A New Rule Importance Measure for Association Rules

Jiye Li[1] and Nick Cercone[2]

[1] School of Computer Science, University of Waterloo
200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
j27li@uwaterloo.ca
[2] Faculty of Computer Science, Dalhousie University
6050 University Avenue, Halifax, Nova Scotia, Canada B3H 1W5
nick@cs.dal.ca

Abstract. When generating association rules often too many rules are generated, and it is difficult to determine which rules are more useful, interesting and important. We introduce a rule importance measure for association rules using rough sets to select the most appropriate rules. We use rough sets to generate reducts. Because reducts are not unique, we use ROSETTA to generate multiple reducts. Apriori association rule algorithm is then used to generate rule sets for each data set based on each reduct. Some rules are generated more frequently than the others among the total rule sets. We consider such rules as more important. We define rule importance as the frequency of an association rule among the rule sets. Rule importance is different from rule interestingness in that it does not consider the predefined knowledge on what kind of information is considered to be interesting. The experimental results show our method reduces the number of rules generated and at the same time provides a measure of how important is a rule.

Key words: Rough Sets, Association Rules, Rule Importance Measure

1 Introduction

Rough sets theory was first presented by Pawlak in the 1980's [1]. He introduced an early application of rough sets theory to knowledge discovery systems, and pointed out that 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 has focused on decision making, data analysis, discovering and characterizing the inter-data relationships, and discovery interesting patterns [2].

Although the rough sets approach is frequently used on attribute selection, little research effort has been expanding on applying this approach to association rules generation. The main problem of association rules algorithm is that there are usually too many rules generated, and it is difficult to process the large amount of rules by hand. In the data preprocessing stage, redundant attributes can be found by a rough sets approach. By removing the redundant attributes, association rules generation will be more efficient and more effective.

Klemettinen introduced the concept of rule templates [3]. Properly defined rule templates can be helpful on generating desired association rules to be used in decision making and collaborative recommender systems [4],[5].

We discuss how the rough sets theory can help with association rules generation to generate important rules. We are interested in applying these rules for making decisions. Therefore, the type of rules we are looking for are rules which have, on the consequent part, the decision attributes, or items that can be of interest for making decisions. We propose a new rule importance measure to evaluate the utilities of the association rules. The new rule importance measure can be applied in both decision making and recommender system applications.

We discuss related work on association rules algorithm, the rough sets theory on rule discovery and recommender system in Section 2. In Section 3 we show our approach to generate reduct sets, and introduce the new rule importance measure. In Section 4, we describe our experiments on an artificial data set and a sanitized geriatric care data set. Finally we summarize our contributions and discuss next step work in Section 5.

2 Related Work

2.1 Association Rules Algorithm

An association rules algorithm helps to find patterns which relate items from transactions. For example, in market basket analysis, by analyzing transaction records from the market, we could use association rules algorithm to discover different shopping behaviors. Association rules can then be used to express these kinds of behaviors, thus helping to increase the number of items sold in the market by arranging related items properly.

An association rule [6] is a rule of the form $\alpha \rightarrow \beta$, where α and β represent itemsets which do not share common items. The association rule $\alpha \rightarrow \beta$ holds in the transaction set L with $c$onfidence c, $c = \frac{|\alpha \cup \beta|}{|\alpha|}$, if c% of transactions in L that contain α also contain β. The rule $\alpha \rightarrow \beta$ has $s$upport s, $s = \frac{|\alpha \cup \beta|}{|L|}$, if s% of transactions in L contain $\alpha \cup \beta$. Here, we call α antecedent, and β consequent. Confidence gives a ratio of the number of transactions that the antecedent and the consequent appear together to the number of transactions the antecedent appears. Support measures how often the antecedent and the consequent appear together in the transaction set.

A problem of using association rules algorithm is that there are usually *too many rules generated* and it is difficult to analyze these rules. Rule interestingness measures have been proposed to reduce the number of rules generated.

2.2 Rough Sets Theory and Rule Discovery

Rough Sets was proposed to classify imprecise and incomplete information. Reduct and core are two important concepts in rough sets theory. A reduct is a subset of attributes that are sufficient to describe the decision attributes. Finding all the reduct sets for a data set is a NP-hard problem [7]. Approximation algorithms are used to obtain the reduct set [8]. All reducts contain core. Core contains the most important information of the original data set. The intersection of all the possible reducts is the core.

Hu et al. [9] introduced core generation and reduct generation algorithms based on the rough sets theory and efficient database operations.

