A Method of Finding Representative Sets of Rules

Jiye Li
York University
Toronto, ON, Canada
jiye@cse.yorku.ca

Nick Cercone
York University
Toronto, ON, Canada
ncercone@yorku.ca

Jianchao Han
California State University
Dominguez Hills, Carson, CA, USA
jhan@csudh.edu

Abstract

The use of rough sets theory to select essential attributes that can represent the original data set is well known. Knowledge discovered from such essential attributes are typically represented as rules, and are therefore representative of the original data. We present three results towards rule evaluation as an extension of the "Rules-as-Attributes measure". First, we present an approach of finding representative sets of rules for a given data set. Secondly, we suggest that the Johnson's Reducer of the ROSETTA software generates a reduct with the minimum number of rules, and can be considered as a minimum representation of the original knowledge. Our third result provides an integrated approach for rule evaluation based on both the Rule Importance Measure and the method of finding representative sets of rules. We argue that this approach can take the representative rules ranking into a further stage. These approaches are proposed to facilitate the rule evaluations and can provide an automatic and complete comprehension of the original data set.

1 Introduction

Knowledge discovery in databases is a process of discovering previously unknown, valid, novel, potentially useful and understandable patterns in large data sets [3]. Rule generation is one of the important processes in the knowledge discovery system. For example, a rule such as "Japanese cars with manual transmission and light weight usually have higher mileage", can be learned by a classification algorithm from a data set of cars which contain mileage of the cars and features such as the manufacturer, the model, the transmission, the weight and so on [6]. Such rules are used for making predictions. A challenging problem in rule generations is that an extensive number of rules are extracted by data mining algorithms over large data sets, and it is infeasible for human beings to select important, useful, and interesting rules manually. How to develop measures to automatically extract and evaluate interesting, relevant, and novel rules becomes an urgent and practical topic in this area. Many existing methods such as rule interestingness measures and rule quality measures from statistics and information theory areas were reported in [2, 5, 16]. We say a measure is a subjective measure if it is defined based on a domain expert's opinions towards the particular application [5]. A measure is an objective measure if it measures the data itself without any predefined opinions. Subjective measures that use real human evaluators are the optimal measures to evaluate rules, although they are sometimes infeasible and expensive. Previous studies on rule evaluations focus mostly on objective measures [5], which do not contain any knowledge from the domain of the data. Therefore such objective measures may not sufficiently evaluate whether a rule is indeed interesting for a certain domain.

In this paper, we study and propose rule evaluation measures for the purpose of facilitating the knowledge understanding process. Our motivation is to design automatic rule evaluation measures that can bring both domain related knowledge (such as what are the important attributes and what are the expected results) and the objective measures together into the rule evaluations. Such measures are proposed to help to extract representative rules from a large number of rules generated by a learning algorithm.

Rough sets theory was introduced by Pawlak in the early 1980's [13]. He introduced an early application of rough sets theory to knowledge discovery systems, and suggested that a rough sets approach can be used to increase the likelihood of correct predictions by identifying and removing redundant variables. Efforts into applying rough sets theory to knowledge discovery in databases have focused on decision making, data analysis, discovering and characterizing the inter-data relationships, and discovering interesting patterns [14]. It has been used previously in attribute selection, rule induction, classification, multi-agent systems, medical diagnosis and other application domains. According to the rough sets theory, an information system can be considered as a decision table, which is used to specify what conditions lead to decisions. A decision table can be defined as

$T = (U, C, D)$, where U is the set of objects in the table, C is the set of the condition attributes and D is the set of the decision attributes. A reduct of a decision table is a subset of the condition attributes that are sufficient to define the decision attributes. Reducts are often used in the attribute selection process at the data preprocessing stage in a knowledge discovery system. A reduct is not unique [13], and there may exist multiple reducts for one decision table. The minimum set of reduct contains the minimum number of condition attributes that are representative of the original data set. The core of a decision table is contained in every reduct, and it can be considered as the essential information of a decision table. Any reduct generated from the original data set cannot exclude the core attributes. Reduct and core are often used in the attribute selection process.

