Using the Apriori Algorithm to Improve Rough Sets Results

María C. Fernández-Baizán[1], Ernestina Menasalvas Ruiz[1]
José M. Peña Sánchez[1], Juan Francisco Martínez Sarrías[1], Socorro Millán[2]
{cfbaizan, emenasalvas, jpena}@fi.upm.es, juanfran@pegaso.ls.fi.upm.es,
millan@borabora.edu.co, ⋆

[1] Departamento de Lenguajes y Sistemas Informáticos e Ingeniería del Software, Facultad de Informática, U.P.M., Campus de Montegancedo, Madrid
[2] Universidad del Valle, Cali. Colombia

Abstract. Ever since Data Mining first appeared, a considerable amount of algorithms, methods and techniques have been developed. As a result of research, most of these algorithms have proved to be more effective and efficient. For solving problems different algorithms are often compared. However, algorithms that use different approaches are not very often applied jointly to obtain better results. An approach based on the joining of a predictive model (rough sets) together with a link analysis model (the Apriori algorithm) is presented in this paper.

Keywords: Data Mining models joining, Rough Sets, Association rules.

1 Introduction

The Rough Set methodology provides a way to generate decision rules. Some condition values may be unnecesary in a decision rule. Thus it is always desirable to reduce the amount of information required to describe a concept. A reduced number of condition attributes results in a set of rules with higher support. On the other hand, this kind of rules are easier to understand. The concept of *reduct* is used when there is a need for reducing the number of attributes, but this is a computational expensive process.
One way to construct a simpler model computed from data, easier to understand and with more predictive power, is to create a set of simplified rules [11]. A simplified rule (also refered to as minimal rule or kernel rule) is one in which the number of conditions in its antecedent is minimal. Thus, when dealing with decision rules, some condition values can be unnecessary and can be dropped to generate a simplified rule preserving essential information. In [7] an approach to simplify decision tables is presented. Such an approach consists of three steps: 1) Computation of reducts of condition attributes; 2) Elimination of duplicate

⋆ This work has been partially supported by UPM under project "Design of a Data Warehouse to be integrated with a Data Mining system"

W. Ziarko and Y. Yao (Eds.): RSCTC 2000, LNAI 2005, pp. 291−295, 2001.

rows; 3) Elimination of superfluous values of attributes. This approach to the problem is not very useful because both the computation of reducts and the superfluous equivalence classes are NP-Hard.
Many algorithms and methods have been proposed and developed to generate minimal decision rules, some based on inductive learning [6], [8], [3] and some other based on Rough Sets theory [11], [10], [12], [2], [9]. Rough sets theory provides a sound basis for the extraction of qualitative knowledge (dependecies) from very large relational databases.
Shan [11] proposes and develops a systematic method for computing all minimal rules, called maximally general rules, based on decision matrices.
Based on rough sets and boolean reasoning, Bazan [2] proposes a method to generate decision rules using dynamic reducts, stable reducts of a given decision table that appear frequently in random samples of a decision table.
Skowron [12] proposes a method that when applied over consistent decisions tables make it possible to obtain minimal decision rules. Based on the relative discernibility matrix notion.
An incremental learning algorithm for computing a set of all minimal decision rules based on the decision matrix method is proposed in [10].
On the other hand, different algorithms have been proposed to calculate reducts based on Rough Sets Theory. However, finding the minimal reduct is a NP-hard problem [13], so its computational complexity makes application in large databases imposible. In [4] a heuristic algorithm to calculate a reduct of the decision table is proposed. The algorithm is based on two matrices that are calculated using information from the Positive Region. In Chen and Lin [5] a modified notion of reducts is introduced.
In this paper we propose to execute prior to Rough Set methodology the Apriori algorithm in order to discover strong dependencies that can, in general, be useful to reduce the original set of attributes.
Observe that this approach wil not generate a minimal reduct. Nevertheless it is important to note that as a side effect it is possible to obtain strong rules to classify the concept that will be refined using the rough set methodology.
The rest of the paper is organized as follows.
Section 2: Rough Sets and association rules introduction.
Section 3: describes the new approach.
Section 4: results discussion and future work.

