Warsaw University of Technology

A Rough Set Approach to Decision Rules Generation

by

Zdzisław Pawlak[1], Andrzej Skowron[2]

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: $\tau \Longrightarrow \tau'$ where τ, τ' 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 $\mathbb{A} = (U, A)$, where U is a non-empty, finite set called the *universe* and A - a non-empty, finite set of *attributes*, i.e. $a: U \longrightarrow V_a$ 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 $\mathbb{A} = (U, A \cup \{d\})$, where $d \notin A$ is a distinguished attribute called *decision*. The elements of A are called *conditions*.

By V is denoted the set $V = \bigcup_{a \in A} V_a \cup V_d \cup \mathbb{P}(V_d)$, where $\mathbb{P}(V_d)$ is the powerset of V_d. The set V_d will be called the *frame of*

[1]

This work has been supported by grant No 8 0570 91 01 from State Committee for Scientific Research. The paper is an extension of the paper prepared for the IJCAI-93 Workshop: The Management of Uncertainty in AI (August 28, 1993; France).

discernment of $\mathbb{A}$ and denoted by $\Theta_{\mathbb{A}}$. To simplify notation we assume $\Theta_{\mathbb{A}} = \{1,\ldots,r(d)\}$. The decision d determines a partition $CLASS_{\mathbb{A}}(d) = \{X_1,\ldots,X_{r(d)}\}$ of the universe U, where $X_k = \{x \in U: d(x)=k\}$ for $1 \le k \le r(d)$.

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

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

Objects x,x' satisfying relation $IND(B)$ are indiscernible by attributes from B.

The sets $\{x \in U: [x]_B \subseteq X\}$ and $\{x \in U: [x]_B \cap X \neq \emptyset\}$ are called *B-lower* and *B-upper approximation* of $X \subseteq U$ in $\mathbb{A}$, and they are denoted by $\underline{B}X$ and $\bar{B}X$, respectively. The *B-boundary* of X is the set $BN_B(X) = \bar{B}X - \underline{B}X$. We write $BN_{\mathbb{A}}(X)$ 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 $IND(B)$, otherwise it is *roughly definable by B*.

An *information function* $Inf_{\mathbb{A}} : U \longrightarrow \mathbb{P}(A \times \bigcup_{a \in A} V_a)$, for a given information system $\mathbb{A}=(U,A)$, is defined by $Inf_{\mathbb{A}}(x) = \{(a,a(x)): a \in A\}$. Any subset of $Inf_{\mathbb{A}}(x)$ is called an *information (vector) about x in* $\mathbb{A}$. The set $\{Inf_{\mathbb{A}}(x) : x \in U\}$ is called the *$\mathbb{A}$-information set* and it is denoted by $\mathrm{INF}(\mathbb{A})$.

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

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

A discernibility *function* $f_{\mathbb{A}}$ for an information system $\mathbb{A}$ is a boolean function of m boolean variables $\bar{a}_1,\ldots,\bar{a}_m$ corresponding to the attributes $a_1,\ldots,a_m$, respectively and defined by

$f_{\mathbf{A}}(\bar{a}_1,..,\bar{a}_m)=\bigwedge\{\bigvee \bar{c}_{ij}: 1\le j<i\le n,\ c_{ij}\ne\emptyset\}$, where $\bar{c}_{ij}=\{\bar{a}: a\in c_{ij}\}$.

We will write *a* instead of $\bar{a}$ 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 $f_{\mathbf{A}}$ 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 $\mathbf{P}(A)\times\{1,...,n\}\times\{1,...,n\}$ computable from M(**A**). $f_{MG(\mathbf{A})}$ denotes a boolean function constructed from MG(**A**) in an analogous way as $f_{\mathbf{A}}$ from M(**A**).

By PRIME_MG(**A**) we denote the set of all prime implicants of $f_{MG(\mathbf{A})}$.

