A Rough Set Approach
to Decision Rules Generation

by

Zdzisław Pawłak¹, Andrzej Skowron²

ICS Research Report 23/93

Institute of Computer Science

Nowowiejska 15/19, 00-665 Warsaw, Poland
Abstract. In the paper we investigate the problem of optimal decision rules generation by applying the rough set and evidence approaches. We assume that the reader is familiar with some basic concepts of the rough set theory [4,5] and evidence theory [7]. Decision rules have the following form: \( \tau \rightarrow \tau' \) where \( \tau, \tau' \) are boolean combinations of descriptors built from some conditions and decision approximating the expert decision [9], respectively. The decision rules are generated with some certainty coefficients expressed by the basic functions of evidence theory (and rough membership functions [6]) computable from a given decision table. These coefficients can be used in decision making. The method of rules generation is based on a construction of some boolean functions from modified discernibility matrices [10]. We construct decision rules in the optimal form with respect to the number of descriptors occurring in any disjunct on the left hand side of rules.

1. Introduction

In this section we recall some basic definitions of the rough set theory [4,5].

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

A decision table is any information system of the form \( \mathbf{A} = (U, A \cup \{d\}) \), where \( d \in A \) is a distinguished attribute called decision. The elements of \( A \) are called conditions.

By \( V \) is denoted the set \( V = \bigcup_{a \in A} V_a \cup V_d \cup \mathcal{P}(V_d) \), where \( \mathcal{P}(V_d) \) is the powerset of \( V_d \). The set \( V_d \) will be called the frame of
discernment of A and denoted by \( \theta_A \). To simplify notation we assume \( \theta_A \) = \( \{1,...,r(d)\} \). The decision \( d \) determines a partition \( \text{CLASS}_A(d) = \{X_1,...,X_r(d)\} \) of the universe \( U \), where \( X_k = \{x \in U : d(x) = k\} \) for \( 1 \leq k \leq r(d) \).

Let \( A = (U,A) \) be an information system. With every subset of attributes \( B \subseteq A \), an equivalence relation, denoted by \( \text{IND}_A(B) \) (or \( \text{IND}(B) \)) called the \( B \)-indiscernibility relation, is associated and defined as follows:

\[
\text{IND}(B) = \{(x,x') \in U^2 : \text{ for every } a \in B, a(x) = a(x')\}
\]

Objects \( x,x' \) satisfying relation \( \text{IND}(B) \) are indiscernible by attributes from \( B \).

The sets \( \{x \in U : [x]_B \subseteq X\} \) and \( \{x \in U : [x]_B \cap X = \emptyset\} \) are called \( B \)-lower and \( B \)-upper approximation of \( X \subseteq U \) in \( A \), and they are denoted by \( \underline{B}X \) and \( \overline{B}X \), respectively. The \( B \)-boundary of \( X \) is the set \( BN_B(X) = \overline{B}X - \underline{B}X \). We write \( BN_A(X) \) instead of \( BN_A(X) \) when \( B = A \). A set \( X \subseteq U \) is definable by \( B \) if it is the union of some classes of the indiscernibility relation \( \text{IND}(B) \), otherwise it is roughly definable by \( B \).

An information function \( \text{Inf}_A : U \rightarrow P(A \cup V_A) \), for a given information system \( A = (U,A) \), is defined by \( \text{Inf}_A(x) = \{(a,a(x)) : a \in A\} \). Any subset of \( \text{Inf}_A(x) \) is called an information (vector) about \( x \) in \( A \). The set \( \{\text{Inf}_A(x) : x \in U\} \) is called the \( A \)-information set and it is denoted by \( \text{INF}(A) \).

Let \( A \) be an information system with \( n \) objects and \( m \) attributes. By \( M(A) \) [10] we denote an \( n \times n \) matrix \( (c_{ij}) \), called the discernibility matrix of \( A \) such that

\[
c_{ij} = \{a \in A : a(x_i) = a(x_j)\} \text{ for } i,j = 1,...,n.
\]

A discernibility function \( f_A \) for an information system \( A \) is a boolean function of \( m \) boolean variables \( \bar{a}_1,...,\bar{a}_m \) corresponding to the attributes \( a_1,...,a_m \), respectively and defined by
\[ f_\bar{a}(\bar{a}_1, ..., \bar{a}_m) = \bigwedge (\bigvee \bar{c}_{ij}: 1 \leq j \leq n, c_{ij} \neq \emptyset), \text{ where } \bar{c}_{ij} = \{a: a \in c_{ij}\}. \]

We will write \(a\) instead of \(a\) because always from the context it will be clear if we consider attributes or corresponding to them boolean variables.