Since one of the usages for rough sets theory is to select the most relevant attributes towards a classification task, and to remove the unimportant attributes, rules generated by a learning algorithm can be considered more important if there exist important attributes in the rules. Therefore this theory can provide a theoretical foundation for rule evaluations. We study the mechanism of rough sets theory, and we first propose a method of finding representative sets of rules, motivated by a Rules-as-Attribute measure [10]. Once the set of rules are generated for a data set by a learning algorithm, a new decision table is constructed by considering rules as the new condition attributes. The Genetic Reducer of the ROSETTA rough set software [12] is then applied on the new decision table for multiple reducts generation. Such reducts contain representative attributes, which are the original rules generated from the original data set. Since the Genetic Reducer obtains multiple reducts, the representative rule set is not unique. Multiple representative set of rules can be used to comprehend the original data from multiple perspective. Secondly, a reduct containing the minimum number of rules can be considered as a minimum representation of the original knowledge. Therefore the minimum representative rule set can be obtained by Johnson's Reducer, which is designed to find a single reduct with minimal length. Our third study provides an integrated approach for rule evaluation based on both the Rule Importance Measure (RIM) [11] and the method of finding a representative set of rules. We argue that this approach can take the representative rules ranking into a further stage. These approaches are proposed to facilitate the rule evaluations and provide an automatic and complete comprehension of the original data set.

Related work on rough sets theory and rule discovery is discussed in Section 2. In Section 3 we discuss the Rules-as-Attributes measure. We propose the new approaches to extract the representative sets of rules in Section 4 and 5. Conclusions are discussed in Section 6.

2 Rough Sets Theory and Rule Discovery

Rough sets theory [13] is commonly used for attribute selection in the decision making process. The decision table consists of condition attributes and decision attributes. A reduct is a subset of condition attributes that can represent the whole data set.

There have been contributions on applying rough sets theory to rule discovery. Rules and decisions generated from the reducts are representative of the data set's knowledge. In [7], two modules were used in the association rules mining procedure for supporting organizational knowledge management and decision making. Self-Organizing Map was applied to cluster sale actions based on the similarities in the characteristics of a given set of customer records. Rough sets theory was used on each cluster to determine rules for association explanations. Hassanien [4] used rough sets to find all the reducts of data that contain the minimal subset of attributes associated with a class label for classification, and classified the data with reduced attributes. Another relevant work is that of Szczuka [15] who proposed a new method of constructing a classification system with a combination of a rule based system and neural networks. In fact, rough sets can be used to determine whether there exists redundant information in the data and whether we can find the essential data needed for our applications. Since the rough sets method can help to generate representative attributes, we expect fewer rules will be generated due to fewer attributes. And the rules will be as significant as the rules generated without using the rough sets approach.

Little effort to date has been expanded on applying rough sets theory to association rules generation. Rules generated from the original data set can be used to represent original knowledge. After a reduct is generated, a rule based on this reduct is generated in the form such that the antecedents of a rule are from the value of condition attributes in the reduct set, and the consequents of a rule are from the value of decision attributes from the original data set. Association rule generations also return rules with certain support and confidence. The association rule algorithm [1] is well known for discovering associations, e.g., shopping behaviours among transaction data. One of the main problems for association rule generations is that the number of rules generated is generally quite large; thus, it is very difficult to evaluate and rank these rules. In order to solve this problem, many novel approaches have been developed to extract more interesting rules. Rule templates [9] as one of the examples of the rule interestingness measures can be applied to extract appropriate rules towards certain applications. They are useful in decision making, recommender systems and other applications. The association rule algorithm can be used to extract rules from the decision table as well.