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 Θ, *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 $\mathbb{A}=(U, A \cup \{d\})$ be a decision table and let $APP\_CLASS_{\mathbb{A}}(d)=\{\underline{A}X_1,\ldots,\underline{A}X_{r(d)}\} \cup \{Bd_{\mathbb{A}}(\theta): \theta\subseteq\Theta_{\mathbb{A}}$ and $|\theta|>1\}$, where $Bd_{\mathbb{A}}(\theta) = \bigcap_{i\in\theta} BN_A(X_i) \cap \bigcap_{i\notin\theta} -BN_A(X_i)$. The family of all non-empty sets from $APP\_CLASS_{\mathbb{A}}(d)$ creates a partition of the universe U. Moreover, the following equality holds:

$$\bigcup_{i\in\theta} \underline{A}X_i \cup \bigcup_{\Delta\subseteq\theta, |\Delta|>1} Bd_{\mathbb{A}}(\Delta) = \underline{A} \bigcup_{i\in\theta} X_i \quad \text{for } \theta\subseteq\Theta_{\mathbb{A}} \text{ with } |\theta|>1.$$

□

The classification (partition) of the universe U described in Proposition 1 is called the *standard classification of U approximating in* $\mathbb{A}$ *the classification* $CLASS_{\mathbb{A}}(d)$ (given by an expert). We have a clear interpretation of this new classification. Any object from the universe U of $\mathbb{A}$ is represented by an information $Inf_{\mathbb{A}}(x) \in INF(\mathbb{A})$. An object x can be classified exactly on the basis of information $Inf_{\mathbb{A}}(x)$ only when the category (i.e. the equivalence class of the indiscernibility relation $IND_{\mathbb{A}}(A)$) corresponding to x is included in X_i for some i. Otherwise, that category is included in a boundary region of the form $Bd_{\mathbb{A}}(\theta)$, for some θ. Then an object x from U represented by the information $Inf_{\mathbb{A}}(x)$ 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 $x\in X_i$, where $i\in\theta$).

There is a natural correspondence between subsets of $\Theta_{\mathbb{A}}$ and elements of $APP\_CLASS_{\mathbb{A}}(d)$, which can be expressed by the following function:

$$F_{\mathbb{A}}(\theta) = \begin{cases} \underline{A}X_i & \text{if } \theta = \{i\} \text{ for some i } (1\le i\le r(d)) \\ \emptyset & \text{if } \theta = \emptyset \\ Bd_{\mathbb{A}}(\theta) & \text{if } |\theta| > 1 \end{cases}$$

Now we can define the injection

$$\partial_{\mathbb{A}}: U \longrightarrow \mathbb{P}(\Theta_{\mathbb{A}})$$

as follows: $\partial_{\mathbb{A}}(x)$ is the unique subset θ of $\Theta_{\mathbb{A}}$ such that $x \in F_{\mathbb{A}}(\theta)$.

In other words

$$\partial_{\mathbb{A}}(x) = \{k: \exists x' \; xIND(A)x' \text{ and } d(x')=k\} \text{ for } x \in U.$$

The function $\partial_{\mathbb{A}}$ can be treated as a new decision attribute (defined by conditions in $\mathbb{A}$) approximating the decision d.

The function $m_{\mathbb{A}} : \mathbb{P}(\Theta_{\mathbb{A}}) \longrightarrow \mathbb{R}_+$, called the *standard basic probability assignment* (defined by $\mathbb{A}$) is defined by

$$m_{\mathbb{A}}(\theta) = \frac{|F_{\mathbb{A}}(\theta)|}{|U|}, \text{ for any } \theta \subseteq \Theta_{\mathbb{A}}.$$

Proposition 2 [9] The function $m_{\mathbb{A}}$ defined above is a basic probability assignment (in the sense of evidence theory). □

Theorem 3. [9] For an arbitrary $\theta \subseteq \Theta_{\mathbb{A}}$ the following equality holds:

$$Bel_{\mathbb{A}}(\theta) = \frac{\left|\underline{A} \bigcup_{i \in \theta} X_i\right|}{|U|}$$