It can be shown [10] that the set of all prime implicants of \(f_\bar{a}\) determines the set of all reducts of \(A\). Here we apply an analogous method for the decision rules generation by a generalization of the discernibility matrix notion. The modified discernibility matrix \(MG(A)\) is a subset of \(P(A) \times \{1, ..., n\} \times \{1, ..., n\}\) computable from \(M(A)\). \(f_{MG(A)}\) denotes a boolean function constructed from \(MG(A)\) in an analogous way as \(f_\bar{a}\) from \(M(A)\).

By \(PRIME_{MG(A)}\) we denote the set of all prime implicants of \(f_{MG(A)}\).

Different forms of decision rules are obtained by construction of some appropriate forms of \(MG(A)\).

2. Rough set theory and evidence theory

The classification problems are central for the rough set approach [5] as well as for the evidence theoretic approach [7]. In the evidence theory [7] information about sets creating a partition is embedded directly in some numerical functions. In the case of rough set approach information about classified sets and objects is included in a decision table. It is possible to compute the basic functions of evidence theory from a given decision table [9] by applying the rough set approach. We use the standard notation \(\Theta, m, Bel\) and \(Pl\) to denote the frame of discernment, basic probability assignment (bpa), belief and plausibility functions, respectively. For a given decision table one can define a new objects classification approximating the classification given by the decision attribute. The approximation is constructed on the basis of conditions in decision table.
Proposition 1. [9] Let $A = (U, A \cup \{d\})$ be a decision table and let $APP\_CLASS_A(d) = \{AX_1, \ldots, AX_r(d)\} \cup \{Bd_A(\theta) : \theta \subseteq A \text{ and } |\theta| > 1\}$, where $Bd_A(\theta) = \bigcap_{i \in \theta} BN_A(X_i) \cap \bigcap_{i \in \theta} \neg BN_A(X_i)$. The family of all non-empty sets from $APP\_CLASS_A(d)$ creates a partition of the universe $U$. Moreover, the following equality holds:

$$\bigcup_{i \in \theta} AX_i \cup \bigcup_{\Delta \subseteq \theta, |\Delta| > 1} Bd_A(\Lambda) = A \bigcup_{i \in \theta} X_i \text{ for } \theta \subseteq A \text{ with } |\theta| > 1.$$

\[ \square \]

The classification (partition) of the universe $U$ described in Proposition 1 is called the standard classification of $U$ approximating in $A$ the classification $CLASS_A(d)$ (given by an expert). We have a clear interpretation of this new classification. Any object from the universe $U$ of $A$ is represented by an information $Inf_A(x) \in INF(A)$. An object $x$ can be classified exactly on the basis of information $Inf_A(x)$ only when the category (i.e. the equivalence class of the indiscernibility relation $IND_A(A)$) corresponding to $x$ is included in $X_i$ for some $i$. Otherwise, that category is included in a boundary region of the form $Bd_A(\theta)$, for some $\theta$. Then an object $x$ from $U$ represented by the information $Inf_A(x)$ can be classified to the boundary region of $\{X_i : i \in \theta\}$ (i.e. it can be classified into $\bigcup_{i \in \theta} X_i$ but there is no enough information about that object $x$ either to decide in which of the sets $X_i$ it is or to eliminate some hypotheses $x \notin X_i$, where $i \in \theta$).

There is a natural correspondence between subsets of $\theta_A$ and elements of $APP\_CLASS_A(d)$, which can be expressed by the following function:

$$F_A(\theta) = \begin{cases} AX_i & \text{if } \theta = \{i\} \text{ for some } i (1 \leq i \leq r(d)) \\ \theta & \text{if } \theta = \emptyset \\ Bd_A(\theta) & \text{if } |\theta| > 1 \end{cases}$$
Now we can define the injection
\[ \delta_A : U \rightarrow \mathcal{P}(\Theta_A) \]
as follows: \( \delta_A(x) \) is the unique subset \( \theta \) of \( \Theta_A \) such that \( x \in F_A(\theta) \).

In other words
\[ \delta_A(x) = \{k: \exists x' \ xIND(A)x' \text{ and } d(x')=k\} \text{ for } x \in U. \]
The function \( \delta_A \) can be treated as a new decision attribute (defined by conditions in \( A \)) approximating the decision \( d \).

The function \( m_A : \mathcal{P}(\Theta_A) \rightarrow \mathbb{R}_+ \), called the standard basic probability assignment (defined by \( A \)) is defined by
\[ m_A(\theta) = \frac{|F_A(\theta)|}{|U|}, \text{ for any } \theta \in \Theta_A. \]

Proposition 2 [9] The function \( m_A \) defined above is a basic probability assignment (in the sense of evidence theory).