We define the rule templates in the following. Because our interest is to make decisions or recommendations based on the condition attributes, we are looking for rules with only decision attributes on the consequent part. Therefore, we specify the following 2 rule templates to extract rules we are interested in as shown by Template 1, and to subsume rules as shown by Template 2.

$$\langle Attribute_1, Attribute_2, \ldots, Attribute_n \rangle \rightarrow \langle DecisionAttribute \rangle \quad (1)$$

Template 1 specifies only decision attributes can be on the consequent part of a rule, and $Attribute_1$, $Attribute_2$, $\ldots$, $Attribute_n$ lead to a decision of $DecisionAttribute$, as shown by Template 1. We specify the rules to be removed or subsumed using Template 2. For example, given rule

$$\langle Attribute_1, Attribute_2 \rangle \rightarrow \langle DecisionAttribute \rangle \quad (2)$$

the following rules

$$\langle Attribute_1, Attribute_2, Attribute_3 \rangle \rightarrow \langle DecisionAttribute \rangle \quad (3)$$

$$\langle Attribute_1, Attribute_2, Attribute_6 \rangle \rightarrow \langle DecisionAttribute \rangle \quad (4)$$

can be removed because they are subsumed by Template 2.

3 Rules-as-Attributes Measure

The Rules-as-Attributes measure was proposed to extract important rules from a decision table [10] based on rough sets theory. In this section, we briefly discuss this approach.

The concept of a reduct is used in a new perspective. A reduct of a decision table contains attributes that can fully represent the original knowledge. When a reduct is given, rules extracted based on this reduct are representative of the original decision table. These representative rules are therefore considered more important than the rules generated without using the reduct. A reduct contains the most representative and important condition attributes of a decision table. Based on this intuition, each of the individual rules among the generated rules sets is considered as a condition attribute in a decision table. The reducts extracted from such decision tables contain representative and important attributes, which are the rules. Since the generation of reduct is an automatic process, we can use this approach to discover important rules from a set of generated rules automatically.

In this approach, Apriori association rules are generated from the original decision table. Each rule is considered as a condition attribute in the new constructed decision table. The decision attributes are the original decision attributes. Therefore, a reduct of such a decision table represents the essential attributes, which are the most important rules that fully describe the decision. Experiments on both artificial data set and real-world data sets demonstrate the effectiveness of the new way of ranking rules.

4 Finding Representative Sets of Rules

In the Rules-as-Attribute measure, Johnson's Reducer [8] is used to generate single reduct that contains the Reduct Rules. Experiments demonstrate the Reduct Rules extracted by the Johnson's Reducer are important and representative of the original data, therefore this approach can be used towards the rule evaluation.

Since reduct is not unique for one decision table, each of the reducts of a decision table is considered to be able to uniquely classify the decisions. Thus, each reduct can be considered as a representation of the classification ability. By obtaining multiple reducts, we can obtain multiple sets of representations on how to classify the original decision table. The Genetic Reducer by ROSETTA software provides the multiple reducts generation for a decision table. Therefore, when using the Genetic Reducer to extract multiple sets of Reduct Rules in the Rules-as-Attribute measure, we can obtain sets of rules that represents the original rules.

We define the **representative sets of rules** as follows.

Definition 1 *Let* $T = (U, C, D)$ *be a decision table. Let* $ReductRulesSet$ *be a set of rules generated based on the Rules-as-Attribute measure. Let* S *be a set containing all the* $ReductRulesSets$ *by Genetic Reducer.* $ReductRulesSet \in S$. *We say set* S *is the representative sets of rules for* T.

We consider these multiple reducts as representative sets of rules. Genetic Reducer obtains multiple reducts, the representative rule set is not unique. Multiple representative sets of rules are obtained for rule evaluation, and they provide multiple interpretation perspectives.

The representative sets of rules provide multiple interpretation perspectives for the original data set. Each element belonging to the representative sets is a set of rules generated from the original data, and it embodies the knowledge of the original data.

