AN APPLICATION OF TEMPLATE ANALYSIS AND REDUCTS FOR ASSOCIATION RULE GENERATION

Nguyen Hung Son

Institute of Mathematics, Warsaw University
ul. Banacha 2, 02097, Warsaw, Poland
Tel. +48 +22 6583449; Fax: +48 +22 6583448;
email: son@mimuw.edu.pl

Abstract. Association rule generation is one of the biggest challenges in Data Mining. In this paper we present some novel methods solving this problem. Our approaches are based on template analysis methods and Rough set methods. We show that problem of searching for optimal association rules from a given template is equivalent to the problem of searching for α-reducts in a corresponding decision table. Hence, association rule generation module can be equipped by easy way in every Rough set based system.

1 Introduction

Many problems in Data Mining can be solved by applying the Rough Set theory (see [4, 6, 11]). Rough Sets theory is based on dealing the discernibility between objects, which are described by finite number of features. In [6, 7], we can convince oneself that Rough Set theory offers many interesting methods for many important Data Mining tasks like: Rule induction, Concept description (granularity), Pattern extraction, Discretization,...

Association rules (see [1]) are one of the important tools for data analysis, in particular in Basket Data Analysis. Association rules describe the relationship between attributes. The example of association rules is as follows:

More than 90% of customers, who bought *article A* and *article B*,
also bought *article C* and *article D* and *article E*...

Many association rule extraction methods has been proposed by (see [1, 10]). Any extraction method consists of two main steps i.e. *template extraction* (or Large item set extraction) and *searching for rules from templates.* In this paper we present the Rough set based methods for Association rule extraction. We show the equivalence between Reduct finding Problem in Rough Set theory and the problem of searching for association rules from template.

2 Basic notions

An *information system* [4] is a pair $\mathbb{A} = (U, A)$, where U is a non-empty, finite set called the *universe* and A is a non-empty, finite set of *attributes*, i.e. $a : U \to V_a$ for $a \in A$, where V_a is called *the value set of a*. Elements of U are called *objects.*

Any information system $\mathbb{A} = (U, A)$ and a non-empty set $B \subseteq A$ define *a B-information function* by $Inf_B(x) = \{(a, a(x)) : a \in B \text{ for } x \in U\}$. The set $\{Inf_A(x) : x \in U\}$ is called the $A-$*information set* and denoted by $INF(\mathbb{A})$.

Any information system of the form $\mathbb{A} = (U, A \cup \{d\})$ is called *decision table* where $d \notin A$ is called *decision* and the elements of A are called *conditions.* Let $V_d = \{1, \dots, r(d)\}$. The decision d determines the partition $\{C_1, ..., C_{r(d)}\}$ of the universe U, where $C_k = \{x \in U : d(x) = k\}$ for $1 \leq k \leq r(d)$. The set C_k is called the $k-th$ *decision class of* $\mathbb{A}$.

2.1 Rough set preliminaries

With any subset of attributes $B \subseteq A$, an equivalence relation called the *B-indiscernibility relation* [4], denoted by $IND(B)$, is defined by

$$IND(B) = \{(x, y) \in U \times U : \forall_{a \in B} (a(x) = a(y))\}$$

Objects x, y satisfying relation $IND(B)$ are indiscernible by attributes from B. By $[x]_{IND(B)}$ we denote the equivalence class of $IND(B)$ defined by x. A minimal subset B of A such that $IND(A) = IND(B)$ is called a *reduct* of $\mathbb{A}$.

If $\mathbb{A} = (U, A)$ is an information system, $B \subseteq A$ is a set of attributes and $X \subseteq U$ is a set of objects then the sets

$$\underline{B}X = \left\{x \in U : [x]_{IND(B)} \subseteq X\right\} \text{ and } \overline{B}X = \left\{x \in U : [x]_{IND(B)} \cap X \neq \emptyset\right\}$$

are called *B-lower* and *B-upper approximation* of X in $\mathbb{A}$, respectively.