The belief function $Bel_{\mathbb{A}}$ is Bayesian iff all sets from $CLASS_{\mathbb{A}}(d)$ are definable by the set A of conditions. In particular the belief function $Bel_{\mathbb{A}'}$, where $\mathbb{A}'=(U, A \cup \{\partial_{\mathbb{A}}\})$, is Bayesian. □

Corollary 4. For an arbitrary $\theta \subseteq \Theta_{\mathbb{A}}$ the following equality holds:

$$Pl_{\mathbb{A}}(\theta) = \frac{\left|\bar{A} \bigcup_{i \in \theta} X_i\right|}{|U|}$$

3. Decision rules

Now we are going to define decision rules.

The atomic formulas over $B \subseteq A \cup \{d\}$ 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 $\mathbb{F}(B,V)$ of formulas over B is the least set containing all atomic formulas over B and closed with respect to the classical propositional connectives $\vee$ (disjunction), $\wedge$ (conjunction) and $\neg$ (negation).

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

1. if τ is of the form $a = v$ then $\tau_{\mathbf{A}}=\{x\in U\colon a(x)=v\}$;
2. $(\tau\wedge\tau')_{\mathbf{A}} = \tau_{\mathbf{A}}\cap\tau'_{\mathbf{A}}$; $(\tau\vee\tau')_{\mathbf{A}} = \tau_{\mathbf{A}}\cup\tau'_{\mathbf{A}}$; $(\neg\tau)_{\mathbf{A}} = U-\tau_{\mathbf{A}}$.

The set $\mathbb{F}(A,V)$ is called the set of *condition formulas* in $\mathbf{A}$ and is denoted by $\mathbb{C}_{\mathbf{A}}$. The set $\mathbb{F}(\{\partial_{\mathbf{A}}\},V)$ is called the set of *decision formulas* in $\mathbf{A}^*=(U,A\cup\{\partial_{\mathbf{A}}\})$ and is denoted by $\mathbb{D}_{\mathbf{A}}$.

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

A *decision rule* for $\mathbf{A}$ is any expression of the form

$$\tau\Longrightarrow\tau' \text{ where } \tau\in\mathbb{C}_{\mathbf{A}} \text{ and } \tau'\in\mathbb{D}_{\mathbf{A}}.$$

The decision rule $\tau\Longrightarrow\tau'$ for $\mathbf{A}$ is *true in* $\mathbf{A}$ iff $\tau_{\mathbf{A}}\subseteq\tau_{\mathbf{A}}{*}$. If $\tau_{\mathbf{A}} = \tau_{\mathbf{A}}{*}$ then we say that the rule $\tau\Longrightarrow\tau'$ is $\mathbf{A}$-*exact*.

An $\mathbf{A}$-exact rule $\tau\Longrightarrow\tau'$ is $\mathbf{A}$-*optimal* iff

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

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

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 $\mathbf{P}(\Theta_{\mathbb{A}})$ defined by:

$$\Delta_1 R_1 \Delta_2 \text{ iff } \Delta_1 = \Delta_2;$$
$$\Delta_1 R_2 \Delta_2 \text{ iff } \Delta_1 \subseteq \Delta_2;$$
and $\Delta_1 R_3 \Delta_2$ iff $\Delta_1 \cap \Delta_2 \neq \emptyset$ for any $\Delta_1, \Delta_2 \subseteq \Theta_{\mathbb{A}}$.