We describe the procedures to obtain the representative sets of rules. First, a set of rules are generated from a learning algorithm for the data. Secondly, a new decision table can be obtained according to the Rules-as-Attribute measure. By considering rules as attributes in this new decision table, we could use rough sets theory to assist with rule evaluations. Thirdly, the Genetic Reducer from ROSETTA is applied on this decision table to extract multiple reducts. Therefore we obtained the representative rule sets by union all these reducts into one set.

4.1 Car Data Set

Let us explain in detail our method of finding representative sets of rules using an artificial data set about cars [6], shown in Table 1, which is used to decide the mileage of different cars. This data set contains 14 records, and 8 condition attributes.

Table 1. Artificial Car Data Set

make_model	cyl	door	displace	compress	power	trans	weight	mileage
USA	6	2	Medium	High	High	Auto	Medium	Medium
USA	6	4	Medium	Medium	Medium	Manual	Medium	Medium
USA	4	2	Small	High	Medium	Auto	Medium	Medium
USA	4	2	Medium	Medium	Medium	Manual	Medium	Medium
USA	4	2	Medium	Medium	High	Manual	Medium	Medium
USA	6	4	Medium	Medium	High	Auto	Medium	Medium
USA	4	2	Medium	Medium	High	Auto	Medium	Medium
USA	4	2	Medium	High	High	Manual	Light	High
Japan	4	2	Small	High	Low	Manual	Light	High
Japan	4	2	Medium	Medium	Medium	Manual	Medium	High
Japan	4	2	Small	High	High	Manual	Medium	High
Japan	4	2	Small	Medium	Low	Manual	Medium	High
Japan	4	2	Small	High	Medium	Manual	Medium	High
USA	4	2	Small	High	Medium	Manual	Medium	High

There is no inconsistent data or incomplete data existing in this data set. Rule templates are applied, e.g., rules with only decision attribute *mileage* on the consequent part are generated; and subsumed rules are removed. There are 19 rules generated by the apriori algorithm with $support = 1\%, confidence = 100\%$[1], as shown in Table 2.

Table 2. Rules Generated by the Car Data Set

No.	Association Rules
0	USACar,Displace_Medium,Weight_Medium → Mileage_Medium
1	USACar, Compress_Medium → Mileage_Medium
2	USACar,Power_High → Mileage_Medium
3	Cyl_6 → Mileage_Medium
4	Door_4 → Mileage_Medium
5	Displace_Medium, Compress_High, Weight_Medium → Mileage_Medium
6	Displace_Medium, Power_High → Mileage_Medium
7	Compress_Medium, Power_High → Mileage_Medium
8	Trans_Auto → Mileage_Medium
9	JapanCar → Mileage_High
10	Cyl_4, Displace_Medium, Compress_High → Mileage_High
11	Cyl_4, Compress_High, Power_High → Mileage_High
12	Displace_Small, Compress_Medium → Mileage_High
13	Displace_Small, Power_High → Mileage_High
14	Displace_Small, Trans_Manual → Mileage_High
15	Displace_Medium, Compress_High, Power_Medium → Mileage_High
16	Compress_High, Trans_Manual → Mileage_High
17	Power_Low → Mileage_High
18	Weight_Light → Mileage_High

Using the Rule Importance Measure [11] ranking the set of rules, we obtained the following Table 3 which indicates the Rule Importance for the car data set. Core attributes from the original data set are generated by the core algorithm. The core for this data set are *make_model*, and *trans*.