If $\mathbb{A} = (U, A \cup \{d\})$ is a decision table and $B \subseteq A$ then we define a function $\partial_B : U \to 2^{\{1,..,r(d)\}}$, called the *generalized decision in* $\mathbb{A}$, by

$$\partial_B(x) = \{i : \exists_{x' \in U} \left[(x' IND(B) x) \wedge (d(x') = i)\right]\} = d\left([x]_{IND(B)}\right)$$

A decision table $\mathbb{A}$ is called *consistent (deterministic)* if $card(\partial_A(x)) = 1$ for any $x \in U$, otherwise $\mathbb{A}$ is *inconsistent (non-deterministic).*

The set of attributes $B \subseteq A$ is called "*relative reduct*" or simply *reduct* of decision table $\mathbb{A}$ if and only if

1. $\partial_B(x) = \partial_A(x)$ for all object $x \in U$.
2. any proper subset of B does not satisfy the previous condition.

i.e. B is minimal subset (with respect to the inclusion relation $\subseteq$) of attributes satisfying the property $\forall_{x \in U} \partial_B(x) = \partial_A(x)$.

There are two problems related to the notion of "reducts", which have been intensively explored in rough set theory by many researchers (see [7, 4, 11]). The first problem is related to looking for "shortest reducts" (i.e. reducts with minimal cardinality). The second problem is related to looking for all reducts. It has been shown that both problems are NP-hard. Some heuristics has been proposed for those problems. Here we present the approach based on Boolean reasoning proposed in [7].

2.2 Boolean reasoning approach

Many problems in Rough set theory (e.g. reduct finding, rule extraction, discretization,..) has been successively solve by Boolean reasoning approach. This simple methods is based on encoding the investigated optimization problem π by corresponding Boolean function f_π in such a way that any prime implicant of f_π states a solution of π. We illustrate this approach by the reduct searching problem (see [8]).

Given a decision table $\mathbb{A} = (U, A \cup \{d\})$, where $U = \{u_1, u_2, ..., u_n\}$, $A = \{a_1, ..., a_k\}$. By discernibility matrix of decision table $\mathbb{A}$ we denote the $(n \times n)$ matrix

$$\mathbf{M}(\mathbb{A}) = [C_{i,j}]_{ij=1}^{n}$$

such that $C_{i,j}$ is the set of attributes discerning x_i and x_j. Formally:

$$C_{i,j} = \begin{cases} \{a_m \in A : a_m(x_i) \neq a_m(x_j)\} & \text{if } d(x_i) \neq d(x_j) \\ \emptyset & \text{otherwise.} \end{cases}$$

One can also define the *discernibility function* $f_{\mathbb{A}}$ as a Boolean function:

$$f_{\mathbb{A}}(a_1^*, ..., a_k^*) = \bigwedge_{i,j} \left(\bigvee_{a_m \in C_{i,j}} a_m^* \right)$$

where $a_1^*, ..., a_k^*$ are Boolean variables corresponding to attributes $a_1, ..., a_k$.

One can show that every prime implicant of $f_{\mathbb{A}}(a_1^*, ..., a_k^*)$ corresponds exactly to one reduct in $\mathbb{A}$. One can see that the set $B \subset A$ is reduct if B has nonempty intersection with any nonempty set $C_{i,j}$ i.e.

$$B \text{ is reduct in } \mathbb{A} \qquad \text{iff} \qquad \forall_{i,j}(C_{i,j} = \emptyset) \vee (B \cap C_{i,j} \neq \emptyset)$$

Moreover, in some applications (see [5]), instead of reducts we prefer to use their approximations called α-reducts, where $\alpha \in [0, 1]$ is a real parameter. The set of attributes $B \subset A$ is called α-reduct if B has nonempty intersection with at least $\alpha \cdot 100\%$ number of nonempty sets $C_{i,j}$ i.e.

$$B \text{ is } \alpha\text{-reduct in } \mathbb{A} \qquad \text{iff} \qquad \frac{|\{C_{i,j} : B \cap C_{i,j} \neq \emptyset\}|}{|\{C_{i,j} : C_{i,j} \neq \emptyset\}|} \geq \alpha$$

One can show that for a given α, the problems of searching for shortest α-reducts and for all α-reducts are also NP-hard.