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

$$MG_k(\mathbb{A}, \Delta, x_i) = \{c_{ij} - (d) : \neg(\partial_{\mathbb{A}}(x_j) R_k \Delta)\} \text{ for } k=1,2,3.$$

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

$$\tau \Longrightarrow \bigvee \{\partial_{\mathbb{A}} = \theta : \theta R_k \Delta\} \text{ , where } \Delta \subseteq \Theta_{\mathbb{A}} \text{ and } \tau \in \mathbf{C}_{\mathbb{A}}$$

Our main result can be formulated as follows:

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

1. $\bigvee \{\tau_u : \exists x, t, u \ (\partial_{\mathbb{A}}(x) R_k \Delta \ \&$
$t \in \text{PRIME\_MG}_k(\mathbb{A}, \Delta, x) \ \&$
$u = \text{INF}(t, \mathbb{A}^*, x))\} \Longrightarrow \bigvee \{\partial_{\mathbb{A}} = \theta : \theta R_k \Delta\}$
where $\text{INF}(t, \mathbb{A}^*, x) = \{(a, a(x)) : a \in A(t)\}$,
is an $\mathbb{A}$-optimal decision rule.

2. The value of $\dfrac{|\{x \in U : \ \partial_{\mathbb{A}}(x) R_k \Delta\}|}{|U|}$ is equal to $m_{\mathbb{A}}(\Delta)$, $Bel_{\mathbb{A}}(\Delta)$ and $Pl_{\mathbb{A}}(\Delta)$ for k=1,2,3, respectively. □

The set $INF_k(t,\mathbf{A}^*,x)=\{(a,a(x)): a\in A(t)\}$ is called the *trace (in* $\mathbf{A}$*) 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 $\underline{A}\bigcup_{i\in\Delta} X_i$ ($\bar{A}\bigcup_{i\in\Delta} X_i$). The value $m_{\mathbf{A}}(\Delta)$ describes a "chance" that an object chosen from $\mathbf{A}$ is classified by $\partial_{\mathbf{A}}$ into Δ. The value $Bel_{\mathbf{A}}(\Delta)$ describes a "chance" that an object chosen from $\mathbf{A}$ is classified with certainty into the union $\bigcup_{i\in\Delta} X_i$ (on the basis of knowledge determined by A). The value $Pl_{\mathbf{A}}(\Delta)$ describes a "chance" that an object chosen from $\mathbf{A}$ is classified as possibly belonging to $\bigcup_{i\in\Delta} X_i$ (on the basis of knowledge determined by 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 $\mathbf{A}=(U,A\cup\{d\})$ presented in Table 1, where we have $U=\{x_1,\ldots,x_8\}$, $A=\{a,b,c\}$ and d is the decision. The values of attributes are presented in Table 1. In the examples we write $\alpha\beta$ and $\alpha+\beta$ instead of $\alpha\wedge\beta$ and $\alpha\vee\beta$, respectively.

	a	b	c	d
x_1	0	0	0	0
x_2	0	1	0	1
x_3	0	0	1	0
x_4	1	1	0	1
x_5	1	1	1	1
x_6	1	1	1	0
x_7	0	0	1	2
x_8	1	0	1	2

Table 1

In Table 2 the values of the "new" attribute $\partial_{\mathbf{A}}$ are presented.

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

	a	b	c	d	$\partial_{\mathbf{A}}$
x_1	0	0	0	0	{0}
x_2	0	1	0	1	{1}
x_3	0	0	1	0	{0,2}
x_4	1	1	0	1	{1}
x_5	1	1	1	1	{0,1}
x_6	1	1	1	0	{0,1}
x_7	0	0	1	2	{0,2}
x_8	1	0	1	2	{2}

Table 2

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

$$\alpha \longrightarrow \partial_{\mathbf{A}}=\{1\}$$

In order to construct α we consider the following discernibility matrices: $MG_1(\mathbf{A},\{1\},x_2)$ and $MG_1(\mathbf{A},\{1\},x_4)$.
$MG_1(\mathbf{A},\{1\},x_2)$ has the following non-empty entries:

b; bc; ac; ac; bc; abc.