Theorem 3. [9] For an arbitrary \( \theta \in \Theta_A \) the following equality holds:
\[ Bel_A(\theta) = \frac{|A \cup \bigcup_{i \in \theta} X_i|}{|U|}. \]
The belief function \( Bel_A \) is Bayesian iff all sets from \( \text{CLASS}_A(d) \) are definable by the set \( A \) of conditions. In particular the belief function \( Bel_{A'} \), where \( A'=(U, A \cup \{\delta_A\}) \), is Bayesian.

Corollary 4. For an arbitrary \( \theta \in \Theta_A \) the following equality holds:
\[ PL_A(\theta) = \frac{|\overline{A} \cup \bigcup_{i \in \theta} X_i|}{|U|}. \]

3. Decision rules

Now we are going to define decision rules.
The atomic formulas over \( BSA \cup \{d\} \) and \( V \) are expressions of
the form $a=v$ (also denoted by $(a,v)$ or $a_v$), called descriptors over $B$, where $a \in B$ and $v \in V_a$. The set $F(B,V)$ of formulas over $B$ is the least set containing all atomic formulas over $B$ and closed with respect to the classical propositional connectives $\lor$ (disjunction), $\land$ (conjunction) and $\neg$ (negation).

Let $\tau \in F(B,V)$ (where $B \cup \{d\}$) then by $\tau_A$ we denote the meaning of $\tau$ in the decision table $A$, i.e. the set of all objects in $U$ with property $\tau$, defined inductively as follows:

1. If $\tau$ is of the form $a=v$ then $\tau_A = \{x \in U: a(x) = v\};$
2. $(\tau_1 \land \tau_2)_A = \tau_1 \land \tau_2 A$; $(\tau_1 \lor \tau_2)_A = \tau_1 \lor \tau_2 A$; $(\neg \tau)_A = U - \tau_A$.

The set $F(A,V)$ is called the set of condition formulas in $A$ and is denoted by $C_A$. The set $F(\{\delta_A\},V)$ is called the set of decision formulas in $A^* = (U,A \cup \{\delta_A\})$ and is denoted by $D_A$.

If $\tau = (a_1=v_1), \ldots, (a_r=v_r)$ then by $\tau_U$ we denote the conjunction $(a_1=v_1) \land \ldots \land (a_r=v_r)$. By $A(\tau)$ we denote the set of attributes occurring in the formula $\tau$ or corresponding to boolean variables occurring in the formula $\tau$.

A decision rule for $A$ is any expression of the form

$$\tau \rightarrow \tau'$$
where $\tau \in C_A$ and $\tau' \in D_A$.

The decision rule $\tau \rightarrow \tau'$ for $A$ is true in $A$ iff $\tau_A \subseteq \tau_A^*$. If $\tau_A = \tau_A^*$ then we say that the rule $\tau \rightarrow \tau'$ is $A$-exact.

An $A$-exact rule $\tau \rightarrow \tau'$ is $A$-optimal iff

(i) every disjunct in $\tau$ has the minimal number of descriptors, i.e. if $\tau''$ is obtained from any disjunct from $\tau$ by an elimination (from that disjunct) of some (but not all) descriptors then $\tau'' \rightarrow \tau'$ is not true in $A$;

and (ii) if the rule $\tau_U \rightarrow \tau'$, where $u$ is an information about
an object in $\mathbf{A}$ is true in $\mathbf{A}$ then there exists a subset $u' \subseteq u$ such that $\tau_{u'} \rightarrow \tau'$ is also true in $\mathbf{A}$ and $\tau_{u'}$ is a disjunct in $\tau$.

We construct the optimal rules by applying the mentioned above method based on the modified discernibility matrices.

We consider three binary relations $R_i (i=1,2,3)$ in $P(\Theta_\mathbf{A})$ defined by:

\begin{align*}
\Delta_1 R_1 \Delta_2 & \text{ iff } \Delta_1 = \Delta_2; \\
\Delta_1 R_2 \Delta_2 & \text{ iff } \Delta_1 \subseteq \Delta_2; \\
\Delta_1 R_3 \Delta_2 & \text{ iff } \Delta_1 \cap \Delta_2 = \emptyset \text{ for any } \Delta_1, \Delta_2 \subseteq \Theta_\mathbf{A}.
\end{align*}

We define for a given decision table $\mathbf{A}$, $\Delta \subseteq \Theta_\mathbf{A}$ and $x_1 \in U$ with a property $\partial_\mathbf{A}(x_1) R_\Delta$ the discernibility matrices:

\[ \text{MG}_k(\mathbf{A}, \Delta, x_1) = \{c_{ij}^{-}(d) : \gamma(\partial_\mathbf{A}(x_j) R_\Delta) \} \text{ for } k=1,2,3. \]

Decision rules for $k=1,2,3$ have the following form:

\[ \tau \rightarrow \sqrt{\{\partial_\mathbf{A} = \theta: \theta R_\Delta\}} \text{, where } \Delta \subseteq \Theta_\mathbf{A} \text{ and } \tau \in \mathbf{C}_\mathbf{A} \]

