

Introducing a Rule Importance Measure

Jiye Li¹ and Nick Cercone²

¹ School of Computer Science, University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
j27li@uwaterloo.ca

² Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia B3H 1W5, Canada
nick@cs.dal.ca

Abstract. Association rule algorithms often generate an excessive number of rules, many of which are not significant. It is difficult to determine which rules are more useful, interesting and important. We introduce a rough set based Rule Importance Measure to select the most important rules. We use ROSETTA software to generate multiple reducts. Apriori association rule algorithm is then applied 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 generated across all the rule sets. Rule importance is different from either rule interestingness measures or rule quality measures because of their application tasks, the processes where the measures are applied and the contents they measure. The experimental results from an artificial data set, UCI machine learning datasets and an actual geriatric care medical data set show that our method reduces the computational cost for rule generation and provides an effective measure of how important is a rule.

Keywords: Rough sets, rule importance measure, association rules.

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 suggested 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 discovering interesting patterns [2].

Although the rough sets approach is frequently used on attribute selection, little research effort has been explored to apply this approach to association rules generation. The main problem of association rule 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 generating important association rules. We propose a new rule importance measure based on rough sets to evaluate the utilities of the association rules. Rules generated from reducts are representative rules extracted from the data set; since a reduct is not unique, rule sets generated from different reducts contain different sets of rules. However, more important rules will appear in most of the rule sets; less important rules will appear less frequently than those more important ones. The frequencies of the rules can therefore represent the importance of the rules.

To test our hypothesis, we first use ROSETTA [6] rough sets toolkit to generate multiple reducts. We then use apriori association rules generation to generate rule sets for each reduct set. 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. Some rules are generated more frequently than the others among the total rule sets. We consider such rules more important. We define the rule importance measure according to the frequency of an association rule among the rule sets. We will show by the experimental results that our method provides diverse measures of how important are the rules, and at the same time reduces the number of rules generated. This method can be applied in both decision making and recommender system applications.

Our method is among the few attempts on applying rough sets theory to association rules generation to improve the utility of an association rule. Rule importance measure is different from either the rule interestingness measures or the rule quality measures, which are the two well-known approaches on evaluating rules. The rule importance measure is different from the rule interestingness measures. Most of the rule interestingness measures are used to evaluate classification rules, and different people have different definition for “interestingness”. Rule importance measure is applied to evaluate association rules. It is a straightforward and objective measure. Rule importance measure is different from rule quality measure as well. Rule quality measure is used to evaluate the quality of an classification rule. Whereas rule importance measure is applied from the process of reduct generation to rule generation, and the rules evaluated are association rules.

In our earlier work [7], we evaluated the rule importance measures on an artificial data set and a geriatric care data set. This paper extends the experiments to include 13 data sets from UCI machine learning repository [8]. Detailed analysis on the differences between rule importance measures and rule interestingness measures, rule importance measures and rule quality measures are provided. We also conduct comparison experiments on the different effects of ranking rules between the rule importance measures and the rule interestingness measures.