		$\{0\}$ x_1	$\{1\}$ x_2	$\{0,2\}$ x_3	$\{1\}$ x_4	$\{0,1\}$ x_5	$\{0,1\}$ x_6	$\{0,2\}$ x_7	$\{2\}$ x_8
$\{0\}$	x_1	ø	*b*	*c*	*ab*	*abc*	*abc*	*c*	*ac*
$\{1\}$	x_2	*b*	ø	*bc*	*a*	*ac*	*ac*	*bc*	*abc*
$\{0,2\}$	x_3	*c*	*bc*	ø	*abc*	*ab*	*ab*	ø	*a*
$\{1\}$	x_4	*ab*	*a*	*abc*	ø	*c*	*c*	*abc*	*bc*
$\{0,1\}$	x_5	*abc*	*ac*	*ab*	*c*	ø	ø	*ab*	*b*
$\{0,1\}$	x_6	*abc*	*ac*	*ab*	*c*	ø	ø	*ab*	*b*
$\{0,2\}$	x_7	*c*	*bc*	ø	*abc*	*ab*	*ab*	ø	*a*
$\{2\}$	x_8	*ac*	*abc*	*a*	*bc*	*b*	*b*	*a*	ø

Table 3

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

$$b(a+c) \equiv ba + bc.$$

We have $PRIME\_MG_1(\mathbf{A},\{1\},x_2)=\{ba,bc\}$. Hence we obtain the following disjuncts of α taking in Table 1 traces of the prime implicants *ba*, *bc* in the row corresponding to x_2:

$$b_1a_0 \;,\; b_1c_0.$$

$MG_1(\mathbf{A},\{1\},x_4)$ 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 colums corresponding to objects x satisfying $\partial_{A}(x)\neq\partial_{A}(x_4)$. The discernibility function corresponding to that matrix has, after the simplification, the following form:

$$c(a+b) = ca + cb.$$

We have $PRIME\_MG_1(A,\{1\},x_4)=\{ca,cb\}$. Hence we obtain the following disjuncts of α taking in Table 1 traces of the prime implicants ca, cb in the row corresponding to x_4 :

$$c_0a_1,\ c_0b_1.$$

Finally we obtain the following A-optimal decision rule:

$$a_0b_1 + b_1c_0 + a_1c_0 \Longrightarrow \partial_{A}=\{1\}$$

Any disjunct in α 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:

$$\beta \Longrightarrow \partial_{A}=\{0\}$$

In order to construct β 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 $\partial_{A}(x)\neq\partial_{A}(x_2)$. The discernibility function corresponding to that matrix has, after the simplification, the following form:

$$bc.$$

We have $PRIME\_MG_1(A,\{0\},x_1)=\{bc\}$. Hence we obtain only one disjunct of β taking in Table 1 the trace of the prime implicant bc in the row corresponding to x_1 :

$$b_0c_0.$$

We obtain the following $\mathbb{A}$-optimal decision rule:

$$b_0c_0 \longrightarrow \partial_{\mathbb{A}}=\{0\}$$

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 $\mathbb{A}$. We have also $m_{\mathbb{A}}(\{0\})=1/8$.

Let us construct now the $\mathbb{A}$-optimal rule of the form:

$$\gamma \Longrightarrow \partial_{\mathbb{A}}=\{0,1\}$$

In order to construct γ we consider the following discernibility matrices: $MG_1(\mathbb{A},\{0,1\},x_5)$ and $MG_1(\mathbb{A},\{0,1\},x_6)$. $MG_1(\mathbb{A},\{0,1\},x_5)$ 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 $\partial_{\mathbb{A}}(x)\neq\partial_{\mathbb{A}}(x_5)$. The discernibility function corresponding to that matrix has, after the simplification, the following form:

$$bc.$$

We have $PRIME\_MG_1(\mathbb{A},\{0,1\},x_5)=\{bc\}$. Hence we obtain the following disjunct of γ taking in Table 1 the trace of the prime implicant bc in the row corresponding to x_5:

$$b_1c_1.$$

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

$$b_1c_1 \Longrightarrow \partial_{\mathbb{A}}=\{0,1\}$$

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 $\mathbb{A}$. We have also $m_{\mathbb{A}}(\{0,1\})=2/8$.