3 Templates as patterns in data

Given an information table $\mathbb{A} = (U, A)$. By *descriptors* (or simple descriptors) we mean the terms of form $(a = v)$, where $a \in A$ is an attribute and $v \in V_a$ is a value in the domain of a (see Hoa...). The notion of descriptor can be generalized by using terms of form $(a \in S)$, where $S \subseteq V_a$ is a set of values. By *template* we mean the conjunction of descriptors:

$$\mathbf{T} = D_1 \wedge D_2 \wedge ... \wedge D_m$$

where $D_1, ...D_m$ are either simple or generalized descriptors. We denote by $length(\mathbf{T})$ the number of descriptors being in $\mathbf{T}$.

For the given template with length m:

$$\mathbf{T} = (a_{i_1} = v_1) \wedge ... \wedge (a_{i_m} = v_m)$$

the object $u \in U$ is said to be satisfy the template $\mathbf{T}$ if and only if $\forall_j a_{i_j}(u) = v_j$. By this way, the template $\mathbf{T}$ describes the set of objects having the common property: "*their values on attributes* $a_{j_1}, ..., a_{j_m}$ are equal to $v_1, ..., v_m$, respectively. In this sense one can use templates to describe the regularity in data, i.e. *patterns - in Data Mining* or *granules - in Soft Computing.*

The support of $\mathbf{T}$ is defined by

$$support(\mathbf{T}) = |\{u \in U : u \text{ satisfies } \mathbf{T}\}|$$

¿From description point of view we prefer the long templates with large support. We considers the following quality functions which can be used to compare templates.

In previous papers (see [2]) we have shown that some optimization problems related to optimal template generation is either NP-hard or are still open problems. We consider the following ones:

1. **Optimal Template Support (OTS) Problem:**
   Instance: Information system $\mathbb{A} = (A, U)$, and positive integer L.
   Question: Find a template $\mathbf{T}$ with the length L and the maximal support.
2. **Optimal Template Quality (OTQ) Problem:**
   Instance: An information system $\mathbb{A} = (U, A)$,
   Question: Find a template for $\mathbb{A}$ with optimal quality.

In [2] we have shown the following theorems

Theorem 1 *Given an information system $\mathbb{A} = (A, U)$ and positive integer L. The optimization searching problem for a template T (if any) of length L and the maximal support is NP-hard.*

The large templates can be found quite efficiently by *Aprori* and *AproriTid* algorithms (see [1, 9]). The number of other methods for large template generation has been proposed in [2].

4 Association rules

Association rules and their generations can be defined by many ways (see [1]). Here, according to the presented notation, association rules can be defined as implications of the form

$$\mathbf{P} \Rightarrow \mathbf{Q}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are different simple templates, i.e. the formulas of the form

$$(a_{i_1} = v_{i_1}) \wedge \ldots \wedge (a_{i_k} = v_{i_k}) \Rightarrow (a_{j_1} = v_{j_1}) \wedge \ldots \wedge (a_{j_l} = v_{j_l}) \tag{1}$$

The presented form can be called *generalized association rules*, because association rules are originally defined by formulas $\mathbf{P} \Rightarrow \mathbf{Q}$ where $\mathbf{P}$ and $\mathbf{Q}$ are the sets of items (i.e. goods or articles in stock market) (see [1]) e.g.

