	[9, 25]
			Continued on next page

Table 2.2 – continued from previous page

No.	Measure	Abbreviation	Ref
39	Leverage	Lev	[25]
40	Lift/interest	Lift	[15, 25]
41	Loevinger	Loe	[44, 25]
42	Normalized mutual information	MutInfo	[25]
43	Odd multiplier	OddMul	[25]
44	Odds ratio	OddR	[49, 25]
45	Piatetsky-Shapiro	PS	[52, 25]
46	Recall/local support	LocSup	[25]
47	Relative risk	RelRisk	[25]
48	Sebag-Schoenauer	SS	[57, 25]
49	Specificity	Spec	[25]
50	Support/global support	GlbSup	[3, 25]
51	Yule's Q	YulQ	[69, 25]
52	Yule's Y	YulY	[70, 25]
53	Zhang	Zhang	[72, 25]

Table 2.3: Objective rule interestingness measures for a rule in the form of $A \rightarrow c$, where $P(X) = \frac{N(X)}{N}$ and $P(X|Y) = \frac{P(XY)}{P(Y)}$

No.	Measure	Formula
1	1waySup	$P(c\|A) \times \log \frac{P(Ac)}{P(A)P(c)}$
2	2waySup	$P(Ac) \times log \frac{P(Ac)}{P(A)P(c)}$
3	2waySupVar (modified)[1]	$P(Ac) \times log \frac{P(Ac)}{P(A)P(c)}$ $+P(\neg A \neg c) \times \log \frac{P(\neg A \neg c)}{P(\neg A)P(\neg c)}$ $-P(A \neg c) \times \log \frac{P(A \neg c)}{P(A)P(\neg c)}$ $-P(\neg A c) \times \log \frac{P(\neg A c)}{P(\neg A)P(c)}$
4	Acc	$P(Ac) + P(\neg A \neg c)$
5	AddVal	$P(c\|A) - P(c)$
6	CerFac	$\frac{P(A\|c) - P(c)}{1 - P(c)}$
7	Chi2	$N \times \left(\frac{P(Ac) - P(A)P(c)}{\sqrt{P(A)P(c)P(\neg A)P(\neg c)}} \right)^2$
8	CCR	$\frac{N(Ac)(N(A \neg c) + N(\neg A \neg c))}{N(A \neg c)(N(Ac) + N(\neg Ac))}$

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 2.3 – continued from previous page

No.	Measure	Formula
9	CollStr	$\frac{P(Ac)+P(\neg c\|\neg A)}{P(A)P(c)+P(\neg A)P(\neg c)} \times \frac{1-P(A)P(c)-P(\neg A)P(\neg c)}{1-P(Ac)-P(\neg c\|\neg A)}$
10	CCS (modified)	$\frac{N(\neg c)}{N(A\neg c)}$
11	Conf	$P(c\|A)$
12	ConfC	$0.5 \times \left(\frac{P(Ac)}{P(A)} + \frac{P(\neg A\neg c)}{P(\neg c)}\right)$
13	CnfrmC	$P(Ac) + P(\neg A\neg c) - 2 \times P(A\neg c)$
14	CnfrmD	$P(Ac) - P(A\neg c)$
15	CCC	$0.5 \times (P(c\|A) + P(\neg A\|\neg c)) - P(\neg c\|A)$
16	CCD	$P(c\|A) - P(\neg c\|A)$
17	Conv	$\frac{P(A)P(\neg c)}{P(A\neg c)}$
18	Corr	$\frac{P(Ac)-P(A)P(c)}{\sqrt{P(A)P(c)P(\neg A)P(\neg c)}}$
19	Cos	$\frac{P(Ac)}{\sqrt{P(A)P(c)}}$
20	DChi2	$\left(\frac{N}{lmax(\chi^2)}\right)^{\alpha} \times$ Chi2 $lmax(\chi^2) = \frac{(n_1 n_2)^2 \times N}{P(A)P(\neg A)P(c)P(\neg c)}$ $n_1 = min(min(P(A), P(\neg A)), min(P(c), P(\neg c)))$ $n_2 = min(max(P(A), P(\neg A)), max(P(c), P(\neg c)))$ $\alpha = 0.5$
21	Ex&Cex	$1 - \frac{P(A\neg c)}{P(Ac)}$
22	FM	$\frac{2\times P(c\|A)P(A\|c)}{P(c\|A)+P(A\|c)}$
23	Gan	$2 \times P(c\|A) - 1$
24	Gini	$P(A)\left(P(c\|A)^2 + P(\neg c\|A)^2\right) + P(\neg A)(P(c\|\neg A)^2 + P(\neg c\|\neg A)^2) - P(c)^2 - P(\neg c)^2$
25	GK	$\frac{\sum_i max_j P(A_i c_j) + \sum_j max_i P(A_i c_j) - max_i P(A_i) - max_j P(c_j)}{2 - max_i P(A_i) - max_j P(c_j)}$ $A_i \in \{A, \neg A\}, c_j \in \{c, \neg c\}$
26	HConf	$P\left(C_{Ac} < N(Ac)\right) = \sum_{i=0}^{N(Ac)-1} P(C_{Ac} = i)$ $P(C_{Ac} = r) = \frac{\binom{N(c)}{r}\binom{N-N(c)}{N(A)-r}}{\binom{N}{N(A)}}$

Continued on next page

Table 2.3 – continued from previous page

No.	Measure	Formula
27	HLift	$\frac{N(Ac)}{Q_\delta(C_{Ac})}$ find the minimum value for $Q_\delta(C_{Ac})$ where $P\left(C_{Ac} < Q_\delta(C_{Ac})\right) < \delta$, $\delta = 0.99$
28	ImpInd	$\sqrt{N}(\frac{P(A)P(\neg c)-P(A\neg c)}{\sqrt{P(A)P(\neg c)}})$
29	InfoGain	$log(\frac{P(Ac)}{P(A)P(c)})$
30	IntImp	$1-\sum_{k=0}^{N(A\neg c)}\frac{\lambda^k}{k!}e^{-\lambda}$ $\lambda = \frac{N(A)(N-N(c))}{N}$
31	IWD	$\left(\left(\frac{P(Ac)}{P(A)P(c)}\right)^k - 1\right) \times P(Ac)^m$ $k = m = 1$
32	Jacc	$\frac{P(Ac)}{P(A)+P(c)-P(Ac)}$
33	JM (modified)	$P(Ac)log\left(\frac{P(c\|A)}{P(c)}\right) -$ $P(A\neg c)log\left(\frac{P(\neg c\|A)}{P(\neg c)}\right)$
34	Kappa	$\frac{P(Ac)+P(\neg A\neg c)-P(A)P(c)-P(\neg A)P(\neg c)}{1-P(A)P(c)-P(\neg A)P(\neg c)}$
35	Klos	$\sqrt{P(Ac)}(P(c\|A) - P(c))$
36	KM	$P(c\|A)log\left(\frac{P(c\|A)}{P(c)}\right) +$ $P(\neg(c\|A))log\left(\frac{P(\neg(c\|A))}{P(\neg c)}\right) -$ $P(c\|A)log\left(\frac{P(c\|A)}{P(\neg c)}\right) -$ $P(\neg(c\|A))log\left(\frac{P(\neg(c\|A))}{P(c)}\right)$
37	Lap	$\frac{N(Ac)+1}{N(A)+2}$
38	LC	$\frac{P(Ac)-P(A\neg c)}{P(c)}$
39	Lev	$P(c\|A) - P(A)P(c)$
40	Lift	$\frac{P(Ac)}{P(A)P(c)}$
41	Loe (modified)	$\frac{P(A)P(\neg c)}{P(A\neg c)} - 1$
42	MutInfo	$\frac{\sum_i\sum_j P(A_ic_j)\times log_2\frac{P(A_ic_j)}{P(A_i)P(c_j)}}{-\sum_i P(A_i)\times log_2 P(A_i)}$ $A_i \in \{A, \neg A\}, c_j \in \{c, \neg c\}$
43	OddMul	$\frac{P(Ac)P(\neg c)}{P(A\neg c)P(c)}$
44	OddR	$\frac{P(Ac)P(\neg A\neg c)}{P(A\neg c)P(\neg Ac)}$
45	PS	$P(Ac) - P(A)P(c)$

Continued on next page

Table 2.3 – continued from previous page

No.	Measure	Formula
46	LocSup	$P(A\|c)$
47	RelRisk	$\frac{P(c\|A)}{P(c\|\neg A)}$
48	SS	$\frac{P(Ac)}{P(A\neg c)}$
49	Spec	$P(\neg c\|\neg A)$
50	GlbSup	$P(Ac)$
51	YulQ	$\frac{P(AB)P(\neg A\neg c)-P(\neg Ac)P(A\neg c)}{P(AB)P(\neg A\neg c)+P(\neg Ac)P(A\neg c)}$
52	YulY	$\frac{\sqrt{P(AB)P(\neg A\neg c)}-\sqrt{P(\neg Ac)P(A\neg c)}}{\sqrt{P(AB)P(\neg A\neg c)}+\sqrt{P(\neg Ac)P(A\neg c)}}$
53	Zhang	$\frac{P(Ac)-P(A)P(c)}{max(P(Ac)P(\neg c),P(A\neg c)P(c))}$

2.2 Properties of Objective Interestingness Measures

To be able to analyze the objective measures, some properties are proposed for these measures in the literature. In this section, four sets of properties are considered for objective interestingness measure M for a rule in the form of $A \rightarrow c$.

Piatetsky-Shapiro [52] has proposed the three main properties which are desirable for any objective interestingness measures. These properties are as follows:

P1. *$M = 0$ if A and c are independent, that is $P(Ac) = P(A)P(c)$.* This property states that if A and c are independent in an association rule, then the measure should show the least interest which is zero. However, some researchers have relaxed this property, hence, if A and c are independent it is enough that the value for M be a constant [61].

P2. *M monotonically increases with $P(Ac)$ when $P(A)$ and $P(c)$ remain the same.* This property states that, the greater the support of Ac, the more interesting the rule while the supports of A and c remain fixed. In other words, the more positive correlation A and c have, the more interesting the rule.

P3. *M monotonically decreases with $P(A)$ (or $P(c)$) when $P(Ac)$ and $P(c)$ (or $P(A)$) remain the same.* This property states that if the supports of Ac and A

(or c) stay the same, the rule is more interesting if the support of c (or A) gets smaller.

Major and Mangano [45] added the forth property to those three above:

P4. *M monotonically increases with $P(A)$ when $P(c)$, $P(\neg c)$ and the confidence of the rule ($P(c|A)$) remain the same.* This property states that if the confidence of a rule remains fixed, then, the greater the support of A, the more interesting the rule.

Tan et al. [61] have proposed five properties that unlike Piatetsky-Shapiro properties are not actually desirable. Instead, these properties can help to categorize interestingness measures. These five properties are proposed based on a 2×2 contingency table (Table 2.1):

T1. *M is symmetric under variable permutation.* This property states that using measure M the value for rule $A \rightarrow c$ and rule $c \rightarrow A$ should be the same. variable permutation of a contingency table is shown in Figure 2.1(a).

T2. *M is the same when we scale any row or column by a positive factor.* This property requires invariant with any row or column scaling which is shown in Figure 2.1(b).

T3. *M becomes $-M$ if either the rows or the columns are permuted.* This property states that by swapping either the rows or the columns in contingency table, the value of the measure should change its sign (i.e., $M(A \rightarrow c) = -M(\neg A \rightarrow c) = -M(A \rightarrow \neg c)$). By having this property, the measure is able to identify both positive and negative correlations. Row and column permutations are shown in Figure 2.1(c) and (d).

T4. *M remains the same if both rows and columns are permuted.* This property is a special case of property T3 which states $M(A \rightarrow c) = M(\neg A \rightarrow \neg c)$. permuting both rows and columns simultaneously is shown in Figure 2.1(e).

T5. *M should remain the same with the count of records that do not contain A and c.* This property states that measure M should not change with $N(\neg A \neg c)$ and should only take into account the records that contain A, c or both. Figure 2.1(f) shows the situation when $N(\neg A \neg c)$ increases.

Lence et al. [39, 40] have proposed five properties for evaluating association measures:

L1. *M is constant if there is no counterexample to the rule.* This property states that if $P(A\neg c) = 0$, then the value of the measure should be constant or infinity. In other words, if the confidence of a rule is one the measure should have the same interestingness value, regardless of support. This property somehow is in contradiction with property P4 which states that if the confidence is fixed, the greater the support the more interesting the rule.

L2. *M decreases with $P(A\neg c)$ in a linear, convex or concave fashion around 0^+.* This property describes the manner that a measure decreases when a few counterexamples are added. the desired manner depends to the problem domain and the user. If a few counterexample records can be tolerated, then a concave decrease is desired. If a strict confidence of 1 is required, the a convex decrease is desired.

L3. *M increases as the total number of records increases.* This property states that the value of the measure increases with N (total number of records), assuming that $P(A)$, $P(c)$ and $P(Ac)$ remain fixed.

L4. *The threshold is easy to fix.* For a measure that has this property, it is easy to find a threshold that can separate the interesting from uninteresting rules.

L5. *The semantics of the measure are easy to express.* This property describes that the semantics of the measure is easily understandable by the user.

Geng and Hamilton [25] have also proposed two properties to evaluate the relationship between a measure and support and confidence:

G1. *M should be an increasing function of support if the margins in the contingency table are fixed.* Assuming that the margins of contingency table are fixed (i.e., $N(A) = a, N(\neg A) = N - a, N(c) = b, N(\neg c) = N - b$), if support is equal to x then $P(Ac) = x$, $P(\neg Ac) = \frac{b}{N} - x$, $P(A\neg c) = \frac{a}{N} - x$ and $P(\neg A\neg c) = 1 - \frac{a+b}{N} + x$. By substituting these formulas in measures, functions with the variable x is obtained. The function should be increasing by x. This property is exactly the same as property P2.

G2. *M should be an increasing function of confidence if the margins in the contingency table are fixed.* Like G1, by assuming the margins of contingency table to be fixed and confidence is equal to y, then $P(Ac) = \frac{ay}{N}$, $P(\neg Ac) = \frac{b-ay}{N}$, $P(A\neg c) = \frac{a(1-y)}{N}$ and $P(\neg A\neg c) = 1 - \frac{a+b}{N} + \frac{ay}{N}$. Again by substituting these formulas in measures, functions with variable y is obtained. The function should be increasing by y.

All these properties are introduced in the context of association rule mining. They can be used for finding similar measures or to find the appropriate measure for a problem domain if the required measure properties for that domain are known.

	c	$\neg c$
A	p	q
$\neg A$	r	s

$\Longrightarrow$

	A	$\neg A$
c	p	r
$\neg c$	q	s

(a) Variable Permutation Operation

	c	$\neg c$
A	p	q
$\neg A$	r	s

$\Longrightarrow$

	c	$\neg c$
A	k_3k_1p	k_4k_1q
$\neg A$	k_3k_2r	k_4k_2s

(b) Row and Column Scaling Operation

	c	$\neg c$
A	p	q
$\neg A$	r	s

$\Longrightarrow$

	c	$\neg c$
A	r	s
$\neg A$	p	q

(c) Row Permutation Operation

	c	$\neg c$
A	p	q
$\neg A$	r	s

$\Longrightarrow$

	c	$\neg c$
A	q	p
$\neg A$	s	r

(d) Column Permutation Operation

	c	$\neg c$
A	p	q
$\neg A$	r	s

$\Longrightarrow$

	c	$\neg c$
A	s	r
$\neg A$	q	p

(e) Inversion Operation

	c	$\neg c$
A	p	q
$\neg A$	r	s

$\Longrightarrow$

	c	$\neg c$
A	p	q
$\neg A$	r	$s+k$

(f) Null Addition Operation

Figure 2.1: Operations on a contingency table [62]

Chapter 3

Related Work

Interestingness measures are used in different aspects of data mining. McGarry et al. [47] have used these measures to evaluate the worth of rules extracted from neural networks to discover their internal operation. Buntine [16] took advantage of these measures in probabilistic graphical model. Romao et al. [55] have used interestingness measures in a genetic algorithm that optimizes expert beliefs to rank the interestingness of fuzzy prediction rules. Hilderman et al. [32] compared the various diversity measures used for ranking data summaries. Kononenko [36] discovered the properties of measures used in decision trees. And Gavrilov et al. [24] and Zhao et al. [73] compared objective functions used in clustering approaches.

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 the interesting ones. The focus of this work is on objective measures for association and classification rules, hence, only the works related to these two areas are introduced in this section.

3.1 Interestingness measures in association rule mining

Bayadro and Agrawel [11] have proposed an algorithm to mine optimized rules under partial ordering of the rules ($\leq_{sc}$) based on support and confidence (instead of the typical total ordering on rules). For rules r_1 and r_2, this partial ordering is defined as follows:

Group	Objective measures
1	Odds ratio, Yule's Q and Yule's Y
2	Cosine and Jaccard
3	Support and Laplace
4	Correlation, Collective strength and Piatetsky-Shapiro
5	Gini index and Goodman-Kruskal
6	Lift, Added value and Klosgen
7	Mutual information, Certainty factor and Kappa

Table 3.1: Groups of objective measures with similar properties [62]

- $r_1 <_{sc} r_2$ if and only if:
  - $sup(r_1) \leq sup(r_2) \wedge conf(r_1) < conf(r_2)$
  - or $sup(r_1) < sup(r_2) \wedge conf(r_1) \leq conf(r_2)$
- and $r_1 =_{sc} r_2$ if and only if: $sup(r_1) = sup(r_2) \wedge conf(r_1) = conf(r_2)$.

Based on this ordering, a sc-optimal rule r is a rule in the upper border for which there is no other rule r' such that $r \leq_{sc} r'$. In this work, it is shown that from a set of rules which the consequent of them are the same, the most interesting rule that is selected using a monotonic interestingness measure in both support and confidence (e.g., support, confidence, conviction, lift, gain), is an sc-optimal rule. However, this property is useful when the consequent of the rules are identical and the user is only interested in a single most interesting rule.

Tan et al. [61, 62] have introduced five key properties that should be considered for selecting the right interestingness measure for a specific application. These properties are based on operations on contingency tables and are described in Section 2.2 (Properties T1-T5). To study these properties, 21 different objective rule interestingness measures have been used. Using these five properties as well as three properties introduced by Piatetsky-Shapiro [52] (properties P1-P3 in Section 2.2), Tan et al. have grouped some of these measures based on the correlation between their property vectors. Table 3.1 shows these groups.

In the same work, two different consequences of using support are described. Support has been widely used in association rule mining because of its anti-monotonic property which makes it efficient to search by pruning the search space. In addition to this property, Tan et al. have illustrated two other effects of using support. First, it is shown that under certain support constraints, many measures become consistent with each other. To show this, a synthetic data set with 10,000 contingency tables were used. These contingency tables are ranked using 21 different measures and then the similarity between the ranking vectors of each measure has been calculated using the Pearson's correlation. The results show that by forcing a tighter bound for support many measures become highly correlated. In this experiment, by having the support to be between 5% and 30%, which the authors believe is a reasonable range of support for most application domains, most of the measures become correlated to each other with correlations higher than 0.85. The second effect of using support that is shown in this work is that by having a minimum support threshold, most of uncorrelated or negatively correlated contingency tables will be eliminated. This is because when a contingency table represented for rule $A \rightarrow B$ has low support, there should be high support for at least one of the rules $A \rightarrow \neg B$, $\neg A \rightarrow B$ or $\neg A \rightarrow \neg B$. If the support of either $A \rightarrow \neg B$ or $\neg A \rightarrow B$ is high, then A and B are weakly or negatively correlated. Another situation where the rankings of all measures become identical is using standardized [49], positively correlated contingency tables.

The work of Tan et al. also addresses finding the best measure for specific application domains using experts. Having a set of patterns, different 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. To find the best interestingness measure for patters of a specific domain, 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 to present a small and still diverse set of patterns to domain experts.

Lenca et al. [39] use a different way to find the best measure for a domain. They rank 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 multi-criteria decision process. The measure with the highest rank can be selected to be used in that specific domain.

In another work, Lenca et al. [40] have compared interestingness measures based on formal definitions (i.e., measure properties) and experimental results. 5 interestingness measure properties have been described in this work that are explained in Section 2.2 (properties L1-L5). Three of these properties along with three other properties (properties P1-P3 in Section 2.2) were used to group 20 different interestingness measures. 5 different groups have been obtained with a hierarchal ascendant clustering using the average linkage and Manhattan distance. Another clustering has been done based on experimental results from 10 different datasets utilizing an experimentation tool called Herbs [63]. A pre-order agreement coefficient, τ_1, which is derived from Kendall's τ [26], is used to find the similarity between two different rankings done by two different measures. These similarities are used to cluster the measures. They have found 5 main clusters using 10 different datasets. Finally, they compare these two clusterings to show that most of the measures are in the same group. The comparison of these two groups are shown in Table 3.2.

Also, there are some application specific research done in this area. Merceron and Yacef [48] have tried 3 different interestingness measure which are cosine, lift and added value on association rules and found the impact of these measures on educational data. The data contains information of 84 students to study the positive impact of additional resources provided for students. In another work, Ohsaki et al. [51] have applied 16 different interestingness measures on mining association rules to examine the usefulness of these measures for finding interesting rules extracted from clinical data. A dataset of the medical test results on viral chronic hepatitis was used for this study and the results were evaluated by a medical expert. They found Chi-square, recall and accuracy to have the highest performances and also observed

		Experimental				
		Class E_1	Class E_2	Class E_3	Class E_4	Class E_5
Formal	Class F_1	Conf, SS, Ex&Cex				
	Class F_2	Lap		GlbSup, LC		
	Class F_3		OddMul, Conv, Loe, Zhang			
	Class F_4				Lift, InfoGain, AddVal	Corr, Kappa, PS
	Class F_5		Entropic IntImp[1]			IntImp, ImpInd, ProbDiscInd [1]

Table 3.2: Comparing clusterings based on measure properties and experimental results [40]

the usefulness of combining interestingness measure for finding interesting rules.

In addition to these related works, there are two surveys [25, 46] having many useful information about interestingness measures in general.

3.2 Interestingness measures for associative classifiers

In an effort to present better alternatives to confidence in Classification Based on Associations (CBA) [43], Lan et al. [38] have proposed two novel interestingness 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. These two measures are used instead of confidence in CBA to build two other classifiers. The results are compared with the original CBA, C4.5 and Naïve Bayes and is based on error rate. The experiments on 16 UCI datasets show the impact of these measures on having a more accurate and more compact set of rules. The error rate was improved between 1% and 4%

[1]Entropic IntImp (entropic intensity of implication) and ProbDiscInd (probabilistic discriminant index) are not included in the list of measures in this dissertation because of their complexity.

and the rules could be pruned up to 90% of the number of rules in CBA.

After showing that even confident rules can have negative correlations, Arunasalam and Chawla [7] propose a new measure called Complement Class Support (CCS) which guarantees rules to be positively correlated. Then, an algorithm called Classification using CCS (CCCS) is described based on the anti-monotonic property of CCS and the fact that "good" rules have low CCS values using a row enumeration algorithm [20]. A rule is considered in the classifier only if it is positively correlated (i.e., $CCS < support$). For classification, the best rule is considered as the rule with the highest Score Strength which is a combination on confidence, local support and CCS. 8 UCI datasets with three different imbalanced versions of each is used for this study. The results are compared with CBA and are based on error rate and true positive rate. The results show that CCCS is a more suitable choice than CBA for imbalanced datasets.

SPACCC [65], an associative classifier, was introduced by Verhein and Chawla. This classifier utilizes the Fisher Exact Test's (FET) 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. The search strategy used for generating the rules is a bottom up item enumeration. To avoid examining all the rules, the antimonotonicity feature of the concept of being potentially interesting is taken into account-i.e., $A \rightarrow c$ is considered potentially interesting if and only if all $\{A' \rightarrow c | A' \subset A\}$ have been found to be potentially interesting. Three different approaches that have this feature are used in this work:

- **Aggressive-S**- This is an approach borrowed from Webb [66] where a rule $A \rightarrow c$ is potentially interesting only if for all immediate generalizations of it (i.e., $A - \{b\} \rightarrow c$), adding $\{b\}$ will make a significant positive contribution.

- **Simple-S**- This approach uses FET and forces to be antimonotonic-i.e., if and only if all rules in the form of $A - \{b\} \rightarrow c$ are statistically significant, then the significance of rule $A \rightarrow c$ is tested.

- **Support**- This approach considers a rule as potentially interesting if the support of that rule is above a minimum support threshold.

In the next phase, interesting rules are selected from potentially interesting rules by checking two criteria. A rule is interesting if it is statistically significant based on FET and if its value for CCR is greater than 1. For classifying, they use a strength score to rank the rules. This score is a combination of ρ-value, confidence and CCR. 6 balanced datasets from UCI and also imbalanced variations of them are used for this study. The results are compared with CBA, Classification based on Multiple Association Rules (CMAR) [42], CCCS and C4.5 and is based on accuracy, true positive rate, number of rules, search space and time. The results show that they can outperform other algorithms when using Aggressive-S in most of the cases.

Azevedo and Jorge [10] have compared 10 different interestingness measures in selecting phase of associative classifier. Two different selecting approaches have been used in this study. One is selecting the *best rule* using an ordering similar to the rule ordering in CMAR [42]. $R_1 > R_2$ if:

- $M(R_1) > M(R_2)$
- or $M(R_1) = M(R_2) \wedge supp(R_1) > supp(R_2)$
- or $M(R_1) = M(R_2) \wedge supp(R_1) = supp(R_2) \wedge length(R_1) < length(R_2)$

where M is an interestingness measure. The other approach is *weighted voting* which selects a rule by assigning a specific weight to the label of each rule. χ^2 has been used for filtering potentially trivial rules. The comparisons were done using 17 different datasets from UCI repository. Each measure has a rank for each dataset based on error rate. It is demonstrated by the results that the "best rule" strategy leads to lower error rate and consequently higher rank in almost all the cases. Results also show that using conviction with "best rule" has the best mean rank. Over all, conviction, confidence and Laplace were the only measures that could produce competitive classifiers. The authors believe that the reason that other measures could not build a strong classifier is that almost all of them are symmetric measures.

In another part of the same work by Azevedo and Jorge, utilizing dataset features as meta features to find the best measure in the selecting phase of associative classification has been studied. Confidence, conviction and Laplace are three measures used for classifying 17 different datasets. The two dataset features selected for this study are number of classes and normalized class entropy, which measures the balance of class distributions. Because of the small set of datasets, there are no obvious patterns to explain the success of each measure. The only observation is that for most unbalanced datasets, conviction is the best measure to be use.

These related works show that there are a wide range of interestingness measures being used in many different studies for association rule mining and different interestingness measures are also being used in associative classifiers in order to improve their performance. However, most of them are studies only about the selection phase of these classifiers. To the best of our knowledge there are no previous studies for comparing different interestingness measures in the pruning phase of associative classifiers.

3.3 Formal Comparison of Measures

In Section 2.2, 16 different properties for probabilistic objective interestingness measures were explained. One way of comparing different measures is to use these formal explanations of measures' properties and find the similar measures based on them. As it was explained in Section 3.1, Tan et al. [62] have clustered 18 different measures using 8 different properties. Lenca et al. [40] have also clustered 20 different measures using six different properties. The aim of this section is to cluster all 53 measures introduced in Table 2.2 in a similar way as Lenca et al. 's work. However, instead of using only six properties, 14 different properties are used.

3.3.1 Properties of Different Measures

Table 3.3 shows 14 different properties mentioned in Section 2.2 for all 53 measures. Properties L4 and L5 are excluded because they are subjective properties and depend on the user and problem domain. For property P3, "A" means the measure monotonically decreases with $P(A)$, "C" means it monotonically decreases with $P(c)$, "B" means it monotonically decreases with both $P(A)$ and $P(c)$ and "N" means it does not decrease with either $P(A)$ or $P(c)$. For property T3, "R" means the sign of the measure changes with row permutation, "C" means the sign of measure changes with column permutation,"B" means the sign of measure changes with both row and column permutations and "N" means the sign of measure does not change with row or column permutations. For property L2, numbers from 0 to 6 are used which represent respectively convex decrease, linear decrease, concave decrease, decrease but the manner depends on parameters,invariant, increase,and depends on the parameters. For properties G1 and G2, numbers from 0 to 3 are used which represent respectively increase, invariant, decrease, and depending on parameters. For other properties, "Y" means the measure has that property and "N" means the measure does not have that property.

Table 3.3: Properties of objective rule interestingness measures

No.	Measure	P1	P2	P3	P4	T1	T2	T3	T4	T5	L1	L2	L3	G1	G2
1	1waySup	Y	N	A	N	N	N	N	N	N	N	6	N	3	3
2	2waySup	Y	N	B	N	Y	N	N	N	N	N	6	N	3	3
3	2waySupVar	Y	N	N	N	Y	N	B	Y	N	N	3	N	3	3
4	Acc	N	Y	B	N	Y	N	N	Y	N	N	1	N	0	0
5	AddVal	Y	Y	B	N	N	N	N	N	N	N	1	N	0	0
6	CerFac	N	Y	B	N	N	N	N	N	N	N	0	N	0	0
7	Chi2	Y	N	N	Y	Y	N	N	Y	N	Y	6	Y	3	3
8	CCR	Y	Y	B	N	N	N	N	N	N	N	3	N	0	0
9	CollStr	N	Y	A	N	N	N	N	N	N	N	6	N	0	0
10	CCS	N	Y	B	N	N	N	N	N	N	Y	0	N	0	0
11	Conf	N	Y	C	N	N	N	N	N	Y	Y	1	N	0	0
12	ConfC	N	Y	B	N	N	N	N	N	N	Y	0	N	0	0
13	CnfrmC	N	Y	B	N	N	N	N	N	N	N	1	N	0	0
14	CnfrmD	N	Y	C	N	N	N	C	N	N	N	1	N	0	0
15	CCC	N	Y	B	N	N	N	N	N	N	Y	0	N	0	0
16	CCD	N	Y	C	N	N	N	C	N	Y	Y	1	N	0	0
17	Conv	Y	Y	B	N	N	N	N	N	N	Y	0	N	0	0
18	Corr	Y	Y	B	N	Y	N	B	Y	N	N	3	N	0	0

Continued on next page

Table 3.3 – continued from previous page

No.	Measure	P1	P2	P3	P4	T1	T2	T3	T4	T5	L1	L2	L3	G1	G2
19	Cos	N	Y	B	Y	Y	N	N	N	Y	N	2	N	0	0
20	DChi2	Y	N	N	N	N	N	N	N	N	N	6	Y	3	3
21	Ex&Cex	N	Y	C	N	N	N	N	N	Y	Y	2	N	0	0
22	FM	N	Y	B	Y	Y	N	N	N	Y	N	2	N	0	0
23	Gan	N	Y	C	N	N	N	N	N	Y	Y	1	N	0	0
24	Gini	Y	N	N	Y	N	N	N	Y	N	N	6	N	3	3
25	GK	Y	Y	N	N	Y	N	N	Y	N	N	6	N	3	3
26	HConf	N	N	N	N	N	N	N	N	N	N	6	Y	3	3
27	HLift	N	N	N	N	N	N	N	N	N	N	6	Y	3	3
28	ImpInd	Y	Y	B	N	N	N	N	N	N	N	0	Y	0	0
29	InfoGain	Y	Y	B	N	Y	N	N	N	N	N	2	N	0	0
30	IntImp	N	N	N	N	N	N	N	N	N	N	6	Y	3	3
31	IWD	Y	N	B	N	Y	N	N	N	N	N	6	N	3	3
32	Jacc	N	Y	B	Y	Y	N	N	N	Y	N	1	N	0	0
33	JM	Y	N	A	N	N	N	C	N	N	N	6	N	3	3
34	Kappa	Y	Y	B	N	Y	N	N	Y	N	N	3	N	0	0
35	Klos	Y	N	B	N	N	N	N	N	N	N	6	N	3	3
36	KM	N	N	N	N	N	N	N	N	N	N	6	N	3	3
37	Lap	N	Y	C	N	N	N	N	N	Y	N	1	Y	0	0
38	LC	N	Y	C	N	N	N	N	N	Y	N	2	N	0	0
39	Lev	N	Y	B	N	N	N	N	N	N	N	1	N	0	0
40	Lift	Y	Y	B	N	Y	N	N	N	N	N	2	N	0	0
41	Loe	Y	Y	B	N	N	N	N	N	N	Y	0	N	0	0
42	MutInfo	Y	N	N	N	N	N	N	Y	N	N	6	N	3	3
43	OddMul	Y	Y	B	N	N	N	N	N	N	Y	3	N	0	0
44	OddR	Y	Y	B	N	Y	Y	N	Y	N	Y	0	N	0	0
45	PS	Y	Y	B	N	Y	N	B	Y	N	N	1	N	0	0
46	LocSup	N	Y	A	Y	N	N	N	N	Y	N	2	N	0	0
47	RelRisk	Y	Y	B	N	N	N	N	N	N	N	1	N	0	0
48	SS	N	Y	C	N	N	N	N	N	Y	Y	0	N	0	0
49	Spec	N	Y	B	N	N	N	N	N	N	N	4	N	0	0
50	GlbSup	N	Y	N	Y	Y	N	N	N	N	N	1	N	0	0
51	YulQ	Y	Y	B	N	Y	Y	B	Y	N	Y	3	N	0	0
52	YulY	Y	Y	B	N	Y	Y	B	Y	N	Y	3	N	0	0
53	Zhang	Y	Y	B	N	N	N	N	N	N	N	3	N	0	0

3.3.2 Clustering Measures Based on Their Properties

Using the properties in Table 3.3, the measures are clustered with an agglomerative hierarchical clustering algorithm [33] with average linkage. Having each measure as a vector of properties, the distance of two measures is based on the Hamming distance [30]. Figure 3.1 shows different levels of this clustering till the maximum distance among measures in each cluster is 0.25. From this figure, measures with similar properties can be inferred. The aim of this clustering is to find out whether the similar measures based on their properties have the same behaviour in different phases of associative classifiers. This will be evaluated in Section 5.7.