Let us construct also the **A**-optimal rule of the form:

$$\delta \Longrightarrow \partial_{\mathbb{A}}=\{0,2\}$$

In order to construct δ we consider the following discernibility matrices: $MG_1(\mathbb{A},\{0,2\},x_3)$ and $MG_1(\mathbb{A},\{0,2\},x_7)$. $MG_1(\mathbb{A},\{0,2\},x_3)$ 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 $\partial_{\mathbb{A}}(x)\neq\partial_{\mathbb{A}}(x_5)$. The discernibility function corresponding to that matrix has, after the simplification, the following form:

ac.

We have $PRIME\_MG_1(\mathbb{A},\{0,2\},x_2)=\{ac\}$. Hence we obtain γ 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 $MG_1(\mathbb{A},\{0,2\},x_7)$ has the same non-empty entries as $MG_1(\mathbb{A},\{0,1\},x_3)$ (in the implementation of indiscernibility matrices one can take only one representat for each indiscernibility class) we obtain the following **A**-optimal decision rule:

$$a_0c_1 \Longrightarrow \partial_{\mathbb{A}}=\{0,2\}$$

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 $m_{\mathbb{A}}(\{0,2\})=1/4$.

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

$$\phi \Longrightarrow \partial_{\mathbb{A}}=\{1\} + \partial_{\mathbb{A}}=\{0\} + \partial_{\mathbb{A}}=\{0,1\}$$

In order to construct ϕ we consider the discernibility matrix for the information system (U,A) with rows and columns labelled by the values of a function d_1: $U \longrightarrow \{0,1\}$ defined by

$d_1(x_j)=0$ for $j=3,7,8$ and $d_1(x_j)=1$ otherwise.

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

Let us now consider the following discernibility matrices:

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

	1 x_1	1 x_2	0 x_3	1 x_4	1 x_5	1 x_6	0 x_7	0 x_8
1 x_1	∅	*b*	*c*	*ab*	*abc*	*abc*	*c*	*ac*
1 x_2	*b*	∅	*bc*	*a*	*ac*	*ac*	*bc*	*abc*
0 x_3	*c*	*bc*	∅	*abc*	*ab*	*ab*	∅	*a*
1 x_4	*ab*	*a*	*abc*	∅	*c*	*c*	*abc*	*bc*
1 x_5	*abc*	*ac*	*ab*	*c*	∅	∅	*ab*	*b*
1 x_6	*abc*	*ac*	*ab*	*c*	∅	∅	*ab*	*b*
0 x_7	*c*	*bc*	∅	*abc*	*ab*	*ab*	∅	*a*
0 x_8	*ac*	*abc*	*a*	*bc*	*b*	*b*	*a*	∅

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 $PRIME\_MG_2(A,\{0,1\},x_1)=\{c\}$,

$$PRIME\_MG_2(\mathbf{A},\{0,1\},x_2)=\{b,c\},$$
$$PRIME\_MG_2(\mathbf{A},\{0,1\},x_4)=\{b,c\},$$
$$PRIME\_MG_2(\mathbf{A},\{0,1\},x_5)=PRIME\_MG_2(\mathbf{A},\{0,1\},x_6)=\{b\}.$$

Now we can compute the disjuncts of ϕ 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:

$$b_1 + c_0 \Longrightarrow \partial_{\mathbf{A}}=\{1\} + \partial_{\mathbf{A}}=\{0\} + \partial_{\mathbf{A}}=\{0,1\}$$

The above rule is $\mathbf{A}$-optimal. We have also

$$Bel_{\mathbf{A}}(\{0,1\}) = |(b_1 + c_0)_{\mathbf{A}}|/|U| = 5/8.$$

The formula $b_1 + c_0$ describes the set of all objects in $\mathbf{A}$ 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 $\mathbf{A}$-optimal rule of the form

$$\phi \Longrightarrow \partial_{\mathbf{A}}=\{1\} + \partial_{\mathbf{A}}=\{0\} + \partial_{\mathbf{A}}=\{0,1\} + \partial_{\mathbf{A}}=\{0,2\}$$

In order to construct ϕ we consider the discernibility matrix for the information system (U,A) with rows and columns labelled by the values of a function d_2: $U \longrightarrow \{0,1\}$ 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 $\partial_{\mathbf{A}}(x_j)\cap\{0,1\}\neq\emptyset$. In this way we obtain Table 5.