Our main result can be formulated as follows:

Theorem 5. Let $\mathbf{A} = (U, A \cup \{d\})$ be a decision table, $\Delta \subseteq \Theta_\mathbf{A}$ and $k \in \{1,2,3\}$. Then we have:

1. $\sqrt{\{\tau_U: \exists x, t, u \ (\partial_\mathbf{A}(x) R_\Delta \&} \ \\
\text{tePRIME}_k \text{MG}_k(\mathbf{A}, \Delta, x) \&} \ \\
\text{u=INF}(t, A^*, x)) \rightarrow \sqrt{\{\partial_\mathbf{A} = \theta: \theta R_\Delta\}} \]

where $\text{INF}(t, A^*, x) = \{(a, a(x)) : a \in A(t)\}$, is an $\mathbf{A}$-optimal decision rule.

2. The value of $\frac{|\{x \in U: \partial_\mathbf{A}(x) R_\Delta\}|}{|U|}$ is equal to $m_\mathbf{A}(\Delta)$, $\text{Bel}_\mathbf{A}(\Delta)$ and $\text{PL}_\mathbf{A}(\Delta)$ for $k=1,2,3$, respectively.
The set \( \text{INF}_k(t, \mathbb{A}^k, x) = \{(a, a(x)): a \in \mathbb{A}(t)\} \) is called the 
trace (in \( \mathbb{A} \)) of the prime implicant \( t \) on \( x \in U \).

For \( k=1 \) the rule describes a minimal information on the 
basis of which we are able to classify objects into the union 
\( \bigcup_{i \in \Delta} X_i \) without possibility to eliminate any hypothesis \( X_i \) for 
\( i \in \Delta \). For \( k=2 \) (\( k=3 \)) the rule describes a minimal information on 
the basis of which we are able to classify objects as certainly 
(possibly) belonging into the union \( \bigcup_{i \in \Delta} X_i \), i.e. as belonging to 
\( A \bigcup_{i \in \Delta} X_i \) \( \bigcup_{i \in \Delta} X_i \). The value \( m_\mathbb{A}(\Delta) \) describes a "chance" that an 
object chosen from \( \mathbb{A} \) is classified by \( \partial_\mathbb{A} \) into \( \Delta \). The value 
\( \text{Bel}_{\mathbb{A}}(\Delta) \) describes a "chance" that an object chosen from \( \mathbb{A} \) is 
classified with certainty into the union \( \bigcup_{i \in \Delta} X_i \) (on the basis of 
knowledge determined by \( \mathbb{A} \)). The value \( \text{Pl}_{\mathbb{A}}(\Delta) \) describes a 
"chance" that an object chosen from \( \mathbb{A} \) is classified as possibly 
belonging to \( \bigcup_{i \in \Delta} X_i \) (on the basis of knowledge determined by \( \mathbb{A} \)).

The numerical coefficients computed as the values of the 
basic probability assignments, belief or plausibility functions 
can be applied in the decision making. The rough membership 
functions [6] can be also used with the same purpose.

4. Examples

Let us consider a decision table \( \mathbb{A}=(U, \mathbb{A} \cup \{d\}) \) presented in 
Table 1, where we have \( U=\{x_1, \ldots, x_8\} \), \( \mathbb{A}=\{a, b, c\} \) and \( d \) is the 
decision. The values of attributes are presented in Table 1. In 
the examples we write \( a \beta \) and \( a+\beta \) instead of \( a \land \beta \) and \( a \lor \beta \),
respectively.
In Table 2 the values of the "new" attribute $\delta_A$ are presented.

In Table 3 we present the discernibility matrix for the information system $(U, \mathcal{A})$ (without the decision attribute $d$). The rows and columns in the table are labelled by values of the function $\delta_A$ for $\mathcal{A}=(U, \mathcal{A}\cup\{d\})$.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>$\delta_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>${0}$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>${1}$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>${0,2}$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>${1}$</td>
</tr>
<tr>
<td>$x_5$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>${0,1}$</td>
</tr>
<tr>
<td>$x_6$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>${0,1}$</td>
</tr>
<tr>
<td>$x_7$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>${0,2}$</td>
</tr>
<tr>
<td>$x_8$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>${2}$</td>
</tr>
</tbody>
</table>

Table 2
We will construct the $\mathbf{A}$-optimal rule of the form:

$$\alpha \rightarrow \delta_{\mathbf{A}}(1)$$

In order to construct $\alpha$ we consider the following discernibility matrices: $MG_1(A,\{1\},x_2)$ and $MG_1(A,\{1\},x_4)$.

$MG_1(A,\{1\},x_2)$ has the following non-empty entries:

- $b; bc; ac; ac; bc; abc$.