The above clustering can be compared with the work done by Tan et al. [62] or Lenca et al. [40]. However, since Tan et al. have not revealed their clustering method, any comparison with their work may not be fair. Hence, we only report the comparison results of our clustering with that of Lenca et al. .

To find out how similar a clustering, V is to another clustering U, three different evaluation measures can be used. The first evaluation measure is Adjusted Rand Index (ARI) [68] which is based on true positive (TP), false positive (FP), true negative (TN) and false negative (FN). TP is equal to the number of item pairs that are in the same cluster in V and also in the same cluster in U. TN is the number of item pairs that are not in the same cluster in V and also not in the same cluster in U. FN is the number of item pairs that are not in the same cluster in V but are in the same cluster in U. FP is the number of item pairs that are in the same cluster in V but not in the same cluster in U. Having these four values, ARI is computed as:

$$ARI = \frac{2(TP \times TN - FP \times FN)}{(TP + FN)(FN + TN) + (TP + FP)(FP + TN)}$$

The second way is to find the f-measure using the same TP, TN, FP, and FN explained above. Based on these values, f-measure is equal to $\frac{2PR}{R+P}$ where $R = \frac{TP}{TP+FN}$ and $P = \frac{TP}{TP+FP}$.

The third way is to use f-measure but with different definitions for TP, TN, FP, and FN. To compare two clusterings based on this evaluation method, first, a mapping should be done between the clusters of two groups. A cluster from V is mapped to a cluster from U when they have the highest similarity. Here, the similarity is defined as the size of the intersection of two clusters. Then, the equivalent clusters (one mapped to the other) are labeled the same. Finally, TP, TN, FP, and FN and consequently the f-measure is calculated based on this mappings.

18 out of 20 measures used by Lenca et al. [40] are included in this clustering. Table 3.4 shows the comparison of the above clustering with the clustering done by Lenca et al. for these 18 measures. The results show that there are some differences. There reason can be because the space is extended , both in term of number of measures and number of properties in our case. Hence, there are some measures that appear to be in different clusters in the two clusterings. In this comparison, $ARI = 0.36$, $Fmeasure_{itemPairs} = 0.46$ and $Fmeasure_{mapping} = 0.73$. If we only choose this 18 measures with only the properties that are used in Lenca et al.'s work, almost the same clusters are obtained. However, there are still some small differences. The reason is because we have split property P3 into two sub properties. The comparison of this clustering with the work done be Lenca et al. is show in Table 3.5. In this comparison, $ARI = 0.67$, $Fmeasure_{itemPairs} = 0.74$ and $Fmeasure_{mapping} = 0.88$ which is reasonable.

		Our clustering							
		C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8
Clustering done by Lenca et al.	C 1	Conf, SS, Ex&Cex							
	C 2	LC	Lap, GlbSup						
	C 3			Conv, Loe			OddMul, Zhang		
	C 4				Kappa, Lift, InfoGain			Corr, PS	AddVal
	C 5					IntImp	ImpInd		

Table 3.4: Comparing our clustering with the clustering done by Lenca et al. [40] based on measure properties.

Clustering done by Lenca et al.	Our sub clustering				
	C 1	C 2	C 3	C 4	C 5
C 1	Conf, SS, Ex&Cex				
C 2		Lap, LC GlbSup			
C 3			OddMul, Conv, Loe	Zhang	
C 4				Kappa, Lift Corr, AddVal InfoGain, PS	
C 5				IntImp	ImpInd

Table 3.5: Comparing our sub clustering using only the properties and measures used by Lenca et al. [40] with the clustering done by them based on measure properties.

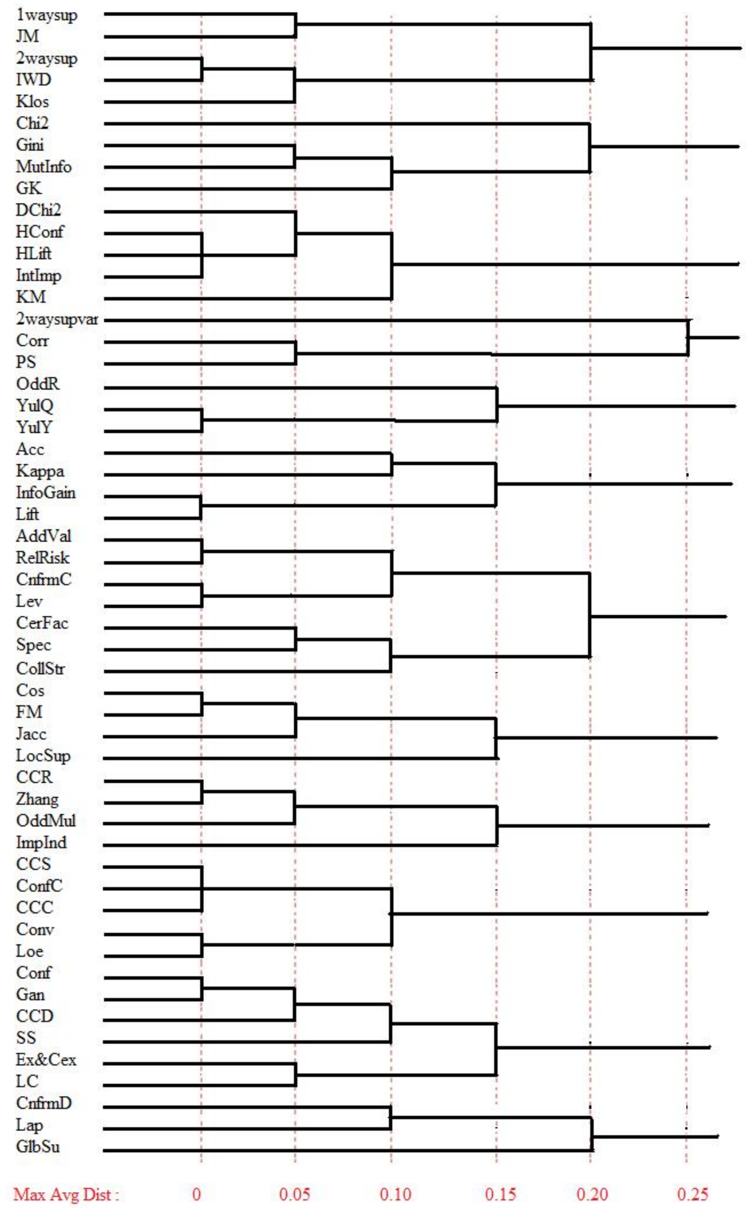

Figure 3.1: An agglomerative hierarchical clustering of measures based on their properties

Chapter 4

Interestingness Measures in Associative Classifiers

As mentioned in Section 1.1.2, support and confidence are not necessarily the ideal measures for associative classifiers. In addition, different interestingness measures have been used to improve association rule mining. Hence, with a high probability, these interestingness measures can also improve associative classifiers if a proper interestingness measure is used.

This chapter addresses the problem of how interestingness measures can be used in three phases of a simple associative classifier: rule generating, rule pruning and rule selection.

4.1 Interestingness Measures in Rule Generation Phase

In the first step of an associative classier, using an association rule mining technique, rules with class labels as consequent are generated. For generating rules, an anti-monotonic 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. For removing the conflicting rules, confidence is the simplest measure that can be used. Here, confidence is used with the lowest possible threshold($\%51$) only to eliminate conflicting rules. Algorithm 1 shows the pseudo code of this phase.

Only measures with anti-monotonic property can be used in this phase. This property for measures listed in Table 2.2 is not known, however, in most of the cases, it is obvious that the measure is not anti-monotone. Hence, in this study, only support (global support) and local support are used for this phase. Local support is usually used for imbalanced datasets. This way, the rules that do not have high support in all the dataset but are frequent in their own class also have the chance to be generated. Although this measure may improve the accuracy of the classifier, it generates more rules than global support.

4.2 Interestingness Measures in Rule Pruning Phase

One of the main problems of using association rule mining is that it generates a huge set of rules even for reasonable minimum support thresholds. In associative classifiers, different approaches are used to prune the rules to make a more compact model with higher accuracy. These pruning approaches are redundancy removal, pruning based on minimum confidence threshold, chi-square test, database coverage and pessimistic error rate [17]. In the first two approaches, an interestingness measure, which is confidence, is used. Hence, redundancy removal and confidence-based pruning are the two pruning approaches that are considered in this study. These two pruning methods can be generalized in order to utilize different interestingness measures.

Redundancy removal pruning

Having a rule set $\mathcal{R}$ and an ordering on $\mathcal{R}$, $<_{or}$, the rules in $\mathcal{R}$ can be pruned as follows: for rules $r_i : A_i \rightarrow c$ and $r_j : A_j \rightarrow c$, if r_i is said to be a general rule of r_j or in other words, $A_i \subset A_j$, and $r_j <_{or} r_i$, r_j is removed as a redundant rule. This redundancy is relative and pruning a rule depends on existent of some other rules in the rule set. The ordering used in this method should be the same as the ordering used in the selection phase to remove the rules that are never used for predicting. This approach can only be safely used when the highest ranked rule is considered in the selection phase. When the prediction is based on the average of the measures used in ordering, this pruning method can change the accuracy because the rules that were pruned are not really redundant in this case. Algorithm

2 shows the pseudo code of this pruning method.

Measure-based pruning

Having a rule set $\mathcal{R}$, an interestingness measure, pm, and a minimum threshold, t_p, rule r_i is removed if $pm(r_i) < t_p$. In fact by defining pm as the interestingness measure, all rules that are not interesting based on this definition are removed. Pruning a rule using this method is independent of other rules. Algorithm 3 shows the pseudo code of this pruning method.

Chiusano and Garza [17] have introduced three properties for pruning techniques: *Idempotency*, *transparency* and *commutativity*. A pruning technique is idempotent if it always provides the same rule set as it is applied one or multiple times on a specific rule set. A pruning approach is said to be transparent if it only removes the redundant rules and does not change the accuracy of the classifier. Having two different pruning approaches, they are said to be commutative if the same rule set is provided independently of the order of pruning methods applied.

Redundancy removal pruning and measure-based pruning both satisfy the idempotency property. Redundancy removal pruning satisfies the transparency property only if the selection phase is based on the highest ranked rule rather than the average of measures of all matchable rules. However, measure-based pruning is not transparent. If the interestingness measure used in measure-based pruning is the same as the one used in selection phase, then the accuracy of the pruned rule set is less or equal to the accuracy of the original rule set. The accuracy decreases here only if for predicting an object, the rule that could match the object is pruned from the original rule set. In this case, the prediction will be the majority class which may not be the same a the object's class. If the measure used for measure-based pruning is not the same as the one for predicting, the accuracy may either decrease or increase in compare to the accuracy of the original rule set. The accuracy may decrease if some rules that are essential in predicting are pruned as uninteresting rules. On the other hand, the accuracy may increase in some cases if some misleading rules that are used for prediction are removed in this pruning.

These two pruning methods can be combined in order to prune even more rules. To study the commutativity property of these two pruning methods, different situations should be considered. Let r_i and r_j be two rules in rule set $\mathcal{R}$ where they have the same consequent and r_i is the general rule of r_j. Using the ordering $<_{or}$ for redundancy removal and measure pm with minimum threshold of t_p for measure-based pruning, four different cases can happen:

- If $pm(r_j) < pm(r_i) \wedge r_j <_{or} r_i$, r_j is removed if redundancy removal pruning is applied. If measure-based pruning is applied, either both r_i and r_j, or only r_j or none of them are removed depending on t_p and there is no way that r_j remains in the rule set but r_i gets eliminated. Hence, applying both redundancy removal pruning and measure-based pruning in any order, does not change the final pruned rule set.

- If $pm(r_i) < pm(r_j) \wedge r_i <_{or} r_j$, none of the rules are removed using redundancy removal. Using measure-based pruning, either both rules, or only r_i, or none of the rules are eliminated depending on t_p. However, because redundancy removal pruning is ineffective in this case, the order of pruning methods does not change the final result.

- If $pm(r_i) > pm(r_j) \wedge r_i <_{or} r_j$, like previous case, the redundancy removal pruning is ineffective. Hence, the order of these two pruning methods does not change the final result.

- If $pm(r_j) > pm(r_i) \wedge r_j <_{or} r_i$, r_j is removed if redundancy removal pruning is applied. Measure-based pruning removes either both rules, or only r_i, or none of the rules depending on t_p. This case, is the only case where the order of pruning application can change the final result. Assuming that measure-based pruning removes only r_i, if redundancy removal pruning is applied first, rule r_j is removed. Then, if the measure-based pruning is applied after that, rule r_i is also eliminated. On the other hand, if measure-based pruning is applied first, rule r_i is removed and then, if redundancy removal pruning is applied because r_j does not have any general rule, will remain in the rule set.

To summarize, redundancy removal and measure-based pruning are commutative only if they use the same interestingness measure or the measure used in measure-based pruning is an increasing function of the ordering used in redundancy removal pruning. Otherwise, the commutativity of these two pruning methods are not guaranteed.

In this study, we only consider one of the orderings, i.e., measure-based pruning and then redundancy removal pruning. The reason is that measure-based pruning removes uninteresting rules and the pruning is independent from other rules in the rule set but redundancy removal pruning eliminates redundant rules depending on the other rules in the rule set. Hence, as the four situations described above, if the last case happens, applying redundancy removal pruning before measure-based pruning may prune an interesting rule because a more general rule for that with a higher rank based on the defined ordering exists in the rule set which is not interesting. This uninteresting rule, that has removed an interesting rule, is also removed by measure-based pruning afterwards.

4.3 Interestingness Measures in Selection Phase

The last phase of an associative classifier is to select a rule or a set of rules for predicting a calss label of an object. For classifying a new object, an ordering should be defined. In this work, an ordering similar to the rule ordering in CMAR [42] is used. However, instead of confidence, any interestingness measure can be used. Let sm be an interestingness measure. Based on this ordering and considering the highest ranked rule, $r_j <_{or} r_i$ (r_i gets a higher rank than r_j), if:

- $sm(r_j) < sm(r_i)$
- or $sm(r_j) = sm(r_i) \wedge support(r_j) < support(r_i)$
- or $sm(r_j) = sm(r_i) \wedge support(r_j) = support(r_i) \wedge length(r_i) < length(r_j)$

Where sm is the selecting measure, $support$ is the support of the rule, and $length$ denotes the length of the rule which is equal to the number of attribute-value pairs in the antecedant.

To take into account the interestingness measure average of all rules that apply, first all rules that apply to the unknown object should be grouped based on their class labels. If $\mathcal{R}_c$ denotes a rule set which all rules have the class label c, based on this ordering, $\mathcal{R}_c <_{orAvg} \mathcal{R'}_{c'}$, if:

- $Avg_{r_j \in \mathcal{R}_c}\{sm(r_j)\} < Avg_{r_i \in \mathcal{R'}_{c'}}\{sm(r_i)\}$
- or $Avg_{r_j \in \mathcal{R}_c}\{sm(r_j)\} = Avg_{r_i \in \mathcal{R'}_{c'}}\{sm(r_i)\} \wedge$
  $Avg_{r_j \in \mathcal{R}_c}\{support(r_j)\} < Avg_{r_i \in \mathcal{R'}_{c'}}\{support(r_i)\}$
- or $Avg_{r_j \in \mathcal{R}_c}\{sm(r_j)\} = Avg_{r_i \in \mathcal{R'}_{c'}}\{sm(r_i)\} \wedge$
  $Avg_{r_j \in \mathcal{R}_c}\{support(r_j)\} = Avg_{r_i \in \mathcal{R'}_{c'}}\{support(r_i)\} \wedge$
  $Avg_{r_i \in \mathcal{R'}_{c'}}\{length(r_i)\} < Avg_{r_j \in \mathcal{R}_c}\{length(r_j)\}$

Both these approaches are used in this study. If no rule can match the object, the dominant class is assigned to it. Algorithms 4 and 5 show the pseudo code of these two selecting approaches.

Algorithm 1: Generating classification rules

```
input  : D, am, t_a.
// D:  a transactional dataset.
// am:  an anti-monotonic measure
// t_a:  a minimum threshold for am
output: A classification rule set.
begin
    // generate frequent itemsets using any frequent
    itemset mining technique.
    I ← FrequentSetMining(D, am, t_a)
    I ← ∅
    // select all classification rules that do not
    conflict.
    foreach rule i in I do
        if i is in the form of {A, c}, where A is a set of attribute-value pair
        and c is a class label then
            r ← (A → c)
            if conf(r) ≥ 0.51 then
                R ← R + {r}
            end
        end
    end
    return R
end
```

Algorithm 2: Redundancy removal pruning

input : $\mathcal{R}$, $<_{or}$.
`// `$\mathcal{R}$`: a classification rule set.`
`// `$<_{or}$`: an ordering definition`
output: A classification rule set.

begin
  $\mathcal{R}' \leftarrow \emptyset$
  foreach *rule* $r : A \rightarrow c$ *in* $\mathcal{R}$ **do**
    $isRedundant \leftarrow false$
    foreach *rule* $r' : A' \rightarrow c'$ *in* $\mathcal{R}'$ **do**
      if $r <_{or} r'$ *and* $A' \subset A$ *and* $c = c'$ **then**
        `// r is redundant`
        $isRedundant \leftarrow true$
        break
      end
      else if $r' <_{or} r$ *and* $A \subset A'$ *and* $c = c'$ **then**
        `// r' is redundant`
        $\mathcal{R}' \leftarrow \mathcal{R}' - \{r'\}$
      end
    end
    `// if r is not redundant so far, it should be added to the new rule set`
    if *not* $isRedundant$ **then**
      $\mathcal{R}' \leftarrow \mathcal{R}' + \{r\}$
    end
  end
  return $\mathcal{R}'$
end

Algorithm 3: Measure-based pruning

input : $\mathcal{R}$, pm, t_p.

// $\mathcal{R}$: a classification rule set.
// pm: a pruning measure measure
// t_p: a minimum threshold for pm

output: A classification rule set.

begin
    foreach *rule* r *in* $\mathcal{R}$ **do**
        if $pm(r) < t_p$ **then**
            $\mathcal{R} \leftarrow \mathcal{R} - \{r\}$
        end
    end
    return $\mathcal{R}$
end

Algorithm 4: Classifying a new object based on the highest ranked rule

input : $\mathcal{R}$, $<_{or}$, obj.

// $\mathcal{R}$: a classification rule set.
// $<_{or}$: an ordering definition
// obj: an object with unknown class label

output: A class label

begin
    $bestRule \leftarrow null$
    foreach *rule* $r : A \rightarrow c$ *in* $\mathcal{R}$ **do**
        if $obj \subseteq A$ *and* $bestRule <_{or} r$ **then**
            $bestRule \leftarrow r$
        end
    end
    if $bestRule = null$ **then**
        return dominant class
    end
    return the class label of $bestRule$
end

Algorithm 5: Classifying a new object based on the average of measures of applied rules

input : $\mathcal{R}$, $<_{orAvg}$, obj.

// $\mathcal{R}$: a classification rule set.
// $<_{orAvg}$: an ordering relation which compares two rule sets based on the average of metrics used in the ordering definition.
// obj: an object with unknown class label, A_obj denotes the attributes of the object.

output: A class label

begin
    foreach *class label* c_i **do**
        $labelToRuleset[c_i] = \emptyset$
    end
    foreach *rule* $r : A \rightarrow c$ *in* $\mathcal{R}$ **do**
        if $A_{obj} \subseteq A$ **then**
            $labelToRuleset[c] \leftarrow labelToRuleset[c] + \{r\}$
        end
    end
    if $labelToRuleset[c_i] = \emptyset \; \forall i$ **then**
        return dominant class
    end
    $bestLabel \leftarrow null$
    foreach *class label* c_i **do**
        if $labelToRuleset[bestLabel] <_{orAvg} labelToRuleset[c_i]$ **then**
            $bestLabel \leftarrow c_i$
        end
    end
    return $bestLabel$
end

Chapter 5

Experimental Results

Experimental results are shown in this chapter to study the impact of interestingness measures based on classification accuracy and number of rules in all three phases of associative classifiers: rule generation, rule pruning and rule selection. Figure 5.1 shows the three different phases in associative classifiers. Any path from start to end point is a matter of discussion in this study such that best possible measures for each phase is to be investigated. First the impact of using different interestingness measures is explored individually for each phase. Then, the combination of the best measures in each phase is also studied.

First, the datasets and the evaluation method used in this study are introduced. Then, the results of each phase are illustrated.

5.1 Datasets

20 datasets having different of characteristics, have been chosen from the UCI repository [8] to observe the impact of using diverse interestingness measures on them. These datasets are commonly used in literature for classification and in particular associative classifiers. To be able to convert the relational datasets into transactional datasets, all numeric attributes are discretized. The same entropy-based discretization method [21] used in CBA [43] is used to categorize the continues attributes. This method is a supervised top-down approach which discritizes the values with no parameters. First, the available values are sorted all being in one interval. Then the potential cut points are selected from the points on boundaries

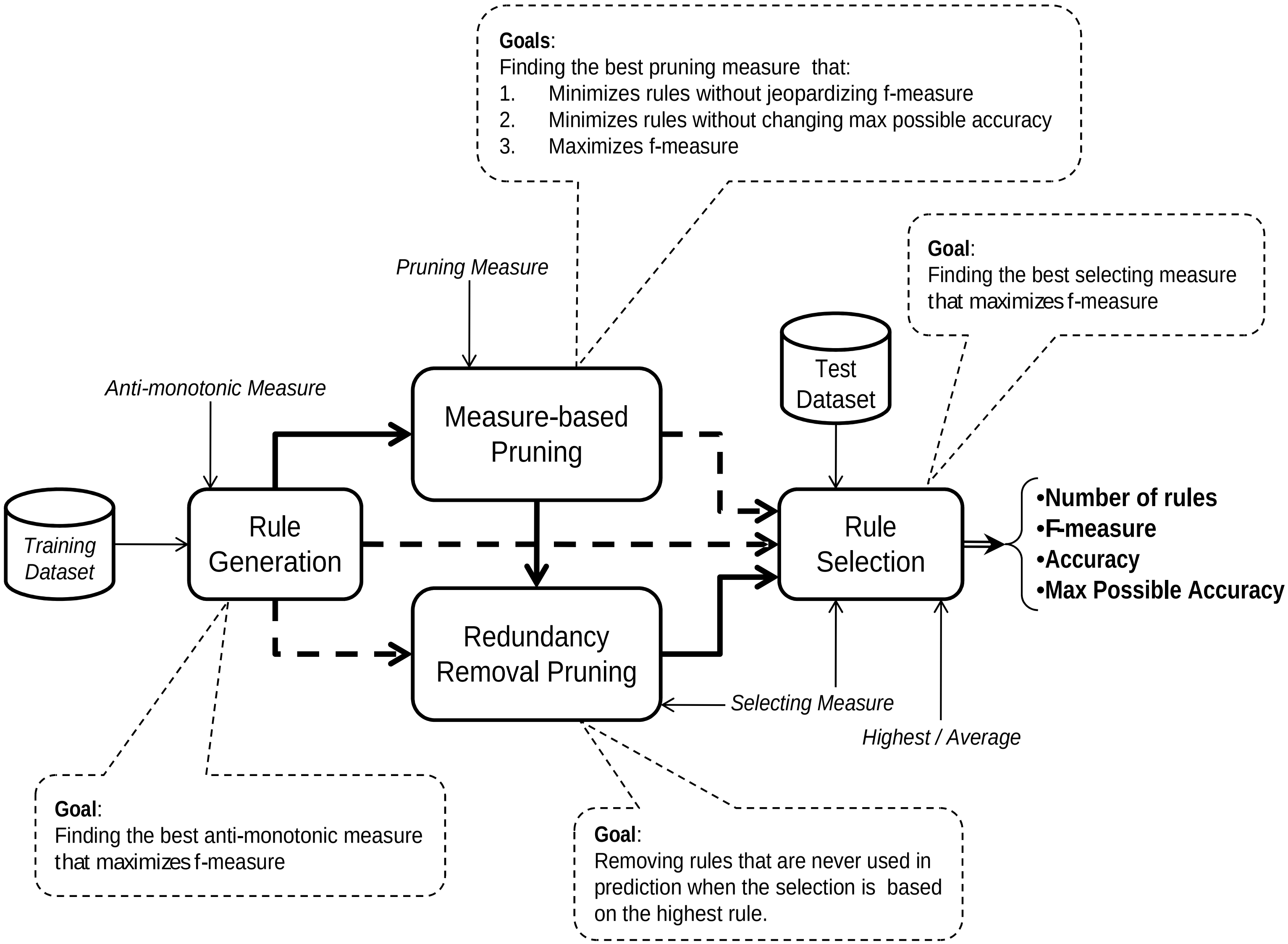

Figure 5.1: Interestingness measures in associative classifiers at a glance.

No.	Datasets	# of att-val	avg # itm	# of trans	# of mxs	avg mxs	max mxs	cls	stdev cls
1	Anneal	70	13.31	898	397	11.74	16	5	0.32
2	Breast	25	8.98	699	257	6.84	10	2	0.22
3	Census	149	12.87	32,561	2,385	7.36	13	2	0.37
4	Colic	61	14.52	368	9,550	7.90	14	2	0.18
5	Credit	54	14.90	690	8,510	8.96	16	2	0.08
6	Diabetes	17	7.00	768	287	6.15	8	2	0.21
7	German	57	15.00	1,000	19,215	8.06	13	2	0.28
8	Glass	20	7.00	214	48	6.52	8	6	0.14
9	Heart	28	10.98	303	1,939	7.27	12	2	0.06
10	Hepatitis	35	16.14	155	1,038	12.39	18	2	0.42
11	Iris	13	4.00	150	23	4.78	5	3	0.00
12	Labor	33	8.40	57	47	9.79	15	2	0.21
13	Led7	14	7.00	3,200	132	4.38	8	10	0.00
14	Pima	17	7.00	768	287	6.15	8	2	0.21
15	Tictactoe	27	9.00	958	2,882	5.27	6	2	0.22
16	Vote	32	15.10	435	8,127	8.95	17	2	0.16
17	Vowel	62	13.00	990	1,022	7.63	11	11	0.00
18	Waveform	107	19.00	5,000	19,245	4.45	9	3	0.00
19	Wine	36	13.00	178	518	9.58	14	3	0.06
20	Zoo	132	17.00	101	53	15.85	17	7	0.13

Table 5.1: 20 different datasets from UCI repository [8]. The columns for each dataset shows name of the dataset, number of attribute-value pairs, average number of items per transactions, number of transactions, number of maximal rules, average size of maximal rules, maximum size of maximal rules, number of classes and standard deviation of class distributions respectively.

where the class label changes. Each time, the best cut point is selected based on entropy and this procedure continues recursively until no "good" cut point is found.

The datasets and some of their properties and characteristics are shown in Table 5.1. These characteristics may help in finding suitable interestingness measures for a dataset. The first column after the name of the dataset shows the total number of attribute-value pairs (items) after discretizing the continues attributes. After that, the average number of items per transaction and total number of transactions are shown. The three next columns show some characteristics of the datasets while generating the maximal rules using global support with minimum support threshold of 1%. The first one shows the total number of maximal rules generated. The second and third columns show the average and the maximum size of these maximal rules respectively. The information about the maximal rules can help to find density-based characteristics of a dataset. The more the number of maximal sets, the higher the number of groups of items that occur frequently, and the higher the average size of the maximals, the larger the groups of items that occur frequently. The last two columns show the number of classes and the standard deviation of class distributions. The more imbalanced the dataset, the higher the value for standard deviation. These characteristics were collected in the hope they could help identify the best characteristics of a dataset that would determine the suitable interestingness measure for that dataset (Hypothesis 3).

5.2 Evaluation Method

Different measures are defined to evaluate the accuracy of a classifier. Some of them are accuracy, micro f_1-measure and macro f_1-measure. The following defines these measures:

- *Accuracy*, also known as overall accuracy, is the fraction of data which is correctly classified: $Acc = \frac{TruePrediction}{N}$, where N is the number of data instances to be classified.

- F_1-*measure*, unlike accuracy, takes into account the prediction of classes separately. For example, a naïve classifier that classifies any data as the majority class can reach the overall accuracy of 90% in an imbalanced data where 90% of the data belongs to the majority class. In spite of a high accuracy, this classifier is not considered "good" because it has never learned other classes. The f_1-measure, however, considers this deficiency by evaluating the prediction power for all classes.