Let us now consider the following discernibility matrices:

$$MG_3(\mathbf{A},\{0,1\},x_1),$$
$$MG_3(\mathbf{A},\{0,1\},x_2),$$
$$MG_3(\mathbf{A},\{0,1\},x_3),$$
$$MG_3(\mathbf{A},\{0,1\},x_4),$$

$MG_3(\mathbf{A},\{0,1\},x_5)$,

$MG_3(\mathbf{A},\{0,1\},x_6)$,

$MG_3(\mathbf{A},\{0,1\},x_7)$.

	1 x_1	1 x_2	1 x_3	1 x_4	1 x_5	1 x_6	1 x_7	0 x_8
1 x_1	ø	*b*	*c*	*ab*	*abc*	*abc*	*c*	*ac*
1 x_2	*b*	ø	*bc*	*a*	*ac*	*ac*	*bc*	*abc*
1 x_3	*c*	*bc*	ø	*abc*	*ab*	*ab*	ø	*a*
1 x_4	*ab*	*a*	*abc*	ø	*c*	*c*	*abc*	*bc*
1 x_5	*abc*	*ac*	*ab*	*c*	ø	ø	*ab*	*b*
1 x_6	*abc*	*ac*	*ab*	*c*	ø	ø	*ab*	*b*
1 x_7	*c*	*bc*	ø	*abc*	*ab*	*ab*	ø	*a*
0 x_8	*ac*	*abc*	*a*	*bc*	*b*	*b*	*a*	ø

Table 5

They have the following non-empty entries:

$MG_3(\mathbf{A},\{0,1\},x_1)$: *ac*;

$MG_3(\mathbf{A},\{0,1\},x_2)$: *abc*;

$MG_3(\mathbf{A},\{0,1\},x_3)$: a;

$MG_3(\mathbf{A},\{0,1\},x_4)$: *bc*;

$MG_3(\mathbf{A},\{0,1\},x_5)$: *b*.

$MG_3(\mathbf{A},\{0,1\},x_6)$: *b*;

$MG_3(\mathbf{A},\{0,1\},x_7)$: a.

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

We have

$PRIME\_MG_3(\mathbf{A},\{0,1\},x_1)=\{a,c\}$,

$PRIME\_MG_3(\mathbf{A},\{0,1\},x_2)=\{a,b,c\}$

$PRIME\_MG_3(\mathbf{A},\{0,1\},x_3)=\{a\}$,

$PRIME\_MG_3(\mathbf{A},\{0,1\},x_4)=\{b,c\}$,

$PRIME\_MG_3(\mathbf{A},\{0,1\},x_5)=PRIME\_MG_3(\mathbf{A},\{0,1\},x_6)=\{b\}$,

$PRIME\_MG_3(\mathbf{A},\{0,1\},x_7)=\{a\}$.