4.2 Rules-as-Attributes

The new decision table $A_{14\times 20}$ as shown in Table 4 is constructed by using the 19 rules as condition attributes, and the original decision on the mileage as the decision attribute. For each rule we check whether it can be applied to the 19 records. For example, $Rule_0$,

$$\text{USACar, Displace\_Medium,Weight\_Medium} \rightarrow \text{Mileage\_Medium} \quad (5)$$

can be applied to the first record, because both the antecedent *USACar*, *Displace_Medium*, *Weight_Medium* and the consequent *Mileage_Medium* appear in the rule. Therefore, we assign $A[0, 0] = 1$. $Rule_0$ can be applied to the second record as well. We assign $A[1, 0] = 1$. However, $Rule_0$ cannot be applied to the third record, because the value for "displace" is "small" instead of "medium". Therefore $A[2, 0] = 0$. Table 4 gives the new constructed decision table for car data set. Note that we set "Mileage_Medium" to be 0, and "Mileage_High" to be 1.

Table 3. Rule Importance for the Car Data Set

No. in Table 2	Rules	Rule Importance
9	JapanCar → Mileage_High	100%
8	Trans_Auto → Mileage_Medium	100%
16	Compress_High, Trans_Manual → Mileage_High	75%
1	USACar, Compress_Medium → Mileage_Medium	75%
14	Displace_Small, Trans_Manual → Mileage_High	50%
3	Cyl_6 → Mileage_Medium	50%
0	USACar, Displace_Medium, Weight_Medium → Mileage_Medium	25%
17	Power_Low → Mileage_High	25%
2	USACar, Power_High → Mileage_Medium	25%
7	Compress_Medium, Power_High → Mileage_Medium	25%
12	Displace_Small, Compress_Medium → Mileage_High	25%
4	Door_4 → Mileage_Medium	25%
18	Weight_Light → Mileage_High	25%

Table 4. New Decision Table for the Car Data

$Rule_0$	$Rule_1$	$Rule_2$	...	$Rule_{15}$	$Rule_{16}$	$Rule_{17}$	$Rule_{18}$	Mileage
1	0	1	...	0	0	0	0	0
1	1	0	...	0	0	0	0	0
0	0	0	...	0	0	0	0	0
1	1	0	...	0	0	0	0	0
1	1	1	...	0	0	0	0	0
1	1	1	...	0	0	0	0	0
1	1	1	...	0	0	0	0	0
0	0	0	...	1	1	0	1	1
0	0	0	...	0	1	1	1	1
0	0	0	...	0	0	0	0	1
0	0	0	...	0	1	0	0	1
0	0	0	...	0	0	1	0	1
0	0	0	...	0	1	0	0	1
0	0	0	...	0	1	0	0	1

4.3 Representative Sets of Rules

There is no inconsistency in this new decision table. ROSETTA's Genetic Reducer is applied on Table 4 for multiple reduct generation. The following Table 5 lists the multiple reducts generated by Genetic Reducer. There are 6 reducts obtained for this car data set.

Since a reduct provides a representation of the original knowledge, each reduct set from Table 5 can be considered

[1]The values of support and confidence can be adjusted to control the number of rules generated. For the rest of our experiments, we set the support and confidence during rule generations for each data set to obtain a certain amount of rules.

Table 5. Multiple Reducts for Table 4

No.	Reduct
1	{Rule 9, Rule 16}
2	{Rule 0, Rule 8}
3	{Rule 1, Rule 8}
4	{Rule 9, Rule 11, Rule 14}
5	{Rule 9, Rule 14, Rule 18}
6	{Rule 9, Rule 10, Rule 14}

as one representation of the decision Table 4. In Table 5, each reduct contains one of the most important rules with a rule importance of 100% such as Rule 8 and Rule 9 both having a rule importance of 100% according to Table 3; and a less importance rules with the rule importance ranging from 25% to 75%. Each reduct set cannot completely cover the original knowledge, therefore it provides a representation of the original data. Such reduct contains representative rules generated from the original data set.

4.4 Minimum Set of Representative Rules

In the Rules-as-Attribute measure, Johnson's Reducer provides a single reduct with the minimum length. Therefore, this reducer can be used to obtain a minimum set of representative rules for a data set.