Procedure 1 Core Generating Algorithm
Input: Decision table $T(C, D)$. Output: Core attributes.
(1) $Core \leftarrow \phi$
(2) For each attribute A
(3) If $Card(\Pi(C - A + D)) \neq Card(\Pi(C - A))$
(4) Then $Core = Core \cup$ A

where C is the set of condition attributes, and D is the set of decision attributes. $Card$ denotes the count operation, and Π denotes the projection operation.

There have been contributions on applying rough sets theory to rule discovery. Rules and decisions generated from the reduct are representative of the data set's knowledge. In [10], 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 [11] 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 classify the data with reduced attributes.

Rough sets can be used to determine whether there is redundant information in the data and whether we can find essential data needed for our applications. We expect fewer rules will be generated due to fewer attributes.

2.3 Recommender Systems

Not many research efforts are found on applying association rules algorithms for collaborative recommender systems, one of the two types of recommender systems of interest. The rule templates [3] can be appropriately defined to extract rules that match the templates in the post processing of the association rules generation. Therefore this method can increase both the efficiency and the accuracy of recommendations. In our experiment, we define rule template, and generate rules with only decision attributes on the consequent part. This type of recommendation rule can be used to make decisions.

3 Rules, Measures and Templates

3.1 Motivation

In medical diagnosis, a doctor requires a list of symptoms in order to make a diagnosis. For different diseases, there are different patient symptoms to examine.

However, there are some routine exams that the doctor must perform for all patients, such as the age of the patient, the blood pressure, the body temperature and so on. There are other symptoms that doctors may take into consideration, such as whether the patients have difficulty walking, whether the patients have bladder problems and so on. We would like to find the most important symptoms for diagnoses. We know that the symptoms that are checked more frequently are more important and essential for making diagnoses than those which are considered less frequently. However, both the symptoms that require frequent checking and the symptoms that are checked less frequently are included in the list of checkup symptoms. In this way, the doctor will make a precise diagnose based on all possible patient information.

3.2 Rule Importance

The medical diagnosis process can be considered as a decision making process. The symptoms can be considered as the condition attributes. The diagnosed diseases can be considered as the decision attributes. Since not all symptoms need to be known to make a diagnosis, the essential symptoms are considered as representative. These symptoms can be selected by a reduct generation algorithm.

All patient information can also be represented in a transaction data set, with each patient's record considered to be an item set. Association rules algorithm can be applied on this transaction data set to generate rules, which have condition attributes on the antecedent part and decision attributes on the consequent part of the rules. Rules generated from different reduct sets can contain different representative information. If only one reduct set is being considered to generate rules, other important information might be omitted. Using multiple reducts, some rules will be generated more frequently than other rules. We consider the rules that are generated more frequently more important.

We propose a new measure, *Rule Importance*, to evaluate the importance of rules. A rule is defined to be important by the following definition.

Definition 1. *If a rule is generated more frequently across different rule sets, we say this rule is more important than other rules.*

Rule importance measure is defined as follows,

Definition 2.

$$\textit{Rule Importance Measure} = \frac{\textit{Number of times a rule appears in all rule sets}}{\textit{Number of reduct sets}}$$

Suppose for a certain data set, there are 3 reducts used for rule generation. For $reduct_1$, the rule set generated is $\{a, b \rightarrow 1; a \rightarrow 0; b, c \rightarrow 1\}$; for $reduct_2$, the rule set generated is $\{b \rightarrow 1; b, c \rightarrow 1; c, d \rightarrow 0\}$; for $reduct_3$, the rule set generated is $\{a, c, d \rightarrow 1; b, c \rightarrow 1; c, d \rightarrow 0\}$. Rule $b, c \rightarrow 1$ is generated from all the 3 reducts, and its rule importance is $3/3 = 100\%$. Rule $c, d \rightarrow 0$ is generated from 2 reducts, therefore its importance is $2/3 = 66.67\%$. The rest rules are only generated once among the 3 rule sets. Their rule importance are $1/3 = 33.33\%$.

Rule importance is different from rule interestingness since it does not require predefined knowledge of what is interesting. Without considering people's interests, rule importance provides diverse choices of how important is a rule.

3.3 Specifying Rule Templates for Wanted and Subsumed Rules

Apriori association rules algorithm is used to generate rules. 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 rule templates for extracting rules we want.

$$\langle Attribute_1, Attribute_2, ..., Attribute_n \rangle \rightarrow \langle DecisionAttribute \rangle$$

This template specifies only decision attributes can be on the consequent of a rule, and $Attribute_1$, $Attribute_2$,..., $Attribute_n$ lead to a decision of $DecisionAttribute$.