$$\{A, B\} \Rightarrow \{C, D, E\}.$$

One can see that this form can be obtained from 1 by replacing values on descriptors by 1 i.e.:

$$(A = 1) \wedge (B = 1) \Rightarrow (C = 1) \wedge (D = 1) \wedge (E = 1).$$

Usually, for a given information table $\mathbb{A}$, the quality of the association rule $\mathcal{R} = \mathbf{P} \Rightarrow \mathbf{Q}$ can be evaluated by two measures called *support and confidence* with respect to $\mathbb{A}$. The support of the rule $\mathcal{R}$ is defined by the number of objects from $\mathbb{A}$ satisfying the condition $(\mathbf{P} \wedge \mathbf{Q})$ i.e.

$$support(\mathcal{R}) = support(\mathbf{P} \wedge \mathbf{Q})$$

The second measure – confidence of $\mathcal{R}$ – is the ratio between the support of $(\mathbf{P} \wedge \mathbf{Q})$ and the support of $\mathbf{P}$ i.e.

$$confidence(\mathcal{R}) = \frac{support(\mathbf{P} \wedge \mathbf{Q})}{support(\mathbf{P})}$$

Many efforts has been investigated for solving the typical task related to associated rules which is formulated as follows:

FOR A GIVEN INFORMATION TABLE $\mathbb{A}$, AN INTEGER s, AND A REAL NUMBER $c \in [0,1]$, FIND AS MUCH AS POSSIBLE ASSOCIATION RULES $\mathcal{R} = \mathbf{P} \Rightarrow \mathbf{Q}$ SUCH THAT $support(\mathcal{R}) \geq s$ AND $confidence(\mathcal{R}) \geq c$

All existing association rule generation methods (see [1, 10]) consists of two main steps:

1. Generate as much as possible templates

$$\mathbf{T} = D_1 \wedge D_2 ... \wedge D_k$$

   such that $support(\mathbf{T}) \geq s$ and $support(\mathbf{T} \wedge D) < s$ for any descriptor D (i.e. maximal templates among those which are supported by more than s objects).
2. For any template $\mathbf{T}$, search for a partition $\mathbf{T} = \mathbf{P} \wedge \mathbf{Q}$ such that:
   (a) $support(\mathbf{P}) < \frac{support(\mathbf{T})}{c}$
   (b) $\mathbf{P}$ is the smallest template satisfying the previous condition

In this paper we show that the second module can be solve by rough set methods.

5 From templates to optimal association rules

Let us assume that the template $\mathbf{T}$, which is supported by as least s objects, has been found by one of the algorithms presented in the previous section. We assume that $\mathbf{T}$ consists of m descriptors i.e.

$$\mathbf{T} = D_1 \wedge D_2 \wedge \ldots \wedge D_m$$

where D_i (for $i = 1, .., m$) is a descriptor of form $(a_i = v_i)$ for some $a_i \in A$ and $v_i \in V_{a_i}$. Any decomposition of $\mathbf{T}$ into two parts $\mathbf{T} = \mathbf{P} \wedge \mathbf{Q}$ define a association rule $\mathcal{R} = (\mathbf{P} \Rightarrow \mathbf{Q})$. The confidence of rule $\mathcal{R}$ is measured by

$$confidence(\mathcal{R}) = \frac{support(\mathbf{T})}{support(\mathbf{P})}$$

For the given confidence threshold $c \in (0;1)$ the decomposition $\mathbf{T} = \mathbf{P} \wedge \mathbf{Q}$ is called c-irreducible if

1. $confidence(\mathbf{P} \Rightarrow \mathbf{Q}) \geq c$.
2. For any decomposition $\mathbf{T} = \mathbf{P}' \wedge \mathbf{Q}'$ such that $\mathbf{P}'$ is a sub-template of $\mathbf{P}$, we have $confidence(\mathbf{P}' \Rightarrow \mathbf{Q}') < c$.

