Chapter 2
Finding Persistent Strong Rules: Using Classification to Improve Association Mining

Anthony Scime
The College at Brockport, State University of New York, USA

Kulathur S. Rajasethupathy
The College at Brockport, State University of New York, USA

Karthik Rajasethupathy
Cornell University, USA

Gregg R. Murray
Texas Tech University, USA

ABSTRACT

Data mining is a collection of algorithms for finding interesting and unknown patterns or rules in data. However, different algorithms can result in different rules from the same data. The process presented here exploits these differences to find particularly robust, consistent, and noteworthy rules among much larger potential rule sets. More specifically, this research focuses on using association rules and classification mining to select the persistently strong association rules. Persistently strong association rules are association rules that are verifiable by classification mining the same data set. The process for finding persistent strong rules was executed against two data sets obtained from the American National Election Studies. Analysis of the first data set resulted in one persistent strong rule and one persistent rule, while analysis of the second data set resulted in 11 persistent strong rules and 10 persistent rules. The persistent strong rule discovery process suggests these rules are the most robust, consistent, and noteworthy among the much larger potential rule sets.

INTRODUCTION

In data mining, there are a number of methodologies used to analyze data. The choice of methodology is an important consideration, which is determined by the goal of the data mining and the type of data. Different methodologies can result in different rules from the same data. Association mining is used to find patterns of data that show conditions where sets of attribute-value pairs occur frequently in the data set. It is often used to determine the relationships among transaction data. Classification mining, on the other hand, is used to find models of data for categorizing

DOI: 10.4018/978-1-4666-2455-9.ch002

instances (e.g., objects, events, or persons). It is typically used for predicting future events from historical data (Han & Kamber, 2001). Because association and classification methodologies or algorithms process data in very different ways, they yield different sets of rules. The process presented here exploits these differences to find particularly robust, consistent, and noteworthy rules among much larger potential rule sets. More specifically, this research focuses on using association rules and classification mining to select the persistently strong association rules, which are association rules that are verifiable by classification mining the same data set.

Decision tree classification algorithms construct models by looking at past performance of input attributes with respect to an outcome class. The model is constructed inductively from records with known values for the outcome class. The input attribute with the strongest association with the outcome class is selected from the training data set using a divide-and-conquer strategy that is driven by an evaluation criterion. The training data are divided based on the values of this attribute, thereby creating subsets of the data. Each subset is evaluated independently to select the attribute with the next strongest association to the outcome class along the subset's edge. The process of dividing the data and selecting the next attribute, which is the one with the next strongest association with the outcome class at that point, continues until a leaf node is constructed (Quinlan, 1993). The rules derived from the decision tree provides insight into how the outcome class's value is, in fact, dependent on the input attributes. A complete decision tree provides for all possible combinations of the input attributes and their allowable values reaching a single, allowable outcome class.

Classification decision trees have a root node. The attribute of this root node is the most predictive attribute of a record's class and is present in the premise of every rule produced by classification. The presence of the root node attribute in the premise of all the rules is a limitation of decision tree classification mining. There may be a domain theory where the root attribute is not relevant and/or another attribute is theoretically relevant and useful for predicting the value of the class attribute. Association mining may find rules in such instances. Further, the class attribute appears in the consequent of every classification rule. This class attribute is the goal of the data mining. It is the attribute that ultimately determines if a record supports a domain theory under consideration.

Association mining evaluates data for relationships among attributes in the data set (Agrawal, Imieliński, & Swami, 1993). The association rule mining algorithm Apriori finds itemsets within the data set at user-specified minimum support levels. An itemset is a collection of attribute-value pairs (items) that occurs in the data set. The support of an itemset is the percent of records that contain all the items in the itemset. The largest supported itemsets are converted into rules where each item implies and is implied by every other item in the itemset.

Given the limitations on decision tree classification rules, association mining may be applied to the classification attributes and data set to find other rules that address the domain. Unlike classification, association mining considers all the attribute combinations in the records. Also unlike classification, it does not have a goal of predicting the value of a specific attribute. As a result, association mining often produces a large number of rules (Bagui, Just, & Bagui, 2008), many of which may not be relevant. The strength of rules is an important consideration in association mining. Generally, a rule's strength is measured by its confidence level. Strong association mined rules are those that meet the minimum confidence level set by the domain expert (Han & Kamber, 2001). The higher the confidence level the stronger the rule and the more likely the rule will be successfully applied to new data.

Measures of interestingness are either subjective or objective (Tan, Steinbach, & Kumar, 2006). Subjective interestingness is based on the