We specify the rules to be removed. For example, given rule

$$\langle Attribute_1, Attribute_2 \rangle \rightarrow \langle DecisionAttribute \rangle$$

the following rules

$$\langle Attribute_1, Attribute_2, Attribute_3 \rangle \rightarrow \langle DecisionAttribute \rangle$$

$$\langle Attribute_1, Attribute_2, Attribute_6 \rangle \rightarrow \langle DecisionAttribute \rangle$$

can be removed because they are subsumed.

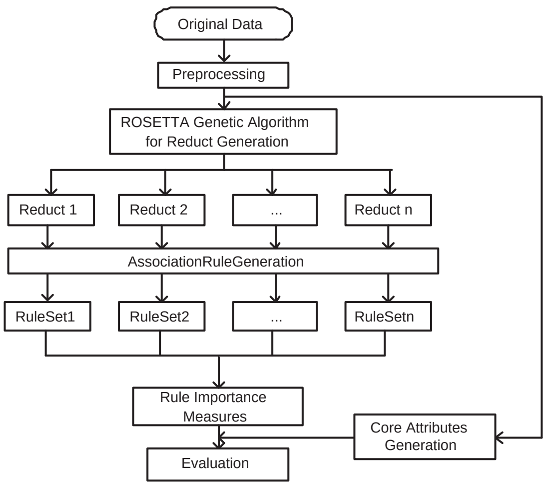

Fig. 1. Experiment Procedure

3.4 Experiment Procedure

In our experiments, we first preprocess the original data set. Then the data is imported to ROSETTA [12] for reduct generation. Association rules algorithm is applied to generate multiple rule sets for multiple reducts. Rule templates are used in the rule generation stage. The rule importance measure is used to rank these rules. Core attributes are generated from the preprocessed data set to help evaluating the rules. The experimental procedure is shown in Figure 1.

4 Experiment

We experiment on two data sets. ROSETTA GUI version 1.4.41[3] is used for reduct generation. The apriori algorithm [13] for large item sets generation and rule generation is performed on Sun Fire V880, four 900Mhz UltraSPARC III processors, with 8GB of main memory.

Table 1. 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
...	...	...	...	...	...	...	...	...
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

Table 2. Reduct sets Generated by Genetic Algorithm for Car Data Set

No.	Reduct Sets
1	{make_model, compress, power, trans}
2	{make_model, cyl, compress, trans}
3	{make_model, displace, compress, trans}
4	{make_model, cyl, door, displace, trans, weight}

4.1 Car Data Set

The first data set we experiment on is an artificial data set about cars [14], as shown in Table 1. It is used to decide the mileage of different cars. The condition attributes are *make_mode, cyl, door, displace, compress, power, trans,*

[3] ROSETTA provides approximation algorithms for reduct generation: Johnson's algorithm, Genetic algorithm and others. Johnson's algorithm returns a single reduct. Genetic algorithm returns multiple reducts. We use genetic algorithm with the option of full discernibility.

weight. *Mileage* is the decision attribute. There are 14 instances. The data set does not contain missing attribute values.

For the Car data set, the core attributes are, *make_model*, and *trans*. ROSETTA generates 4 reducts as shown in Table 2. We then generate the rule sets based on these 4 reduct sets with $support = 1\%, confidence = 100\%$, and we also rank their rule importance, as shown in Table 3.

Discussion From Table 3, the first 2 rules have an importance of 100%. This matches our experiences on cars. The auto transmission cars usually have a lower mileage than the manual cars. Japanese cars are well known for using less gas and higher mileage. The rule "Door_4 → Mileage_Medium" has a lower importance because the number of doors belonging to a car does not affect car mileage. We noticed that two rules with importance of 100% contain core attributes and only core attributes to make a decision of mileage. For the rest of the rules with importance less than 100%, the attributes on the left hand side of a rule contains non-core attributes. This observation implies that core attributes are important when evaluating the importance of the rules. Our method of generating rules with reduct sets is efficient. There are 6327 rules generated from the original data without using reducts or rule templates. 13 rules are generated using reducts and rule templates.

4.2 Experiment on a medical data set

In this experiment, a sanitized geriatric care data set is used as our test data set. This data set contains 8547 patient records with 44 symptoms and their survival status. The data set is used to determine the survival status of a patient giving all the symptoms he or she shows. We use *survival status* as the decision attribute, and the 44 symptoms of a patient as condition attributes, which includes *education level, the eyesight, the age of the patient at investigation* and so on. [4] There is no missing value in this data set. Table 4 gives selected data records of this data set.