<table>
<thead>
<tr>
<th></th>
<th>{0}</th>
<th>{1}</th>
<th>{0,2}</th>
<th>{1}</th>
<th>{0,1}</th>
<th>{0,2}</th>
<th>{2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$\emptyset$</td>
<td>$b$</td>
<td>$c$</td>
<td>$ab$</td>
<td>$abc$</td>
<td>$abc$</td>
<td>$c$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$b$</td>
<td>$\emptyset$</td>
<td>$bc$</td>
<td>$a$</td>
<td>$ac$</td>
<td>$ac$</td>
<td>$bc$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>$c$</td>
<td>$bc$</td>
<td>$\emptyset$</td>
<td>$abc$</td>
<td>$ab$</td>
<td>$ab$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>$ab$</td>
<td>$a$</td>
<td>$abc$</td>
<td>$\emptyset$</td>
<td>$c$</td>
<td>$c$</td>
<td>$abc$</td>
</tr>
<tr>
<td>$x_5$</td>
<td>$abc$</td>
<td>$ac$</td>
<td>$ab$</td>
<td>$c$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$abc$</td>
</tr>
<tr>
<td>$x_6$</td>
<td>$abc$</td>
<td>$ac$</td>
<td>$ab$</td>
<td>$c$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$ab$</td>
</tr>
<tr>
<td>$x_7$</td>
<td>$c$</td>
<td>$bc$</td>
<td>$\emptyset$</td>
<td>$abc$</td>
<td>$ab$</td>
<td>$ab$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>$x_8$</td>
<td>$ac$</td>
<td>$abc$</td>
<td>$a$</td>
<td>$bc$</td>
<td>$b$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
</tbody>
</table>

Table 3

They are in Table 3 in the row corresponding to $x_2$ and in the columns corresponding to objects $x$ satisfying $\delta_{\mathbf{A}}(x) \neq \delta_{\mathbf{A}}(x_2)$. The discernibility function corresponding to that matrix has, after the simplification, the following form:

$$b(a+c) = ba + bc.$$  

We have $PRIME_{MG_1}(A,\{1\},x_2) = (ba, bc)$. Hence we obtain the following disjuncts of $\alpha$ taking in Table 1 traces of the prime implicants $ba$, $bc$ in the row corresponding to $x_2$:

$$b_1a_0, b_1c_0.$$  

$MG_1(A,\{1\},x_4)$ has the following non-empty entries:

- $ab; abc; c; c; abc; bc$.

They are in Table 3 in the row corresponding to $x_2$ and in the
columns corresponding to objects \( x \) satisfying \( \delta_A(x) \neq \delta_A(x) \). The discernibility function corresponding to that matrix has, after the simplification, the following form:

\[
c(a+b) = ca + cb.
\]

We have \( PRIME_{MG}(A,\{1\},x_4) = \{ca, cb\} \). Hence we obtain the following disjuncts of \( \alpha \) taking in Table 1 traces of the prime implicants \( ca, cb \) in the row corresponding to \( x_4 \):

\[
c_0 a_1, c_0 b_1.
\]

Finally we obtain the following \( A \)-optimal decision rule:

\[
a_0 b_1 + b_1 c_0 + a_1 c_0 \rightarrow \delta_A = \{1\}
\]

Any disjunct in \( \alpha \) has the minimal number of descriptors: if any descriptor is deleted from any disjunct then obtained rule is no longer true in \( A \). We have also \( m_A(\{1\}) = 2/8 \).

Now we will construct the \( A \) - optimal rule of the form:

\[
\beta \rightarrow \delta_A = \{0\}
\]

In order to construct \( \beta \) we consider the following discernibility matrix: \( MG_1(A,\{0\},x_1) \).

\( MG_1(A,\{0\},x_1) \) has the following non-empty entries:

\[
b; c; ab; abc; abc; c; ac.
\]

They are in Table 3 in the row corresponding to \( x_1 \) and in the columns corresponding to objects \( x \) satisfying \( \delta_A(x) \neq \delta_A(x_2) \). The discernibility function corresponding to that matrix has, after the simplification, the following form:

\[
bc.
\]

We have \( PRIME_{MG}(A,\{0\},x_1) = \{bc\} \). Hence we obtain only one disjunct of \( \beta \) taking in Table 1 the trace of the prime implicant \( bc \) in the row corresponding to \( x_1 \):

\[
b_0 c_0.
\]
We obtain the following A-optimal decision rule:

\[ b_0c_0 \rightarrow \delta_A = \{0\} \]

The disjunct \( b_0c_0 \) has the minimal number of descriptors: if any descriptor is deleted from it then the obtained rule is no longer true in \( A \). We have also \( m_A(\{0\}) = 1/8 \).