  The f_1-measure with respect to the positive class is based on the confusion matrix which is shown in Table 5.2. Precision, recall and f_1-measure is defined as:

$$Precision(P) = \frac{TP}{TP + FP}$$

$$Recall(R) = \frac{TP}{TP + FN}$$

$$F_1measure(FM) = \frac{2PR}{P+R} = \frac{2TP}{2TP+FP+FN}$$

The definition above for the f_1-measure can only be used for binary class classifications. For an m-class classifier where $m > 2$, macro and micro average f_1-measures are used. These two measures take into account the confusion matrix of the classifier with respect to each class i ($1 \leq i \leq m$), where class i is the positive class and all other classes are put in the negative class. Using Table 5.3,

$$MacroF_1measure = average_{1\leq i\leq m}FM_i$$

$$MicroF_1measure = \frac{2\sum_{(i=1)}^{m} TP_i}{2\sum_{(i=1)}^{m} TP_i + \sum_{(i=1)}^{m} FP_i + \sum_{(i=1)}^{m} FN_i}$$

		Predicted	
		Positive	Negative
Real	Positive	TP	FN
	Negative	FP	TN

Table 5.2: Confusion matrix

Class	Confusion Matrix				F-measure
1	TP_1	TN_1	FP_1	FN_1	FM_1
$\ldots$	$\ldots$				$\ldots$
i	TP_i	TN_i	FP_i	FN_i	FM_i
$\ldots$	$\ldots$				$\ldots$
m	TP_m	TN_m	FP_m	FN_m	FM_m
Total	$\sum_{i=1}^{m} TP_i$	$\sum_{i=1}^{m} TN_i$	$\sum_{i=1}^{m} FP_i$	$\sum_{i=1}^{m} FN_i$	

Table 5.3: Confusion matrix of all classes

Theorem 5.1 *Micro f_1-measure is equal to accuracy.*

Proof Assume that the real label of a data instance is C_i and its predicted label is C_j. This is a true prediction (TP or TN) with respect to all classes iff $i = j$, but when $i \neq j$, it is a false positive (FP) with respect to class j, and a false negative (FN) with respect to class i. With respect to other classes, both i and j become negative and it is considered a true negative (TN). In total, per each prediction,

$\sum_{i=1}^{m} FP_i$ and $\sum_{i=1}^{m} FN_i$ either remain unchanged or both are incremented by 1. Hence, $\sum_{i=1}^{m} FP_i = \sum_{i=1}^{m} FN_i$ and

$$MicroF_1measure = \frac{2\sum_{i=1}^{m} TP_i}{2\sum_{i=1}^{m} TP_i + \sum_{i=1}^{m} FP_i + \sum_{i=1}^{m} FN_i}$$
$$= \frac{\sum_{i=1}^{m} TP_i}{\sum_{i=1}^{m} TP_i + \sum_{i=1}^{m} FP_i} = \frac{TruePrediction}{True\&FalsePrediction}$$
$$= \frac{TruePrediction}{N} = Accuracy$$

To evaluate our work, each classifier is evaluated based on the number of rules its model contains, macro average f_1-measure, accuracy and maximum possible accuracy. Henceforth, f-measure refers to macro average f_1-measure.

The maximum possible accuracy shows the maximum accuracy that is achievable if for each test object, the right rule is selected from the set of available rules. Hence, if for a test object there exists at least one rule that applies to that object with the same class label, that object is considered as a correct classification, otherwise, it is a misclassification. This evaluation measure is useful to evaluate the pruning and see whether the essential rules are pruned or preserved.

All the results in this chapter are based on 10-fold cross validation and the folds used for all classifiers are the same for each dataset.

5.3 Global vs. Local Support in Rule Generation Phase

To generate the rules, Borgelt's implementation [12] of Eclat [71] is used with some modifications in order to only generate classification rules. Local and global supports with a threshold of 1% are used as anti-monotonic measures to prune the search space. To remove the conflicting rules, a minimum confidence threshold of 51% is used. No pruning method is used here and the measure used in the selection phase is confidence with two different approaches, selecting based on the "highest ranked rule" and based on the "average of rules". Rule sets generated only using local/global support and confidence are called "original rule sets". All other results are compared with the results of these rule sets. The results of the original rule sets using global support is shown in Table 5.4 and the same results using local support

			Highest		Average	
Datasets	**# of rules**	**Max acc %**	**FM%**	**Acc%**	**FM%**	**Acc%**
Anneal	309,828	98.11	64.32	88.98	66.64	89.76
Breast	6,936	100.00	94.55	95.12	96.08	96.42
Census	63,226	98.35	65.11	81.69	73.50	83.89
Colic	188,278	98.63	62.48	72.57	80.36	82.57
Credit	299,311	99.42	74.55	76.68	87.64	87.97
Diabetes	923	97.79	66.87	73.83	68.54	74.09
German	223,508	99.00	48.80	71.50	43.62	70.10
Glass	1,599	88.51	55.35	69.31	54.85	66.97
Heart	41,096	100.00	66.05	70.27	80.37	80.85
Hepatitis	1,150,690	100.00	44.26	79.42	67.17	83.65
Iris	108	99.33	95.19	95.33	91.06	91.33
Labor	44,203	100.00	50.10	68.67	82.02	82.33
Led7	473	86.98	73.26	74.00	70.97	72.00
Pima	988	97.53	66.96	74.35	68.64	74.48
Tictactoe	7,398	100.00	70.30	78.72	88.06	90.19
Vote	955,659	99.77	85.17	87.12	95.69	95.87
Vowel	18,501	87.88	61.00	62.73	56.24	58.38
Waveform	35,626	100.00	79.98	80.32	75.57	76.54
Wine	185,942	100.00	76.76	79.87	95.60	95.48
Zoo	971,581	100.00	66.00	81.16	91.26	94.99
Average	225,294	97.56	68.35	78.08	76.69	82.39

Table 5.4: Results on 20 datasets using global support with threshold of 1%, with selecting based on the highest ranked rule and the rules' average of measures. "Max acc" denotes the maximum possible accuracy, "FM" denotes the macro average f-measure and "Acc" denotes the accuracy of the classifier.

is shown in Table 5.5 in terms of number of rules, maximum possible accuracy, f-measure and accuracy.

The results of the original rule sets show that using local support yields a very large number of generated rules, especially when the class labels are imbalanced (i.e., when the standard deviation of class distributions is high), but it also creates more accurate models for this kind of datasets as it also finds frequent patterns in small classes. Figure 5.2 shows the f-measure using local and global supports with selecting either the highest ranked rule or the rules' average of measures for prediction. An observation from this figure is that using "rules' average of measures" leads to more accurate classifiers in most datasets.

			Highest		Average	
Datasets	**# of rules**	**Max acc %**	**FM%**	**Acc%**	**FM%**	**Acc%**
Anneal	791,998	99.78	82.28	91.65	88.59	93.33
Breast	11,338	100.00	94.22	94.84	95.74	96.13
Census	139,508	98.89	72.14	84.00	75.44	84.74
Colic	794,762	99.72	61.49	72.02	79.46	80.98
Credit	868,705	100.00	65.83	71.04	86.92	87.25
Diabetes	1,133	97.79	67.20	73.70	68.26	73.70
German	512,204	99.90	51.65	72.20	61.89	73.00
Glass	2,779	88.51	53.08	69.82	50.64	63.19
Heart	62,833	100.00	64.74	69.94	79.40	79.87
Hepatitis	1,150,690	100.00	44.26	79.42	67.17	83.65
Iris	132	99.33	94.48	94.67	91.73	92.00
Labor	44,203	100.00	50.10	68.67	82.02	82.33
Led7	989	87.38	73.25	74.01	71.49	72.44
Pima	1,201	97.53	67.64	74.47	68.50	74.22
Tictactoe	19,534	100.00	53.44	70.46	97.21	97.49
Vote	1,539,000	99.77	84.82	86.89	95.69	95.87
Vowel	1,176,990	99.70	74.67	75.25	72.28	72.32
Waveform	489,169	100.00	72.90	74.10	79.49	80.14
Wine	479,564	100.00	74.79	77.58	92.51	92.60
Zoo	971,581	100.00	66.00	81.16	91.26	94.99
Average	452,916	98.41	68.45	77.79	79.78	83.51

Table 5.5: Results on 20 datasets using local support with threshold of 1%, with selecting based on the highest ranked rule and the rules' average of measures. "Max acc" denotes the maximum possible accuracy, "FM" denotes the macro average f-measure and "Acc" denotes the accuracy of the classifier.

From now on, only the results for rule sets generated with global support are shown in this chapter. The results for local support can be seen in Appendix B.

5.4 Using Redundancy Removal Pruning

Redundancy removal pruning can be used in order to eliminate the redundant rules. While using the highest ranked rule in selection phase, this pruning can be used to remove the rules that are never used in predicting. Hence, the f-measure and accuracy does not change. Table 5.6 shows a huge percentage of rule reduction while using the redundancy removal pruning on original rule sets generated with global support. On the other hand, using this pruning method while predicting

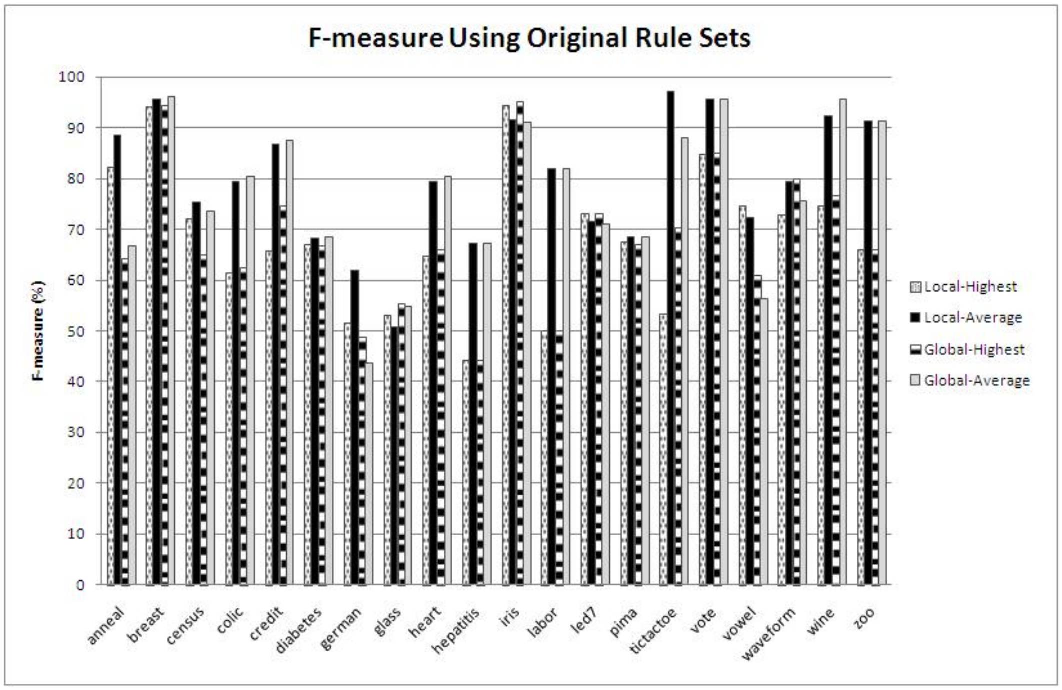

Figure 5.2: F-measure using original rule sets.

based on the rules' average of measures, changes the number of rules as well as f-measure and accuracy. The percentage of change in f-measure and accuracy are also shown in Table 5.6. The results show large reduction of f-measure in some datasets. Hence, although redundancy removal pruning can reduce a large number of rules, it is not safe to be used while predicting is based on the rules' average of measures because it may decrease the f-measure as well.

5.5 Using Different Measures for Measure-based Pruning

Measure-based pruning can be used to filter uninteresting rules. Three different experiments are conducted to find the impact of using 53 different measures in measure-based pruning. Two of these experiments are based on rule reduction. In the first experiment, the goal is to find the minimum number of rules without jeopardizing the f-measure. In the second experiment, the aim is to find the minimum number of rules without changing the maximum possible accuracy. The goal of the last experiment is to eliminate the misleading rules in order to improve the

Datasets	RR %	FC%	AC%
Anneal	99.86	-38.98	-9.78
Breast	95.23	-1.18	-1.04
Census	92.49	-22.09	-5.43
Colic	96.69	-8.23	-4.24
Credit	97.62	-1.68	-1.64
Diabetes	83.42	-2.51	-0.88
German	95.36	-5.61	-0.14
Glass	82.47	-12.11	-10.86
Heart	96.72	-0.03	+0.33
Hepatitis	99.74	-34.10	-5.06
Iris	69.65	+0.90	+0.73
Labor	99.23	-8.19	-6.88
Led7	31.85	-0.30	-0.78
Pima	83.47	-2.20	-0.88
Tictactoe	69.79	-49.92	-25.81
Vote	99.53	-6.21	-5.29
Vowel	88.52	-6.23	-6.75
Waveform	71.58	-3.57	-2.85
Wine	99.14	-13.73	-11.69
Zoo	99.77	-32.45	-21.31

Table 5.6: Percentage of rule reduction while using redundancy removal pruning on rule sets generated with global support as well as the change of f-measure and accuracy while the rules' average of measures are used for prediction. RR, FC and AC are short forms for rule reduction, f-measure change and accuracy change respectively.

f-measure. The selecting measure used for these experiments is confidence.

5.5.1 Rule reduction without jeopardizing the f-measure while using measure-based pruning

The first experiment is to show the impact of different measures on reducing the number of rules while f-measure is not jeopardized. For this experiment, the best measures are found for each dataset that can have the most rule reduction while keeping the f-measure above 95% of its original f-measure. As it is shown in Tables 5.7 and 5.9, in all datasets the number of rules decreases significantly and even in some datasets, the f-measure also improves. Another observation from these tables are that the maximum possible accuracy decreases significantly as well, but the f-

Datasets	RR %	FC%	AC%	MPAC%	Measure
Anneal	-93.04	7.91	1.77	-21.05	Klos (0.1)
Breast	-99.94	-2.22	-1.95	-69.54	2WaySup (0.4)
Census	-99.95	10.74	-1.58	-85.81	IWD (0.1)
Colic	-99.99	-0.63	-0.04	-25.06	CF (0.7)
Credit	-99.99	14.65	11.50	-58.60	CCR (9)
Diabetes	-99.73	-0.14	-8.12	-74.31	Kappa (0.3)
German	-99.99	-2.16	-0.28	-89.90	GK (0.1)
Glass	-80.22	-3.07	-2.59	-15.89	LC (0.3)
Heart	-99.99	19.88	14.64	-65.78	CnfrmC (0.7)
Hepatitis	-99.99	0.00	0.00	-25.75	CnfrmD (0.6)
Iris	-92.90	-1.50	-1.40	-6.04	Acc (0.95)
Labor	-99.99	2.68	-0.97	-80.00	GK (0.5)
Led7	-57.56	-1.43	-1.15	-0.40	CF (0.6)
Pima	-99.66	-0.34	-9.11	-72.36	Kappa (0.3)
Tictactoe	-99.99	-4.84	-11.18	-79.96	CF (0.01)
Vote	-99.99	4.80	5.34	-66.31	Kappa (0.9)
Vowel	-54.63	-3.85	-6.60	-30.92	Gan (0.6)
Waveform	-98.74	-4.47	-5.08	-31.84	SS (13)
Wine	-99.64	5.59	1.19	-40.48	IWD (0.5)
Zoo	-99.09	3.81	4.29	-13.69	CnfrmC (0.99)

Table 5.7: Percentage of rule reduction, f-measure change, accuracy change, maximum possible accuracy change and the measure with the minimum threshold used to get the minimum number of rules with measure-based pruning without jeopardizing the f-measure. Global support is used for rule generation and the selection phase is based on the highest ranked rule. RR, FC, AC and MPAC are short forms for rule reduction, f-measure change, accuracy change and maximum possible accuracy change respectively.

measure does not change that much. The reason is that by this pruning, most of the majority class rules are eliminated. Hence, no rules are available to be applied to the test objects with the majority class, and they are classified as the majority class, which is correct.

Using the same results, the measures are ranked for each dataset based on the percentage of rule reduction. Two measures get the same rank if their percentage of rule reduction is equal up to three decimal places. The rankings are available in Appendix A. Then, the measures are clustered using theses rankings. To cluster the measures, an agglomerative hierarchical clustering algorithm [33] with average linkage is used. Having each measure as a vector of rankings, the correlation be-

Group	Objective measures
1	1WaySup, AddVal, InfoGain, Lift
2	2WaySup, Klos, Corr, Kappa, IWD, ImpInd, Gini, Acc
3	2WaySupVar, JM, HLift
4	CF, MutInfo, LC, Cos, FM, Jacc, Spec
5	Chi2, IntImp
6	CCR, RelRisk, CnfrmC
7	CollStr
8	ConfC, Lev, Lap
9	CnfrmD, HConf
10	CCC, CCD, Gan, Conf, Ex&Cex, SS, Zhang, CCS, Conv, Loe, LocSup, KM, OddMul, OddR, YulQ, YulY
11	DChi2
12	PS, GK
13	GlbSup

Table 5.8: Clusters of measures with similar behaviour in finding the most rule reduction without jeopardizing the f-measure using measure based pruning.Global support is used for rule generation and the selection phase is based on the highest ranked rule.

tween two measures is based on the Spearman's rank correlation coefficient [50]. The clusters are shown in Tables 5.8 and 5.10 for using the highest ranked rule and the average of rules in selection phase respectively. To find the measures that can have the most impact in rule reduction, the number of times a measure ranked between 1 and 3 are counted. Based on these counts, IWD, Kappa, GK, Corr, Klos, 2WaySup, CF, Gini, and Spec are measures that have the most high ranks and 2WaySupVar, ConfC, CCD, Ex&Cex, JM, Lev, OddMul, OddR, YulQ, YulY, Zhang, LocSup, GlbSup, Conf, MutInfo, and CCS are the measures that could not get a top rank in even one dataset.

5.5.2 Rule reduction without changing the maximum possible accuracy while using measure-based pruning

The second experiment is to see how much measure-based pruning can reduce the rules before the maximum possible accuracy decreases. Tables 5.11 and 5.12 show the maximum rule reduction before the maximum possible accuracy changes along with the percentage of change in f-measure and accuracy. The results show that a

Datasets	RR %	FC%	AC%	MPAC%	Measure
Anneal	-93.04	1.34	0.02	-21.05	Klos (0.1)
Breast	-99.94	-3.78	-3.27	-69.54	2WaySup (0.4)
Census	-99.95	-1.90	-4.17	-85.81	IWD (0.1)
Colic	-99.99	4.70	3.98	-73.27	Gini (0.2)
Credit	-99.99	-2.47	-2.81	-58.60	CCR (9)
Diabetes	-99.73	-2.57	-8.44	-74.31	Kappa (0.3)
German	-99.99	9.46	1.71	-89.90	GK (0.1)
Glass	-80.22	-1.01	-3.02	-15.89	LC (0.3)
Heart	-99.99	-1.49	-0.37	-65.78	CnfrmC (0.7)
Hepatitis	-99.99	-4.04	-3.75	-92.95	Gini (0.1)
Iris	-92.90	2.97	2.92	-6.04	Acc (0.95)
Labor	-99.99	-0.49	4.86	-75.00	IWD (0.3)
Led7	-57.56	-1.24	-1.26	-0.40	CF (0.6)
Pima	-99.66	-2.78	-9.27	-72.36	Kappa (0.3)
Tictactoe	-99.57	3.01	1.01	-38.04	Klos (0.1)
Vote	-99.99	-0.24	-0.24	-57.20	Acc (0.95)
Vowel	-65.77	-4.08	-10.38	-42.99	AddVal (0.8)
Waveform	-99.66	-4.81	-4.91	-15.91	Klos (0.1)
Wine	-98.27	-4.55	-4.69	-6.21	Jacc (0.6)
Zoo	-89.80	-4.44	-1.35	0.00	CF (0.6)

Table 5.9: Percentage of rule reduction, f-measure change, accuracy change, maximum possible accuracy change and the measure with the minimum threshold used to get the minimum number of rules with measure-based pruning without jeopardizing the f-measure. Global support is used for rule generation and the selection phase is based on the rules' average of measures. RR, FC, AC and MPAC are short forms for rule reduction, f-measure change, accuracy change and maximum possible accuracy change respectively.

huge number of rules can be eliminated before the maximum possible accuracy is reduced. While using the highest ranked rule for predicting, the f-measures also have significant increase in some of the datasets. However, when the rules' average of measures is used for predicting, there are some cases with significant decrease in f-measure. The reason might be because some of the rules that have positive impact on predicting are eliminated but there still exist some misleading rules in the rule set. In this case, a better selection measure is needed in order to select the right rules for predicting.

The measures are ranked and then clustered based on the percentage of rule reduction similar to what was explained in Section 5.5.1. The clusters are shown

Group	Objective measures
1	1WaySup, AddVal, InfoGain, KM, Lift
2	2WaySup, ImpInd, Corr
3	2WaySupVar
4	CF, Spec
5	Chi2, IntImp
6	CCR, RelRisk
7	CollStr
8	ConfC, Lev, MutInfo
9	CnfrmC
10	CnfrmD, PS
11	CCC, CCD, Gan, Ex&Cex, SS, Conv, Loe, Conf, Lap
12	Cos, FM, Jacc, Acc
13	DChi2, LocSup, CCS
14	Gini, Klos
15	HConf
16	HLift, GK
17	IWD, Kappa
18	JM
19	LC
20	OddMul, OddR
21	YulQ, YulY
22	Zhang
23	GlbSup

Table 5.10: Clusters of measures with similar behaviour in finding the most rule reduction without jeopardizing the f-measure using measure based pruning.Global support is used for rule generation and the selection phase is based on the rules' average of measures.

in Tables 5.13. It is obvious that the amount of rule reduction without changing the maximum possible accuracy is independent from the strategy used in selection phase. Hence, predicting based on the highest ranked rule or the rules' average of measures has no effect on the measures' rankings. The rankings are available in Appendix A. Based on these rankings, the measures that achieved top ranks (between 1 and 3) in more datasets than the others are FM, Cos, Jacc, CollStr, Acc, and Spec and there are 29 measures that could not get a top rank even in one dataset.

Datasets	RR %	FC%	AC%	Measure
Anneal	-89.75	10.24	2.01	GK (0.01)
Breast	-87.60	-1.97	-1.65	FM (0.4)
Ccensus	-87.93	9.34	1.78	CCR (1.5), RelRisk (1.5), Spec (0.5)
Colic	-61.39	4.46	-9.72	Acc (0.5)
Credit	-58.31	0.16	-2.30	Spec (0.5)
Diabetes	-89.84	-5.79	-1.42	FM (0.3)
German	-93.62	35.34	-2.38	Acc (0.5)
Glass	-44.74	11.37	2.71	FM (0.2)
Heart	-99.94	8.47	4.65	CF (0.3)
Hepatitis	-95.16	43.48	0.65	CollStr (1.5)
Iris	-69.37	-0.75	-0.70	CF (0.2)
Labor	-99.89	73.12	31.07	LC (0.3)
Led7	-4.31	0.00	0.00	CCS (17)
Pima	-89.15	-5.60	-1.92	FM (0.3)
Tictactoe	-98.97	-2.70	-4.67	CollStr (1.5)
Vote	-99.16	11.45	9.24	FM (0.6)
Vowel	-9.24	-0.48	-0.48	Zhang (0.9)
Waveform	-89.96	-6.21	-5.33	HConf (0.2)
Wine	-96.41	12.85	9.02	PS (0.1)
Zoo	-98.43	14.41	9.46	Spec (1)

Table 5.11: Percentage of rule reduction, f-measure change, accuracy change and the measure with the minimum threshold used to get the minimum number of rules while the maximum possible accuracy does not change at all. Global support is used for rule generation and the selection phase is based on the highest ranked rule. RR, FC and AC are short forms for rule reduction, f-measure change and accuracy change respectively.

5.5.3 F-measure improvement while using measure-based pruning

The last experiment shows the impact of using measures in measure-based pruning on improving the f-measure. Each row in Tables 5.14 and 5.16 shows the maximum percentage of f-measure improvement and the measure used for this achievement. The results show that there are some significant improvements in f-measure specially when the selection phase is based on the highest ranked rule. In addition, there are some significant rule reductions as well.

The measures are ranked based on their f-measure improvements for each dataset in Appendix A, and are clustered based on these rankings in Tables 5.15 and 5.17.

Datasets	RR %	FC%	AC%	Measure
Anneal	-89.75	-5.48	-2.10	GK (0.01)
Breast	-87.60	-1.71	-1.48	FM (0.4)
Census	-87.93	-31.00	-7.44	RelRisk (1.5), CCR (1.5), Spec (0.5)
Colic	-61.39	1.17	1.00	Acc (0.5)
Credit	-58.31	-3.55	-3.12	Spec (0.5)
Diabetes	-89.84	-4.59	-0.53	FM (0.3)
German	-93.62	-5.61	-0.14	Acc (0.5)
Glass	-44.74	2.19	-3.25	FM (0.2)
Heart	-99.94	-19.93	-15.11	CF (0.3)
Hepatitis	-95.16	-30.13	-4.26	CollStr (1.5)
Iris	-69.37	2.39	2.19	CF (0.2)
Labor	-99.89	6.55	6.88	LC (0.3)
Led7	-4.31	-0.54	-0.26	CCS (17)
Pima	-89.15	-4.68	-0.88	FM (0.3)
Tictactoe	-98.97	-53.34	-26.97	CollStr (1.5)
Vote	-99.16	-2.67	-2.41	FM (0.6)
Vowel	-9.24	-0.56	-0.69	Zhang (0.9)
Waveform	-89.96	-8.26	-7.21	HConf (0.2)
Wine	-96.41	-1.16	-1.06	PS (0.1)
Zoo	-98.43	-14.35	-5.31	Spec (1)

Table 5.12: Percentage of rule reduction, f-measure change, accuracy change and the measure with the minimum threshold used to get the minimum number of rules while the maximum possible accuracy does not change at all. Global support is used for rule generation and the selection phase is based on the average of rules. RR, FC and AC are short forms for rule reduction, f-measure change and accuracy change respectively.

The measures that achieved top ranks more than the others are Lev, Kappa, Zhnag, Acc, GK, 1WaySup, CF, Cos, FM, LC, and Spec when the selection phase is based on the highest ranked rule and Chi2, Ex&Cex, FM and SS when the selection phase is based on the rules' average of measures. LocSup and GlbSup are the only measures that did not achieve any top ranks.

5.6 Using Different Measures in Rule Selection Phase

The effect of using different selecting measures, in the third phase of the associative classifier, are also studied. This effect is only on the improvement of the

Group	Objective measures
1	1WaySup, AddVal, InfoGain, YulQ, YulY, JM
2	2WaySup, IWD, Corr, Kappa, ImpInd
3	2WaySupVar
4	CF, MutInfo
5	Chi2, FM
6	CCR, RelRisk, Acc
7	CollStr, IntImp
8	ConfC, Lev
9	CnfrmC
10	CnfrmD, Gini, LC
11	CCC, CCS, Ex&Cex, LocSup, GlbSup, KM, CCD, Gan, SS
12	Conv
13	Cos, Jacc, PS
14	DChi2
15	HConf
16	HLift, OddMul, OddR
17	Klos, Spec
18	Lap, Conf
19	Lift
20	Loe, Zhang
21	GK

Table 5.13: Clusters of measures with similar behaviour in finding the minimum number of rules while the maximum possible accuracy does not change using measure-based pruning.Global support is used for rule generation and the selection phase is based on the highest ranked rule.

f-measure. There is no change in the number of rules per-se. Table 5.18 shows the best measures for f-measure improvement for each dataset. From the results, it can be inferred that there are some significant improvements in f-measure, specially when predicting is based on the highest ranked rule.

The measures are ranked based on the f-measure improvements in each dataset. The ranks can be found in Appendix A. The measures are clustered based on this rankings in Tables 5.19 and 5.20. OddMul, CCS, CnfrmC, Conv, Lap, Loe, and Zhang are the measures with the most top ranks when the highest ranked rule is used for selecting and ConfC, CCC, Lev, Conv, CCS, Loe and Ex&Cex are the measures with the most top ranks when the rules' average of measures are used in selection phase. There are 5 measures that did not achieve any top ranks with

Datasets	RR %	FC%	AC%	MPAC%	Measure
Anneal	-16.51	21.64	7.27	-1.81	KM (0.1)
Breast	-9.67	1.84	1.52	-1.43	Lev (0.8)
Census	-55.70	18.64	3.03	-17.99	Zhang (0.8)
Colic	-99.99	34.81	18.32	-72.98	2WaySup (0.3)
Credit	-69.82	17.63	14.54	-5.38	Lap (0.9)
Diabetes	-40.43	8.44	1.40	-19.04	AddVal (0.2)
German	-93.62	35.34	-2.38	0.00	Acc (0.5)
Glass	-44.74	11.37	2.71	0.00	FM (0.2)
Heart	-99.87	28.01	20.68	-0.67	CollStr (9)
Hepatitis	-99.86	62.66	1.68	-0.59	CF (0.05)
Iris	-58.67	0.80	0.70	-0.67	Klos (0.2)
Labor	-1.11	85.92	35.92	-1.67	Lev (0.9)
Led7	-16.86	0.70	0.76	-0.07	HConf (0.9)
Pima	-40.13	9.00	1.05	-18.97	AddVal (0.2)
Tictactoe	-98.91	40.91	25.97	-0.84	Corr (0.2)
Vote	-99.91	12.35	10.02	-0.92	CF (0.4)
Vowel	-8.90	0.15	0.16	-2.76	CCD (0.1), Gan (0.1)
Waveform	-33.55	0.15	0.15	-0.02	Zhang (0.7)
Wine	-10.57	19.42	14.71	-0.53	Lev (0.95)
Zoo	-94.30	19.44	10.50	0.00	LC (0.7)

Table 5.14: Percentage of rule reduction, f-measure change, accuracy change, maximum possible accuracy change and the measure with the minimum threshold used to get the maximum f-measure using measure-based pruning. Global support is used for rule generation and the selection phase is based on the highest of rules. RR, FC, AC and MPAC are short forms for rule reduction, f-measure change, accuracy change and maximum possible accuracy change respectively.

selecting based on the highest ranked rule. However, this number goes up to 32 when selection is based on the average of the rules.

5.7 Comparing Different Clusterings

In previous sections, many measures were introduced that can have impact on improving the f-measure or reducing the number of rules. These results show that there is no one measure that wins in all datasets. Some measures are winners for some datasets and do not have any impact for other datasets. The question is how to find suitable measures for a dataset? One way is to find patterns that show the relation between dataset characteristics and their suitable measures. Some experiments

Group	Objective measures
1	1WaySup, AddVal, InfoGain, OddMul, OddR, Lift
2	2WaySup, Klos
3	2WaySupVar, MutInfo
4	CF, Cos, Jacc, LC, FM
5	Chi2, IntImp, CnfrmC
6	CCR, RelRisk, KM
7	CollStr
8	ConfC, Conf, CCC, CCD, Gan, Ex&Cex, SS, Lap, Conv, Loe
9	CnfrmD, HConf
10	Corr, Kappa, Acc
11	DChi2, HLift
12	Gini, IWD
13	JM
14	Lev
15	PS
16	Spec, GK
17	YulQ, YulY
18	Zhang
19	LocSup
20	GlbSup
21	ImpInd
22	CCS

Table 5.15: Clusters of measures with similar behaviour in finding the maximum f-measure using measure-based pruning.Global support is used for rule generation and the selection phase is based on the highest ranked rule.

were studied to find the relationship between the characteristics of datasets shown in Table 5.1 and the best measures found for each dataset. However, it seems that these characteristics are not proper or maybe not enough for this purpose. Hence, further characteristics of datasets should be studied.