In this section we show the following theorem:

Theorem 2 *For any fixed real number $c \in [0;1]$, the problem of searching for shortest association rule for a given table $\mathbb{A}$ from the template $\mathbf{T}$ with confidence limited by c (Optimal c-Association Rules Problem) is NP-hard.*

For solving the presented problem, we show that the problem of searching for optimal association rules from the given template is equivalent to the problem of searching for local α-reducts for a decision table, which is well known problem in Rough set theory.

We construct the new decision table $\mathbb{A}|_{\mathbf{T}} = (U, A|_{\mathbf{T}} \cup d)$ from the original information table $\mathbb{A}$ and the template $\mathbf{T}$ as follows:

– $A|_{\mathbf{T}} = \{a_{D_1}, a_{D_2}, ..., a_{D_m}\}$ is a set of attributes corresponding to the descriptors of the template $\mathbf{T}$

$$a_{D_i}(u) = \begin{cases} 1 \text{ if the object } u \text{ satisfies } D_i, \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

– the decision attribute d determines if the given object satisfies template $\mathbf{T}$ i.e.

$$d(u) = \begin{cases} 1 \text{ if the object } u \text{ satisfies } \mathbf{T}, \\ 0 \text{ otherwise.} \end{cases} \tag{3}$$

The following theorems describe the relationship between association rules problem and reduct searching problem.

Theorem 3 *For the given information table $\mathbb{A} = (U, A)$ and the template $\mathbf{T}$, the set of descriptors $\mathbf{P}$ is reduct in $\mathbb{A}|_{\mathbf{T}}$ if and only if the rule*

$$\bigwedge_{D_i \in \mathbf{P}} D_i \Rightarrow \bigwedge_{D_j \notin \mathbf{P}} D_j$$

is 100%-association rule from $\mathbf{T}$.

Proof. Any set of descriptors $\mathbf{P}$ is reduct in the decision table $\mathbb{A}|_{\mathbf{T}}$ if and only if every object u with decision 0 is discerned from objects with decision 1 by one of the descriptors from $\mathbf{P}$ (i.e. there is at least one 0 in the information vector $inf_{\mathbf{P}}(u)$). Thus u does not satisfy the template $\bigwedge_{D_i \in \mathbf{P}} D_i$. Hence

$$support(\bigwedge_{D_i \in \mathbf{P}} D_i) = support(\mathbf{T})$$

The last equalization means that

$$\bigwedge_{D_i \in \mathbf{P}} D_i \Rightarrow \bigwedge_{D_j \notin \mathbf{P}} D_j$$

is 100%-confidence association rule for table $\mathbb{A}$. □

Analogously, one can show the following

Theorem 4 *For the given information table $\mathbb{A} = (U, A)$ and the template $\mathbf{T}$, the set of descriptors $\mathbf{P}$ the rule*

$$\bigwedge_{D_i \in \mathbf{P}} D_i \Rightarrow \bigwedge_{D_j \notin \mathbf{P}} D_j$$

is c-irreducible association rule from $\mathbf{T}$ if and only if it is α-reduct of $\mathbb{A}|_{\mathbf{T}}$, where

$$\alpha = 1 - \frac{\frac{1}{c} - 1}{\frac{n}{s} - 1}$$

and n is the total number of objects from U and $s = support(\mathbf{T})$. In particular, the problem of searching for optimal association rules can be solved by α-reduct finding problem.