Let us construct now the \( A \)-optimal rule of the form:

\[ \gamma \rightarrow \delta_A = \{0,1\} \]

In order to construct \( \gamma \) we consider the following discernibility matrices: \( MG_1(A,\{0,1\},x_5) \) and \( MG_1(A,\{0,1\},x_6) \). \( MG_1(A,\{0,1\},x_5) \) has the following non-empty entries:

\[ abc; ac; ab; c; ab; b \]

They are in Table 3 in the row corresponding to \( x_5 \) and in the columns corresponding to objects \( x \) satisfying \( \delta_A(x) = \delta_A(x_5) \). The discernibility function corresponding to that matrix has, after the simplification, the following form:

\[ bc. \]

We have \( PRIME_{MG_1}(A,\{0,1\},x_5) = \{bc\} \). Hence we obtain the following disjunct of \( \gamma \) taking in Table 1 the trace of the prime implicant \( bc \) in the row corresponding to \( x_5 \):

\[ b_1c_1. \]

Since \( MG_1(A,\{0,1\},x_6) \) has the same non-empty entries as \( MG_1(A,\{0,1\},x_6) \) we obtain the following \( A \)-optimal decision rule:

\[ b_1c_1 \rightarrow \delta_A = \{0,1\} \]

The disjunct \( b_1c_1 \) has the minimal number of descriptors: if any descriptor is deleted from it then the obtained rule is no longer true in \( A \). We have also \( m_A(\{0,1\}) = 2/8 \).
Let us construct also the $A$-optimal rule of the form:

$$
\delta \quad \rightarrow \quad \delta_A = \{0,2\}
$$

In order to construct $\delta$ we consider the following discernibility matrices: $MG_1(A,\{0,2\},x_3)$ and $MG_1(A,\{0,2\},x_7)$.

$MG_1(A,\{0,2\},x_3)$ has the following non-empty entries:

$$
c; \ bc; \ abc; \ ab; \ ab; \ a
$$

They are in Table 3 in the row corresponding to $x_5$ and in the columns corresponding to objects $x$ satisfying $\delta_A(x) \neq \delta_A(x_5)$. The discernibility function corresponding to that matrix has, after the simplification, the following form:

$$
ac.
$$

We have $PRIME_{MG_1}(A,\{0,2\},x_2) = \{ac\}$. Hence we obtain $\gamma$ with only one disjunct taking in Table 1 the trace of the prime implicant $ac$ in the row corresponding to $x_5$:

$$
a_0c_1.
$$

Since $MG_1(A,\{0,2\},x_7)$ has the same non-empty entries as $MG_1(A,\{0,1\},x_3)$ (in the implementation of indiscernibility matrices one can take only one representative for each indiscernibility class) we obtain the following $A$-optimal decision rule:

$$
a_0c_1 \quad \rightarrow \quad \delta_A = \{0,2\}
$$

The disjunct $a_0c_1$ has the minimal number of descriptors: if any descriptor is deleted from it then the obtained rule is no longer true in $A$. We have also $m_A(\{0,2\}) = 1/4$.

Let us now construct the $A$-optimal rule of the form

$$
\phi \quad \rightarrow \quad \delta_A = \{1\} + \delta_A = \{0\} + \delta_A = \{0,1\}
$$

In order to construct $\phi$ we consider the discernibility matrix for the information system $(U,A)$ with rows and columns labelled by the values of a function $d_1: U \rightarrow \{0,1\}$ defined by
\[ d_1(x_j) = 0 \text{ for } j = 3, 7, 8 \text{ and } d_1(x_j) = 1 \text{ otherwise.} \]

Hence the value \( d_1(x_j) = 1 \) iff \( \delta_A(x_j) \subseteq \{0, 1\} \). In this way we obtain Table 4.

Let us now consider the following discernibility matrices:

\[
MG_2(A, \{0, 1\}, x_1), \\
MG_2(A, \{0, 1\}, x_2), \\
MG_2(A, \{0, 1\}, x_4), \\
MG_2(A, \{0, 1\}, x_5), \\
MG_2(A, \{0, 1\}, x_6).
\]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
\begin{array}{cccccccc}
1 & x_1 & b & c & ab & abc & abc & c & ac \\
1 & x_2 & b & c & bc & a & ac & ac & bc \\
0 & x_3 & c & bc & 0 & abc & ab & ab & c \\
1 & x_4 & ab & a & abc & 0 & c & c & abc \\
1 & x_5 & abc & ac & ab & c & 0 & 0 & ab \\
1 & x_6 & abc & ac & ab & c & 0 & 0 & ab \\
0 & x_7 & abc & ac & ab & c & 0 & 0 & ab \\
0 & x_8 & abc & ac & ab & c & 0 & 0 & ab \\
\end{array}
\]