Table 6 shows the reducts generated by the Johnson's Reducer for Table 4, which contains Rule 9 and Rule 16. As discussed earlier in Table 5, there are another two reducts containing the same number of rules as the reduct returned by Johnson's Reducer. We can therefore consider either of these three a minimum set of representative rules. However, in the situation when there is only one reduct containing the minimum number of rules, such reduct is an unique representative rule set with minimum length.

Table 6. Reduct Rules for the Car Data Set

No. in Table 2	Reduct Rules	Rule Importance
9	JapanCar → Mileage_High	100%
16	Compress_High, Trans_Manual → Mileage_High	75%

Johnson's Reduct generation algorithm generates one reduct, $\{Rule_9, Rule_{16}\}$. The *Reduct Rules* for the car data set is shown in Table 6.

From Table 3 we can see that $Rule_9$ and $Rule_{16}$ have the rule importance of 100%, and 75% respectively. We also observe that, in $Rule_9$, *JapanCar* is the core attribute value, in $Rule_{16}$, *Trans_Manual* is the core attribute value.The *Reduct Rules* all contain core attributes.

5 Further Ranking by Integrating with the Rule Importance Measure

The representative sets of rules discussed in Section 4 are obtained by the Genetic Reducer from ROSETTA. The motivation behind this approach is to have a diverse interpretation for the original knowledge. This is similar to the Rule Importance Measure [11], which provides a diverse ranking of the association rules generated from an original data set. By considering the multiple reducts of a data set, the frequencies of the attributes are taking into consideration. More frequently occurred attributes in the multiple reducts are considered more important, therefore rules generated more frequently across multiple reducts are considered more important.

The Rule importance measure are designed to extract important rules among the rule sets. However, such rules are not necessary representative. Rules with a lower importance may be representative of the original data.

Based on the motivation of the Rule Importance Measure, in Table 5, rules contained in multiple reduct rule sets are considered more **representative**.

We say a rule is more representative if a rule occurs across multiple representative rule sets. For n reduct rule sets, the **rule representative degree** is defined to be the percentage of a rule occurs in multiple representative rule sets.

Given the representative sets of rules obtained by the approach introduced in Section 4, we count the number of times a rule occurs across all the representative rule sets, and divide by the total number of representative rule sets. Therefore we can obtain list of rules ranked by their rule representative degree.

Let us take Table 5 as an example. The rules as well as their representative degrees and the rule importance are listed in Table 7. Rule 9 has a representative degree of 66.67%, because among the 6 rule sets in Table 5, this rule is included in 4 representative rule set.

Table 7. Rule Representative Degree for Car Data Set

No.	Rules	Representative Degree	RIM
9	JapanCar → Mileage_High	4/6=66.67%	100%
14	Displace_Small, Trans_Manual → Mileage_High	3/6=50.00%	50%
8	Trans_Auto → Mileage_Medium	2/6=33.33%	100%
0	USACar, Displace_Medium, Weight_Medium → Mileage_Medium	1/6=16.67%	25%
1	USACar, Compress_Medium → Mileage_Medium	1/6=16.67%	75%
16	Compress_High, Trans_Manual → Mileage_High	1/6=16.67%	75%
18	Weight_Light → Mileage_High	1/6=16.67%	25%
10	Cyl_4, Displace_Medium, Compress_High → Mileage_High	1/6=16.67%	N/A
11	Cyl_4, Compress_High, Power_High → Mileage_High	1/6=16.67%	N/A

5.1 Discussions

Table 7 indicates that the Rule Representative Degree is different from the Rule Importance. The rule importance measures how important the rules are by considering whether the rule contains more important condition attributes. For example, Rule 9 has an importance of 100% because the make_model is the core attribute, which is the essential of the original knowledge. The representative degree is not designed to describe how important the rules are, but it is designed to evaluate whether a rule can represent the original knowledge. The rule representative degree is designed to evaluate how representative the rules are, because the measure is designed based on the representative rule sets. For example, Rule 9 is considered more representative than Rule 8, although they have the same high rule importance measures. We also observed Rule 10 and Rule 11 are considered as representative rules although they are not considered as important, because the attribute "cyl" is not contained by a reduct, therefore rules containing attribute "cyl" are not generated. However, such rules are indeed representative of the original knowledge. In the car data set, only focusing on extracting important rules may omit some hidden information from the original data.