2 Preliminaries

2.1 Rough Sets Theory

The original Rough Set model was proposed by Pawlak [7]. This model is concerned with the analysis of deterministic data dependencies. According to Ziarko [14] Rough Set Theory is the discovery representation and analysis of data regularities. In this model, the objects are classified into indiscernibility classes based on pairs (attribute, values).
The following are the basic concepts of the rough set model.

Let OB be a non-empty set called the universe, and let IND be an equivalence relation over the universe OB, called an indiscernibility relation which represents a classification of the universe into classes of objects which are indiscernible or identical in terms of the knowledge provided by the given attributes. The main notion in Rough Sets Theory is that of the approximation space which is formally defined as $A = (OB, IND)$.
Equivalence classes of the relation are also called elementary sets. Any finite union of elementary sets is refered to as a definable set. Let's take $X \subseteq OB$ which represents a concept. It is not always the case that X can be defined exactly as the union of some elementary sets. That is why two new sets are defined: $\underline{Apr}(X) = \{o \in OB/[o] \subseteq X\}$ will be called the **lower approximation** $\overline{Apr}(X) = \{o \in OB/[o] \cap X \neq \emptyset\}$ will be called the **upper approximation**. Any set defined in terms of its lower and upper approximations is called a **rough set**.

2.2 Information Systems

The main computational effort in the process of data analysis in rough set theory is associated with the determination of attribute relationships in information systems. An Information System is a quadruple: $S = (OB, AT, V, f)$ where:

- OB is a set of objects
- AT is a set of attributes
- $V = \bigcup V_a$ being V_a the values of attribute a
- $f : OB \times AT \to V$

2.3 Decision Tables

Formally, a decision table S is a quadruple $S = (OB, C, D, V, f)$. All the concepts are defined similarly to those of information systems; the only difference is that the set of attributes has been divided into two sets, C and D, which are conditions and decision respectively.
Let P be a non empty subset of $C \cup D$, and let x, y be members of OB. x, y are indiscernible by P in S if and only if $f(x, p) = f(y, p)$ for all $p \in P$. Thus P defines a partition on OB. This partition is called a classification of OB generated by P. Then for any subset P of $C \cup D$, we can define an approximation space, and for any $X \subset OB$ the lower approximation of X in S and the upper approximation of X in S will be denoted as $\underline{P}(X)$ and $\overline{P}(X)$, respectively.

2.4 Association Rules

The purpose of association discovery is to find items that imply the presence of other items. An association rule is formally described as follows:
Let $I = \{i_1, i_2 \dots, i_n\}$ be a set of literals called items.
Let D be a set of transactions, each transaction $T \subset I$.
An association rule is an implication of the form $X \to Y$ where $X \subset I$ and

$Y \subset I$ and $X \cap Y = \emptyset$. The rule $X \rightarrow Y$ holds in the transaction set D with confidence c if $c\%$ of transactions in D that contain X also contain Y. The rule $X \rightarrow Y$ holds in the transaction set D with support s if $s\%$ of transactions in D contain $X \cup Y$.
Given a set of transaction D the problem of mining association rules is to generate all association rules that have support and confidence greater that the user-specified minimum support (*minsup*) and minimum confidence (*minconf*). In order to derive the association rules two steps are required: 1) Find the large itemsets for a given *minsup*; 2) Compute rules for a given *minconf* based on the itemsets obtained before.

3 The Cooperative Algorithm

Algorithm input:

- T the input decision data table. Note that the set of attributes AT will be divided in condition (eventually antecedent) and decision (described attribute)
- $minconf$ the minimun confidence for rules
- n maximun number of condition allowed in the antecendent of the rules

We will assume that input table (T) is one that is either discrete or has been discretized in a pre-processing stage. We will also assume that the input table has been binarized. Formally expressed:
Let AT be a set of attributes, let $A \in AT$ be an attribute. Let $V = \{a_1, a_2, \ldots, a_n\}$ be the set of values of A.
The binarization of A will yield n attributes $A_1, A_2, \ldots, A_n$ such that for any $o \in OB$ if $f(o, A) = a_i$ then $f(o, A_j) = 1$ if $i = j$ $f(o, A_j) = 0$ if $i \neq j$
The process that will be performed is as follows:

1. Execute the Apriori algorithm being T the input data table. Internally the algorithm will calculate the best *minsupport* and k (size of large itemsets) depending on the nature of T. We will call this new set of rules T_{Assoc}.
2. Delete all those rules in T_{Assoc} in which the decision attribute occurs as part of their antecedent.
3. If $\forall i \exists k$ such that $A_k \rightarrow D_i$ then A_k is a superfluos attribute so that it can be removed from AT. Remove also such rules from T_{Assoc}.
4. Analysis of the rules in the new T_{Assoc}. This set of rules contains two kind of rules that we have called *strong classification rules* and *meta-rules*. The former is composed by all those rules in T_{Assoc} in which the consequent is some D_i with confidence $\geq minconf$. The latter is a set containing rules that will allow us to reduce the number of condition attributes.
   - Include *strong classification rules* in the *output data mining rule set*
   - Those associations rules in *meta-rules* that only contain condition attributes have to be taken into account as they highlight dependencies among condition attributes. We will call this set of rules T_{red}
5. Reduce AT taking into account the rules in T_{red}
6. Execute the Positive Region algorithm to obtain a set of rules that will be included in *output data mining rule set*.

4 Conclusions

This approach provides a sound basis for the definition of a new cooperative algorithm to obtain comprehensible rules, while avoiding the computational complexity of *classical* predictive methods.

5 Acknowledgments

This work has been partially supported by the Universidad Politécnica de Madrid under grant A9913 for the project entitled "Design of a Data Warehouse to be integrated with a Data Mining system" and under grant AL200-1005-2.38 for project of collaboration with the University of Valle (Colombia).

References

1. Agrawal R., Imielinski T., Swami A. Mining association rules between sets of Item in large Databases, Proceedings of ACM SIGMOD, pages 207-216, May 1993.
2. J. Bazan Dynamic Reducts and statistical inference IPMU'96 1147-1151
3. Grzymala-Busse J. W. LERS: A system for learning from examples based on Rough Sets In Slowinski, R. (ed.) Intelligent Decision Support: Handbook of applicactions and advances of Rough Set theory. Kluwer Netherlands pp. 3-18
4. Fernandez C.,Menasalvas E., J.Pena J., Castano M. A new approach for efficient calculus of reducts in large databases Joint Conference of Information Sciences, 1997, pp.340-343
5. Chen R., Lin T. Y. Finding Reducts in Very Large DataBases Joint Conference of Information Sciences, 1997, pp.350-352
6. Michalski R., Carbonell J. Mitchell T. M. Machine Learning: An Artificial Intelligence Approach, vol. 1, Tioga Publishing, Palo Alto CA. 1986
7. Z. Pawlak, Rough Sets - Theoretical Aspects of Reasoning about Data, Kluwer, 1991.
8. J.R. Quinlan *Induction of decision Trees* Machine Learning 1: 81-106, 1986, Kluwer Academic Publisher
9. Shan N. W. Ziarko, An incremental learning algorithm for constructing decision rules Rough Sets, Fuzzy Sets and Knowledge Discovery. W. Ziarko ed. Springer Verlag 1994 pp. 326-334
10. Shan N. W. Ziarko, Data-Base acquisition an incremental modification of classification rules pp. 357-368 Computational Intelligence, vol. 11 n. 2, 1995.pp. 357-369.
11. Shan N., Rule Discovery from Data using decision matrices Master degree Thesis. University of Regina. 1995
12. Skowron A., Extracting laws from decision tables: A Rough Set approach Computational Intelligence, vol. 11 n. 2, 1995.pp. 371-387
13. Skowron A., The discernibility matrices and functions in Information Systems Decision Support by Experience, R. Slowinski (Ed.) Kluwer Academic Publishers, 1992
14. W. Ziarko The discovery, analysis, and representation of data dependencies in databases KDD-93 pp. 195-209
15. W. Ziarko Rough Sets and Knowledge Discovery: An Overview In Rough Sets, Fuzzy Sets and knowledge Discovery. W. Ziarko (ed.). Springer-Verlag 1994 pp.11-15