Another way is to find similar measures by clustering them based on their behaviour on different datasets. Using this information, if a suitable measure is known for a dataset, other possible suitable measures can be found. The clusterings shown in previous sections can be used for this purpose. By comparing these clusterings some interesting information can be found. The following measures are found to be in the same cluster in all the above clusterings:

- 1WaySup, AddVall, and InfoGain

Datasets	RR %	FC%	AC%	MPAC%	Measure
Anneal	-16.51	10.78	4.34	-1.81	KM (0.1)
Breast	-2.05	0.00	0.00	-0.14	CollStr (1)
Census	-98.19	4.50	-0.93	-5.81	GK (0.05)
Colic	-99.99	4.81	3.98	-72.98	2WaySup (0.3)
Credit	-0.88	0.01	0.00	0.00	IntImp (0.2)
Diabetes	-40.43	5.34	0.87	-19.04	AddVal (0.2)
German	-99.56	45.78	-5.71	-79.90	1WaySup (1.1)
Glass	-8.37	9.46	5.85	-2.11	Ex&Cex (0.3)
Heart	-98.30	2.83	2.79	0.00	Cos (0.5)
Hepatitis	-99.01	10.27	-2.76	-1.25	GK (0.1)
Iris	-58.67	5.38	5.11	-0.67	Klos (0.2)
Labor	-99.40	13.16	13.36	0.00	Jacc (0.2)
Led7	-33.69	3.11	2.52	-4.39	Lift (7)
Pima	-60.72	5.09	0.01	-35.38	1WaySup (0.5)
Tictactoe	-98.91	12.49	9.96	-0.84	Corr (0.2)
Vote	-4.46	0.43	0.49	0.00	Loe (0.9)
Vowel	-36.19	5.26	4.33	-16.32	1WaySup (2)
Waveform	-95.87	3.48	2.32	-13.18	OddMul (15)
Wine	-92.96	1.17	1.27	0.00	GK (0.2)
Zoo	-60.02	1.88	0.88	0.00	CF (0.2)

Table 5.16: Percentage of rule reduction, f-measure change, accuracy change, maximum possible accuracy change and the measure with the minimum threshold used to get the maximum f-measure using measure-based pruning. Global support is used for rule generation and the selection phase is based on the rules' average of measures. RR, FC, AC and MPAC are short forms for rule reduction, f-measure change, accuracy change and maximum possible accuracy change respectively.

- Cos and Jacc
- CCR and RelRisk
- CCD, Gan, and Ex&Cex
- YulQ and YulY

There are also some measures that appear together only when either the highest ranked rule or the rules' average of measures are used in the selection phase. When the highest ranked rule is used, the measures are:

- Corr and Kappa

Group	Objective measures
1	1WaySup, AddVal, InfoGain, KM, Zhang
2	2WaySup, Corr, JM
3	2WaySupVar, MutInfo
4	CF, HConf
5	Chi2, PS, Gini
6	CCR, RelRisk, CnfrmC, HLift
7	CollStr, Cos, FM, Jacc, Acc
8	ConfC, CCC, Lev
9	CnfrmD, IntImp
10	CCD, Gan, Ex&Cex, SS, Conf, Lap
11	Conv, OddR
12	DChi2, Spec, LocSup
13	IWD, Kappa, ImpInd
14	Klos, GK, LC
15	Lift, GlbSup
16	Loe, YulesQ, YulesY
17	OddMul, CCS

Table 5.17: Clusters of measures with similar behaviour in finding the maximum f-measure using measure-based pruning. Global support is used for rule generation and the selection phase is based on the average of rules.

- CCS and SS (These two measures appear in the same cluster as CCD, Gan, and Ex&Cex)

and the measures in the same cluster, while using the rules' average of measures, are:

- ConfC and Lev
- IWD and Kappa.

Among these clusters, {Cos, Jacc} , {CCD, Gan, and Ex&Cex} and {YulQ and YulY} are the only measures that also appear in the same cluster shown in Figure 3.1 which is based on measure properties. All the properties used in this clustering were introduced in the context of association rule mining. Having only 7 measures from 53 measures in the same clusters shows that there may exist some properties related to associative classifiers that are different from that of association rule mining that should be studied.

5.8 Using Interestingness Measures in Both Pruning and Selection Phases

The impact of using different interestingness measures on each individual phase of the associative classifier is studies in previous sections. In this section, the goal is to study the impact of using different interestingness measure both in pruning and selection phases together. For this reason, the best measures found in measure-based pruning are combined with the best measure found in selection phase for each dataset. For cases where the highest ranked rule is used for predicting, redundancy removal is also used after the measure-based pruning. The results for f-measure changes are shown in Tables 5.21- 5.26. In these tables, the percentage of f-measure changes using measure-based pruning and using different measures in selection phase are compared with that of the combination of these two phases. However, the results show that not only combining the best interestingness measure of each phase does not improve the f-measure, but also, there are some cases with significant decrease in f-measure. Hence, a suitable selecting measure based on an original rule set is not necessarily a suitable selecting measure for a pruned version of that rule set. The tables for rule reduction are not shown for simplicity. However, the redundancy removal could even prune more rules from rule sets already pruned by measure-based pruning.

	Highest			Average		
Datasets	**FC%**	**AC%**	**Measure**	**FC%**	**AC%**	**Measure**
Anneal	20.25	6.02	Klos,	11.45	2.85	ConfC
Breast	0.72	0.47	Lev,	0.68	0.59	ConfC
Census	19.99	3.74	CCS,	5.21	-0.08	ConfC
Colic	30.74	14.21	IntImp,	0.89	0.35	ConfC, Lev
Credit	15.77	12.84	Lap,	0.00	0.00	CCD, Conf, Gan
Diabetes	8.64	0.86	DChi2,	6.45	0.86	CCS
German	35.97	-0.56	Klos,	32.87	3.71	Conv, Loe
Glass	6.84	-1.45	Lev,	8.25	4.01	ConfC
Heart	25.21	17.91	IntImp,	1.57	1.23	ImpInd
Hepatitis	63.74	-0.89	DChi2,	6.95	-1.45	ConfC
Iris	0.80	0.70	1WaySup, Loe, CCC, CCD, Conv, ConfC, InfoGain, Lift, OddMul, Gan,	4.54	4.38	Conv, Lev, Loe, OddMul, SS
Labor	77.68	30.58	IntImp,	1.48	5.26	Lap
Led7	0.31	0.33	CnfrmC,	1.73	1.60	SS
Pima	8.37	-0.36	OddMul,	7.09	1.22	Lev
Tictactoe	40.91	25.97	IntImp,	8.93	6.83	ConfC
Vote	12.65	10.30	CnfrmC,	0.00	0.00	ConfC, CCC, CCD, Lev, Conf, Gan
Vowel	0.90	0.64	CCS,	6.98	5.36	CCS
Waveform	0.02	0.02	Lev,	4.36	3.55	CCS
Wine	14.35	10.50	Lap,	0.04	0.07	Ex&Cex
Zoo	19.49	11.53	CnfrmC,	0.14	0.00	Ex&Cex

Table 5.18: Percentage of f-measure change, accuracy change and the measure used in selection phase to get the maximum f-measure. Global support is used for rule generation and the selection phase is based on both the highest ranked rule and rules' average of measures. FC and AC are the short forms for f-measure change and accuracy change respectively.

Group	Objective measures
1	1WaySup, AddVal, InfoGain, Lift, KM
2	2WaySup, Gini, 2WaySupVar, JM, IntImp, Klos, ImpInd, IWD, Chi2, Corr, Kappa, PS, GK
3	CF, LocSup
4	CCR, RelRisk, Spec
5	CollStr, MutInfo
6	ConfC, CCC, CCD, Gan, Ex&Cex, SS, Conf, Conv, Loe, OddMul, CCS, Zhang
7	CnfrmC, Acc, Cos, FM, Jacc, LC
8	CnfrmD
9	DChi2
10	HConf
11	HLift
12	Lap
13	Lev
14	OddR, YulQ, YulY
15	GlbSup

Table 5.19: Clusters of measures with similar behaviour in selection phase in finding the maximum f-measure .Global support is used for rule generation and the selection phase is based on the highest ranked rule.

Group	Objective measures
1	1WaySup, AddVal, InfoGain, Lift, Zhang, HLift
2	2WaySup, PS, ImpInd
3	2WaySupVar, CF, MutInfo, Cos, FM, Jacc, LocSup, GlbSup
4	Chi2, Gini
5	CCR, RelRisk
6	CollStr
7	ConfC, CCC, Lev
8	CnfrmC, Acc, GK, Spec
9	CnfrmD
10	CCD, Conf, Gan, Ex&Cex, Lap
11	Conv, Loe, CCS, OddMul, SS, OddR
12	Corr, IWD, Kappa
13	DChi2
14	HConf
15	IntImp
16	JM, KM
17	Klos
18	LC
19	YulQ, YulY

Table 5.20: Clusters of measures with similar behaviour in selection phase in finding the maximum f-measure .Global support is used for rule generation and the selection phase is based on the rules' average of measures.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	Klos	Klos	7.91	20.25	6.86
Breast	2waySup	Lev	-2.22	0.72	-2.22
Census	IWD	ccs	10.74	19.99	10.74
Colic	CF	IntImp	-0.63	30.74	-34.89
Credit	CCR	Lap	14.65	15.77	14.65
Diabetes	Kappa	DChi2	-0.14	8.64	-0.14
German	GK	Klos	-2.16	35.97	-2.16
Glass	LC	Lev	-3.07	6.84	-6.28
Heart	CnfrmC	IntImp	19.88	25.21	19.88
Hepatitis	CnfrmD	DChi2	0.00	63.74	0.00
Iris	Acc	ConfC	-1.50	0.80	-1.50
Labor	GK	IntImp	2.68	77.68	2.68
Led7	CF	CnfrmC	-1.43	0.31	-2.02
Pima	Kappa	OddMul	-0.34	8.37	-0.34
Tictactoe	CF	IntImp	-4.84	40.91	-4.84
Vote	Kappa	CnfrmC	4.80	12.65	4.80
Vowel	Gan	ccs	-3.85	0.90	-7.82
Waveform	SS	Lev	-4.47	0.02	-4.49
Wine	IWD	Lap	5.59	14.35	5.59
Zoo	CnfrmC	CnfrmC	3.81	19.49	3.81

Table 5.21: Comparing the changes of f-measure with the best measure used in measure-based pruning for rule reduction without jeopardizing the f-measure, the best measure used in selection phase, and the combination of these two measures. Global support is used for rule generation and the selection phase is based on the highest ranked rule. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	Klos	ConfC	1.34	11.45	2.52
Breast	2waySup	ConfC	-3.78	0.68	-3.78
Census	IWD	ConfC	-1.90	5.21	-1.90
Colic	Gini	ConfC	4.70	0.89	4.70
Credit	CCR	Gan	-2.47	0.00	-2.47
Diabetes	Kappa	ccs	-2.57	6.45	-2.57
German	GK	Conv	9.46	32.87	9.46
Glass	LC	ConfC	-1.01	8.25	-2.31
Heart	CnfrmC	ImpInd	-1.49	1.57	-2.13
Hepatitis	Gini	ConfC	-4.04	6.95	-4.04
Iris	Acc	Lev	-1.91	4.54	-1.13
Labor	IWD	Lap	-0.49	1.48	-0.49
Led7	CF	SS	-1.24	1.73	-0.59
Pima	Kappa	Lev	-2.78	7.09	-2.78
Tictactoe	Klos	ConfC	3.01	8.93	3.01
Vote	Acc	ConfC	-0.24	0.00	-0.24
Vowel	AddVal	ccs	-4.08	6.98	-6.86
Waveform	Klos	ccs	-4.81	4.36	-5.44
Wine	Jacc	Ex&Cex	-4.55	0.04	-4.55
Zoo	CF	Ex&Cex	-4.44	0.14	-6.52

Table 5.22: Comparing the changes of f-measure with the best measure used in measure-based pruning for rule reduction without jeopardizing the f-measure, the best measure used in selection phase, and the combination of these two measures. Global support is used for rule generation and the selection phase is based on the average of rules. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	GK	Klos	10.24	20.25	-7.43
Breast	FM	Lev	-1.97	0.72	0.27
Census	CCR	ccs	9.34	19.99	-0.23
Colic	Acc	IntImp	4.46	30.74	30.74
Credit	Spec	Lap	0.16	15.77	13.10
Diabetes	FM	DChi2	-5.79	8.64	5.27
German	Acc	Klos	35.34	35.97	34.78
Glass	FM	Lev	11.37	6.84	4.36
Heart	CF	IntImp	8.47	25.21	-32.86
Hepatitis	CollStr	DChi2	43.48	63.74	63.36
Iris	CF	1WaySup	-0.75	0.80	0.05
Labor	LC	IntImp	73.12	77.68	81.10
Led7	ccs	CnfrmC	0.00	0.31	0.31
Pima	FM	OddMul	-5.60	8.37	7.72
Tictactoe	CollStr	IntImp	-2.70	40.91	-5.75
Vote	FM	CnfrmC	11.45	12.65	12.65
Vowel	Zhang	ccs	-0.48	0.90	0.38
Waveform	HConf	Lev	-6.21	0.02	-6.22
Wine	PS	Lap	12.85	14.35	14.35
Zoo	Spec	CnfrmC	14.41	19.49	13.72

Table 5.23: Comparing the changes of f-measure with the best measure used in measure-based pruning for rule reduction without changing the maximum possible accuracy, the best measure used in selection phase, and the combination of these two measures. Global support is used for rule generation and the selection phase is based on the highest ranked rule. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	GK	ConfC	-5.48	11.45	-16.90
Breast	FM	ConfC	-1.71	0.68	-0.13
Census	CCR	ConfC	-31.00	5.21	-11.62
Colic	Acc	ConfC	1.17	0.89	-2.07
Credit	Spec	CCD	-3.55	0.00	-3.55
Diabetes	FM	ccs	-4.59	6.45	1.14
German	Acc	Conv	-5.61	32.87	32.43
Glass	FM	ConfC	2.19	8.25	7.41
Heart	CF	ImpInd	-19.93	1.57	-13.54
Hepatitis	CollStr	ConfC	-30.13	6.95	-4.55
Iris	CF	Lev	2.39	4.54	2.39
Labor	LC	Lap	6.55	1.48	4.34
Led7	ccs	SS	-0.54	1.73	2.16
Pima	FM	Lev	-4.68	7.09	3.47
Tictactoe	CollStr	ConfC	-53.34	8.93	-22.03
Vote	FM	ConfC	-2.67	0.00	-2.33
Vowel	Zhang	ccs	-0.56	6.98	4.60
Waveform	HConf	ccs	-8.26	4.36	-5.74
Wine	PS	Ex&Cex	-1.16	0.04	-3.05
Zoo	Spec	Ex&Cex	-14.35	0.14	-16.75

Table 5.24: Comparing the changes of f-measure with the best measure used in measure-based pruning for rule reduction without changing the maximum possible accuracy, the best measure used in selection phase, and the combination of these two measures. Global support is used for rule generation and the selection phase is based on the rules' average of measures. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	KM	Klos	21.64	20.25	17.62
Breast	Lev	Lev	1.84	0.72	0.72
Census	Zhang	ccs	18.64	19.99	19.77
Colic	2waySup	IntImp	34.81	30.74	24.82
Credit	Lap	Lap	17.63	15.77	15.58
Diabetes	AddVal	DChi2	8.44	8.64	9.23
German	Acc	Klos	35.34	35.97	34.78
Glass	FM	Lev	11.37	6.84	4.36
Heart	CollStr	IntImp	28.01	25.21	22.64
Hepatitis	CF	DChi2	62.65	63.74	39.37
Iris	Klos	1WaySup	0.80	0.80	0.80
Labor	Lev	IntImp	85.92	77.68	91.54
Led7	HConf	CnfrmC	0.70	0.31	0.31
Pima	AddVal	OddMul	9.00	8.37	10.16
Tictactoe	Corr	IntImp	40.91	40.91	40.91
Vote	CF	CnfrmC	12.35	12.65	12.08
Vowel	CCD	ccs	0.15	0.90	1.09
Waveform	Zhang	Lev	0.15	0.02	0.09
Wine	Lev	Lap	19.41	14.35	18.69
Zoo	LC	CnfrmC	19.44	19.49	19.49

Table 5.25: Comparing the changes of f-measure with the best measure used in measure-based pruning for f-measure improvement, the best measure used in selection phase, and the combination of these two measures. Global support is used for rule generation and the selection phase is based on the highest ranked rules. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	KM	ConfC	10.78	11.45	10.40
Breast	CollStr	ConfC	0.00	0.68	0.35
Census	GK	ConfC	4.50	5.21	3.53
Colic	2waySup	ConfC	4.81	0.89	4.81
Credit	IntImp	Conf	0.01	0.00	0.01
Diabetes	AddVal	ccs	5.34	6.45	7.17
German	1WaySup	Conv	45.78	32.87	45.78
Glass	Ex&Cex	ConfC	9.46	8.25	7.94
Heart	Cos	ImpInd	2.83	1.57	-0.52
Hepatitis	GK	ConfC	10.27	6.95	2.19
Iris	Klos	Lev	5.38	4.54	4.66
Labor	Jacc	Lap	13.16	1.48	7.75
Led7	Lift	SS	3.11	1.73	2.76
Pima	1WaySup	Lev	5.09	7.09	5.09
Tictactoe	Corr	ConfC	12.49	8.93	12.49
Vote	Loe	ConfC	0.43	0.00	3.53
Vowel	1WaySup	ccs	5.26	6.98	70.85
Waveform	OddMul	ccs	3.48	4.36	4.05
Wine	GK	Ex&Cex	1.17	0.04	0.09
Zoo	CF	Ex&Cex	1.88	0.14	1.88

Table 5.26: Comparing the changes of f-measure with the best measure used in measure-based pruning for f-measure improvement, the best measure used in selection phase, and the combination of these two measures. Global support is used for rule generation and the selection phase is based on the rules' average of measures. FC is the short form for f-measure change.

Chapter 6

Conclusion

Associative classification is a relatively 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 or rankers 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 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 drastic reduction in the number of rules) and a more accurate classifier. However, there was no single measure that was consistently impacting the rule set for all datasets tested, even though for each dataset, some interestingness measure was successful in reducing the rule set or improving the effectiveness of the classifier. These measures are introduced for each individual phase. The results show that the measures that are the best in one phase are not necessarily the best measures for the other phase.

Another observation is that using the combination of the best measures in pruning and selection phases does not improve the accuracy of the classifier which means that the best selecting measure for an original rule set is not the best for the pruned version of that rule set. This observation shows that there might exist some rule set characteristics that have effect on selecting the best measure. Hence, for each pruned rule set, the appropriate selecting measure should be investigated separately.

All the measures were clustered in different experiments. Some of the measures behave similarly in all the cases. Hence, in future work, selecting only one measure from each group as a representative, should be sufficient.

Some experiments were conducted to find the relationships between the dataset characteristics and the suitable interestingness measures. However, no evident feature was found that could provide discriminant power to distinguish winning measures for specific datasets. An interesting future study would be to identify the relevant features of a dataset or a rule set 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.

Bibliography

[1] J.M. Adamo. *Data mining for association rules and sequential patterns: sequential and parallel algorithms*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.

[2] C.C. Aggarwal and PS. Yu. A new framework for itemset generation. In *PODS '98: Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 18–24, New York, NY, USA, 1998. ACM.

[3] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international Conference on Management of Data*, pages 207–216, 1993.

[4] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *The International Conference on Very Large Databases*, pages 487–499, 1994.

[5] A. Agresti. *Categorical Data Analysis (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2002.

[6] M.L. Antonie and O.R. Zaïane. Text document categorization by term association. In *Proc. of the IEEE 2002 International Conference on Data Mining*, pages 19–26, Maebashi City, Japan, 2002.

[7] B. Arunasalam and S. Chawla. 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.

[8] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007.

[9] J. Az and Y. Kodratoff. A study of the effect of noisy data in rule extraction systems. In *Sixteenth European Meeting on Cybernetics and Systems Research*, pages 781–786, 2002.

[10] P. J. Azevedo and A. M. Jorge. Comparing rule measures for predictive association rules. In *ECML '07: Proceedings of the 18th European conference on Machine Learning*, pages 510–517, Berlin, Heidelberg, 2007. Springer-Verlag.

[11] R. Bayardo and R. Agrawal. 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.

[12] C. Borgelt. Efficient implementations of apriori and eclat. In *Workshop of Frequent Item Set Mining Implementations (FIMI 2003)*, Melbourne, FL, USA, 2003.

[13] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.

[14] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: Generalizing association rules to correlations. In *SIGMOD '97: Proceedings of the 1997 ACM SIGMOD international conference on Management of data*, pages 265–276, New York, NY, USA, 1997. ACM.

[15] S .Brin, R. Motwani, J.D. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. In *SIGMOD '97: Proceedings of the 1997 ACM SIGMOD international conference on Management of data*, pages 255–264, New York, NY, USA, 1997. ACM.

[16] W. Buntine. Graphical models for discovering knowledge. *Advances in knowledge discovery and data mining*, pages 59–82, 1996.

[17] S. Chiusano and P. Garza. Selection of high quality rules in associative classification. In C. Zhang Y. Zhao and L. Cao, editors, *Post-Mining of Association RUles: Techniques for Effective Knowledge Extraction*. Information Science Reference, Hershey, NY, USA, 2009.

[18] P. Clark and R. Boswell. Rule induction with cn2: Some recent improvements. In *EWSL-91: Proceedings of the European working session on learning on Machine learning*, pages 151–163, New York, NY, USA, 1991. Springer-Verlag New York, Inc.

[19] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46, 1960.

[20] G. Cong, A.K.H. Tung, X. Xu, F. Pan, and J. Yang. 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.

[21] U.M. Fayyad and K.B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the Thirteenth International Joint Conference on AI*, pages 1022–1027, 1993.

[22] A. A. Freitas. On rule interestingness measures. *Knowledge-Based Systems*, 12:309–315, 1999.

[23] J. G. Ganascia. Deriving the learning bias from rule properties. *Machine intelligence 12: towards an automated logic of human thought*, pages 151–167, 1991.

[24] M. Gavrilov, D. Anguelov, P. Indyk, and R. Motwani. Mining the stock market: Which measure is best? In *proceedings of the 6 th ACM Int'l Conference on Knowledge Discovery and Data Mining*, pages 487–496, 2000.

[25] L. Geng and H. J. Hamilton. Interestingness measures for data mining: A survey. *ACM Comput. Surv.*, 38(3):9, 2006.

[26] V. Giakoumakis and B. Monjardet. Coefficients d'accord entre deux prordres totaux. *Statistique et analyse des donnes*, 12(1 et 2):46–99, 1987.

[27] L.A. Goodman and W.H. Kruskal. Measures of association for cross-classification. *Journal of the American Statistical Association*, 49:732–764, 1968.

[28] B .Gray and M. E. Orlowska. Ccaiia: Clustering categorial attributed into interseting accociation rules. In *PAKDD '98: Proceedings of the Second Pacific-Asia Conference on Research and Development in Knowledge Discovery and Data Mining*, pages 132–143, London, UK, 1998. Springer-Verlag.

[29] M. Hahsler and K. Hornik. New probabilistic interest measures for association rules. *Intell. Data Anal.*, 11(5):437–455, 2007.

[30] R. W. Hamming. Error-detecting and error-correcting codes. *Bell Systems Technical Journal*, 29(2):147–160, 1950.

[31] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 1–12, New York, NY, USA, 2000. ACM.

[32] R.J. Hilderman, H.J. Hamilton, and B. Barber. Ranking the interestingness of summaries from data mining systems. In *In Proceedings of the 12th Annual Florida Artificial Intelligence Research Symposium (FLAIRS'99*, pages 100–106, 1999.

[33] L. Kaufman and P.J. Rousseeuw. *Finding Groups in Data*. John Wiley & Sons, New York, 1990.

[34] W. Klösgen. Explora: a multipattern and multistrategy discovery assistant. *Advances in knowledge discovery and data mining*, pages 249–271, 1996.

[35] Y. 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.

[36] I. Kononenko. On biases in estimating multi-valued attributes. In *in Proc. 14th Int. Joint Conf Artificial Intelligence*, pages 1034–1040. Morgan Kaufmann, 1995.

[37] S. Lallich, O. Teytaud, and E. Prudhomme. Association rule interestingness: Measure and statistical validation. In *Quality Measures in Data Mining*, pages 251–275. Springer, 2007.

[38] Y. Lan, D. Janssens, G. Chen, and G. Wets. 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.

[39] P. Lenca, P. Meyer, B. Vaillant, and S. Lallich. A multicriteria decision aid for interestingness measure selection. Technical Report LUSSI-TR-2004-01-EN, LUSSI Department, GET/ENST, Bretagne, France, 2004.

[40] P. Lenca, B. Vaillant, P. Meyer, and S. Lallich. Association rule interestingness measures: Experimental and theoretical studies. In *Quality Measures in Data Mining*, pages 51–76. Springer, 2007.

[41] I.C. Lerman, R. Gras, and H. Rostam. laboration et valuation d'un indice d'implication pour des donnes binaires. *Mathmatiques et sciences humaines*, 75:5–47, 1981.

[42] W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on multiple class-association rules. In *IEEE International Conference on Data Mining (ICDM'01)*, San Jose, California, November 29-December 2 2001.

[43] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In *KDD*, pages 80–86, 1998.

[44] J. Loevinger. A systematic approach to the construction and evaluation of tests of ability. *American Psychological Assn*, 61, 1947.

[45] J. A. Major and J. J. Mangano. Selecting among rules induced from a hurricane database. *Journal of Intelligent Information systems*, 4:39–52, 1995.

[46] K. McGarry. A survey of interestingness measures for knowledge discovery. *Knowl. Eng. Rev.*, 20(1):39–61, 2005.

[47] K. McGarry and J. Malone. Analysis of rules discovered by the data mining process. In *Applications and Science in Soft Computing Series: Advances in Soft Computing.*, pages 219–224. Springer, 2004.

[48] A. Merceron and K. Yacef. Interestingness measures for association rules in educational data. In *EDM '08: Proceedings of Educational Data Mining Conference*, pages 57–66, 2008.

[49] F. Mosteller. Association and estimation in contingency tables. *Journal of the American Statistical Association*, 63:1–28, 1968.

[50] J.L. Myers and A. Well. *Research Design And Statistical Analysis*. Lawrence Erlbaum, 2nd edition, 2003.

[51] M. Ohsaki, S. Kitaguchi, H. Yokoi, and T. Yamaguchi. Investigation of rule interestingness in medical data mining. In *Active Mining*, pages 174–189, 2003.

[52] G. Piatetsky-Shapiro. Discovery, analysis, and presentation of strong rules. In G. Piatetsky-Shapiro and W.J. Frawley, editors, *Knowledge Discovery in Databases*. AAAI/MIT Press, Cambridge, MA, 1991.

[53] R.J. Quinlan. *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*. Morgan Kaufmann, 1993.

[54] C.J. Van Rijsbergen. *Information Retrieval*. Butterworth, London, 1979.

[55] W. Romão, A. Freitas, and I. Gimenes. Discovering interesting knowledge from a science and technology database with a genetic algorithm. *Appl. Soft Comput.*, 4(2):121–137, 2004.

[56] S. Sahar and Y. Mansour. An empirical evaluation of objective interestingness criteria. In *SPIE Conference on Data Mining and Knowledge*, pages 63–74, 1999.

[57] M. Sebag and M. Schoenauer. Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In *Proc. of the European Knowledge Acquisition Workshop, EKAW'88*, pages 28–1 – 28–20, 1988.

[58] E.H. Shortliffe and B.G. Buchanan. A model of inexact reasoning in medicine. pages 259–275, 1990.

[59] P. Smyth and R.M. Goodman. Rule induction using information theory. In W. Frawley, editor, *Knowledge Discovery in Databases*, pages 159–176. MIT Press, Cambridge, MA, 1991.

[60] P. Tan and V. Kumar. Interestingness measures for association patterns: A perspective. Technical Report 00-036, Department of Computer Sciences, University of Minnesota, 2000.

[61] P. Tan, V. Kumar, and J. Srivastava. 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.

[62] P .Tan, V. Kumar, and J. Srivastava. Selecting the right objective measure for association analysis. *Inf. Syst.*, 29(4):293–313, 2004.

[63] B. Vaillant. Evaluation de connaissances : le problme du choix d'une mesure de qualit en extraction de connaissances partir des donnes. Master's thesis, The cole Nationale Suprieure des Tlcommunications de Bretagne, 2002.

[64] A. Veloso, W. Meira Jr., and M.J. Zaki. Lazy associative classification. In *ICDM '06: Proceedings of the Sixth International Conference on Data Mining*, pages 645–654, Washington, DC, USA, 2006. IEEE Computer Society.

[65] F. Verhein and S. Chawla. 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.

[66] G.I. Webb. Discovering significant rules. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 434–443, New York, NY, USA, 2006. ACM.

[67] Y. Y. Yao and N. Zhong. An analysis of quantitative measures associated with rules. In *PAKDD '99: Proceedings of the Third Pacific-Asia Conference on Methodologies for Knowledge Discovery and Data Mining*, pages 479–488, London, UK, 1999. Springer-Verlag.

[68] K.Y. Yeung and W.L. Ruzzo. Principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763–774, 2001.

[69] G.U. Yule. On the association of attributes in statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 194:257–319, 1900.

[70] G.U. Yule. On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, 75(6):579–652, 1912.

[71] M. Zaki, S. Parthasarathy, M. Ogihara, and W. Li. 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.

[72] T. Zhang. Association rules. In *PADKK '00: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications*, pages 245–256, London, UK, 2000. Springer-Verlag.

[73] Y. Zhao and G. Karypis. Criterion functions for document clustering: Experiments and analysis. Technical report, Department of Computer Science, University of Minnesota, 2002.