There are 12 inconsistent data entries in the medical data set. After removing these instances, the data contains 8535 records. [5]

There are 14 core attributes generated for this data set. They are *eartroub, livealone, heart, hbp, eyetroub, hearing, sex, health, edulevel, chest, housewk, diabetes, dental, studyage*. Table 5 shows selected reduct sets among the 86 reducts generated by ROSETTA. All of these reducts contain the core attributes. For each reduct set, association rules are generated with $support = 30\%, confidence = 80\%$. [6]

Discussion There are 218 rules generated and ranked according to their rule importance as shown in Table 6. We noticed there are 8 rules having importance of 100%. All attributes contained in these 8 rules are core attributes.

[4] Refer to [15] for details about this data set.

[5] Notice from our previous experiments that core generation algorithm can not return correct core attributes when the data set contains inconsistent data entries.

[6] Note that the value of support and confidence can be adjusted to generate as many or as few rules as required.

Table 3. The rule importance for the Car data set

No.	Selected Rules	Rule Importance
1	Trans_Auto → Mileage_Medium	100%
2	JapanCar → Mileage_High	100%
3	USACar, Compress_Medium → Mileage_Medium	75%
4	Compress_High, Trans_Manual → Mileage_High	75%
5	Displace_Small, Trans_Manual → Mileage_High	50%
6	Cyl_6 → Mileage_Medium	50%
...	...	...
12	Door_4 → Mileage_Medium	25%
13	Weight_Light → Mileage_High	25%

Table 4. Geriatric Care Data Set

edulevel	eyesight	...	health	trouble	livealone	cough	hbp	heart	...	studyage	sex	livedead
0.6364	0.25	...	0.25	0.00	0.00	0.00	0.00	0.00	...	73.00	1.00	0
0.7273	0.50	...	0.25	0.50	0.00	0.00	0.00	0.00	...	70.00	2.00	0
0.9091	0.25	...	0.00	0.00	0.00	0.00	1.00	1.00	...	76.00	1.00	0
0.5455	0.25	...	0.50	0.00	1.00	1.00	0.00	0.00	...	81.00	2.00	0
0.4545	0.25	...	0.25	0.00	1.00	0.00	1.00	0.00	...	86.00	2.00	0
...	...	...	...	...	...	...	...	...	...	...	...	...

These 8 rules are more important when compared to other rules. For example, consider rule No.5 and No.11. Rule No.11 has an importance measure of 95.35%. The difference between these two rules is that rule No.5 contains attribute *Livealone, HavingDiabetes, HighBloodPressure*, and rule No. 11 contains the first 2 attributes, and instead of *HighBloodPressure*, *SeriousNerveProblem* is considered to decide whether the patient will survive. Generally high blood pressure does affect people's health condition more than nerve problem in combination with the other 2 symptoms. Rule No.11 are more important than rule No.218 because in addition to the *NerveProblem*, whether a patient is able to take medicine by himself or herself is not as fatal as whether he or she has diabetes, or lives alone without care. With the same support and confidence, $2,626,392$ rules are generated from the original medical data set without considering reduct sets or rule templates. Our method efficiently extracts important rules, and at the same time provides a ranking for important rules.

Johnson's reduct generation algorithm [12] generates one reduct with the minimum attributes. 16 rules are generated using this reduct [15]. The 8 rules with 100% importance in Table 6 are also generated. Although the reduct generated by Johnson's algorithm can provide all the 100% importance rules, the result does not cover other important rules. A doctor may be interested to know a patient is not in a good condition, if he is living alone, has diabetes and also coughs often. This information is more important than whether a patient has

Table 5. Reduct sets for the medical data set after preprocessing

No.	Reduct Sets
1	{edulevel,eyesight,hearing,shopping,housewk,health,trouble,livealone, cough,sneeze,hbp,heart,arthriti,eyetroub,eartroub,dental, chest,kidney,diabetes,feet,nerves,skin,studyage,sex}
2	{edulevel,eyesight,hearing,phoneuse,meal,housewk,health,trouble, livealone,cough,sneeze,hbp,heart,arthriti,eyetroub,eartroub,dental, chest,bladder,diabetes,feet,nerves,skin,studyage,sex}
...	...
86	{edulevel,eyesight,hearing,shopping,meal,housewk,takemed,health, trouble,livealone,cough,tired,sneeze,hbp,heart,stroke,arthriti, eyetroub,eartroub,dental,chest,stomach,kidney,bladder,diabetes, feet,fracture,studyage,sex}