Table 4

They have the following non-empty entries:

\[
MG_2(A, \{0, 1\}, x_1): c, c, ac; \\
MG_2(A, \{0, 1\}, x_2): bc, bc, abc; \\
MG_2(A, \{0, 1\}, x_4): abc, abc, bc; \\
MG_2(A, \{0, 1\}, x_5): ab, ab, b; \\
MG_2(A, \{0, 1\}, x_6): ab, ab, b.
\]

Hence \( PRIME\_MG_2(A, \{0, 1\}, x_1) = \{c\} \).
PRIME\_MG\_2(A,\{0,1\},x_2)={b,c},
PRIME\_MG\_2(A,\{0,1\},x_4)={b,c},
PRIME\_MG\_2(A,\{0,1\},x_5)=PRIME\_MG\_2(A,\{0,1\},x_6)={b}.

Now we can compute the disjuncts of $\phi$ by taking in Table 1 the traces of all prime implicants of those discernibility functions in rows 1,2,4,5 and 6. In this way we obtain:

$$b_1 + c_0 \implies \delta_A={1} + \delta_A={0} + \delta_A={0,1}$$

The above rule is $A$-optimal. We have also

$$Bel_A(\{0,1\}) = |(b_1 + c_0)_A|/|U| = 5/8.$$  

The formula $b_1 + c_0$ describes the set of all objects in $A$ which can be classified on the basis of knowledge included in conditions $a,b,c$ as belonging with certainty to the set of objects with decision 0 or 1.

Let us now construct the $A$-optimal rule of the form

$$\phi \implies \delta_A={1} + \delta_A={0} + \delta_A={0,1} + \delta_A={0,2}$$

In order to construct $\phi$ we consider the discernibility matrix for the information system $(U,A)$ with rows and columns labelled by the values of a function $d_2: U \rightarrow \{0,1\}$ defined by

$$d_2(x_j) = 0 \text{ for } j=8 \text{ and } d_2(x_j) = 1 \text{ otherwise.}$$

Hence the value $d_2(x_j) = 1$ iff $\delta_A(x_j) \cap \{0,1\} \neq \emptyset$. In this way we obtain Table 5.

Let us now consider the following discernibility matrices:

$$MG_3(A,\{0,1\},x_1),$$
$$MG_3(A,\{0,1\},x_2),$$
$$MG_3(A,\{0,1\},x_3),$$
$$MG_3(A,\{0,1\},x_4),$$
\[ MG_3(\mathbb{A}, \{0, 1\}, x_5), \]
\[ MG_3(\mathbb{A}, \{0, 1\}, x_6), \]
\[ MG_3(\mathbb{A}, \{0, 1\}, x_7). \]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
<td>$x_5$</td>
<td>$x_6$</td>
<td>$x_7$</td>
<td>$x_8$</td>
</tr>
<tr>
<td>1</td>
<td>$\emptyset$</td>
<td>$b$</td>
<td>$c$</td>
<td>$ab$</td>
<td>$abc$</td>
<td>$abc$</td>
<td>$c$</td>
<td>$ac$</td>
</tr>
<tr>
<td>1</td>
<td>$x_2$</td>
<td>$b$</td>
<td>$\emptyset$</td>
<td>$bc$</td>
<td>$a$</td>
<td>$ac$</td>
<td>$ac$</td>
<td>$bc$</td>
</tr>
<tr>
<td>1</td>
<td>$x_3$</td>
<td>$c$</td>
<td>$bc$</td>
<td>$\emptyset$</td>
<td>$abc$</td>
<td>$ab$</td>
<td>$ab$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>1</td>
<td>$x_4$</td>
<td>$ab$</td>
<td>$a$</td>
<td>$abc$</td>
<td>$\emptyset$</td>
<td>$c$</td>
<td>$c$</td>
<td>$abc$</td>
</tr>
<tr>
<td>1</td>
<td>$x_5$</td>
<td>$abc$</td>
<td>$ac$</td>
<td>$ab$</td>
<td>$c$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$ab$</td>
</tr>
<tr>
<td>1</td>
<td>$x_6$</td>
<td>$abc$</td>
<td>$ac$</td>
<td>$ab$</td>
<td>$c$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$ab$</td>
</tr>
<tr>
<td>1</td>
<td>$x_7$</td>
<td>$c$</td>
<td>$bc$</td>
<td>$\emptyset$</td>
<td>$abc$</td>
<td>$ab$</td>
<td>$ab$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>0</td>
<td>$x_8$</td>
<td>$ac$</td>
<td>$abc$</td>
<td>$a$</td>
<td>$bc$</td>
<td>$b$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
</tbody>
</table>

Table 5

They have the following non-empty entries:

\[ MG_3(\mathbb{A}, \{0, 1\}, x_1): ac; \]
\[ MG_3(\mathbb{A}, \{0, 1\}, x_2): abc; \]
\[ MG_3(\mathbb{A}, \{0, 1\}, x_3): a; \]
\[ MG_3(\mathbb{A}, \{0, 1\}, x_4): bc; \]
\[ MG_3(\mathbb{A}, \{0, 1\}, x_5): b. \]
\[ MG_3(\mathbb{A}, \{0, 1\}, x_6): b; \]
\[ MG_3(\mathbb{A}, \{0, 1\}, x_7): a. \]

Hence the discernibility functions corresponding to these discernibility matrices are defined by \( a+c, a+b+c, a, b+c, b, b, a, a \), respectively.