Appendix A

Measures Rankings

A.1 Measures Rankings With Global Support

Table A.1: Measure rankings based on percentage of rule reduction without jeopardizing the f-measure while using measure-based pruning with global support in generation phase and the highest ranked rule in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
IWD	6	4	1	1	1	3	2	2	1	1	9	1	24	3	1	1	29	33	1	5	13
Kappa	9	1	8	1	1	1	1	3	1	1	14	1	9	1	1	1	30	25	2	4	13
GK	5	2	6	1	1	6	1	13	1	1	4	1	3	8	1	1	15	11	3	1	12
Corr	8	1	7	1	1	6	2	10	1	1	15	1	12	5	1	1	30	32	3	2	11
Klos	1	2	2	1	1	9	1	14	1	1	17	2	29	10	2	1	27	26	9	12	11
2WaySup	2	1	5	1	1	5	1	18	1	1	13	1	29	6	1	1	30	34	13	8	10
CF	10	2	42	1	1	45	29	9	1	1	3	1	1	43	1	1	30	39	6	6	10
Gini	14	1	12	1	1	2	3	43	2	1	10	1	28	2	12	1	26	18	14	9	10
Spec	11	5	1	1	1	26	6	4	1	1	2	1	14	28	1	1	30	39	17	3	10
RelRisk	7	3	10	1	1	18	2	11	1	1	16	1	35	20	3	1	8	31	10	18	9
CCR	7	3	10	1	1	18	2	11	1	1	11	1	29	20	4	1	8	31	10	17	8
CnfrmC	13	8	17	1	1	24	3	16	1	10	7	1	41	26	19	1	12	38	2	1	8
LC	15	9	15	1	1	14	27	1	2	1	6	1	10	30	5	1	17	14	4	1	8
PS	21	12	16	1	1	7	1	41	2	3	20	1	5	9	1	1	30	19	15	13	8
CollStr	12	5	25	1	1	8	18	6	1	1	44	1	42	7	1	3	16	35	7	14	7
Cos	19	7	21	1	1	20	11	8	1	1	5	1	23	23	6	1	14	20	10	1	7
FM	16	6	19	1	1	21	9	5	1	1	4	1	20	24	7	1	24	21	8	1	7
Jacc	18	10	22	1	1	23	12	7	2	1	8	1	26	26	11	1	30	24	5	1	7
Acc	13	2	18	1	1	4	6	25	3	12	1	1	39	4	18	1	30	39	5	11	7

Continued on next page

Table A.1 – continued from previous page

	Datasets																				
Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
ImpInd	20	11	14	1	1	17	5	22	1	1	23	1	21	19	9	2	28	22	18	15	6
CnfrmD	38	14	38	1	1	43	26	47	3	1	18	4	32	41	23	1	30	23	12	16	5
addedvalue	3	17	3	6	8	15	1	32	11	8	37	7	31	16	10	19	9	16	29	31	3
Chi2	17	20	39	3	5	19	4	17	6	2	24	1	42	21	13	7	30	36	19	19	3
HConf	40	13	42	2	2	45	29	49	4	1	12	13	18	11	26	5	30	39	11	7	3
HLift	26	33	42	5	4	10	3	37	8	5	19	9	2	13	1	4	21	39	22	29	3
Lap	36	18	36	2	3	39	24	33	5	6	22	1	30	38	8	6	4	4	16	20	3
1WaySup	4	15	4	7	6	11	1	34	9	8	40	7	33	12	10	17	10	17	32	33	1
CCC	28	24	31	12	12	36	23	39	16	16	30	11	10	32	17	12	2	15	24	25	1
Conv	32	28	29	12	14	33	17	31	18	17	33	11	37	31	17	14	7	2	27	26	1
DChi2	25	34	30	9	19	41	25	44	22	11	41	3	40	39	24	20	30	29	33	23	1
InfoGain	35	16	11	16	7	15	2	42	10	9	36	7	42	16	10	18	30	27	34	35	1
IntImp	24	25	41	4	11	27	7	12	7	4	26	1	41	27	15	8	19	28	20	10	1
KM	29	23	13	12	12	22	14	19	16	17	27	11	11	18	17	12	3	10	24	25	1
Lift	37	32	9	17	18	16	2	46	12	13	43	8	6	17	22	21	13	30	35	36	1
Loe	31	27	28	12	13	32	16	20	17	17	31	11	16	31	17	13	7	3	26	25	1
SS	33	29	33	13	15	35	21	35	19	20	35	11	13	36	21	15	11	1	28	27	1
Gan	29	24	32	12	12	37	22	24	16	17	34	11	15	33	17	12	1	9	24	25	1
2WaySupVar	29	23	17	12	12	12	14	38	16	17	29	11	4	14	17	12	30	39	25	28	0
ConfC	27	22	35	11	12	42	19	45	15	15	27	10	17	40	17	11	20	7	24	24	0
CCD	29	23	32	12	12	35	22	24	16	17	34	11	15	33	17	12	7	9	24	25	0
Ex&Cex	29	23	33	12	12	34	20	30	16	17	27	11	13	35	17	12	11	5	24	25	0
JM	29	23	28	12	12	13	14	36	16	17	27	11	7	15	17	12	30	39	24	25	0
Lev	22	21	34	10	9	38	10	40	13	7	25	6	27	37	16	10	6	6	23	21	0
OddMul	33	31	27	15	17	25	16	21	20	18	38	11	22	22	20	16	22	13	30	30	0
OddR	34	30	24	14	16	28	15	26	21	19	39	11	38	25	20	16	23	11	31	32	0
YulQ	29	23	23	12	12	30	15	23	16	17	28	11	25	25	17	12	18	8	25	28	0
YulY	29	23	26	12	12	31	14	29	16	17	32	11	34	25	17	12	25	12	25	28	0
Zhang	23	19	20	8	10	29	8	28	14	14	21	5	8	29	14	9	5	5	21	22	0
LocSup	40	36	42	19	21	45	29	49	24	22	44	13	42	43	26	23	30	39	37	37	0
GlbSup	40	36	42	19	21	45	13	49	24	22	44	13	42	43	26	23	30	39	37	37	0
Conf	30	26	37	12	12	40	22	27	16	17	34	11	30	34	17	12	7	9	24	25	0
MutInfo	29	23	42	12	12	45	29	15	16	17	28	11	19	43	17	12	30	39	25	28	0
CCS	39	35	40	18	20	44	28	48	23	21	42	12	36	42	25	22	30	37	36	34	0

Table A.2: Measure rankings based on percentage of rule reduction without jeopardizing the f-measure while using measure-based pruning with global support in generation phase and the rules' average of measures in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Klos	1	2	2	1	1	8	1	44	1	1	17	2	32	9	1	1	34	1	14	30	12
Gini	8	1	8	1	1	2	3	38	2	1	10	1	31	2	11	1	33	21	13	29	10
IWD	32	3	1	1	1	3	2	8	1	24	9	1	27	3	26	1	13	7	10	12	10
GK	4	2	3	1	1	6	1	7	5	1	4	1	3	7	26	1	26	17	1	16	10
2WaySup	32	1	32	1	1	5	1	46	1	24	13	1	32	6	2	1	36	31	2	10	9
CF	32	4	32	1	1	44	22	3	2	24	3	1	1	46	26	1	36	36	11	1	9
Kappa	32	1	32	1	1	1	1	12	2	24	14	1	14	1	26	1	17	16	4	9	9
Corr	6	1	32	1	1	6	1	13	1	24	15	1	17	5	5	1	16	8	7	14	7
FM	13	6	12	1	1	10	17	18	3	10	4	1	23	13	17	1	32	24	3	2	7
LC	9	9	10	1	1	13	7	1	2	4	6	1	15	17	6	1	8	2	5	17	7
Spec	7	5	1	1	1	43	8	2	8	24	2	1	19	45	26	2	36	36	10	4	7
Acc	7	2	32	1	1	4	22	19	3	7	1	1	39	4	26	1	36	36	12	7	7
Jacc	10	10	14	1	1	21	22	20	2	5	8	1	29	24	18	1	36	26	1	5	6
PS	32	20	11	1	1	7	1	47	2	9	20	1	5	8	13	1	21	22	8	31	6
CCR	5	4	29	1	1	27	2	4	11	22	11	12	32	32	4	1	20	30	9	3	5
CnfrmC	7	8	31	1	1	23	3	10	1	24	7	7	41	24	20	3	23	35	19	33	5
RelRisk	5	4	29	1	1	27	2	4	11	22	16	12	36	32	3	1	20	30	9	8	5
ImpInd	32	13	32	1	1	18	5	40	1	24	22	1	24	20	9	2	35	13	15	18	5
Chi2	12	18	22	3	3	19	4	11	9	2	24	1	42	21	12	7	36	33	17	11	4
CollStr	30	5	29	1	1	33	1	47	1	19	23	5	42	39	26	4	27	32	16	6	4
CnfrmD	28	12	21	1	1	16	19	43	3	8	18	4	33	19	23	1	8	25	13	33	4
Cos	11	7	13	1	1	20	21	6	4	6	5	1	26	23	7	1	25	23	6	13	4
addedvalue	2	16	4	12	24	14	1	36	23	24	37	12	8	16	10	23	1	6	31	32	3
HLift	17	35	32	10	16	9	1	27	17	3	19	8	2	11	26	8	29	36	27	28	3
1WaySup	3	14	6	11	23	15	1	24	24	24	29	12	34	12	10	24	22	20	30	32	2
HConf	32	11	32	2	2	44	22	47	6	24	12	12	21	10	26	5	36	36	39	33	2
Conv	22	30	19	11	20	31	12	42	20	22	33	11	9	33	19	17	2	4	29	32	1
DChi2	15	34	20	9	22	37	18	39	16	6	41	3	40	41	24	18	36	28	22	19	1
InfoGain	27	15	7	11	25	14	2	37	25	24	36	12	42	16	10	25	15	5	31	30	1
IntImp	31	24	25	4	4	24	9	9	10	11	26	2	41	27	15	9	28	27	18	15	1
KM	20	23	9	6	28	22	12	33	15	24	27	11	16	22	19	28	9	3	33	32	1
Lift	32	33	5	13	26	17	2	41	27	24	27	12	6	18	26	26	24	29	38	32	1
Loe	21	29	18	11	13	30	12	35	18	24	31	11	9	31	19	17	4	3	24	31	1
YulY	20	22	32	6	17	25	12	21	15	24	32	11	35	30	19	19	3	4	32	28	1
2WaySupVar	20	22	32	6	5	11	12	28	15	12	28	11	4	14	19	28	18	15	21	24	0

Continued on next page

Table A.2 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
ConfC	18	26	30	5	11	39	12	15	13	17	38	10	20	42	19	17	9	12	34	22	0
CCC	19	28	27	6	12	36	12	16	15	14	30	11	15	35	19	13	9	6	25	25	0
CCD	20	27	24	6	7	34	12	25	15	13	27	11	10	34	19	14	10	14	27	23	0
Ex&Cex	20	25	26	6	8	38	12	29	15	16	27	11	18	38	19	15	5	10	27	22	0
JM	20	22	32	6	6	12	12	5	15	24	27	11	12	15	19	11	14	18	23	29	0
Lap	26	17	24	8	9	42	16	23	7	19	21	12	7	37	8	6	6	9	21	26	0
Lev	14	19	28	6	10	40	10	14	14	18	25	9	30	43	16	16	9	11	26	25	0
OddMul	24	36	15	11	18	32	14	31	19	22	39	11	25	25	22	21	30	19	36	33	0
OddR	25	32	16	11	19	26	13	22	21	21	40	12	38	26	26	22	31	17	37	28	0
SS	24	31	24	7	8	38	15	25	22	13	35	12	18	34	21	12	5	5	33	22	0
YulQ	20	22	32	6	14	28	12	30	15	24	28	11	28	29	19	22	12	13	26	28	0
Zhang	16	21	17	15	21	29	6	34	12	24	27	6	13	28	14	10	19	10	20	20	0
LocSup	32	38	32	15	28	44	22	47	28	24	43	12	42	46	26	28	36	36	39	33	0
GlbSup	32	38	32	15	28	44	11	47	28	24	43	12	42	46	26	28	36	36	39	33	0
Conf	23	25	24	6	15	35	12	32	15	23	34	11	11	40	19	20	11	14	28	27	0
MutInfo	20	22	32	6	5	44	12	17	15	24	28	11	22	46	19	28	7	36	39	21	0
Gan	20	27	24	6	7	34	12	26	15	15	27	11	10	36	19	14	4	14	27	23	0
CCS	29	37	23	14	27	41	20	45	26	20	42	12	37	44	25	27	36	34	35	27	0

Table A.3: Measure rankings based on percentage of rule reduction without changing the maximum possible accuracy while using measure-based pruning with global support in generation phase and the highest ranked rule in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
FM	7	1	8	2	4	1	5	1	2	2	2	2	12	1	3	1	6	4	2	8	12
Cos	10	3	23	11	30	2	28	5	3	18	4	2	12	2	2	3	6	3	7	9	8
Jacc	5	2	27	11	30	4	32	2	3	3	3	1	12	4	4	2	6	5	4	5	7
CollStr	32	16	5	11	3	16	2	17	36	1	40	4	12	19	1	39	5	17	3	39	5
Acc	2	7	1	1	2	16	1	17	9	10	19	12	12	7	9	18	6	27	19	11	5
Spec	2	21	1	11	1	15	26	17	5	11	15	12	12	15	43	17	6	27	17	1	4
CF	32	21	27	11	30	16	32	17	1	39	1	26	12	19	43	39	6	27	6	3	3
Chi2	4	5	20	3	7	3	6	8	11	7	7	8	12	3	6	13	6	22	15	17	3
PS	32	21	27	11	30	6	32	3	7	14	6	17	12	6	7	7	6	2	1	38	3
CCR	2	6	1	8	19	15	23	17	16	10	34	22	12	18	23	27	6	27	21	12	2
Kappa	3	21	7	11	10	16	3	6	8	8	5	7	12	19	5	5	6	6	8	4	2

Continued on next page

Table A.3 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
LC	27	4	27	11	30	8	32	17	13	16	8	1	8	10	41	2	6	27	10	4	2
RelRisk	2	6	1	8	19	15	23	17	16	10	34	22	12	18	23	27	6	27	20	14	2
GK	1	21	27	11	30	16	32	17	10	4	16	26	8	5	43	16	6	27	13	2	2
CnfrmC	6	21	2	11	26	14	24	17	36	10	36	14	4	14	10	19	6	26	18	13	1
CnfrmD	32	11	27	11	30	16	32	17	14	31	12	3	12	19	31	6	6	27	9	39	1
CCC	29	21	25	11	30	15	32	17	36	33	35	18	2	15	36	38	6	19	30	25	1
DChi2	12	12	21	4	9	7	19	17	23	5	23	13	3	8	37	25	6	8	38	23	1
HConf	32	21	27	11	30	16	32	17	4	39	40	26	12	19	43	4	6	1	42	7	1
IWD	9	21	3	11	30	16	32	11	12	6	11	6	12	19	14	9	6	10	11	10	1
KM	32	21	27	11	30	16	32	13	36	39	32	26	5	19	15	39	3	16	37	25	1
Loe	22	21	15	11	14	16	20	17	29	28	28	18	6	19	27	28	2	14	27	25	1
Zhang	13	21	14	11	8	16	8	10	27	19	29	11	11	19	19	22	1	14	22	20	1
CCS	31	21	24	11	25	11	27	16	36	32	39	25	1	13	42	39	6	25	41	35	1
1WaySup	25	21	17	11	16	16	14	15	30	27	33	22	12	19	16	26	6	15	36	34	0
2waySup	11	21	4	11	30	16	32	11	6	17	9	6	12	19	18	8	6	11	5	16	0
2waySupVar	15	9	27	11	30	16	32	17	20	20	18	21	12	19	11	21	6	27	24	28	0
AddVal	24	21	16	11	20	16	18	17	28	29	28	23	12	19	17	31	6	20	35	32	0
ConfC	28	21	26	11	29	15	31	16	36	37	37	19	12	17	39	39	6	27	32	24	0
CCD	31	17	27	11	23	16	32	13	34	35	32	18	7	19	34	34	6	16	28	25	0
Conv	22	20	19	7	21	16	25	17	32	28	38	19	12	11	30	29	6	14	28	26	0
Corr	8	21	9	11	22	16	7	9	15	9	14	8	12	19	8	11	6	9	8	6	0
Ex&Cex	30	18	27	10	26	13	32	12	33	36	29	18	7	16	33	33	4	13	28	25	0
Gini	32	8	27	11	30	5	32	17	17	12	10	15	12	4	43	10	6	27	9	15	0
HLift	19	16	27	6	17	9	13	14	24	22	27	19	10	9	25	24	6	27	23	30	0
InfoGain	21	21	18	11	21	16	17	17	29	28	30	22	12	19	21	28	6	18	39	36	0
IntImp	32	10	22	5	6	16	9	17	36	4	17	5	12	19	20	15	6	7	16	19	0
JM	18	21	6	11	5	16	4	4	21	21	22	20	12	19	12	20	6	27	26	25	0
Klos	20	21	12	11	15	16	21	17	22	13	21	16	12	19	38	14	6	27	11	22	0
Lap	32	19	27	11	30	16	32	17	36	34	31	10	12	19	40	32	6	27	14	29	0
Lev	31	21	26	11	27	12	30	17	36	38	35	21	12	14	32	39	6	26	25	21	0
Lift	26	20	19	11	18	16	22	14	31	30	40	24	9	19	22	37	6	27	40	37	0
OddMul	23	20	19	7	24	10	15	16	28	24	32	20	12	12	28	30	6	27	33	31	0
OddR	16	15	10	9	11	15	16	17	26	23	26	19	12	15	29	23	6	27	34	33	0
SS	31	19	27	11	28	16	29	13	35	35	32	18	8	19	38	36	6	23	29	27	0
YulQ	17	13	10	11	11	16	10	17	26	25	24	20	12	19	24	23	6	21	31	28	0
YulY	17	14	11	11	12	16	11	17	25	26	25	20	12	19	26	23	6	24	30	28	0
LocSup	32	21	27	11	30	16	32	17	36	39	40	26	12	19	43	39	6	27	42	39	0
GlbSup	32	21	27	11	30	16	32	17	36	39	40	26	12	19	43	39	6	27	42	39	0
Conf	32	20	27	11	30	16	32	17	36	36	32	18	12	19	35	35	6	27	28	25	0
MutInfo	32	21	27	11	30	16	32	17	19	20	20	26	12	19	43	39	6	27	27	28	0
Gan	31	17	27	11	23	16	32	13	34	35	32	18	7	19	34	34	6	16	28	25	0
ImpInd	14	21	13	11	13	16	12	7	18	15	13	9	12	19	13	12	6	12	12	18	0

Table A.4: Measure rankings based on percentage of rule reduction while using measure-based pruning with global support in generation phase and the rules' average of measures in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
FM	7	1	8	2	4	1	5	1	2	2	2	2	12	1	3	1	6	4	2	8	12
Cos	10	3	23	11	30	2	28	5	3	18	4	2	12	2	2	3	6	3	7	9	8
Jacc	5	2	27	11	30	4	32	2	3	3	3	1	12	4	4	2	6	5	4	5	7
CollStr	32	16	5	11	3	16	2	17	36	1	40	4	12	19	1	39	5	17	3	39	5
Acc	2	7	1	1	2	16	1	17	9	10	19	12	12	7	9	18	6	27	19	11	5
Spec	2	21	1	11	1	15	26	17	5	11	15	12	12	15	43	17	6	27	17	1	4
CF	32	21	27	11	30	16	32	17	1	39	1	26	12	19	43	39	6	27	6	3	3
Chi2	4	5	20	3	7	3	6	8	11	7	7	8	12	3	6	13	6	22	15	17	3
PS	32	21	27	11	30	6	32	3	7	14	6	17	12	6	7	7	6	2	1	38	3
CCR	2	6	1	8	19	15	23	17	16	10	34	22	12	18	23	27	6	27	21	12	2
Kappa	3	21	7	11	10	16	3	6	8	8	5	7	12	19	5	5	6	6	8	4	2
LC	27	4	27	11	30	8	32	17	13	16	8	1	8	10	41	2	6	27	10	4	2
RelRisk	2	6	1	8	19	15	23	17	16	10	34	22	12	18	23	27	6	27	20	14	2
GK	1	21	27	11	30	16	32	17	10	4	16	26	8	5	43	16	6	27	13	2	2
CnfrmC	6	21	2	11	26	14	24	17	36	10	36	14	4	14	10	19	6	26	18	13	1
CnfrmD	32	11	27	11	30	16	32	17	14	31	12	3	12	19	31	6	6	27	9	39	1
CCC	29	21	25	11	30	15	32	17	36	33	35	18	2	15	36	38	6	19	30	25	1
DChi2	12	12	21	4	9	7	19	17	23	5	23	13	3	8	37	25	6	8	38	23	1
HConf	32	21	27	11	30	16	32	17	4	39	40	26	12	19	43	4	6	1	42	7	1
IWD	9	21	3	11	30	16	32	11	12	6	11	6	12	19	14	9	6	10	11	10	1
KM	32	21	27	11	30	16	32	13	36	39	32	26	5	19	15	39	3	16	37	25	1
Loe	22	21	15	11	14	16	20	17	29	28	28	18	6	19	27	28	2	14	27	25	1
Zhang	13	21	14	11	8	16	8	10	27	19	29	11	11	19	19	22	1	14	22	20	1
CCS	31	21	24	11	25	11	27	16	36	32	39	25	1	13	42	39	6	25	41	35	1
1WaySup	25	21	17	11	16	16	14	15	30	27	33	22	12	19	16	26	6	15	36	34	0
2waySup	11	21	4	11	30	16	32	11	6	17	9	6	12	19	18	8	6	11	5	16	0
2waySupVar	15	9	27	11	30	16	32	17	20	20	18	21	12	19	11	21	6	27	24	28	0
AddVal	24	21	16	11	20	16	18	17	28	29	28	23	12	19	17	31	6	20	35	32	0
ConfC	28	21	26	11	29	15	31	16	36	37	37	19	12	17	39	39	6	27	32	24	0
CCD	31	17	27	11	23	16	32	13	34	35	32	18	7	19	34	34	6	16	28	25	0
Conv	22	20	19	7	21	16	25	17	32	28	38	19	12	11	30	29	6	14	28	26	0
Corr	8	21	9	11	22	16	7	9	15	9	14	8	12	19	8	11	6	9	8	6	0
Ex&Cex	30	18	27	10	26	13	32	12	33	36	29	18	7	16	33	33	4	13	28	25	0
Gini	32	8	27	11	30	5	32	17	17	12	10	15	12	4	43	10	6	27	9	15	0

Continued on next page

Table A.4 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
HLift	19	16	27	6	17	9	13	14	24	22	27	19	10	9	25	24	6	27	23	30	0
InfoGain	21	21	18	11	21	16	17	17	29	28	30	22	12	19	21	28	6	18	39	36	0
IntImp	32	10	22	5	6	16	9	17	36	4	17	5	12	19	20	15	6	7	16	19	0
JM	18	21	6	11	5	16	4	4	21	21	22	20	12	19	12	20	6	27	26	25	0
Klos	20	21	12	11	15	16	21	17	22	13	21	16	12	19	38	14	6	27	11	22	0
Lap	32	19	27	11	30	16	32	17	36	34	31	10	12	19	40	32	6	27	14	29	0
Lev	31	21	26	11	27	12	30	17	36	38	35	21	12	14	32	39	6	26	25	21	0
Lift	26	20	19	11	18	16	22	14	31	30	40	24	9	19	22	37	6	27	40	37	0
OddMul	23	20	19	7	24	10	15	16	28	24	32	20	12	12	28	30	6	27	33	31	0
OddR	16	15	10	9	11	15	16	17	26	23	26	19	12	15	29	23	6	27	34	33	0
SS	31	19	27	11	28	16	29	13	35	35	32	18	8	19	38	36	6	23	29	27	0
YulQ	17	13	10	11	11	16	10	17	26	25	24	20	12	19	24	23	6	21	31	28	0
YulY	17	14	11	11	12	16	11	17	25	26	25	20	12	19	26	23	6	24	30	28	0
LocSup	32	21	27	11	30	16	32	17	36	39	40	26	12	19	43	39	6	27	42	39	0
GlbSup	32	21	27	11	30	16	32	17	36	39	40	26	12	19	43	39	6	27	42	39	0
Conf	32	20	27	11	30	16	32	17	36	36	32	18	12	19	35	35	6	27	28	25	0
MutInfo	32	21	27	11	30	16	32	17	19	20	20	26	12	19	43	39	6	27	27	28	0
Gan	31	17	27	11	23	16	32	13	34	35	32	18	7	19	34	34	6	16	28	25	0
ImpInd	14	21	13	11	13	16	12	7	18	15	13	9	12	19	13	12	6	12	12	18	0

Table A.5: Measure rankings based on percentage of f-measure improvement while using measure-based pruning with global support in generation phase and the highest ranked rule in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Lev	31	1	18	12	6	30	33	21	13	3	1	1	4	21	3	3	2	1	1	8	9
Kappa	12	6	22	22	15	21	13	5	3	4	1	11	2	16	1	2	2	2	9	2	8
Zhang	3	11	1	18	4	2	18	2	16	27	1	14	4	3	8	7	4	1	3	14	8
Acc	10	3	15	24	19	23	1	3	9	11	1	15	4	10	1	2	2	2	17	6	8
GK	16	2	19	22	19	22	5	3	9	5	1	16	2	13	20	2	2	2	7	2	8
1WaySup	5	6	2	26	8	3	4	13	27	17	1	2	4	2	6	14	2	2	24	24	7
CF	26	12	35	22	19	36	40	8	6	1	2	3	1	31	23	1	2	9	11	3	7
Cos	18	8	29	7	11	33	32	6	2	16	1	8	2	29	2	2	2	2	7	4	7
FM	21	8	28	13	12	32	24	1	4	16	1	2	2	29	6	2	2	2	12	4	7
LC	20	8	29	15	13	33	36	3	10	22	1	13	2	29	7	2	2	2	8	1	7
Spec	9	2	11	24	19	26	2	3	29	14	1	19	1	23	22	5	2	2	20	6	7

Continued on next page

Table A.5 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
2WaySup	28	10	26	1	16	9	21	32	10	13	1	6	2	7	18	3	2	2	11	16	6
AddVal	4	9	4	26	8	1	6	17	31	15	1	2	4	1	13	22	2	2	22	22	6
CnfrmC	15	3	17	13	13	5	3	11	12	11	1	4	4	11	14	3	2	2	16	6	6
Corr	14	9	24	22	19	10	12	7	4	6	1	11	2	5	1	2	2	2	14	5	6
IWD	17	10	12	1	17	16	14	5	7	7	1	11	2	22	17	2	2	2	16	6	6
Jacc	20	8	29	12	11	32	35	4	5	20	1	13	2	29	7	2	2	2	8	1	6
KM	1	2	5	26	20	7	9	18	30	15	1	2	4	9	4	9	1	2	4	11	6
Chi2	11	4	29	25	16	19	7	3	13	2	1	9	4	23	12	15	2	2	23	12	5
CCR	7	2	10	24	19	11	10	10	28	18	1	21	2	5	5	5	2	2	20	7	5
OddMul	2	11	1	16	4	7	22	12	20	23	1	4	4	6	11	13	2	2	22	21	5
YulQ	23	11	14	21	10	6	21	19	25	28	1	7	4	3	7	1	2	1	25	26	5
2WaySupVar	30	11	31	20	12	14	35	29	30	33	1	7	2	14	7	2	2	8	25	26	4
CollStr	25	7	27	21	7	30	19	25	1	8	4	15	5	24	16	8	3	3	2	23	4
CCC	33	11	21	6	3	31	34	23	21	29	1	7	4	26	7	10	1	1	5	13	4
CCD	35	11	29	1	4	34	34	24	22	29	1	7	4	29	7	10	1	2	5	13	4
Conv	19	11	9	11	3	17	27	27	17	29	1	4	4	12	7	13	1	1	10	10	4
Gini	13	5	23	2	19	20	13	16	8	9	1	9	2	17	12	2	5	5	15	16	4
InfoGain	8	7	3	23	8	4	11	9	25	17	1	5	4	4	14	18	2	2	26	26	4
JM	34	7	28	20	10	15	28	29	30	21	1	4	2	25	7	5	2	2	5	13	4
Klos	18	7	8	12	18	18	8	32	15	10	1	17	2	18	17	6	2	2	18	25	4
Lap	35	11	29	8	1	34	34	32	18	32	1	20	4	29	9	17	2	1	5	20	4
Loe	18	11	7	10	2	13	25	22	13	26	1	4	4	12	7	12	1	2	6	10	4
OddR	22	11	13	20	9	8	20	8	26	24	1	15	4	3	10	23	2	1	26	26	4
RelRisk	7	2	10	24	19	11	10	10	28	18	1	21	4	5	5	5	2	2	20	17	4
SS	35	11	29	5	3	34	34	25	17	30	1	7	4	29	15	19	2	2	13	15	4
YulY	23	11	14	19	9	7	21	15	19	25	1	4	4	8	7	1	2	1	25	26	4
Gan	35	11	29	1	4	34	34	24	22	29	1	7	4	29	7	10	1	2	5	13	4
ImpInd	32	10	22	24	14	16	23	31	11	20	1	6	2	19	6	11	2	2	20	19	4
CCS	27	11	16	3	4	13	31	5	14	29	1	4	4	13	21	16	2	2	22	18	4
ConfC	33	11	20	4	6	27	34	23	23	29	1	4	4	20	7	4	2	1	5	13	3
CnfrmD	36	11	32	2	10	34	37	32	6	30	1	18	3	29	22	7	8	4	19	26	3
DChi2	31	11	25	31	23	28	29	32	33	11	1	12	4	27	22	24	2	2	27	26	3
Ex&Cex	35	11	29	6	5	34	34	20	22	29	1	7	4	29	7	4	2	2	5	13	3
HLift	6	11	36	29	22	12	15	14	32	12	1	10	4	10	19	20	2	2	26	26	3
IntImp	24	7	30	27	21	24	17	30	24	19	1	6	4	28	18	21	2	3	21	19	3
Lift	35	11	6	30	8	25	16	9	26	31	1	5	4	15	14	20	2	2	26	26	3
Conf	35	11	29	14	4	34	34	26	22	29	1	7	4	29	7	4	3	2	5	13	3
MutInfo	37	11	34	28	20	37	38	28	30	33	1	7	2	32	7	1	6	10	25	26	3
HConf	38	13	36	9	13	35	39	34	18	33	3	22	1	30	24	6	9	7	19	9	2
PS	29	10	33	17	14	29	26	33	10	20	1	4	4	28	7	2	7	6	10	26	2
LocSup	39	14	36	32	24	38	41	35	34	34	5	23	6	33	25	25	10	12	28	27	0
GlbSup	39	14	36	32	24	38	30	35	34	34	5	23	6	33	25	25	10	11	28	27	0

Table A.6: Measure rankings based on percentage of f-measure improvement while using measure-based pruning with global support in generation phase and the rules' average of measures in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Chi2	24	2	11	3	2	13	5	16	3	2	2	9	15	5	2	3	26	13	4	5	8
Ex&Cex	4	1	16	11	1	22	15	1	17	14	3	18	3	19	2	3	4	7	6	6	7
FM	27	1	19	9	1	22	38	18	2	13	3	2	10	21	18	3	23	15	1	5	7
SS	5	1	16	18	1	22	15	6	17	14	3	23	3	23	3	3	7	2	6	6	7
1WaySup	20	9	3	20	11	2	1	7	15	20	5	25	12	1	15	4	1	10	6	3	6
CnfrmC	28	2	21	5	1	7	3	9	18	14	2	23	15	16	19	3	19	15	3	5	6
DChi2	20	1	8	17	3	18	36	14	6	3	3	16	15	12	17	3	26	14	6	3	6
Jacc	27	1	19	9	1	22	38	25	4	9	3	1	15	20	20	3	26	15	1	5	6
CCC	10	2	20	13	1	24	20	10	21	14	7	18	2	27	6	3	7	10	3	5	5
CCD	5	1	15	18	1	22	16	3	17	14	5	18	5	23	5	3	10	3	6	6	5
Conv	23	1	14	18	1	20	29	10	7	11	3	18	14	23	10	2	11	1	6	5	5
Cos	27	1	19	4	1	22	38	28	1	10	3	4	10	23	20	3	22	15	6	5	5
KM	1	10	3	24	17	3	11	12	26	23	5	18	3	4	10	12	4	2	6	5	5
Lev	11	2	18	16	1	21	25	10	21	14	10	12	13	23	8	3	7	2	3	5	5
OddMul	5	1	2	18	1	8	17	9	16	14	9	20	8	9	20	3	5	1	6	5	5
OddR	26	1	19	18	1	10	18	15	17	11	2	23	15	14	20	3	10	3	6	5	5
Acc	27	1	19	12	1	22	38	21	8	14	3	21	15	8	20	3	26	15	1	6	5
GK	20	18	1	14	4	25	27	24	8	1	2	22	10	15	23	5	3	9	1	6	5
Gan	6	1	15	15	1	22	16	3	17	14	5	18	5	23	5	3	14	3	6	6	5
CCR	13	1	19	16	1	15	10	8	12	11	3	23	15	19	4	3	20	15	6	4	4
CnfrmD	8	4	12	2	3	23	24	19	19	6	2	14	17	25	11	2	29	12	6	6	4
IntImp	30	3	17	11	1	27	22	12	14	12	5	8	19	24	3	3	19	8	6	6	4
Lap	3	1	19	15	1	22	6	13	19	8	5	23	12	23	6	3	13	7	6	6	4
Loe	18	7	13	21	11	16	26	8	7	22	3	18	8	28	10	1	2	4	6	3	4
RelRisk	13	1	19	16	1	15	10	8	12	11	3	23	15	19	4	3	20	15	6	6	4
Spec	27	2	19	16	1	22	35	27	9	14	4	17	15	23	20	3	26	15	6	2	4
YulY	12	2	25	20	10	11	18	11	12	23	2	18	10	19	10	4	9	2	6	3	4
CCS	16	2	19	13	1	15	37	4	21	14	11	23	13	22	20	3	25	15	3	5	4
2WaySup	37	15	29	1	7	6	28	28	3	27	3	10	15	11	21	7	26	11	6	6	3
AddVal	9	9	7	20	10	1	2	6	15	20	7	25	8	2	15	4	9	12	6	5	3
ConfC	15	1	19	13	1	22	23	11	21	14	9	12	6	23	7	3	6	5	6	5	3
Corr	29	11	27	7	8	9	13	28	10	19	2	10	14	3	1	5	26	4	5	5	3
Gini	32	8	6	2	7	12	13	26	18	5	2	9	16	17	2	5	16	11	6	5	3
HLift	21	1	35	18	1	17	8	20	11	14	6	23	9	8	20	3	17	15	6	6	3