Table 6. The rule importance for the medical data set

No.	Selected Rules	Rule Importance
1	SeriousChestProblem $\rightarrow$ *Dead*	100%
2	SeriousHearingProblem, HavingDiabetes $\rightarrow$ *Dead*	100%
3	SeriousEarTrouble $\rightarrow$ *Dead*	100%
4	SeriousHeartProblem $\rightarrow$ *Dead*	100%
5	Livealone, HavingDiabetes, HighBloodPressure $\rightarrow$ *Dead*	100%
...	...	...
11	Livealone, HavingDiabetes, NerveProblem $\rightarrow$ *Dead*	95.35%
...	...	...
217	SeriousHearingProblem, ProblemUsePhone$\rightarrow$ *Dead*	1.16%
218	TakeMedicineProblem, NerveProblem $\rightarrow$ *Dead*	1.16%

diabetes and loses control of the bladder. The experimental results show that considering multiple reducts gives us more diverse view of the data set, the rule importance measure provides a ranking of how important is a rule. Important rules consist of only core attributes.

5 Conclusions

We introduced a rule importance measure which provides an automatic and efficient way of ranking rules for decision making applications. We observed that all important rules contain core attributes. Therefore, the core attributes should be taken into consideration while choosing important and useful rules.

We enjoy the advantage of not specifying the criteria of selecting the reduct set. We consider as many reduct sets as possible, and we try to cover all representative subsets of the original data set.

Our rule importance measure provides an objective measure to evaluate how important is a rule. Our method provides a diverse evaluation by giving the various percentage. This measure can be combined with other human specified interestingness measures to evaluate rules.

In the future, we are interested in studying this rule importance measure on large recommender system to improve the quality of the recommendations.

References

1. Pawlak, Z.: Rough Sets. In Theoretical Aspects of Reasoning about Data. Kluwer, Netherlands, 1991.
2. Pawlak, Z., Grzymala-Busse, J., Slowinshi, R., Ziarko, W.: Rough Sets. Communications of the ACM, Vol.38, No. 11, November (1995)
3. Klemettinen, M., Mannila, H., Ronkainen, R., Toivonen, H., Verkamo, A.I.: Finding interesting rules from large sets of discovered association rules. CIKM'94,401–407
4. Li, J., Tang, B., Cercone, N.: Applying Association Rules for Interesting Recommendations Using Rule Templates. PAKDD 2004, pp. 166 – 170.
5. Lin, W., Alvarez, S., Ruiz, C.: Efficient adaptive-support association rule mining for recommender systems, Data Mining and Knowledge Discovery (2002) 6:83–105
6. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. Proc. of 20th VLDB, Santiago de Chile, Chile, Morgan Kaufmann, (1994) 487–499
7. Kryszkiewicz, M., Rybinski, H.: Finding Reducts in Composed Information Systems, Rough Sets, Fuzzy Sets Knowldege Discovery. In W.P. Ziarko (Ed.), Proceedings of the International Workshop on Rough Sets, Knowledge Discovery pp. 261 – 273. Heidelberg/Berlin: Springer-Verlag, 1994.
8. Bazan, J.,Nguyen, H.S.,Nguyen,S.H.,Synak,P.,and Wroblewski,J.: Rough set algorithms in classification problems. In Polkowski, L., Lin, T.Y., and Tsumoto, S., editors, Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems, Vol. 56 of Studies in Fuzziness and Soft Computing, pages 49-88. Physica-Verlag, Heidelberg, Germany, 2000
9. Hu, X., Lin, T., Han, J.: A New Rough Sets Model Based on Database Systems. Fundamenta Informaticae XX(2004) 1-18.
10. Huang, Z, Hu, Y.Q.: Applying AI Technology and Rough Set Theory to Mine Association Rules for Supporting Knowledge Management. Proc. of the 2nd Int'l Conference on Machine Learning and Cybernetics, Xi'an. November (2003).
11. Hassanien, A.E.: Rough Set Approach for Attribute Reduction and Rule Generation: A Case of Patients with Suspected Breast Cancer. Journal of The American Society for Information Science and Technology, 55(11):954-962. (2004)
12. Aleksander Ohrn: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, NTNU report 1999:133, IDI report 1999:14, ISBN 82-7984-014-1, 239 pages. 1999.
13. Borgelt, C.: Efficient Implementations of Apriori and Eclat. Proceedings of the FIMI'03 Workshop on Frequent Itemset Mining Implementations. CEUR Workshop Proceedings (2003) 1613-0073
14. Hu, X.: Knowledge Discovery in Databases: an Attribute-Oriented Rough Set Approach. PhD Thesis, University of Regina, 1995.
15. Li, J. and Cercone, N.: Empirical Analysis on the Geriatric Care Data Set Using Rough Sets Theory. Technical Report, School of Computer Science, University of Waterloo, 2005.