Proof. Assume that $support(\bigwedge_{D_i \in \mathbf{P}} D_i) = s + e$, where $s = support(\mathbf{T})$. Then we have

$$confidence(\bigwedge_{D_i \in \mathbf{P}} D_i \Rightarrow \bigwedge_{D_j \notin \mathbf{P}} D_j) = \frac{s}{s+e} \geq c \tag{4}$$

This condition is equivalent to

$$e \leq \left(\frac{1}{c} - 1\right) s$$

Hence one can evaluate the discernibility degree of $\mathbf{P}$ by

$$disc\_degree(\mathbf{P}) = \frac{e}{n-s} \leq \frac{\left(\frac{1}{c} - 1\right) s}{n-s} = \frac{\frac{1}{c} - 1}{\frac{n}{s} - 1} = 1 - \alpha$$

Thus

$$\alpha = 1 - \frac{\frac{1}{c} - 1}{\frac{n}{s} - 1}$$

□

Searching for minimal α-reducts is well known problem in Rough Sets theory. One can show, that the problem of searching for the all α-reducts as well as the problem of searching for shortest α-reducts is NP-hard. Many great efforts has been involved to solve those problems. In the next papers we present the Rough set based algorithms of association rule generation for large data table using SQL queries.

6 The Example

The following example illustrates the main idea of our method. Let us consider the following information table $\mathbb{A}$ with 18 objects and 9 attributes.

$\mathbb{A}$	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
u_1	0	1	1	1	80	2	2	2	3
u_2	0	1	2	1	81	0	aa	1	aa
u_3	0	2	2	1	82	0	aa	1	aa
u_4	0	1	2	1	80	0	aa	1	aa
u_5	1	1	2	2	81	1	aa	1	aa
u_6	0	2	1	2	81	1	aa	1	aa
u_7	1	2	1	2	83	1	aa	1	aa
u_8	0	2	2	1	81	0	aa	1	aa
u_9	0	1	2	1	82	0	aa	1	aa
u_{10}	0	3	2	1	84	0	aa	1	aa
u_{11}	0	1	3	1	80	0	aa	2	aa
u_{12}	0	2	2	2	82	0	aa	2	aa
u_{13}	0	2	2	1	81	0	aa	1	aa
u_{14}	0	3	2	2	81	2	aa	2	aa
u_{15}	0	4	2	1	82	0	aa	1	aa
u_{16}	0	3	2	1	83	0	aa	1	aa
u_{17}	0	1	2	1	84	0	aa	1	aa
u_{18}	1	2	2	1	82	0	aa	2	aa

The template **T**

T	0	*	2	1	*	0	*	1	*

$\mathbb{A}\vert_{\mathbf{T}}$	D_1 $a_1 = 0$	D_2 $a_3 = 2$	D_3 $a_4 = 1$	D_4 $a_6 = 0$	D_5 $a_8 = 1$	d
u_1	1	0	1	0	0	
u_2	1	1	1	1	1	1
u_3	1	1	1	1	1	1
u_4	1	1	1	1	1	1
u_5	0	1	0	0	1	
u_6	1	0	0	0	1	
u_7	0	0	0	0	1	
u_8	1	1	1	1	1	1
u_9	1	1	1	1	1	1
u_{10}	1	1	1	1	1	1
u_{11}	1	0	1	1	0	
u_{12}	1	0	0	1	0	
u_{13}	1	1	1	1	1	1
u_{14}	1	1	0	0	0	
u_{15}	1	1	1	1	1	1
u_{16}	1	1	1	1	1	1
u_{17}	1	1	1	1	1	1
u_{18}	0	1	1	1	0	

Table 1. The example of information table $\mathbb{A}$ and template **T** support by 10 objects; the new decision table $\mathbb{A}|_{\mathbf{T}}$ constructed from $\mathbb{A}$ and template **T**.

Assume that the template

$$\mathbf{T} = (a_1 = 0) \wedge (a_3 = 2) \wedge (a_4 = 1) \wedge (a_6 = 0) \wedge (a_8 = 1)$$

has been extracted from the information table $\mathbb{A}$. One can see that $support(\mathbf{T}) = 10$ and $length(\mathbf{T}) = 5$. The new constructed decision table $\mathcal{A}|_{\mathbf{T}}$ is presented in Table 1.