Continued on next page

Table A.6 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
IWD	34	14	23	1	6	20	12	28	9	21	2	15	8	29	22	5	26	9	1	5	3
Kappa	35	17	28	10	4	30	4	28	12	18	3	3	14	31	25	5	26	15	3	5	3
YulQ	14	5	26	21	11	8	18	9	12	23	2	18	4	18	10	4	12	1	6	3	3
Conf	7	7	24	19	6	30	19	2	20	14	5	18	12	30	9	3	9	2	7	7	3
CollStr	22	1	19	17	1	21	38	29	11	14	13	6	21	25	20	5	27	15	6	8	2
Lift	39	7	5	22	13	26	9	11	18	28	3	26	1	13	24	4	15	15	6	5	2
Zhang	2	7	4	25	12	5	14	5	18	25	7	5	6	7	14	5	12	6	2	5	2
2WaySupVar	25	4	32	24	11	28	33	23	24	15	2	18	15	33	10	12	26	17	8	5	1
CF	36	13	34	14	14	32	40	23	23	24	4	24	17	35	26	8	26	19	9	1	1
InfoGain	25	7	7	21	11	4	7	9	7	20	3	25	15	6	15	4	21	6	6	5	1
JM	31	10	31	24	9	14	34	28	13	23	3	18	15	26	10	9	26	15	6	6	1
Klos	19	12	10	4	7	29	21	26	11	16	1	19	15	25	16	6	18	9	6	5	1
LC	17	6	22	8	5	28	39	28	11	7	3	7	9	23	12	5	8	9	6	6	1
PS	33	14	9	5	3	19	31	30	5	4	8	11	18	10	18	5	30	16	6	6	1
MutInfo	27	3	33	23	16	33	32	22	25	17	4	18	11	36	10	11	28	20	10	9	1
ImpInd	38	7	30	6	4	23	32	17	6	26	2	13	7	34	13	4	24	5	6	5	1
HConf	40	16	35	11	15	31	40	31	22	29	12	27	20	32	27	10	31	18	11	10	0
LocSup	41	19	35	26	18	34	41	32	27	30	14	28	22	37	28	13	32	22	12	11	0
GlbSup	41	19	35	26	18	34	30	32	27	30	14	28	22	37	28	13	32	21	12	11	0

Table A.7: Measure rankings based on percentage of f-measure improvement while using different measures in selection phase with global support in generation phase and the highest ranked rule in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
OddMul	3	3	2	9	13	2	11	11	24	20	1	26	5	1	10	12	2	3	7	10	8
CCS	5	3	1	11	14	3	12	3	27	20	2	26	11	2	11	12	1	5	7	10	7
CnfrmC	26	3	16	6	3	25	28	4	9	19	5	23	1	19	29	1	8	13	3	1	6
Conv	6	3	2	12	12	4	14	9	26	20	1	26	3	5	10	12	2	2	7	10	6
Lap	25	9	16	5	1	19	21	16	2	18	4	6	3	17	2	5	7	2	1	13	6
Loe	6	3	2	12	12	4	14	9	26	20	1	26	3	5	10	12	2	2	7	10	6
Zhang	2	2	2	8	7	2	10	10	25	20	4	20	5	2	6	10	5	3	5	7	6
ConfC	8	2	11	12	11	11	19	7	26	20	1	26	3	11	12	12	2	2	7	10	5
CCC	9	2	13	12	11	15	19	6	26	20	1	26	3	12	12	12	2	2	7	10	5
Cos	27	16	24	2	3	36	26	25	30	20	3	14	18	32	24	2	21	20	17	3	5
FM	27	18	24	2	3	38	26	26	29	20	3	16	18	32	23	2	25	23	16	3	5

Continued on next page

Table A.7 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
IntImp	7	13	19	1	2	9	9	32	1	11	5	1	25	7	1	8	16	10	5	17	5
Jacc	27	18	24	2	3	38	26	26	29	20	3	16	18	32	23	2	25	23	18	3	5
Kappa	16	7	14	4	3	23	3	22	10	14	3	12	16	17	13	3	23	22	13	3	5
LC	27	14	24	2	3	37	26	19	28	20	3	10	12	31	25	2	14	16	9	3	5
Lev	4	1	10	7	9	12	15	1	18	9	7	22	3	10	3	15	4	1	8	8	5
GK	14	6	23	3	3	20	13	26	16	7	3	8	18	13	20	3	13	15	18	3	5
ImpInd	13	6	5	3	3	6	8	31	3	11	5	3	19	4	5	2	20	16	6	18	5
CCD	10	4	15	12	11	17	20	6	26	20	1	26	3	13	12	12	2	2	7	10	4
Corr	18	7	17	4	3	14	5	21	5	4	3	4	17	12	7	3	18	18	11	2	4
Ex&Cex	10	4	15	12	11	17	20	6	26	20	2	26	3	13	12	12	2	2	7	10	4
Klos	1	10	4	3	3	7	1	28	4	15	5	5	10	4	4	9	17	14	4	15	4
PS	28	6	24	3	3	21	16	36	7	16	3	12	27	16	18	3	26	18	29	24	4
SS	10	4	15	12	11	17	20	6	26	20	2	26	3	13	12	12	2	2	7	10	4
Conf	10	4	15	12	11	17	20	6	26	20	2	26	3	13	12	12	2	1	7	10	4
Gan	10	4	15	12	11	17	20	6	26	20	1	26	3	13	12	12	2	2	7	10	4
2WaySup	12	10	6	4	3	16	4	33	8	12	3	2	23	14	8	7	24	21	22	19	3
AddVal	21	15	16	17	15	5	22	2	20	19	2	25	4	6	21	24	2	4	15	11	3
Chi2	17	7	22	4	3	13	7	21	5	4	3	4	17	12	7	3	18	18	11	5	3
Gini	11	7	22	4	3	13	7	35	5	4	3	4	21	12	7	3	18	18	19	22	3
Acc	27	8	18	4	3	30	27	12	11	13	3	11	8	25	29	3	15	17	11	4	3
1WaySup	21	19	16	18	16	10	25	5	19	19	1	28	6	9	22	25	2	6	24	12	2
2WaySupVar	35	17	21	3	3	24	18	38	21	5	10	12	28	20	14	16	31	24	28	30	2
CCR	35	26	3	5	3	32	28	20	15	19	13	15	24	26	27	18	12	12	32	29	2
CollStr	29	5	20	13	8	26	24	23	23	3	16	13	22	22	17	11	22	28	2	27	2
DChi2	19	11	9	4	8	1	17	14	12	1	5	7	9	8	16	4	11	11	12	5	2
InfoGain	23	24	16	22	17	29	28	14	22	19	1	29	7	24	28	27	2	8	27	20	2
IWD	20	12	8	4	3	22	2	24	6	17	5	9	15	15	9	14	19	18	10	6	2
Lift	23	24	16	22	17	29	28	14	22	19	1	29	7	24	28	27	3	8	27	20	2
OddR	35	25	7	10	13	8	11	13	24	20	11	26	9	3	10	12	7	3	25	29	2
RelRisk	35	26	3	5	3	32	28	20	15	19	13	15	24	26	27	18	12	12	32	29	2
YulQ	15	15	7	10	13	8	11	13	24	20	12	26	9	3	10	12	7	3	25	29	2
YulY	15	15	7	10	13	8	11	13	24	20	12	26	9	3	10	12	7	3	25	29	2
CnfrmD	35	29	24	16	4	39	26	34	31	20	3	31	13	33	26	13	14	15	20	25	1
HLift	24	23	25	14	10	28	23	15	14	2	4	27	10	23	19	26	10	9	14	16	1
JM	30	21	12	4	3	18	6	37	13	10	9	8	26	18	15	20	30	19	31	23	1
KM	22	24	16	24	21	33	28	8	37	19	6	24	2	27	29	28	9	7	26	9	1
Spec	31	20	16	15	3	33	28	27	17	19	13	21	31	27	29	23	26	25	33	28	1
CF	32	27	24	20	5	27	26	18	33	6	15	19	29	21	18	22	28	30	33	26	0
HConf	34	22	25	19	18	31	26	29	32	20	14	30	14	29	26	19	6	27	21	21	0
LocSup	35	28	24	21	6	34	26	17	35	20	13	18	29	28	18	17	27	29	33	29	0
GlbSup	35	29	24	23	20	40	26	39	36	20	13	32	30	33	26	21	28	27	30	30	0
MutInfo	33	15	24	21	19	35	26	30	34	8	8	17	20	30	18	6	29	26	23	14	0

Table A.8: Measure rankings based on percentage of f-measure improvement while using different measures in selection phase with global support in generation phase and the rules' average of measures in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
ConfC	1	1	1	1	2	2	3	1	7	1	3	5	5	2	1	1	6	6	1	3	15
CCC	7	3	4	2	3	5	10	2	5	3	5	3	5	5	3	1	6	8	1	2	10
Lev	2	5	3	1	2	3	8	3	5	5	1	6	7	1	2	1	6	7	4	4	10
Conv	3	7	5	13	21	4	1	5	16	20	1	18	2	3	7	10	3	3	13	11	7
CCS	4	7	2	13	21	1	2	4	18	20	2	18	3	4	7	10	1	1	13	11	7
Ex&Cex	10	3	7	5	4	7	13	9	2	6	6	2	7	8	6	2	11	12	1	1	6
Loe	6	7	6	13	21	4	1	5	16	20	1	18	2	3	7	10	3	3	13	11	6
CCD	8	4	8	4	1	8	14	8	4	7	7	2	6	7	4	1	7	9	1	2	5
Lap	11	6	9	3	3	9	16	16	2	8	4	1	6	8	5	4	8	10	3	5	5
Conf	8	4	8	4	1	8	14	8	4	7	7	2	6	7	4	1	7	9	1	2	5
Gan	8	4	8	4	1	8	14	8	4	7	7	2	6	7	4	1	7	9	1	2	5
SS	9	8	10	14	21	11	11	6	19	20	1	18	1	9	8	10	3	2	13	11	4
OddMul	5	7	11	12	21	6	2	7	16	20	1	18	4	6	7	10	3	4	13	11	3
Zhang	18	2	17	9	7	13	26	13	6	14	6	8	8	11	29	6	11	16	2	6	2
Cos	27	28	17	23	25	33	1	29	27	9	24	17	25	24	10	24	24	33	31	25	1
HConf	37	19	40	21	28	15	18	25	26	19	12	23	39	23	22	13	2	44	14	24	1
Klos	12	10	22	18	10	17	25	31	3	17	11	20	16	16	21	11	19	22	16	9	1
LC	19	27	32	10	15	29	18	20	20	2	13	9	14	26	9	15	15	25	21	10	1
OddR	38	26	12	12	21	10	2	10	16	20	9	18	6	7	7	10	4	5	22	38	1
YulY	31	24	13	6	5	12	9	14	8	4	16	6	13	10	11	3	6	15	6	8	1
ImpInd	14	9	23	16	9	18	15	32	1	12	14	4	21	15	18	8	20	26	19	17	1
1WaySup	20	12	18	17	11	16	31	12	10	17	5	20	6	14	33	10	5	13	8	12	0
2WaySup	18	16	25	24	22	19	24	36	22	15	19	10	29	17	20	23	25	32	28	30	0
2WaySupVar	42	41	38	34	37	41	7	41	35	24	33	31	36	35	31	34	37	41	37	40	0
AddVal	13	11	19	15	8	16	30	7	9	17	7	16	6	14	32	7	7	14	7	7	0
CF	35	39	37	26	34	40	19	39	37	21	28	30	35	34	25	29	31	42	35	36	0
Chi2	34	34	21	22	13	26	22	26	14	17	13	12	19	20	17	21	22	29	24	16	0
CCR	38	29	15	37	18	25	31	22	15	17	21	22	30	19	33	27	14	20	23	38	0
CollStr	36	30	14	19	26	28	17	37	28	16	29	28	37	21	13	18	29	36	12	35	0
CnfrmC	21	13	15	34	27	22	31	15	23	17	4	22	11	13	33	19	12	21	15	15	0
CnfrmD	38	31	34	7	19	35	18	37	21	19	13	24	15	30	22	9	17	24	27	31	0
Corr	25	14	24	28	14	21	28	27	13	17	17	14	20	17	28	20	21	27	20	21	0
DChi2	40	40	37	20	31	38	20	21	8	17	11	20	38	33	26	14	34	43	18	13	0

Continued on next page

Table A.8 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
FM	28	32	27	30	29	30	6	34	29	13	25	26	28	25	16	31	28	39	36	29	0
Gini	33	34	21	22	13	26	22	35	14	17	13	12	24	20	17	21	22	28	25	18	0
HLift	22	18	40	11	12	20	29	18	4	17	10	21	9	18	27	12	18	11	9	23	0
InfoGain	16	20	20	35	17	25	31	11	12	17	8	22	10	19	33	16	9	19	10	22	0
IntImp	39	37	31	29	32	37	23	43	38	25	27	13	40	31	35	26	32	42	30	34	0
IWD	24	15	25	27	20	20	27	24	17	17	14	19	18	17	24	24	23	30	21	20	0
Jacc	29	33	28	31	35	32	4	28	30	11	23	25	26	26	14	30	27	38	34	26	0
JM	41	42	35	39	40	34	32	42	34	23	31	31	32	28	36	35	36	37	40	40	0
Kappa	26	17	26	32	24	23	31	30	25	17	20	15	22	17	30	25	26	33	29	27	0
KM	43	41	39	40	41	42	32	38	36	24	30	31	12	36	36	36	33	25	39	40	0
Lift	21	20	15	33	16	25	31	11	11	17	7	22	9	19	33	17	7	18	11	22	0
PS	15	23	30	25	23	27	12	40	24	13	22	11	31	22	12	22	30	34	33	32	0
RelRisk	38	29	15	37	18	25	31	22	15	17	21	22	30	19	33	27	14	20	23	38	0
Spec	23	21	15	38	36	25	31	23	31	17	26	22	35	19	33	27	30	35	32	33	0
YulQ	32	25	16	8	6	14	21	19	4	10	18	7	17	12	19	5	10	17	5	14	0
Acc	23	22	15	38	36	25	31	17	31	17	13	22	16	19	33	27	16	31	17	28	0
LocSup	30	36	29	36	39	31	5	33	32	12	28	27	34	27	15	33	31	40	38	37	0
GlbSup	38	35	33	41	38	36	18	44	33	18	28	29	33	29	23	32	31	40	37	39	0
MutInfo	44	38	36	26	33	39	19	45	37	22	32	30	27	32	25	28	35	42	35	40	0
GK	17	22	14	38	30	24	31	23	31	17	15	20	23	19	34	27	13	23	26	19	0

A.2 Measures Rankings With Local Support

Table A.9: Measure rankings based on percentage of rule reduction without jeopardizing the f-measure while using measure-based pruning with local support in generation phase and the highest ranked rule in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
IWD	6	4	1	1	1	3	2	2	1	1	9	1	24	3	1	1	29	33	1	5	13
Kappa	9	1	8	1	1	1	1	3	1	1	14	1	9	1	1	1	30	25	2	4	13
GK	5	2	6	1	1	6	1	13	1	1	4	1	3	8	1	1	15	11	3	1	12
Corr	8	1	7	1	1	6	2	10	1	1	15	1	12	5	1	1	30	32	3	2	11
Klos	1	2	2	1	1	9	1	14	1	1	17	2	29	10	2	1	27	26	9	12	11
2WaySup	2	1	5	1	1	5	1	18	1	1	13	1	29	6	1	1	30	34	13	8	10

Continued on next page

Table A.9 – continued from previous page

	Datasets																				
Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
CF	10	2	42	1	1	45	29	9	1	1	3	1	1	43	1	1	30	39	6	6	10
Gini	14	1	12	1	1	2	3	43	2	1	10	1	28	2	12	1	26	18	14	9	10
Spec	11	5	1	1	1	26	6	4	1	1	2	1	14	28	1	1	30	39	17	3	10
RelRisk	7	3	10	1	1	18	2	11	1	1	16	1	35	20	3	1	8	31	10	18	9
CCR	7	3	10	1	1	18	2	11	1	1	11	1	29	20	4	1	8	31	10	17	8
CnfrmC	13	8	17	1	1	24	3	16	1	10	7	1	41	26	19	1	12	38	2	1	8
LC	15	9	15	1	1	14	27	1	2	1	6	1	10	30	5	1	17	14	4	1	8
PS	21	12	16	1	1	7	1	41	2	3	20	1	5	9	1	1	30	19	15	13	8
CollStr	12	5	25	1	1	8	18	6	1	1	44	1	42	7	1	3	16	35	7	14	7
Cos	19	7	21	1	1	20	11	8	1	1	5	1	23	23	6	1	14	20	10	1	7
FM	16	6	19	1	1	21	9	5	1	1	4	1	20	24	7	1	24	21	8	1	7
Jacc	18	10	22	1	1	23	12	7	2	1	8	1	26	26	11	1	30	24	5	1	7
Acc	13	2	18	1	1	4	6	25	3	12	1	1	39	4	18	1	30	39	5	11	7
ImpInd	20	11	14	1	1	17	5	22	1	1	23	1	21	19	9	2	28	22	18	15	6
CnfrmD	38	14	38	1	1	43	26	47	3	1	18	4	32	41	23	1	30	23	12	16	5
addedvalue	3	17	3	6	8	15	1	32	11	8	37	7	31	16	10	19	9	16	29	31	3
Chi2	17	20	39	3	5	19	4	17	6	2	24	1	42	21	13	7	30	36	19	19	3
HConf	40	13	42	2	2	45	29	49	4	1	12	13	18	11	26	5	30	39	11	7	3
HLift	26	33	42	5	4	10	3	37	8	5	19	9	2	13	1	4	21	39	22	29	3
Lap	36	18	36	2	3	39	24	33	5	6	22	1	30	38	8	6	4	4	16	20	3
1WaySup	4	15	4	7	6	11	1	34	9	8	40	7	33	12	10	17	10	17	32	33	1
CCC	28	24	31	12	12	36	23	39	16	16	30	11	10	32	17	12	2	15	24	25	1
Conv	32	28	29	12	14	33	17	31	18	17	33	11	37	31	17	14	7	2	27	26	1
DChi2	25	34	30	9	19	41	25	44	22	11	41	3	40	39	24	20	30	29	33	23	1
InfoGain	35	16	11	16	7	15	2	42	10	9	36	7	42	16	10	18	30	27	34	35	1
IntImp	24	25	41	4	11	27	7	12	7	4	26	1	41	27	15	8	19	28	20	10	1
KM	29	23	13	12	12	22	14	19	16	17	27	11	11	18	17	12	3	10	24	25	1
Lift	37	32	9	17	18	16	2	46	12	13	43	8	6	17	22	21	13	30	35	36	1
Loe	31	27	28	12	13	32	16	20	17	17	31	11	16	31	17	13	7	3	26	25	1
SS	33	29	33	13	15	35	21	35	19	20	35	11	13	36	21	15	11	1	28	27	1
Gan	29	24	32	12	12	37	22	24	16	17	34	11	15	33	17	12	1	9	24	25	1
2WaySupVar	29	23	17	12	12	12	14	38	16	17	29	11	4	14	17	12	30	39	25	28	0
ConfC	27	22	35	11	12	42	19	45	15	15	27	10	17	40	17	11	20	7	24	24	0
CCD	29	23	32	12	12	35	22	24	16	17	34	11	15	33	17	12	7	9	24	25	0
Ex&Cex	29	23	33	12	12	34	20	30	16	17	27	11	13	35	17	12	11	5	24	25	0
JM	29	23	28	12	12	13	14	36	16	17	27	11	7	15	17	12	30	39	24	25	0
Lev	22	21	34	10	9	38	10	40	13	7	25	6	27	37	16	10	6	6	23	21	0
OddMul	33	31	27	15	17	25	16	21	20	18	38	11	22	22	20	16	22	13	30	30	0
OddR	34	30	24	14	16	28	15	26	21	19	39	11	38	25	20	16	23	11	31	32	0
YulQ	29	23	23	12	12	30	15	23	16	17	28	11	25	25	17	12	18	8	25	28	0
YulY	29	23	26	12	12	31	14	29	16	17	32	11	34	25	17	12	25	12	25	28	0
Zhang	23	19	20	8	10	29	8	28	14	14	21	5	8	29	14	9	5	5	21	22	0

Continued on next page

Table A.9 – continued from previous page

	Datasets																				
Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
LocSup	40	36	42	19	21	45	29	49	24	22	44	13	42	43	26	23	30	39	37	37	0
GlbSup	40	36	42	19	21	45	13	49	24	22	44	13	42	43	26	23	30	39	37	37	0
Conf	30	26	37	12	12	40	22	27	16	17	34	11	30	34	17	12	7	9	24	25	0
MutInfo	29	23	42	12	12	45	29	15	16	17	28	11	19	43	17	12	30	39	25	28	0
CCS	39	35	40	18	20	44	28	48	23	21	42	12	36	42	25	22	30	37	36	34	0

Table A.10: Measure rankings based on percentage of rule reduction without jeopardizing the f-measure while using measure-based pruning with local support in generation phase and the rules' average of measures in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Gini	34	1	4	1	1	2	22	17	2	1	10	1	15	2	6	1	23	3	6	29	10
Kappa	11	1	38	1	1	1	1	1	2	24	14	1	4	1	23	1	5	6	4	9	10
Klos	24	1	2	1	1	7	22	10	1	1	17	2	16	9	8	1	1	8	3	30	10
LC	1	7	6	1	1	11	5	3	2	4	6	1	6	14	3	1	23	2	1	17	10
GK	34	2	3	1	1	5	22	13	4	1	4	1	2	7	23	1	23	1	2	16	10
IWD	34	3	1	1	1	3	22	4	1	24	9	1	13	3	23	1	23	14	5	12	9
CF	34	4	38	1	1	42	22	5	2	24	3	1	1	44	23	1	23	31	7	1	8
FM	2	5	8	1	1	9	22	6	2	10	4	1	11	12	23	1	5	13	3	2	8
Jacc	3	8	9	1	1	19	22	2	1	5	8	1	14	23	19	1	23	30	2	5	8
Corr	26	1	38	1	1	5	22	8	1	24	15	1	7	5	2	1	23	11	5	14	7
Acc	8	1	38	1	1	4	22	20	2	7	1	1	36	4	23	1	23	31	12	7	7
2WaySup	34	1	38	1	1	5	22	40	1	24	13	1	16	6	23	1	4	12	6	10	6
CCR	7	4	35	1	1	30	18	21	30	22	11	12	18	31	1	1	17	29	3	3	6
Cos	4	6	9	1	1	8	12	7	3	6	5	1	12	21	16	1	23	19	3	13	6
PS	34	14	7	1	1	6	22	46	1	9	19	1	3	8	23	1	23	31	8	31	6
ImpInd	21	11	38	1	1	13	22	14	1	24	21	1	11	18	5	1	2	4	13	18	6
CnfrmC	9	7	34	1	1	23	3	19	1	24	7	7	43	28	22	2	21	31	11	33	5
RelRisk	7	4	35	1	1	30	18	21	30	22	16	12	20	31	1	1	17	29	3	8	5
Spec	8	4	37	1	1	41	22	12	7	24	2	1	9	43	23	1	23	31	9	4	5
Chi2	5	13	32	2	4	16	2	15	8	2	24	1	42	20	6	6	7	25	14	11	4
CollStr	32	4	36	1	1	29	19	11	1	19	22	5	44	32	23	3	5	10	10	6	4
CnfrmD	34	10	20	1	1	12	22	18	2	8	18	4	17	15	9	1	23	4	4	33	4
HConf	34	9	38	1	2	42	22	46	5	24	12	12	10	10	23	4	23	31	25	33	2
IntImp	12	17	31	3	6	21	17	9	9	11	25	2	38	26	7	7	4	26	15	15	2

Continued on next page

Table A.10 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
DChi2	10	34	24	5	10	36	6	35	10	6	36	3	40	40	13	9	23	24	17	19	1
HLift	23	36	38	6	9	10	22	38	14	3	23	8	29	11	23	23	23	31	20	28	1
Lap	31	12	28	9	3	35	15	16	6	19	20	12	22	37	4	5	6	20	16	26	1
GlbSup	34	39	5	4	5	42	4	46	31	24	40	12	44	44	10	28	3	9	25	33	1
1WaySup	18	32	11	10	22	18	22	37	24	24	28	12	37	13	23	20	20	18	21	32	0
2WaySupVar	34	19	38	15	7	14	22	39	11	12	27	11	5	16	23	28	8	7	25	24	0
addedvalue	17	16	10	9	26	17	22	36	22	24	35	12	24	19	23	25	19	17	25	32	0
ConfC	27	31	26	8	24	39	14	28	27	17	27	10	32	41	18	24	14	7	22	22	0
CCC	21	24	27	9	19	33	8	29	21	14	29	11	27	34	12	11	15	7	23	25	0
CCD	23	22	29	7	11	37	7	24	16	13	32	11	30	33	11	13	11	7	22	23	0
Conv	15	26	23	12	17	32	22	23	25	22	31	11	39	30	23	17	22	28	25	32	0
Ex&Cex	22	27	22	8	13	38	11	27	19	16	27	11	28	38	17	15	9	7	21	22	0
InfoGain	26	15	17	9	25	17	22	43	26	24	34	12	44	19	23	22	23	22	24	30	0
JM	19	19	38	15	8	15	22	22	12	24	27	11	8	17	23	10	18	7	19	29	0
KM	13	20	12	15	29	22	22	30	31	24	27	11	27	24	23	28	11	7	21	32	0
Lev	20	18	25	8	20	34	10	25	18	18	26	9	33	36	14	18	15	7	22	25	0
Lift	30	33	14	10	27	20	22	44	29	24	27	12	23	22	23	26	12	27	25	32	0
Loe	14	25	21	8	14	31	22	42	18	24	30	11	31	30	23	12	12	16	23	31	0
OddMul	16	37	19	9	23	26	21	32	25	22	37	11	32	25	22	21	19	21	24	33	0
OddR	28	30	16	9	18	28	20	33	15	21	38	12	34	27	21	21	15	23	24	28	0
SS	29	29	22	11	15	38	7	27	16	13	33	12	28	33	11	16	9	15	21	22	0
YulQ	25	21	15	9	18	25	22	31	15	24	27	11	21	27	23	19	13	7	24	28	0
YulY	25	23	18	9	21	27	22	34	23	24	27	11	26	27	23	19	16	7	25	28	0
Zhang	6	35	13	15	16	24	22	23	13	24	27	6	25	29	23	8	10	5	18	20	0
LocSup	34	39	38	15	29	42	22	46	31	24	40	12	44	44	23	28	23	31	25	33	0
Conf	23	28	30	13	12	37	13	41	20	23	32	11	35	39	15	14	19	7	24	27	0
MutInfo	34	19	38	15	29	42	22	26	11	24	27	11	19	44	23	28	23	7	25	21	0
Gan	23	22	29	7	11	37	9	24	17	15	27	11	30	35	12	13	11	7	22	23	0
CCS	33	38	33	14	28	40	16	45	28	20	39	12	41	42	20	27	23	31	23	27	0

Table A.11: Measure rankings based on percentage of rule reduction without changing the maximum possible accuracy while using measure-based pruning with local support in generation phase and the highest ranked rule in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
FM	25	1	23	8	20	1	31	1	2	2	2	2	1	1	3	1	4	4	2	8	12
Jacc	25	2	21	8	20	4	31	2	3	3	3	1	2	4	4	1	4	5	3	5	9
Cos	21	3	23	6	20	2	31	4	3	18	4	2	10	2	2	2	4	3	5	9	8
Acc	2	10	1	1	2	15	1	17	9	10	17	12	16	7	21	15	4	34	16	11	5
Spec	2	24	1	8	1	14	23	17	5	11	14	12	10	15	33	15	4	34	14	1	4
CF	25	24	23	8	20	15	31	17	1	39	1	26	16	19	33	37	4	34	5	3	3
Chi2	4	5	17	3	6	3	7	8	11	7	7	8	12	3	6	11	4	19	12	17	3
CollStr	25	20	18	8	20	15	26	17	37	1	39	4	16	19	1	37	4	8	2	39	3
Kappa	8	24	3	8	20	15	2	5	8	8	5	7	5	19	5	3	4	6	7	4	3
PS	25	24	23	8	20	6	31	3	7	14	6	17	16	6	7	5	4	2	1	38	3
CCR	2	9	1	7	11	14	21	17	18	10	33	22	16	18	28	27	4	27	17	12	2
CnfrmC	3	24	2	8	18	12	22	17	37	10	35	14	13	14	21	16	4	34	15	13	2
HConf	25	24	23	8	20	15	31	17	4	39	39	26	16	19	33	2	4	1	32	7	2
LC	20	4	23	8	20	7	31	17	13	16	7	1	16	8	30	1	4	34	8	4	2
RelRisk	2	9	1	7	11	14	21	17	18	10	33	22	16	18	28	27	4	27	17	14	2
Zhang	25	24	10	8	3	15	5	12	26	19	28	11	14	19	14	17	3	30	19	20	2
GK	1	24	23	8	20	15	31	17	10	4	15	26	16	5	33	14	4	34	10	2	2
2waySupVar	25	11	23	8	20	15	31	17	21	20	20	21	3	19	18	20	4	34	22	28	1
CnfrmD	25	7	23	8	20	15	31	17	14	31	11	3	16	19	15	4	4	34	7	39	1
Corr	6	24	11	8	13	15	3	9	15	9	13	8	8	19	9	9	4	9	6	6	1
DChi2	7	13	4	2	5	8	10	17	20	5	19	13	11	9	29	18	4	15	18	23	1
JM	25	24	23	8	20	15	31	10	22	21	22	20	7	19	18	19	1	12	22	25	1
Loe	15	24	12	8	8	15	17	17	30	28	27	18	16	19	18	24	3	21	23	25	1
OddR	10	15	5	8	7	14	14	17	25	23	25	19	16	15	19	21	2	33	27	33	1
1WaySup	18	24	14	8	10	15	12	16	31	27	32	22	16	19	24	26	4	22	28	34	0
2waySup	25	24	23	8	20	15	31	7	6	17	8	6	4	19	10	6	4	10	4	16	0
AddVal	17	24	13	8	11	15	16	17	27	29	27	23	16	19	25	33	4	28	25	32	0
ConfC	22	24	20	8	20	13	30	17	37	37	36	19	16	17	17	37	4	31	22	24	0
CCC	23	24	22	8	20	13	25	17	37	33	34	18	16	15	18	35	4	23	22	25	0
CCD	24	16	23	7	15	15	28	14	35	35	31	18	16	19	18	29	4	17	22	25	0
Conv	15	23	16	5	12	15	17	17	33	28	37	19	16	11	18	25	4	25	23	26	0
Ex&Cex	24	17	23	8	18	12	27	13	34	36	28	18	16	16	18	28	4	25	22	25	0
Gini	25	6	23	8	20	5	31	17	16	12	9	15	16	4	33	8	4	34	7	15	0
HLift	12	18	23	4	16	9	11	15	23	22	26	19	15	10	23	22	4	34	20	30	0
InfoGain	14	24	15	8	12	15	15	17	29	28	29	22	16	19	26	32	4	24	29	36	0