We have

\[ \text{PRIME}_MG_3(\mathbb{A}, \{0, 1\}, x_1) = \{a, c\}, \]
\[ \text{PRIME}_MG_3(\mathbb{A}, \{0, 1\}, x_2) = \{a, b, c\} \]
\[ \text{PRIME}_MG_3(\mathbb{A}, \{0, 1\}, x_3) = \{a\}, \]

16
PRIME\_MG_3(A,\{0,1\},x_4) = \{b, c\},  
PRIME\_MG_3(A,\{0,1\},x_5) = PRIME\_MG_3(A,\{0,1\},x_6) = \{b\},  
PRIME\_MG_3(A,\{0,1\},x_7) = \{a\}.

Now we can compute the disjuncts of \( \phi \) by taking in Table 1 the traces of all prime implicants of those discernibility functions in rows 1, 2, 3, 4, 5, 6 and 7, respectively. In this way we obtain:

\[ a_0 + b_1 + c_0 \implies \delta_A = \{1\} + \delta_A = \{0\} + \delta_A = \{0, 1\} + \delta_A = \{0, 2\} \]

The above rule is \( A \)-optimal. We have also

\[ Pl_A(\{0,1\}) = |(a_0 + b_1 + c_0)_A|/|U| = 7/8. \]

The formula \( a_0 + b_1 + c_0 \) describes the set of all objects in \( A \) which can be classified on the basis of knowledge included in conditions \( a, b, c \) as possibly belonging to the set of objects with decision 0 or 1.

**Conclusions**

Let us observe that the introduced numerical coefficients are computable from a given decision table. The discussed methods of decision rules generation are implemented in our system for classifying objects. The complexity of the method is of the same order as for the reduct set generation. In many tested practical applications the method was successful despite of that the complexity of the reduct set generation (in worst case analysis) is high (exponential with respect to the number of objects). Moreover, the large number of computed reducts can be treated as a signal that our conditions are inappropriate for defining a given classification. In fact, this corresponds to a situation when the generated rules would have many disjuncts on the left hand side of the decision rule with each disjunct supported only by a few examples. This implies that the
attributes chosen for decision taking are not suitable for expressing the characteristic properties of the decision classes and a searching process for some new, more appropriate, attributes (classifiers) is necessary. We investigate procedures for classifiers searching from formulae sets of modal and temporal logics [11].

We propose to investigate also logics with belief functions. The semantics of these logics is based on so called decision table maps. These are some kinds of Kripke models with worlds indexed by information vectors defined by a given decision table $A$, the accessibility relation between worlds defined by the inclusion relation between information vectors and with a special structure attached to any information vector. Any such a structure is defined by the restriction of $A$ to the information labelling that structure and contains restricted to that table belief functions. We will investigate this kind of logics as candidates for expressing new classifiers, i.e. we would like to verify a hypothesis that formulas of those logics can be often more suitable for expressing characteristic properties of object classes than those given in decision table.

The well known rule of evidence combination from independent sources of information is the Dempster-Shafer rule [7]. It was shown in [9] that this rule is related to an independent product of decision tables. In the case of that product two descriptions seen as contradictory (on the basis of evidence from independent sources) are eliminated. Quite often one can not assume that the evidence sources are independent. If an object is classified by two sources into $A_2=\{\theta_1, \theta_2\}$ and $A_2=\{\theta_3, \theta_4\}$ and the first source eliminates the hypothesis $\theta_3$ because of not sufficient knowledge to judge if $\theta_3$ holds and the knowledge of the second source to judge about $\theta_3$ is enough deep then the hypothesis $\theta_3$ should not be eliminated. Hence, in general, the combination rule should be based not only on the bpa functions but also on properties of knowledge embedded in
both sources. We will investigate special logics for this kind of reasoning.

References

Recently published Research Reports
of the Institute of Computer Science, W.U.T.


15/93  Moumine Ballo, Modelowanie terenu z wykorzystaniem map topograficznych, marzec 1993.

16/93  Moumine Ballo, Nowy algorytm wyznaczania części wspólnej dwóch wielokątów prostych, marzec 1993.