This approach of extracting representative rules motivated by the Rule Importance Measure, can be used while the purpose of the rule evaluation is to extract a representation of the original knowledge. In the situation when the purpose for the rule evaluation is to obtain only important rules towards the classification task, the Rule Important Measure should be used instead of the representative rules sets.

6 Concluding Remarks and Future Work

In this paper we demonstrate three results on rule evaluation and finding representative sets of rules, as an extension of the "Rules-as-Attribute" measure. The first result provides multiple representative rule sets, which when used together can suggest a multiple interpretation perspectives of the rules. The second result discusses how to obtain the minimum length of representative rules by using the Johnson's Reducer. While integrating the "Rules-as-Attribute" and the Rule Importance Measure, we propose another measure on ranking such representative rules. These three approaches are designed for rule evaluation and rule ranking in order to help understanding the original knowledge in large data sets. Our future work is to conduct empirical experiments to explore the applicabilities of these approaches.

Acknowledgements

We gratefully acknowledge the financial support of the NSERC Canada. Thanks to Professor Yiyu Yao who provided the authors helpful suggestions.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, pages 487–499, 12–15 1994.

[2] I. Bruha. Quality of decision rules: Definitions and classification schemes for multiple rules. In *Machine Learning and Statistics.*, pages 107 – 131. John Wiley & Sons, Inc., 1997.

[3] U. Fayyad, G. Piatesky-Shapiro, and P. Smyth. From data mining to knowledge discovery: an overview. *Advances in Knowledge discovery and data mining (AAAI/MIT Press)*, pages 1–34, 1996.

[4] A.-E. Hassanien. Rough set approach for attribute reduction and rule generation: a case of patients with suspected breast cancer. *J. Am. Soc. Inf. Sci. Technol.*, 55(11):954–962, 2004.

[5] R. Hilderman and H. Hamilton. Knowledge discovery and interestingness measures: A survey. Technical Report CS-99-04, Department of Computer Science, University of Regina, 1999.

[6] X. Hu. *Knowledge Discovery in Databases: an Attribute-Oriented Rough Set Approach.* PhD thesis, University of Regina, 1995.

[7] Z. Huang and Y.-Q. Hu. Applying AI technology and rough set theory to mine association rules for supporting knowledge management. In *ICMLC*, pages 1820–1825, 2003.

[8] D. S. Johnson. Approximation algorithms for combinatorial problems. *J. Comput. Syst. Sci.*, 9(3):256–278, 1974.

[9] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo. Finding interesting rules from large sets of discovered association rules. In *(CIKM'94)*, pages 401–407, 1994.

[10] J. Li and N. Cercone. Discovering and ranking important rules. In *IEEE Granular Computing*, volume 2, pages 506–511, Beijing, China, July 2005.

[11] J. Li and N. Cercone. Introducing a rule importance measure. In *Transactions on Rough Sets V*, volume 4100, pages 167–189, 2006.

[12] A. Øhrn. *Discernibility and Rough Sets in Medicine: Tools and Applications.* PhD thesis, Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim Norway, 1999.

[13] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Norwell, MA, USA, 1992.

[14] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, and W. Ziarko. Rough sets. *Commun. ACM*, 38(11):88–95, 1995.

[15] M. S. Szczuka. Rules as attributes in classifier construction. In *RSFDGrC*, volume 1711 of *Lecture Notes in Computer Science*, pages 492–499. Springer, 1999.

[16] P.-N. Tan and V. Kumar. Interestingness measures for association patterns: A perspective. *KDD Workshop on Post-processing in Machine Learning and Data Mining, Boston, MA.*, 2000.