Continued on next page

Table A.11 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
IntImp	25	8	9	5	17	15	19	17	37	4	16	5	16	19	12	13	4	20	13	19	0
IWD	5	24	23	8	20	15	31	11	12	6	10	6	9	19	11	7	4	11	8	10	0
Klos	13	24	7	8	4	15	18	17	19	13	18	16	16	19	31	12	4	34	8	22	0
KM	25	24	23	8	20	15	31	14	37	39	31	26	16	19	18	37	4	18	22	25	0
Lap	25	19	23	8	20	15	31	17	37	34	30	10	16	19	22	23	4	14	11	29	0
Lev	24	24	22	8	19	11	29	17	37	38	34	21	16	14	16	37	4	16	21	21	0
Lift	19	23	16	8	11	15	20	14	32	30	39	24	16	19	27	36	4	24	30	37	0
OddMul	16	23	16	5	14	10	13	17	28	24	31	20	16	12	19	31	4	32	26	31	0
SS	24	21	23	8	19	15	29	14	36	35	31	18	16	19	20	34	4	25	24	27	0
YulQ	11	12	5	8	7	15	8	17	25	25	23	20	16	19	18	21	4	26	22	28	0
YulY	11	14	6	8	7	15	9	17	24	26	24	20	16	19	18	21	4	29	22	28	0
LocSup	25	24	23	8	20	15	31	17	37	39	39	26	16	19	33	37	4	34	32	39	0
GlbSup	25	24	23	8	20	15	4	17	37	39	39	26	16	19	13	37	4	7	32	39	0
Conf	25	22	23	8	20	15	31	17	37	36	31	18	16	19	18	30	4	18	22	25	0
MutInfo	25	24	23	8	20	15	31	17	21	20	21	26	16	19	18	37	4	34	22	28	0
Gan	24	16	23	7	15	15	28	14	35	35	31	18	16	19	18	29	4	17	22	25	0
ImpInd	9	24	8	8	9	15	6	6	17	15	12	9	6	19	8	10	4	13	9	18	0
CCS	24	24	19	8	16	11	24	16	37	32	38	25	12	13	32	37	4	34	31	35	0

Table A.12: Measure rankings based on percentage of rule reduction while using measure-based pruning with local support in generation phase and the rules' average of measures in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
FM	25	1	23	8	20	1	31	1	2	2	2	2	1	1	3	1	4	4	2	8	12
Jacc	25	2	21	8	20	4	31	2	3	3	3	1	2	4	4	1	4	5	3	5	9
Cos	21	3	23	6	20	2	31	4	3	18	4	2	10	2	2	2	4	3	5	9	8
Acc	2	10	1	1	2	15	1	17	9	10	17	12	16	7	21	15	4	34	16	11	5
Spec	2	24	1	8	1	14	23	17	5	11	14	12	10	15	33	15	4	34	14	1	4
CF	25	24	23	8	20	15	31	17	1	39	1	26	16	19	33	37	4	34	5	3	3
Chi2	4	5	17	3	6	3	7	8	11	7	7	8	12	3	6	11	4	19	12	17	3
CollStr	25	20	18	8	20	15	26	17	37	1	39	4	16	19	1	37	4	8	2	39	3
Kappa	8	24	3	8	20	15	2	5	8	8	5	7	5	19	5	3	4	6	7	4	3
PS	25	24	23	8	20	6	31	3	7	14	6	17	16	6	7	5	4	2	1	38	3
CCR	2	9	1	7	11	14	21	17	18	10	33	22	16	18	28	27	4	27	17	12	2
CnfrmC	3	24	2	8	18	12	22	17	37	10	35	14	13	14	21	16	4	34	15	13	2

Continued on next page

Table A.12 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
HConf	25	24	23	8	20	15	31	17	4	39	39	26	16	19	33	2	4	1	32	7	2
LC	20	4	23	8	20	7	31	17	13	16	7	1	16	8	30	1	4	34	8	4	2
RelRisk	2	9	1	7	11	14	21	17	18	10	33	22	16	18	28	27	4	27	17	14	2
Zhang	25	24	10	8	3	15	5	12	26	19	28	11	14	19	14	17	3	30	19	20	2
GK	1	24	23	8	20	15	31	17	10	4	15	26	16	5	33	14	4	34	10	2	2
2waySupVar	25	11	23	8	20	15	31	17	21	20	20	21	3	19	18	20	4	34	22	28	1
CnfrmD	25	7	23	8	20	15	31	17	14	31	11	3	16	19	15	4	4	34	7	39	1
Corr	6	24	11	8	13	15	3	9	15	9	13	8	8	19	9	9	4	9	6	6	1
DChi2	7	13	4	2	5	8	10	17	20	5	19	13	11	9	29	18	4	15	18	23	1
JM	25	24	23	8	20	15	31	10	22	21	22	20	7	19	18	19	1	12	22	25	1
Loe	15	24	12	8	8	15	17	17	30	28	27	18	16	19	18	24	3	21	23	25	1
OddR	10	15	5	8	7	14	14	17	25	23	25	19	16	15	19	21	2	33	27	33	1
1WaySup	18	24	14	8	10	15	12	16	31	27	32	22	16	19	24	26	4	22	28	34	0
2waySup	25	24	23	8	20	15	31	7	6	17	8	6	4	19	10	6	4	10	4	16	0
AddVal	17	24	13	8	11	15	16	17	27	29	27	23	16	19	25	33	4	28	25	32	0
ConfC	22	24	20	8	20	13	30	17	37	37	36	19	16	17	17	37	4	31	22	24	0
CCC	23	24	22	8	20	13	25	17	37	33	34	18	16	15	18	35	4	23	22	25	0
CCD	24	16	23	7	15	15	28	14	35	35	31	18	16	19	18	29	4	17	22	25	0
Conv	15	23	16	5	12	15	17	17	33	28	37	19	16	11	18	25	4	25	23	26	0
Ex&Cex	24	17	23	8	18	12	27	13	34	36	28	18	16	16	18	28	4	25	22	25	0
Gini	25	6	23	8	20	5	31	17	16	12	9	15	16	4	33	8	4	34	7	15	0
HLift	12	18	23	4	16	9	11	15	23	22	26	19	15	10	23	22	4	34	20	30	0
InfoGain	14	24	15	8	12	15	15	17	29	28	29	22	16	19	26	32	4	24	29	36	0
IntImp	25	8	9	5	17	15	19	17	37	4	16	5	16	19	12	13	4	20	13	19	0
IWD	5	24	23	8	20	15	31	11	12	6	10	6	9	19	11	7	4	11	8	10	0
Klos	13	24	7	8	4	15	18	17	19	13	18	16	16	19	31	12	4	34	8	22	0
KM	25	24	23	8	20	15	31	14	37	39	31	26	16	19	18	37	4	18	22	25	0
Lap	25	19	23	8	20	15	31	17	37	34	30	10	16	19	22	23	4	14	11	29	0
Lev	24	24	22	8	19	11	29	17	37	38	34	21	16	14	16	37	4	16	21	21	0
Lift	19	23	16	8	11	15	20	14	32	30	39	24	16	19	27	36	4	24	30	37	0
OddMul	16	23	16	5	14	10	13	17	28	24	31	20	16	12	19	31	4	32	26	31	0
SS	24	21	23	8	19	15	29	14	36	35	31	18	16	19	20	34	4	25	24	27	0
YulQ	11	12	5	8	7	15	8	17	25	25	23	20	16	19	18	21	4	26	22	28	0
YulY	11	14	6	8	7	15	9	17	24	26	24	20	16	19	18	21	4	29	22	28	0
LocSup	25	24	23	8	20	15	31	17	37	39	39	26	16	19	33	37	4	34	32	39	0
GlbSup	25	24	23	8	20	15	4	17	37	39	39	26	16	19	13	37	4	7	32	39	0
Conf	25	22	23	8	20	15	31	17	37	36	31	18	16	19	18	30	4	18	22	25	0
MutInfo	25	24	23	8	20	15	31	17	21	20	21	26	16	19	18	37	4	34	22	28	0
Gan	24	16	23	7	15	15	28	14	35	35	31	18	16	19	18	29	4	17	22	25	0
ImpInd	9	24	8	8	9	15	6	6	17	15	12	9	6	19	8	10	4	13	9	18	0
CCS	24	24	19	8	16	11	24	16	37	32	38	25	12	13	32	37	4	34	31	35	0

Table A.13: Measure rankings based on percentage of f-measure improvement while using measure-based pruning with local support in generation phase and the highest ranked rule in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
KM	1	3	3	21	21	10	1	29	30	15	2	2	5	9	2	11	6	10	4	11	7
CF	32	17	30	25	18	35	35	6	6	1	2	3	2	33	27	1	26	32	10	3	6
Kappa	9	9	19	25	14	23	14	5	3	4	1	11	3	17	3	2	21	20	8	2	6
Lev	7	1	15	10	10	28	23	35	10	3	2	1	5	21	3	4	7	6	1	8	6
1WaySup	5	8	2	26	19	3	4	23	27	17	2	2	5	1	6	15	11	14	26	24	5
Corr	10	12	20	25	18	12	2	7	4	6	1	11	3	7	3	2	21	17	12	5	5
Cos	14	12	20	12	10	32	32	2	2	16	1	8	3	29	4	2	22	21	6	4	5
FM	9	12	20	17	11	32	15	1	4	16	1	2	3	29	8	2	22	14	11	4	5
Jacc	11	12	20	16	9	32	32	3	5	20	1	13	3	29	9	2	23	22	7	1	5
Zhang	3	14	1	15	12	2	9	8	15	27	2	14	5	4	6	9	6	7	3	14	5
AddVal	4	16	4	26	19	1	5	31	28	15	2	2	5	1	14	22	10	5	24	22	4
InfoGain	8	10	2	24	19	4	3	16	23	17	2	5	5	3	22	19	15	13	27	26	4
IWD	13	12	21	6	17	17	2	5	7	7	1	11	3	24	17	2	23	18	15	6	4
LC	11	12	22	19	12	32	32	4	9	22	1	13	3	29	9	2	21	24	7	1	4
Acc	6	2	11	27	18	22	32	10	8	11	1	15	5	12	1	2	21	31	17	6	4
MutInfo	28	15	29	30	21	37	27	27	30	33	3	7	3	35	5	1	3	10	29	26	4
GK	27	6	13	25	18	24	22	4	8	5	1	16	3	10	23	2	19	26	6	2	4
2WaySupVar	26	16	25	22	20	15	25	36	30	33	2	7	3	14	5	3	4	10	28	26	3
Chi2	7	5	20	28	16	20	2	4	12	2	1	9	5	26	13	16	21	31	22	12	3
CollStr	17	10	20	23	3	29	20	11	1	8	4	15	6	27	16	10	18	12	2	23	3
CCD	19	15	20	2	3	32	26	33	21	29	2	7	5	29	5	12	8	5	5	13	3
Conv	16	16	5	9	3	16	21	38	16	29	2	4	5	14	5	15	8	2	13	10	3
Gini	31	7	26	7	18	21	18	9	7	9	1	9	3	18	13	2	25	19	15	16	3
OddMul	2	15	1	13	10	7	10	21	19	23	2	4	5	6	12	15	13	16	25	21	3
SS	25	15	20	4	2	32	29	38	16	30	2	7	5	29	15	20	7	3	14	15	3
YulQ	23	16	10	18	19	6	7	25	25	28	2	7	5	2	5	1	8	7	29	26	3
Gan	20	15	20	2	3	32	26	33	21	29	2	7	5	29	5	12	6	5	5	13	3
2WaySup	25	13	24	6	15	11	11	28	9	13	1	6	3	8	19	5	24	10	11	16	2
CCR	21	6	8	27	18	14	32	12	31	18	2	21	3	5	7	7	16	31	20	7	2
ConfC	19	16	16	1	7	27	27	38	22	29	2	4	5	16	5	6	8	8	5	13	2
CnfrmC	6	2	14	17	12	5	32	19	11	11	1	4	5	12	21	4	12	25	16	6	2
CCC	20	16	17	5	6	31	27	32	20	29	2	7	5	23	5	12	8	2	5	13	2
HConf	35	18	31	14	12	34	34	39	17	33	3	22	1	32	29	8	29	29	19	9	2
HLift	12	16	31	32	24	12	16	13	33	12	2	10	5	11	24	5	2	31	29	26	2
JM	15	10	23	22	19	18	24	36	30	21	2	4	3	22	5	7	6	10	5	13	2
Klos	18	9	12	16	17	19	8	34	14	10	1	17	3	19	18	7	22	11	18	25	2

Continued on next page

Table A.13 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Lap	29	14	20	11	1	32	32	24	17	32	1	20	5	29	11	18	14	7	5	20	2
Loe	16	15	4	8	5	13	19	33	12	26	2	4	5	13	5	14	9	1	8	10	2
OddR	23	16	9	22	19	9	12	14	26	24	2	15	5	2	11	23	18	15	29	26	2
PS	33	13	28	20	13	30	30	26	9	20	1	4	5	30	9	2	27	23	9	26	2
Spec	20	4	10	27	18	25	32	10	29	14	1	19	2	21	27	7	21	31	20	6	2
YulY	23	15	10	16	19	8	6	20	18	25	2	4	5	8	5	1	5	4	29	26	2
GlbSup	24	19	7	31	23	36	13	40	32	34	5	23	7	34	10	25	1	3	31	27	2
Conf	20	15	20	12	3	32	27	33	21	29	2	7	5	29	5	6	5	6	5	13	2
ImpInd	29	13	20	27	13	17	15	18	10	20	1	6	3	20	8	13	22	9	20	19	2
CCS	22	16	14	3	13	10	28	22	13	29	2	4	5	11	25	17	20	27	23	18	2
CnfrmD	34	14	27	7	8	33	33	21	6	30	1	18	4	31	26	9	28	13	19	26	1
DChi2	18	16	18	34	25	25	32	37	34	11	1	12	5	25	28	24	21	30	30	26	1
Ex&Cex	19	16	20	5	4	32	27	30	21	29	2	7	5	29	5	6	7	6	5	13	1
IntImp	30	11	21	29	22	26	17	15	24	19	1	6	5	28	20	21	17	31	21	19	1
Lift	29	14	6	33	19	26	31	17	23	31	2	5	5	15	22	21	14	28	27	26	1
RelRisk	21	6	8	27	18	14	32	12	31	18	2	21	4	5	7	7	17	31	20	17	1
LocSup	36	20	31	35	26	38	36	41	35	34	6	23	8	36	30	26	30	33	32	27	0

Table A.14: Measure rankings based on percentage of f-measure improvement while using measure-based pruning with local support in generation phase and the rules' average of measures in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Chi2	3	3	11	3	1	12	9	6	10	2	2	9	11	6	3	2	2	1	3	5	11
DChi2	2	2	8	16	5	18	6	29	9	3	3	16	10	8	7	3	8	12	11	3	6
FM	9	2	17	11	6	25	16	2	2	13	2	2	11	19	11	7	4	14	1	5	6
IntImp	21	3	16	14	1	28	12	1	17	12	4	8	12	21	4	5	2	3	5	6	5
Jacc	9	2	18	10	6	24	16	3	4	9	2	1	11	18	10	7	8	17	1	5	5
InfoGain	18	8	10	19	11	3	30	22	11	20	3	25	11	7	25	1	8	2	11	5	4
GK	22	17	1	15	5	26	26	24	5	1	2	22	11	12	21	8	7	8	1	6	4
1WaySup	18	8	7	19	11	2	30	15	17	20	5	25	9	1	26	4	8	4	11	3	3
2WaySup	26	15	31	1	7	8	31	36	3	27	2	10	13	11	19	6	5	18	8	6	3
CnfrmD	28	4	23	2	2	22	27	5	17	6	2	14	17	23	12	5	16	20	6	6	3
Corr	17	9	28	8	9	10	21	6	13	19	2	10	11	3	1	4	8	6	4	5	3
Cos	6	2	17	4	6	21	15	8	1	10	2	4	11	21	11	7	8	11	7	5	3
Ex&Cex	5	2	13	21	5	25	1	9	22	14	3	18	5	19	8	7	7	14	11	6	3

Continued on next page

Table A.14 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Gini	29	8	12	2	8	11	21	7	16	5	2	9	15	17	3	8	13	19	6	5	3
IWD	23	13	26	1	7	20	20	12	7	21	2	15	9	28	20	6	8	15	1	5	3
Kappa	18	16	29	12	4	31	18	11	9	18	2	3	11	29	22	6	4	12	2	5	3
Loe	10	7	17	19	11	13	25	20	15	22	2	18	10	25	18	2	6	9	11	3	3
OddMul	9	2	2	16	5	7	16	20	19	14	9	20	7	11	11	1	7	11	11	5	3
OddR	9	2	17	16	5	9	16	23	24	11	1	23	11	13	11	1	8	10	11	5	3
SS	5	2	14	22	6	25	3	16	22	14	2	23	5	22	9	7	7	13	11	6	3
YulQ	16	8	25	20	11	5	29	28	17	23	1	18	2	14	18	4	8	7	11	3	3
ImpInd	17	7	32	7	5	22	33	4	14	26	2	13	9	31	14	4	2	2	6	5	3
AddVal	15	8	4	19	11	1	30	25	14	20	6	25	9	2	26	4	8	8	11	5	2
CCR	9	2	17	16	5	16	16	31	17	11	2	23	11	16	5	4	8	14	11	4	2
CollStr	14	1	18	16	5	23	14	30	8	14	11	6	19	24	11	8	3	14	6	8	2
ConfC	7	2	17	22	6	25	10	22	22	14	4	12	11	22	2	7	8	14	11	5	2
CnfrmC	9	3	19	6	6	6	14	13	23	14	2	23	10	15	11	7	7	14	11	5	2
CCD	5	2	13	22	6	25	2	9	22	14	5	18	10	22	9	7	7	13	11	6	2
Conv	9	2	17	18	5	21	16	32	20	11	3	18	10	22	11	4	7	9	11	5	2
HLift	8	2	34	17	5	14	16	25	20	14	5	23	2	7	11	4	8	14	11	6	2
KM	12	10	3	27	18	4	29	24	30	23	5	18	3	4	18	13	7	11	11	5	2
Lift	16	7	6	19	10	27	34	26	18	28	3	26	1	10	24	4	8	14	11	5	2
RelRisk	9	2	17	16	5	16	16	31	17	11	2	23	11	16	5	4	8	14	11	6	2
Spec	9	3	17	16	6	24	16	35	7	14	4	17	11	22	11	7	8	14	11	2	2
YulY	16	6	24	20	11	10	29	28	17	23	1	18	8	16	18	4	8	4	11	3	2
Acc	9	2	17	16	6	25	16	35	6	14	2	21	11	8	11	6	8	14	9	6	2
Gan	5	2	13	22	6	25	2	9	22	14	5	18	10	22	9	7	7	13	11	6	2
CCS	1	1	18	16	5	17	13	10	22	14	10	23	8	13	11	7	7	14	11	5	2
2WaySupVar	25	5	27	27	16	22	22	34	26	15	2	18	13	25	18	13	11	8	14	5	1
CF	27	12	33	15	14	33	35	17	27	24	4	24	16	32	23	10	14	24	13	1	1
CCC	9	1	18	21	6	26	5	19	25	14	6	18	4	26	8	7	8	14	11	5	1
JM	20	11	27	26	13	15	24	14	12	23	2	18	13	19	18	9	9	9	11	6	1
Klos	21	11	22	5	8	30	19	33	13	16	1	19	11	23	15	4	8	14	10	5	1
Lap	9	2	17	22	6	25	7	16	16	8	5	23	11	22	11	7	8	14	11	6	1
LC	11	4	15	9	8	29	17	13	8	7	2	7	10	20	13	5	8	5	7	6	1
Lev	4	3	17	22	6	22	8	21	25	14	8	12	11	22	6	7	8	14	11	5	1
PS	30	14	21	5	3	19	32	37	4	4	7	11	17	9	16	8	15	22	6	6	1
Zhang	19	7	5	25	10	4	23	18	23	25	6	5	6	5	17	2	8	7	12	5	1
GlbSup	22	18	9	24	12	34	11	39	31	30	13	28	20	33	12	14	1	16	17	11	1
Conf	13	3	20	23	9	31	4	12	29	14	5	18	14	27	8	7	10	17	13	7	1
MutInfo	24	5	30	26	17	35	28	27	28	17	2	18	11	34	18	12	12	21	16	9	1
HConf	31	16	34	13	15	32	35	38	21	29	12	27	18	30	27	11	17	23	15	10	0
LocSup	32	19	34	28	19	36	36	40	32	30	14	28	21	35	28	15	18	25	18	11	0

Table A.15: Measure rankings based on percentage of f-measure improvement while using different measures in selection phase with local support in generation phase and the highest ranked rule in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
CCS	3	4	2	10	12	2	13	3	27	20	2	26	1	1	17	12	5	9	8	10	7
Lap	30	9	15	5	1	19	15	15	2	18	2	6	5	19	2	5	9	1	1	13	6
Zhang	1	4	3	7	9	1	10	8	25	20	2	20	7	3	13	11	3	5	6	7	6
CnfrmC	23	1	19	6	3	24	25	5	9	19	4	23	4	21	27	1	10	12	3	1	5
Cos	31	18	25	2	3	35	18	25	30	20	2	14	18	34	20	2	21	19	19	3	5
FM	31	20	25	2	3	37	18	27	29	20	2	16	17	34	16	2	24	22	18	3	5
IntImp	10	13	12	1	2	8	8	32	1	11	4	1	11	8	1	8	17	4	5	17	5
Jacc	31	20	25	2	3	37	18	27	29	20	2	16	17	34	16	2	24	22	19	3	5
Kappa	16	7	17	4	3	22	3	22	10	14	2	12	15	19	8	3	23	21	15	3	5
LC	31	14	25	2	3	36	18	18	28	20	2	10	12	33	21	2	15	15	10	3	5
OddMul	2	4	3	8	11	1	13	10	23	20	1	26	7	2	17	12	8	8	8	10	5
GK	12	5	24	3	3	18	9	27	15	7	2	8	17	15	14	3	14	14	19	3	5
ImpInd	26	6	8	3	3	5	7	31	3	11	4	3	19	6	4	2	20	14	7	18	5
2WaySup	24	10	9	4	3	16	3	33	8	12	2	2	23	16	6	7	24	20	22	19	4
Corr	14	7	18	4	3	13	4	21	5	4	2	4	16	14	5	3	18	17	12	2	4
HLift	18	25	26	15	8	27	21	19	24	2	3	27	3	25	19	26	4	3	9	16	4
Klos	5	10	7	3	3	6	1	28	4	15	4	5	11	6	3	9	17	11	4	15	4
PS	31	5	25	3	3	20	11	36	7	16	2	12	26	18	12	3	25	17	30	24	4
Chi2	13	7	23	4	3	12	6	21	5	4	2	4	16	14	5	3	18	17	12	5	3
CollStr	27	2	21	11	7	25	17	23	21	3	13	13	21	24	11	10	22	29	2	27	3
DChi2	17	11	14	4	7	3	20	13	12	1	4	7	9	7	18	4	13	2	14	5	3
Gini	21	7	23	4	3	12	6	35	5	4	2	4	22	14	5	3	18	17	20	22	3
Acc	25	8	20	4	3	30	23	6	11	13	2	11	10	27	27	3	16	16	13	4	3
2WaySupVar	36	19	22	3	3	23	14	38	18	5	8	12	27	22	9	15	28	25	28	30	2
AddVal	19	17	19	19	13	6	22	2	19	19	2	25	6	5	22	24	6	23	23	11	2
ConfC	6	3	8	9	11	10	12	9	26	20	2	26	5	11	17	12	7	8	8	10	2
CCC	7	3	10	9	11	14	12	8	26	20	1	26	5	12	17	12	7	8	8	10	2
Conv	4	4	1	9	11	4	13	11	26	20	1	26	5	4	17	12	7	8	8	10	2
HConf	35	24	26	16	16	29	18	30	32	20	7	30	2	30	23	18	1	28	16	21	2
IWD	15	12	11	4	3	21	2	24	6	17	4	9	14	17	7	14	19	17	11	6	2
Lev	8	6	6	14	10	11	16	1	17	9	4	22	5	10	15	20	2	10	17	8	2
Loe	4	4	1	9	11	4	13	11	26	20	1	26	5	4	17	12	7	8	8	10	2
1WaySup	20	21	19	20	14	9	24	7	20	19	1	28	7	9	24	25	7	24	25	12	1
CCR	36	28	5	5	3	31	25	20	14	19	10	15	24	28	25	17	11	6	33	29	1
CnfrmD	36	31	25	13	4	38	18	34	31	20	2	31	13	35	23	13	15	14	21	25	1
CCD	9	4	13	9	11	15	12	8	26	20	1	26	5	13	17	12	7	7	8	10	1

Continued on next page

Table A.15 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Ex&Cex	9	4	13	9	11	15	12	8	26	20	2	26	5	13	17	12	6	7	8	10	1
InfoGain	22	26	19	22	15	28	25	12	22	19	1	29	8	26	26	27	8	24	29	20	1
JM	32	23	16	4	3	17	5	37	13	10	8	8	25	20	10	19	27	18	32	23	1
Lift	22	26	19	22	15	28	25	12	22	19	1	29	8	26	26	27	8	24	29	20	1
RelRisk	36	28	5	5	3	31	25	20	14	19	10	15	24	28	25	17	11	6	33	29	1
SS	9	4	13	9	11	15	12	8	26	20	2	26	5	13	17	12	7	7	8	10	1
Spec	29	22	19	12	3	32	25	26	16	19	10	21	30	29	27	23	25	26	34	28	1
Conf	9	4	13	9	11	15	12	8	26	20	2	26	5	13	17	12	6	7	8	10	1
Gan	9	4	13	9	11	15	12	8	26	20	1	26	5	13	17	12	8	7	8	10	1
CF	33	29	25	17	5	26	19	17	33	6	12	19	28	23	12	22	25	31	34	26	0
KM	28	26	19	23	19	32	25	4	37	19	5	24	6	29	27	28	12	13	27	9	0
OddR	36	27	4	9	11	7	13	14	23	20	9	26	10	4	17	12	8	8	26	29	0
YulQ	11	16	4	9	11	7	13	14	23	20	11	26	10	4	17	12	7	8	26	29	0
YulY	11	16	4	9	11	7	13	14	23	20	11	26	10	4	17	12	7	8	26	29	0
LocSup	36	30	25	18	6	33	18	16	35	20	10	18	28	31	12	16	25	30	34	29	0
GlbSup	36	31	25	21	18	39	18	39	36	20	10	32	29	35	23	21	25	28	31	30	0
MutInfo	34	15	25	18	17	34	18	29	34	8	6	17	20	32	12	6	26	27	24	14	0

Table A.16: Measure rankings based on percentage of f-measure improvement while using different measures in selection phase with local support in generation phase and the rules' average of measures in selection phase. Instead of datasets' names, datasets' numbers are shown in the table. Tot shows the total number of high ranks for each dataset. The measures are sorted based on this number.

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
ConfC	2	1	1	1	3	2	1	2	4	1	2	5	2	1	4	2	7	1	2	3	16
CCC	4	1	3	2	1	5	3	7	7	3	6	3	5	4	1	2	8	3	2	2	12
Lev	5	2	2	3	4	3	2	6	6	5	1	6	6	2	2	3	8	2	3	4	11
Ex&Cex	3	3	7	1	2	9	6	9	5	6	6	2	5	10	5	1	3	6	1	1	9
CCD	1	4	5	4	1	8	5	11	7	7	6	2	7	7	3	1	9	4	1	2	7
Conf	1	4	5	4	1	8	5	11	7	7	6	2	7	7	3	1	9	4	1	2	7
Gan	1	4	5	4	1	8	5	11	7	7	6	2	7	7	3	1	9	4	1	2	7
CCS	7	6	9	12	24	1	10	1	20	20	1	18	1	3	10	13	13	20	15	11	5
Lap	9	5	6	2	3	10	8	16	1	8	5	1	6	9	6	5	10	5	7	5	4
SS	8	7	4	14	25	7	9	3	21	20	1	18	3	8	10	13	15	21	15	11	3
CnfrmC	16	12	14	38	28	21	31	14	26	17	3	22	11	14	29	22	1	27	13	15	2
Conv	6	6	6	12	25	4	10	4	18	20	1	18	4	3	10	13	15	22	15	11	2

Continued on next page

Table A.16 – continued from previous page

	Datasets																				
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Tot
Loe	6	6	8	12	25	4	10	4	18	20	1	18	4	3	10	13	15	22	15	11	2
OddMul	7	6	10	13	25	6	10	5	18	20	1	18	2	5	10	13	15	22	15	11	2
Zhang	12	3	16	10	8	13	30	10	12	14	6	8	8	12	26	7	3	12	4	6	2
CCR	36	27	14	39	17	24	31	22	17	17	19	22	22	23	29	30	2	15	21	38	1
Klos	21	9	20	15	9	18	18	30	3	17	9	20	20	18	17	10	19	16	17	9	1
LC	18	26	32	6	20	28	16	20	22	2	12	9	17	31	12	17	18	23	25	10	1
RelRisk	36	27	14	39	17	24	31	22	17	17	19	22	22	23	29	30	2	15	21	38	1
ImpInd	11	8	17	9	7	18	12	31	2	12	11	4	26	17	13	8	23	18	20	17	1
1WaySup	13	11	18	35	12	17	31	15	14	17	7	20	6	16	29	12	9	10	9	12	0
2WaySup	22	14	22	19	19	20	11	39	24	15	18	10	28	21	14	24	27	30	30	30	0
2WaySupVar	42	40	39	29	36	38	23	43	37	24	32	31	35	40	22	37	35	40	35	40	0
AddVal	10	10	19	28	11	16	31	8	13	17	6	16	5	15	27	9	9	9	8	7	0
CF	34	39	38	30	33	37	22	41	40	21	27	30	32	39	19	34	34	40	31	35	0
Chi2	31	33	26	18	14	25	13	28	16	17	12	12	23	24	11	23	21	26	22	16	0
CollStr	33	28	12	20	26	26	7	35	30	16	28	28	36	26	9	20	26	35	11	36	0
CnfrmD	36	29	34	7	22	33	16	38	23	19	12	24	16	35	18	11	17	19	26	31	0
Corr	20	13	24	21	13	20	27	27	16	17	16	14	24	20	25	21	24	24	24	21	0
Cos	24	30	28	11	27	31	21	32	29	9	23	17	27	28	15	28	25	31	35	25	0
DChi2	39	38	38	26	30	36	19	21	8	17	10	20	37	38	23	15	20	41	18	13	0
FM	26	31	31	24	32	30	21	37	31	13	24	26	26	29	20	32	31	37	37	29	0
Gini	32	33	26	18	14	25	13	36	16	17	12	12	26	24	11	23	21	25	27	18	0
HConf	35	15	41	23	29	15	16	25	28	19	4	23	38	25	18	16	11	43	10	24	0
HLift	14	16	41	17	10	19	29	19	11	17	8	21	12	22	24	14	12	7	14	23	0
InfoGain	16	18	21	39	15	24	31	9	15	17	6	22	9	23	29	18	5	14	12	22	0
IntImp	38	36	36	33	30	35	24	44	41	25	26	13	39	36	28	29	33	42	29	34	0
IWD	19	12	23	22	18	22	14	26	19	17	13	19	23	21	16	26	22	28	23	20	0
Jacc	27	32	30	25	31	30	19	29	32	11	22	25	26	32	21	31	30	36	36	26	0
JM	40	41	35	32	39	32	25	45	36	23	31	31	30	33	31	38	35	45	40	40	0
Kappa	23	17	25	27	23	22	26	34	27	17	20	15	25	19	23	27	28	33	33	27	0
KM	41	40	40	34	39	39	25	40	38	24	29	31	14	41	31	39	32	44	40	40	0
Lift	16	18	14	39	16	24	31	9	15	17	6	22	10	23	29	19	9	13	12	22	0
OddR	36	25	11	13	25	11	10	12	18	20	8	18	13	6	10	13	14	22	19	38	0
PS	28	24	27	16	21	27	4	42	25	13	21	11	29	27	7	25	28	32	32	32	0
Spec	17	19	14	39	35	24	31	23	33	17	25	22	34	23	29	30	29	34	34	33	0
YulQ	30	23	15	8	6	14	28	17	9	10	17	7	19	13	23	6	4	11	5	14	0
YulY	29	22	13	5	5	12	15	13	10	4	14	6	15	11	8	4	6	8	6	8	0
Acc	17	20	14	39	34	24	31	18	33	17	12	22	18	23	29	30	17	29	16	28	0
LocSup	25	34	29	37	38	29	20	33	34	12	27	27	33	30	22	36	31	38	38	37	0
GlbSup	37	35	33	36	37	34	17	46	35	18	27	29	31	34	18	35	31	39	39	39	0
MutInfo	43	37	37	31	33	36	22	47	39	22	30	30	25	37	19	33	36	40	34	40	0
GK	15	21	16	39	32	23	31	24	33	17	15	20	21	23	30	30	16	17	28	19	0

Appendix B

Impact of Measures While Using Local Support

B.1 Rule Reduction While Using Redundancy Removal Pruning

Datasets	RR %	FC change%	AC change%
Anneal	99.90	-28.57	-10.85
Breast	96.01	-0.28	-0.30
Census	94.05	-15.82	-4.38
Colic	98.79	-4.00	-0.33
Credit	98.83	-1.98	-1.99
Diabetes	84.84	-1.62	-0.17
German	95.42	-33.47	-4.11
Glass	86.65	-5.65	-6.02
Heart	97.47	+0.80	+1.16
Hepatitis	99.74	-34.10	-5.06
Iris	71.58	+0.16	0.00
Labor	99.23	-8.19	-6.88
Led7	53.96	-0.74	-1.11
Pima	84.77	-1.74	-0.54
Tictactoe	77.76	-52.37	-30.73
Vote	99.65	-6.74	-5.77
Vowel	98.94	-9.34	-9.22
Waveform	78.34	-4.04	-3.29
Wine	99.60	-12.00	-10.32
Zoo	99.77	-32.45	-21.31

Table B.1: Percentage of Rule reduction while using redundancy removal pruning on rule sets generated with local support as well as change of f-measure and accuracy while using redundancy removal pruning with average of rules for prediction using the same rule sets. RR, FC and AC are short forms for rule reduction, f-measure change and accuracy change respectively.