Now we can compute the disjuncts of ϕ 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:

$$a_0 + b_1 + c_0 \Longrightarrow \partial_{\mathbf{A}}=\{1\} + \partial_{\mathbf{A}}=\{0\} + \partial_{\mathbf{A}}=\{0,1\} + \partial_{\mathbf{A}}=\{0,2\}$$

The above rule is $\mathbf{A}$-optimal. We have also

$$Pl_{\mathbf{A}}(\{0,1\}) = |(a_0 + b_1 + c_0)_{\mathbf{A}}|/|U| = 7/8.$$

The formula $a_0 + b_1 + c_0$ describes the set of all objects in $\mathbf{A}$ 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 $\Delta_2=\{\theta_1,\theta_2\}$ and $\Delta_2=\{\theta_3,\theta_4\}$ and the first source eliminates the hypothesis θ_3 because of not sufficient knowledge to judge if θ_3 holds and the knowledge of the second source to judge about θ_3 is enough deep then the hypothesis θ_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

1. R.K. Bhatnager, L.N. Kanal: Handling uncertain information: A review of numeric and non-numeric methods. In: .N. Kanal, J.F. Lemmer (eds.): Uncertainty in Artificial Intelligence. Amsterdam: North - Holland 1986
2. F.M. Brown: Boolean reasoning. Dordrecht: Kluwer 1990
3. Y. Kodratoff, R. Michalski: Machine learning: An Artificial Intelligence approach, vol.3. San Mateo: Morgan Kaufmann 1990
4. Z. Pawlak: Rough sets. International Journal of Information and Computer Science 11, 344-356 (1982)
5. Z. Pawlak: Rough sets: Theoretical aspects of reasoning about data. Dordrecht: Kluwer 1991
6. Z. Pawlak, A. Skowron: Rough membership functions. to appear In: M. Federizzi, J. Kacprzyk and R.R. Yager (eds.): Advances in the Dempster-Shafer Theory of Evidence. New York: John Wiley and Sons
7. G. Shafer: Mathematical Theory of Evidence. Princeton: University Press 1976
8. G. Shafer, J. Pearl: Readings in Uncertain Reasoning. San Mateo: Morgan Kaufmann 1990
9. A. Skowron, J.W. Grzymała-Busse: From the rough set theory to the evidence theory. to appear In: M. Federizzi, J. Kacprzyk and R.R. Yager (eds.): Advances in the Dempster-Shafer Theory of Evidence. New York: John Wiley and Sons
10. A. Skowron, C. Rauszer: The Discernibility Matrices and Functions in Information Systems. In: R. Słowiński (ed.): Decision Support by Experience - Applications of the Rough Sets Theory. Dordrecht: Kluwer 1992, pp.331-362
11. A. Skowron, J. Stepaniuk: Searching for Classifiers. In: M. De Glas, D. Gabbay: Proc. of First World Conference on Foundations of Artificial Intelligence, Paris, July 1-5, 1991. Paris: Angkor 1991, pp. 447-460

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

10/93	Wacław Struk, Koncepcja nawigacji hipertekstowej w wielomedialnych bazach danych, luty 1993.
11/93	Lech T. Polkowski, Maria Semeniuk-Polkowska. A Formal Approach to Anaphora within Montague Grammar. Anaphoric Sets and Stable Situations, February 1993.
12/93	J. A. Pomykała, J.M. Pomykała, On Regularity of Hypergraph Sequences, February 1993.
13/93	Jarosław Chudziak, Henryk Rybiński, James Vorbach, On Mapping Semantic Data Model Schemata into First-Order Logic, February 1993.
14/93	Maciej Modrzejewski, Feature Selection Using Rough Sets Theory, March 1993.
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.
17/93	Piotr Sapiecha, Boltzmann Machine for Maximal Queens Problem, March 1993.
18/93	Andrzej Skowron, Zbigniew Suraj, A Rough Set Approach to Real-Time State, Identification for Decision Making, March 1993.
19/93	Mieczysław Muraszkiewicz, Michał Ostrowski, Henryk Rybiński, Handling Amorphous Data Structures in Unified Information Systems, March 1993.
20/93	Zdzisław Pawlak, Rough Sets Present State and the Future, March 1993.
21/93	Zdzisław Pawlak, Roman Słowiński, Decision Analysis Using Rough Sets, March 1993.
22/93	Jerzy Mieścicki, Roman Podraza, Proceedings of M.Sc. Seminar (winter 1992/93), vol. I, II, March 1993.