The discernibility function for decision table $\mathbb{A}|_{\mathbf{T}}$ can be explained as follows

$$\begin{aligned} f(D_1, D_2, D_3, D_4, D_5) = & (D_2 \vee D_4 \vee D_5) \wedge (D_1 \vee D_3 \vee D_4) \wedge (D_2 \vee D_3 \vee D_4) \\ & \wedge (D_1 \vee D_2 \vee D_3 \vee D_4) \wedge (D_1 \vee D_3 \vee D_5) \\ & \wedge (D_2 \vee D_3 \vee D_5) \wedge (D_3 \vee D_4 \vee D_5) \wedge (D_1 \vee D_5) \end{aligned}$$

After simplification the condition presented in Table 1 we obtain six reducts for the decision table $\mathbb{A}|_{\mathbf{T}}$.

$$\begin{aligned} f(D_1, D_2, D_3, D_4, D_5) = & (D_3 \wedge D_5) \vee (D_4 \wedge D_5) \vee (D_1 \wedge D_2 \wedge D_3) \vee \\ & (D_1 \wedge D_2 \wedge D_4) \vee (D_1 \wedge D_2 \wedge D_5) \vee (D_1 \wedge D_3 \wedge D_4) \end{aligned}$$

Thus, we have found from template **T** six association rules with (100%)-confidence.

For $c = 90\%$, we would like to find α-reducts for the decision table $\mathbb{A}|_{\mathbf{T}}$, where

$$\alpha = 1 - \frac{\frac{1}{c} - 1}{\frac{n}{s} - 1} = 0.86$$

$\mathbb{M}(\mathbb{A}\vert_{\mathbf{T}})$	u_2, u_3, u_4, u_8, u_9 $u_{10}, u_{13}, u_{15}, u_{16}, u_{17}$
u_1	$D_2 \vee D_4 \vee D_5$
u_5	$D_1 \vee D_3 \vee D_4$
u_6	$D_2 \vee D_3 \vee D_4$
u_7	$D_1 \vee D_2 \vee D_3 \vee D_4$
u_{11}	$D_1 \vee D_3 \vee D_5$
u_{12}	$D_2 \vee D_3 \vee D_5$
u_{14}	$D_3 \vee D_4 \vee D_5$
u_{18}	$D_1 \vee D_5$

$\Rightarrow$

100% confidence rules
$D_3 \wedge D_5 \Rightarrow D_1 \wedge D_2 \wedge D_4$
$D_4 \wedge D_5 \Rightarrow D_1 \wedge D_2 \wedge D_3$
$D_1 \wedge D_2 \wedge D_3 \Rightarrow D_4 \wedge D_5$
$D_1 \wedge D_2 \wedge D_4 \Rightarrow D_3 \wedge D_5$
$D_1 \wedge D_2 \wedge D_5 \Rightarrow D_3 \wedge D_4$
$D_1 \wedge D_3 \wedge D_4 \Rightarrow D_2 \wedge D_5$

$\Rightarrow$

90% confidence rules
$D_1 \wedge D_2 \Rightarrow D_3 \wedge D_4 \wedge D_5$
$D_1 \wedge D_3 \Rightarrow D_3 \wedge D_4 \wedge D_5$
$D_1 \wedge D_4 \Rightarrow D_2 \wedge D_3 \wedge D_5$
$D_1 \wedge D_5 \Rightarrow D_2 \wedge D_3 \wedge D_4$
$D_2 \wedge D_3 \Rightarrow D_1 \wedge D_4 \wedge D_5$
$D_2 \wedge D_5 \Rightarrow D_1 \wedge D_3 \wedge D_4$
$D_3 \wedge D_4 \Rightarrow D_1 \wedge D_2 \wedge D_5$

Table 2. Simplified version of discernibility matrix $\mathbb{M}(\mathbb{A}|_{\mathbf{T}})$. Association rules with (100%)-confidence. Association rules with at least (90%)-confidence.