B.2 Rule Reduction Without Jeopardizing F-measure While Using Measure-based Pruning

Datasets	RR %	FC%	AC%	MPAC%	Measure
Anneal	-91.39	0.20	-1.46	-5.98	Corr (0.3)
Breast	-99.96	-1.88	-1.65	-69.54	2WaySup (0.4)
Census	-99.98	-0.06	-4.29	-85.89	IWD (0.1)
Colic	-99.99	0.96	0.72	-25.88	CF (0.7)
Credit	-99.99	29.85	20.35	-58.84	CCR (9)
Diabetes	-99.78	-0.63	-7.96	-74.31	Kappa (0.3)
German	-99.99	-0.52	-6.23	-58.26	2WaySup (0.1)
Glass	-94.19	-2.86	-11.26	-30.10	CollStr (7)
Heart	-99.99	22.30	15.17	-65.78	CnfrmC (0.7)
Hepatitis	-99.99	0.00	0.00	-25.75	CnfrmD (0.6)
Iris	-96.73	-3.18	-2.11	-10.74	CF (0.9)
Labor	-99.99	2.68	-0.97	-80.00	GK (0.5)
Led7	-79.68	-1.42	-1.15	-0.86	CF (0.6)
Pima	-99.72	-1.34	-9.27	-72.36	Kappa (0.3)
Tictactoe	-99.99	25.19	-0.76	-79.96	CF (0.01)
Vote	-99.99	5.23	5.62	-66.31	Kappa (0.9)
Vowel	-82.74	-4.95	-4.30	-5.27	IntImp (0.95)
Waveform	-99.99	-1.35	-2.48	-26.08	LC (0.2)
Wine	-99.86	8.37	4.18	-40.48	IWD (0.5)
Zoo	-99.09	3.81	4.29	-13.69	CnfrmC (0.99)

Table B.2: Percentage of rule reduction, f-measure change, accuracy change, maximum possible accuracy change and the measure with the minimum threshold used to get the minimum number of rules with measure-based pruning without jeopardizing the f-measure. Local support is used for rule generation and the selection phase is based on the highest ranked rule. RR, FC, AC and MPAC are short forms for rule reduction, f-measure change, accuracy change and maximum possible accuracy change respectively.

Group	Objective measures
1	1WaySup, InfoGain, AddVal, 2WaySup, Lift
2	2WaySupVar, JM
3	CF, CnfrmD, CCR, RelRisk, Spec
4	Chi2, IntImp, CollStr
5	ConfC, Lev, MutInfo
6	CnfrmC, Acc
7	CCC, CCD, Gan, Ex&Cex, Conf, Zhang, LocSup, GlbSup, Conv, Loe, KM, YulQ,YulY, OddMul, OddR, SS, CCS
8	Corr, GK, IWD
9	Cos, Jacc, LC, FM
10	DChi2
11	Gini, PS, Kappa
12	HConf
13	HLift
14	Klos
15	Lap
16	ImpInd

Table B.3: Clusters of measures with similar behaviour in finding the most rule reduction without jeopardizing the f-measure using measure based pruning.Local support is used for rule generation and the selection phase is based on the highest ranked rule.

Datasets	RR %	FC%	AC%	MPAC%	Measure
Anneal	-74.38	-2.63	-0.74	-0.78	LC (0.05)
Breast	-99.96	-3.44	-2.97	-69.54	2WaySup (0.4)
Census	-99.98	-4.43	-5.13	-85.89	IWD (0.1)
Colic	-99.99	5.88	6.03	-73.56	Gini (0.2)
Credit	-99.99	-1.66	-2.01	-58.84	CCR (9)
Diabetes	-99.78	-2.18	-7.96	-74.31	Kappa (0.3)
German	-99.99	-2.96	-5.34	-46.35	Kappa (0.2)
Glass	-89.59	-2.48	-5.04	-37.70	Kappa (0.5)
Heart	-99.99	-0.29	0.85	-65.78	CnfrmC (0.7)
Hepatitis	-99.99	-4.04	-3.75	-92.95	Gini (0.1)
Iris	-94.15	-2.62	-1.45	-6.04	Acc (0.95)
Labor	-99.99	-0.49	4.86	-75.00	IWD (0.3)
Led7	-79.68	-1.95	-1.86	-0.86	CF (0.6)
Pima	-99.72	-2.58	-8.96	-72.36	Kappa (0.3)
Tictactoe	-99.74	0.93	0.86	-67.02	RelRisk (3)
Vote	-99.99	-0.24	-0.24	-57.20	Acc (0.95)
Vowel	-72.23	-0.62	-0.28	-3.75	Klos (0.05)
Waveform	-99.75	-4.59	-5.19	-1.94	GK (0.05)
Wine	-99.75	-4.88	-5.25	-6.77	LC (0.7)
Zoo	-89.80	-4.44	-1.35	0.00	CF (0.6)

Table B.4: Percentage of rule reduction, f-measure change, accuracy change, maximum possible accuracy change and the measure with the minimum threshold used to get the minimum number of rules with measure-based pruning without jeopardizing the f-measure. Local support is used for rule generation and the selection phase is based on the rules' average of measures. RR, FC, AC and MPAC are short forms for rule reduction, f-measure change, accuracy change and maximum possible accuracy change respectively.

Group	Objective measures
1	1WaySup, AddVal, InfoGain, KM, YulQ, YulY, Lift
2	2WaySup, Corr, ImpInd
3	2WaySupVar
4	CF, Spec
5	Chi2, IntImp, DChi2
6	CCR, RelRisk
7	CollStr
8	ConfC, Lev, CCC, Conf, CCD, Gan, Ex&Cex, SS, LocSup, CCS, GlbSup
9	CnfrmC
10	CnfrmD, Gini, IWD, GK, PS
11	Conv
12	Cos, Jacc, LC
13	FM
14	HConf
15	HLift, Acc
16	JM, Kappa
17	Klos
18	Lap
19	Loe, Zhang
20	OddMul, OddR
21	MutInfo

Table B.5: Clusters of measures with similar behaviour in finding the most rule reduction without jeopardizing the f-measure using measure based pruning.Local support is used for rule generation and the selection phase is based on the rules' average of measures.

B.3 Rule Reduction Without Changing The Maximum Possible Accuracy While Using Measure-based Pruning

Datasets	**RR %**	**FC%**	**AC%**	**Measure**
Anneal	-49.61	-0.45	-3.74	GK (0.01)
Breast	-92.42	-1.63	-1.36	FM (0.4)
Census	-58.14	4.96	0.43	RelRisk (1.5), CCR (1.5), Spec (0.5)
Colic	-32.46	-23.27	-30.57	Acc (0.5)
Credit	-47.32	0.03	-5.37	Spec (0.5)
Diabetes	-91.73	-6.25	-1.25	FM (0.3)
German	-60.70	-8.25	-33.80	Acc (0.5)
Glass	-61.98	14.78	0.71	FM (0.2)
Heart	-99.96	10.65	5.14	CF (0.3)
Hepatitis	-95.16	43.48	0.65	CollStr (1.5)
Iris	-74.77	0.00	0.00	cf (0.2)
Labor	-99.89	73.12	31.07	LC (0.3)
Led7	-14.93	0.01	0.00	FM (0.1)
Pima	-91.07	-6.55	-2.09	FM (0.3)
Tictactoe	-99.61	28.01	6.50	CollStr (1.5)
Vote	-99.48	11.92	9.52	FM (0.6)
Vowel	-2.62	-0.27	-0.27	JM (0.01)
Waveform	-99.27	2.90	2.62	HConf (0.2)
Wine	-98.61	15.82	12.24	PS (0.1)
Zoo	-98.43	14.41	9.46	Spec (1)

Table B.6: Percentage of rule reduction, f-measure change, accuracy change and the measure with the minimum threshold used to get the minimum number of rules while the maximum possible accuracy does not change at all. Local support is used for rule generation and the selection phase is based on the highest ranked rule. RR, FC and AC are short forms for rule reduction, f-measure change and accuracy change respectively.

Datasets	RR %	FC%	AC%	Measure
Anneal	-49.61	-5.29	-3.68	GK (0.01)
Breast	-92.42	-1.36	-1.18	FM (0.4)
Census	-58.14	-21.93	-5.73	RelRisk (1.5), CCR (1.5), Spec (0.5)
Colic	-32.46	-1.36	-1.64	Acc (0.5)
Credit	-47.32	-1.19	-1.00	Spec (0.5)
Diabetes	-91.73	-4.20	0.00	FM (0.3)
German	-60.70	-23.57	-3.97	Acc (0.5)
Glass	-61.98	11.60	1.95	FM (0.2)
Heart	-99.96	-18.95	-14.07	CF (0.3)
Hepatitis	-95.16	-30.13	-4.26	CollStr (1.5)
Iris	-74.77	1.65	1.45	CF (0.2)
Labor	-99.89	6.55	6.88	LC (0.3)
Led7	-14.93	0.10	0.04	FM (0.1)
Pima	-91.07	-4.48	-0.53	FM (0.3)
Tictactoe	-99.61	-57.73	-32.44	CollStr (1.5)
Vote	-99.48	-2.67	-2.41	FM (0.6)
Vowel	-2.62	-1.91	-2.09	JM (0.01)
Waveform	-99.27	-12.78	-11.38	HConf (0.2)
Wine	-98.61	2.14	2.03	PS (0.1)
Zoo	-98.43	-14.35	-5.31	Spec (1)

Table B.7: Percentage of rule reduction, f-measure change, accuracy change and the measure with the minimum threshold used to get the minimum number of rules while the maximum possible accuracy does not change at all. Local support is used for rule generation and the selection phase is based on the average of rules. RR, FC and AC are short forms for rule reduction, f-measure change and accuracy change respectively.

Group	Objective measures
1	1WaySup, Lift, AddVal, InfoGain, YulQ, YulY, OddR, Loe
2	2WaySup, IWD, JM, Cos, FM, Jacc, PS
3	2WaySupVar, CnfrmD
4	CF, MutInfo
5	Chi2, HLift
6	CCR, RelRisk, Acc, DChi2, Conv, OddMul
7	CollStr, IntImp
8	ConfC, CCC, Lev
9	CnfrmC
10	CCD, Gan, SS, Ex&Cex, KM
11	Corr, ImpInd, Kappa
12	Gini, LC, GK
13	HConf
14	Klos, Spec
15	Lap, Conf
16	Zhang
17	LocSup, CCS
18	GlbSup

Table B.8: Clusters of measures with similar behaviour in finding the minimum number of rules while the maximum possible accuracy does not change using measure-based pruning.Local support is used for rule generation.

B.4 F-measure Improvement While Using Measure-based Pruning

Datasets	RR %	FC%	AC%	MPAC%	Measure
Anneal	-7.66	13.66	6.09	-1.34	KM (0.1)
Breast	-13.01	2.05	1.67	-1.43	Lev (0.8)
Census	-43.39	8.97	0.64	-16.08	Zhang (0.8)
Colic	-16.09	39.28	20.75	-2.70	ConfC (0.9)
Credit	-89.60	33.22	23.62	-5.93	Lap (0.9)
Diabetes	-36.82	8.69	2.11	-18.91	AddVal (0.2)
German	-59.69	31.65	-0.28	-1.90	KM (0.4)
Glass	-61.98	14.78	0.71	0.00	FM (0.2)
Heart	-99.92	30.58	21.24	-0.67	CollStr (9)
Hepatitis	-99.86	62.66	1.68	-0.59	CF (0.05)
Iris	-65.96	1.57	1.41	-0.67	Klos (0.2)
Labor	-1.11	85.92	35.92	-1.67	Lev (0.9)
Led7	-59.65	0.71	0.75	-0.39	HConf (0.9)
Pima	-36.67	8.12	0.87	-18.70	AddVal (0.2)
Tictactoe	-57.10	87.14	41.93	0.00	Acc (0.4)
Vote	-99.95	12.81	10.31	-0.92	CF (0.4)
Vowel	-1.57	10.62	9.93	-0.51	GlbSup (0.05)
Waveform	-92.82	10.89	9.48	-2.60	Loe (9)
Wine	-4.10	22.56	18.10	0.00	Lev (0.95)
Zoo	-94.30	19.44	10.50	0.00	LC (0.7)

Table B.9: Percentage of rule reduction, f-measure change, accuracy change, maximum possible accuracy change and the measure with the minimum threshold used to get the maximum f-measure while using measure-based pruning. Local support is used for rule generation and the selection phase is based on the highest ranked rule. RR, FC, AC and MPAC are short forms for rule reduction, f-measure change, accuracy change and maximum possible accuracy change respectively.

Group	Objective measures
1	1WaySup, InfoGain, AddVal, OddMul, OddR, Lift
2	2WaySup, Gini, IWD
3	2WaySupVar, MutInfo, YulQ, YulY
4	CF, Cos, Jacc, LC, FM
5	Chi2, IntImp
6	CCR, RelRisk, Spec, GK
7	CollStr
8	ConfC, CCC, CCD, Gan, Ex&Cex, Conf, SS, Conv, Loe
9	CnfrmC
10	CnfrmD, Lap
11	Corr, Kappa, Acc
12	DChi2
13	HConf
14	HLift
15	JM, KM
16	Klos
17	Lev
18	PS
19	Zhang
20	LocSup
21	GlbSup
22	ImpInd
23	CCS

Table B.10: Clusters of measures with similar behaviour in finding the maximum f-measure using measure-based pruning.Local support is used for rule generation and the selection phase is based on the highest ranked rule.

Datasets	**RR %**	**FC%**	**AC%**	**MPAC%**	**Measure**
Anneal	-1.01	1.68	1.31	-0.22	CCS (13)
Breast	-1.26	0.18	0.15	-0.14	CollStr (1)
Census	-77.31	3.24	-1.05	-0.69	GK (0.01)
Colic	-100.00	5.99	6.03	-73.28	2WaySup (0.3)
Credit	-90.38	0.79	0.65	-0.73	IntImp (0.99)
Diabetes	-36.82	5.94	1.59	-18.91	AddVal (0.2)
German	-30.07	8.78	-3.70	-0.70	Ex&Cex (0.6)
Glass	-45.13	12.31	11.31	-0.56	IntImp (0.5)
Heart	-98.89	4.08	4.05	0.00	Cos (0.5)
Hepatitis	-99.01	10.27	-2.76	-1.25	GK (0.1)
Iris	-65.96	4.62	4.35	-0.67	Klos (0.2)
Labor	-99.40	13.16	13.36	0.00	Jacc (0.2)
Led7	-35.66	2.02	1.61	-4.80	Lift (7)
Pima	-56.36	5.49	0.53	-34.98	1WaySup (0.5)
Tictactoe	-99.59	1.90	1.71	-0.84	Corr (0.2)
Vote	-1.03	0.97	0.97	0.00	InfoGain (0.05), OddMul (1.1)
Vowel	-42.86	5.19	5.45	-1.11	GlbSup (0.2)
Waveform	-45.83	1.96	1.50	0.00	Chi2 (19)
Wine	-97.27	4.54	4.43	0.00	GK (0.2)
Zoo	-60.02	1.88	0.88	0.00	CF (0.2)

Table B.11: Percentage of rule reduction, f-measure change, accuracy change, maximum possible accuracy change and the measure with the minimum threshold used to get the maximum f-measure while using measure-based pruning. Local support is used for rule generation and the selection phase is based on the rules' average of measures. RR, FC, AC and MPAC are short forms for rule reduction, f-measure change, accuracy change and maximum possible accuracy change respectively.

Group	Objective measures
1	1WaySup, AddVal, InfoGain, Lift, KM, Zhang
2	2WaySup, HConf
3	2WaySupVar
4	CF
5	Chi2, IntImp
6	CCR, RelRisk, OddR
7	CollStr
8	ConfC, Lev, CCC, Conf, CCD, Gan, Ex&Cex, SS, Lap
9	CnfrmC
10	CnfrmD, Gini, PS
11	Conv
12	Corr, JM
13	Cos, FM, Jacc
14	DChi2, Spec
15	HLift, CCS
16	IWD, Kappa, ImpInd
17	Klos
18	LC
19	Loe, YulQ, YulY
20	OddMul
21	Acc
22	LocSup
23	GlbSup
24	MutInfo
25	GK

Table B.12: Clusters of measures with similar behaviour in finding the maximum f-measure using measure-based pruning.Local support is used for rule generation and the selection phase is based on the rules' average of measures.

B.5 F-measure Improvement While Using Different Measures in Selection Phase

	Highest			Average		
Datasets	**FC%**	**AC%**	**Measure**	**FC%**	**AC%**	**Measure**
Anneal	9.02	4.26	Zhang	0.00	0.00	CCD, Conf, Gan
Breast	0.48	0.31	CnfrmC	0.70	0.60	ConfC
Census	9.64	1.09	Conv, Loe	4.44	-0.26	ConfC
Colic	32.83	15.08	IntImp	1.79	1.70	Ex&Cex
Credit	31.12	21.79	Lap	0.01	0.00	CCC
Diabetes	7.61	0.69	oddMul, Zhang	6.27	0.87	CCS
German	29.44	-1.25	Klos	8.53	-0.55	ConfC
Glass	8.41	-4.14	Lev	9.62	12.75	CCS
Heart	27.73	18.46	IntImp	2.02	2.07	Lap
Hepatitis	63.74	-0.89	DChi2	6.95	-1.45	ConfC
Iris	0.81	0.70	1WaySup, CCC, CCD, Conv, InfoGain, Lift, Loe, oddMul, Gan	3.78	3.62	Conv, Lev, Loe, oddMul, SS, CCS
Labor	77.68	30.58	IntImp	1.48	5.26	Lap
Led7	2.08	1.43	CCS	2.31	1.43	CCS
Pima	7.82	0.88	CCS	7.53	1.40	ConfC
Tictactoe	85.39	40.74	IntImp	0.50	0.43	CCC
Vote	13.12	10.59	CnfrmC	0.00	0.00	Conf, Gan CCD, Ex&Cex,
Vowel	3.42	2.95	HConf	1.88	2.23	CnfrmC
Waveform	9.67	8.18	Lap	0.31	0.27	ConfC
Wine	17.36	13.76	Lap	0.00	0.00	Conf, Gan CCD, Ex&Cex,
Zoo	19.49	11.53	CnfrmC	0.14	0.00	Ex&Cex

Table B.13: Percentage of f-measure change, accuracy change and the measure used in selection phase to get the maximum f-measure in selection phase. Local support is used for rule generation and the selection phase is based on both the highest ranked rule and rules' average of measures. FC and AC are short forms for f-measure change and accuracy change respectively.

Group	Objective measures
1	1WaySup, AddVal, InfoGain, Lift, KM, Lev
2	2WaySup, 2WaySupVar, JM, Gini, PS, Chi2, Corr, Kappa, GK, IWD, IntImp, Klos,ImpInd
3	CF, LocSup
4	CCR, RelRisk, Spec
5	CollStr, MutInfo
6	ConfC, Ex&Cex, Conf, SS, CCC, CCD, Gan, Conv, Loe, OddMul, CCS, Zhang
7	CnfrmC, Acc
8	CnfrmD
9	Cos, LC, FM, Jacc
10	DChi2
11	HConf, HLift
12	Lap
13	OddR, YulQ, YulY
14	GlbSup

Table B.14: Clusters of measures with similar behaviour in selection phase in finding the maximum f-measure .Local support is used for rule generation and the selection phase is based on the highest ranked rule.

Group	Objective measures
1	1WaySup, AddVal, InfoGain, Lift, Zhang, HLift, CCR, RelRisk
2	2WaySup, PS, IWD
3	2WaySupVar, CF, MutInfo, GlbSup, CollStr
4	Chi2, Gini, IntImp
5	ConfC, Lev
6	CnfrmC, Acc, GK, Spec
7	CnfrmD
8	CCC, CCD, Conf, Gan, Ex&Cex
9	Conv, Loe, OddMul, CCS, SS
10	Corr, Kappa
11	Cos, LC, FM, Jacc, LocSup
12	DChi2
13	HConf
14	JM, KM
15	Klos, Lap, ImpInd
16	OddR
17	YulQ, YulY

Table B.15: Clusters of measures with similar behaviour in selection phase in finding the maximum f-measure .Local support is used for rule generation and the selection phase is based on the rules' average of measures.

B.6 Using Interesting Measures in Both Pruning and Selection Phases

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	Corr	Zhang	0.20	9.02	-3.49
Breast	2waySup	CnfrmC	-1.88	0.48	-1.88
Census	IWD	Conv	-0.06	9.64	-0.06
Colic	CF	IntImp	0.96	32.83	-33.85
Credit	CCR	Lap	29.85	31.12	29.85
Diabetes	Kappa	OddMul	-0.63	7.61	-0.63
German	2waySup	Klos	-0.52	29.44	-0.52
Glass	CollStr	Lev	-2.86	8.41	-4.27
Heart	CnfrmC	IntImp	22.30	27.73	22.30
Hepatitis	CnfrmD	DChi2	0.00	63.74	0.00
Iris	CF	CCC	-3.18	0.81	0.81
Labor	GK	IntImp	2.68	77.68	2.68
Led7	CF	ccs	-1.42	2.08	-2.07
Pima	Kappa	ccs	-1.34	7.82	-1.34
Tictactoe	CF	IntImp	25.19	85.39	25.19
Vote	Kappa	CnfrmC	5.23	13.11	5.23
Vowel	IntImp	HConf	-4.95	3.42	-3.06
Waveform	LC	Lap	-1.35	9.67	-1.35
Wine	IWD	Lap	8.37	17.36	8.37
Zoo	CnfrmC	CnfrmC	3.81	19.49	3.81

Table B.16: Comparing the changes of f-measure with the best measure used in measure-based pruning for rule reduction without jeopardizing the f-measure, the best measure used in selection phase, and the combination of these two measures. Local support is used for rule generation and the selection phase is based on the highest ranked rule. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	LC	CCD	-2.63	0.00	-2.63
Breast	2waySup	ConfC	-3.44	0.70	-3.44
Census	IWD	ConfC	-4.43	4.44	-4.43
Colic	Gini	Ex&Cex	5.88	1.79	5.88
Credit	CCR	CCC	-1.66	0.01	-1.66
Diabetes	Kappa	ccs	-2.18	6.27	-2.18
German	Kappa	ConfC	-2.96	8.53	2.38
Glass	Kappa	ccs	-2.48	9.62	-0.22
Heart	CnfrmC	Lap	-0.29	2.02	-0.29
Hepatitis	Gini	ConfC	-4.04	6.95	-4.04
Iris	Acc	Lev	-2.62	3.78	-1.84
Labor	IWD	Lap	-0.49	1.48	-0.49
Led7	CF	ccs	-1.95	2.31	-0.97
Pima	Kappa	ConfC	-2.58	7.53	-2.58
Tictactoe	RelRisk	CCC	0.93	0.50	0.93
Vote	Acc	CCD	-0.24	0.00	-0.24
Vowel	Klos	CnfrmC	-0.62	1.88	-7.23
Waveform	GK	ConfC	-4.59	0.31	-4.61
Wine	LC	CCD	-4.88	0.00	-4.88
Zoo	CF	Ex&Cex	-4.44	0.14	-6.52

Table B.17: Comparing the changes of f-measure with the best measure used in measure-based pruning for rule reduction without jeopardizing the f-measure, the best measure used in selection phase, and the combination of these two measures. Local support is used for rule generation and the selection phase is based on the average of rules. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	GK	Zhang	-0.45	9.02	-23.95
Breast	FM	CnfrmC	-1.63	0.48	0.48
Census	CCR	Conv	4.96	9.64	-12.06
Colic	Acc	IntImp	-23.27	32.83	32.83
Credit	Spec	Lap	0.03	31.12	28.09
Diabetes	FM	Zhang	-6.25	7.61	6.80
German	Acc	Klos	-8.25	29.44	27.76
Glass	FM	Lev	14.78	8.41	7.66
Heart	CF	IntImp	10.65	27.73	-31.51
Hepatitis	CollStr	DChi2	43.48	63.74	63.36
Iris	CF	Conv	0.00	0.81	0.00
Labor	LC	IntImp	73.12	77.68	81.10
Led7	FM	ccs	0.01	2.08	1.72
Pima	FM	ccs	-6.55	7.82	8.82
Tictactoe	CollStr	IntImp	28.01	85.39	24.01
Vote	FM	CnfrmC	11.92	13.11	13.11
Vowel	JM	HConf	-0.27	3.42	2.75
Waveform	HConf	Lap	2.90	9.67	2.82
Wine	PS	Lap	15.82	17.36	17.36
Zoo	Spec	CnfrmC	14.41	19.49	13.72

Table B.18: Comparing the changes of f-measure with the best measure used in measure-based pruning for rule reduction without changing the maximum possible accuracy, the best measure used in selection phase, and the combination of these two measures. Local support is used for rule generation and the selection phase is based on the highest ranked rule. FC is the short form for f-measure change.

Datasets	Pruninng measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	GK	CCD	-5.29	0.00	-5.29
Breast	FM	ConfC	-1.36	0.70	0.22
Census	CCR	ConfC	-21.93	4.44	-15.91
Colic	Acc	Ex&Cex	-1.36	1.79	0.28
Credit	Spec	CCC	-1.19	0.01	-0.42
Diabetes	FM	ccs	-4.20	6.27	1.54
German	Acc	ConfC	-23.57	8.53	-56.57
Glass	FM	ccs	11.60	9.62	10.03
Heart	CF	Lap	-18.95	2.02	-18.64
Hepatitis	CollStr	ConfC	-30.13	6.95	-4.55
Iris	CF	Lev	1.65	3.78	1.65
Labor	LC	Lap	6.55	1.48	4.34
Led7	FM	ccs	0.10	2.31	1.57
Pima	FM	ConfC	-4.48	7.53	1.00
Tictactoe	CollStr	CCC	-57.73	0.50	-36.66
Vote	FM	CCD	-2.67	0.00	-2.67
Vowel	JM	CnfrmC	-1.91	1.88	0.12
Waveform	HConf	ConfC	-12.78	0.31	-11.02
Wine	PS	Ex&Cex	2.14	0.00	0.18
Zoo	Spec	Ex&Cex	-14.35	0.14	-16.75

Table B.19: Comparing the changes of f-measure with the best measure used in measure-based pruning for rule reduction without changing the maximum possible accuracy, the best measure used in selection phase, and the combination of these two measures. Local support is used for rule generation and the selection phase is based on the rules' average of measures. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	KM	Zhang	13.66	9.02	11.61
Breast	Lev	CnfrmC	2.05	0.48	-4.96
Census	Zhang	Loe	8.97	9.64	9.62
Colic	ConfC	IntImp	39.28	32.83	28.61
Credit	Lap	Lap	33.22	31.12	30.90
Diabetes	AddVal	OddMul	8.69	7.61	8.07
German	KM	Klos	31.65	29.44	17.30
Glass	FM	Lev	14.78	8.41	7.66
Heart	CollStr	IntImp	30.58	27.73	25.11
Hepatitis	CF	DChi2	62.65	63.74	39.37
Iris	Klos	1WaySup	1.57	0.81	1.57
Labor	Lev	IntImp	85.92	77.68	91.54
Led7	HConf	ccs	0.71	2.08	-1.52
Pima	AddVal	ccs	8.12	7.82	9.14
Tictactoe	Acc	IntImp	87.14	85.39	38.15
Vote	CF	CnfrmC	12.81	13.11	12.55
Vowel	GlbSup	HConf	10.62	3.42	3.79
Waveform	Loe	Lap	10.89	9.67	9.97
Wine	Lev	Lap	22.56	17.36	22.56
Zoo	LC	CnfrmC	19.44	19.49	19.49

Table B.20: Comparing the changes of f-measure with the best measure used in measure-based pruning for f-measure improvement, the best measure used in selection phase, and the combination of these two measures. Local support is used for rule generation and the selection phase is based on the highest ranked rules. FC is the short form for f-measure change.

Datasets	Pruning measure	Selecting measure	FC % prune	FC % select	FC % combine
Anneal	ccs	Conf	1.68	0.00	1.68
Breast	CollStr	ConfC	0.18	0.70	1.04
Census	GK	ConfC	3.24	4.44	-11.96
Colic	2waySup	Ex&Cex	5.99	1.79	5.99
Credit	IntImp	CCC	0.79	0.01	0.33
Diabetes	AddVal	ccs	5.94	6.27	7.58
German	Ex&Cex	ConfC	8.78	8.53	4.14
Glass	IntImp	ccs	12.31	9.62	15.28
Heart	Cos	Lap	4.08	2.02	4.08
Hepatitis	GK	ConfC	10.27	6.95	2.19
Iris	Klos	ccs	4.62	3.78	2.33
Labor	Jacc	Lap	13.16	1.48	7.75
Led7	Lift	ccs	2.02	2.31	3.03
Pima	1WaySup	ConfC	5.49	7.53	5.49
Tictactoe	Corr	CCC	1.90	0.50	1.90
Vote	OddMul	Conf	0.97	0.00	0.97
Vowel	GlbSup	CnfrmC	5.19	1.88	2.05
Waveform	Chi2	ConfC	1.96	0.31	2.03
Wine	GK	Gan	4.54	0.00	4.54
Zoo	CF	Ex&Cex	1.88	0.14	1.88

Table B.21: Comparing the changes of f-measure with the best measure used in measure-based pruning for f-measure improvement, the best measure used in selection phase, and the combination of these two measures. Local support is used for rule generation and the selection phase is based on the rules' average of measures. FC is the short form for f-measure change.