Hence we would like to search for a set of descriptors that covers at least

$$\lceil (n-s)(\alpha) \rceil = \lceil 8 \cdot 0.86 \rceil = 7$$

elements of discernibility matrix $\mathbb{M}(\mathbb{A}|_{\mathbf{T}})$. One can see that the following sets of descriptors:

$$\{D_1, D_2\}, \{D_1, D_3\}, \{D_1, D_4\}, \{D_1, D_5\}, \{D_2, D_3\}, \{D_2, D_5\}, \{D_3, D_4\}$$

have nonempty intersection with exactly 7 members of the discernibility matrix $\mathbb{M}(\mathbb{A}|_{\mathbf{T}})$. In Table 2 we present all association rules achieved from those sets.

7 Conclusions

We present the equivalence between optimal association rule generation problem and reduct finding problem in Rough Set theory. This result states that one can apply the well known rough set methods, in particular the approximate algorithms for reduct problems, to solve the corresponding association rule problem. In the next papers we will describe the realization of those methods for large data bases by using SQL queries.

AcknowledgementThe author acknowledges the help of the Polish-Japanese Institute of Information Technology

References

1. Agrawal R., Mannila H., Srikant R., Toivonen H., Verkamo A.I., 1996. Fast discovery of assocation rules. In V.M. Fayad, G.Piatetsky Shapiro, P. Smyth, R. Uthurusamy (eds): *Advanced in Knowledge Discovery and Data Mining*, AAAI/MIT Press, pp.307-328.
2. Nguyen S. Hoa, A. Skowron, P. Synak, 1998. Discovery of data pattern with applications to Decomposition and classification problems. In L. Polkowski, A. Skowron (eds.): Rough Sets in Knowledge Discovery **2**. Physica-Verlag, Heidelberg 1998, pp. 55–97.
3. Nguyen H.Son, Skowron A., 1997. Boolean reasoning for feature extraction problems. In: Z.W. Raś and A.Skowron (Eds.): Proceedings of Tenth International Symposium on Foundation of Intelligent Systems, ISMIS'97, LNAI 1325, Springer Verlag, Oct. 1997, NC, USA,pp. 117-126.
4. Pawlak Z., 1991. Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht.
5. Skowron A., 1995. Synthesis of adaptive decision systems from experimental data. In A. Aamodt, J. Komorowski (eds), Proc. of the Fifth Scandinavian Conference on Artificial Intelligence (SCAI'95), May 1995, Trondheim, Norway, IOS Press, Amsterdam, 220–238.
6. Komorowski J., Pawlak Z., Polkowski L., Skowron A., 1998. Rough Set: A Tutorial. In S.K. Pal and A. Skowron(eds): *Rough fuzzy hibridization: A new trend in decision–making*, Springer-Verlag, Singapore, pp.3–98.
7. Skowron A., Rauszer C., 1992. The discernibility matrices and functions in information systems. In: R. Słowiński (ed.): Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory, pp.331 - 362, Kluwer, Dordrecht.
8. I. Wegener. *The complexity of boolean functions.* Wiley, Stuttgart, 1987.
9. Mohammed Javeed Zaki, Srinivasan Parthasarathy, Mitsunori Ogihara, Wei Li, 1997. New Parallel Algorithms for Fast Discovery of Association Rules. In Data Mining and Knowledge Discovery : An International Journal, special issue on Scalable High-Performance Computing for KDD, Vol. 1, No. 4, December 1997, pp 343-373.

10. Mohammed Javeed Zaki, Mitsunori Ogihara, 1998. Theoretical Foundations of Association Rules. In: 3rd SIGMOD'98 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD), Seattle, WA, June 1998, pp.
11. Ziarko W., 1998. Rough Sets as a Methodology for Data Mining. In L. Polkowski, A. Skowron (eds.): Rough Sets in Knowledge Discovery **1**. Physica-Verlag, Heidelberg 1998, pp. 554–576.