# Using Rough Set to Find the Factors That Negate the Typical Dependency of a Decision Attribute on Some Condition Attributes

Feng Honghai[1,2], Xu Hao[3], Liu Baoyan[4], Yang Bingru[2], Gao Zhuye[3], and Li Yueli[1]

[1] Hebei Agricultural University, 071001 Baoding, Hebei, China
[2] University of Science and Technology Beijing, 100083 Beijing, China
[3] Beijing Sino-Japen Friendship Hospital, 100029 Beijing, China
[4] China Academy of Traditional Chinese Medicine, 100700 Beijing, China
liuby@tcmcec.com

**Abstract.** In real world, there are a lot of knowledge such as the following: most human beings that are infected by a kind of virus suffer from a corresponding disease, but a small number human beings do not. Which are the factors that negate the effects of the virus? Standard rough set method can induce simplified rules for classification, but cannot generate this kind of knowledge directly. In this paper, we propose two algorithms to find the factors. In the first algorithm, the typical rough set method is used to generate all the variable precision rules firstly; secondly reduce attributes and generate all the non-variable precision rules; lastly compare the variable precision rules and non-variable precision rules to generate the factors that negate the variable precision rules. In the second algorithm, firstly, induce all the variable precision rules; secondly, select the examples corresponding to the variable precision rules to build decernibility matrixes; thirdly, generate the factors that negate the variable precision rules. Three experimental results show that using the two algorithms can get the same results and the computational complexity of the second algorithm is largely less than the firs one.

## 1 Introduction

In real world, there are a lot of knowledge such as the following: most people suffer from hyperpyrexia when they take a heavy cold, whereas some people do not; most earthquakes in the sea cannot cause a ground sea, but in 2005 the earthquake in Indian Ocean cause a ground sea and cause thousands upon thousands people death. Which are the factors that negate the effects of virus, or a heavy cold? And which factors make an earthquake cause the Indian Ocean ground sea? Standard rough set method [1] can induce simplified rules for classification, but cannot generate this kind of knowledge directly. Other machine learning theories such as SVM [2], ANN [3] and Bayesian networks have not been found that they can be used to induce this kind of knowledge.

All the disasters such as floods, dam collapses, terror events, epidemics etc are exceptional cases. The factors that cause these disasters are significant to us. Additionally,

the exceptional students education, exceptional customers service, exceptional patients therapy and nurse etc need the factors that cause exceptional rules.

The decernibility matrix in rough set theory is a valid method for attribute reduction and rule generation whose main idea is to compare the examples that are not in the same class. However, generally, the decernibility matrix is used to the whole data set [4], can it be used to partial data to induce the knowledge with which we can find the factors that negate the typical dependency of a decision attribute on some condition attributes? The answer is affirmative.

## 2   Basic Concepts of Rough Set Theory

A decision table is composed of a 4-tuple DT=$\langle U, A, V, f \rangle$, where $U = \{x_1, x_2, \cdots, x_n\}$, is a nonempty, finite set called the universe; $A$ is a nonempty, finite set of attributes; $A = C \cup D$, in which $C$ is a finite set of condition attributes and $D$ is a finite set of decision attributes; $V = \bigcup_{a \in A} V_a$, where $V_a$ is a domain (value) of the attribute a, and $f : U \times A \rightarrow V$ is called the information function such that $f(x, a) \in V_a$ for every $a \in A, x_i \in U$.

For every set of attributes $B \subset A$, an indiscernibility relation $IND(B)$ is defined in the following way: two objects, $x_i$ and $x_j$, are indiscernible by the set of attributes $B$ in $A$, if $f(b, x_i) = f(b, x_j)$ for every $b \subseteq B$. The equivalence class of $IND(B)$ is called elementary set in $B$ because it presents the smallest discernible groups of objects. For any element $x_i$ of $U$, the equivalence class of $x_i$ in relation $IND(B)$ is represented as $[x_i]_{IND(B)}$. The construction of elementary sets is the first step in classification with rough sets.

By a discernibility matrix of $B \subseteq A$ denoted $M(B)$ a $n \times n$ matrix is defined as

$$(c_{ij}) = \{a \in B : f(a, x_i) \neq f(a, x_j)\} \text{ for } i, j = 1, 2, \ldots, n.$$

Thus entry $c_{ij}$ is the set of all attributes that discern objects $x_i$ and $x_j$.

It is easily seen that the core is the set of all single element entries of the discernibility matrix $M(B)$, i.e.,

$$CORE(B) = \{a \in B : c_{ij} = \{a\}, \text{ for some } i, j\}$$

Every discernibility matrix $M(B)$ defines uniquely a discernibility (boolean) function $f(B)$ defined as follows.

Let us assign to each attribute $a \in B$ a binary Boolean variable $\bar{a}$, and let $\Sigma \delta(x, y)$ denote Boolean sum of all Boolean variables assigned to the set of attributes $\delta(x, y)$. Then the discernibility function can be defined by the formula

$$f(B) = \prod_{(x,y) \in U^2} \{\Sigma \delta(x, y) : (x, y) \in U^2 \text{ and } \delta(x, y) \neq \varnothing\}.$$

Where $\prod$ denotes the Boolean multiplication.

The following property establishes the relationship between disjunctive normal form of the function $f(B)$ and the set of all reducts of $B$.

All constituents in the minimal disjunctive normal form of the function $f(B)$ are all reducts of B.

In order to compute the value core and value reducts for $x$ we can also use the discernibility matrix as defined before and the discernibility function, which must be slightly modified:

$$f^x(B) = \prod_{y \in U} \{\Sigma \delta(x, y) : y \in U \text{ and } \delta(x, y) \neq \varnothing\}.$$

## 3  Algorithms

(1) Algorithm 1

  (a)  Generate all the variable precision rules.
  (b)  Attribute reduction.
  (c)  Select the reduced attribute set with the attributes that are in the variable precision rules.
  (d)  Generate the non-variable precision rules.
  (e)  Select the non-variable precision rules whose preconditions are the same as the variable precision rules whereas the postcondition is not the same as the variable precision rules.
  (f)  Compare the variable precision rules and their corresponding to generate the factors that negate the variable precision rules.

(2) Algorithm 2

  (a)  For every condition attribute $A$, (or select anyone among all the condition attributes) calculate its equivalence classes or partition
       $IND(A) = \{A_1, A_2, \ldots, A_n,\}$. And for the decision attribute $D$, calculate $IND(D) = \{D_1, D_2, \ldots, D_m,\}$.
  (b)  For ($i=0$; $i<n$; $i$ ++)// $n$: amount of equivalence classes of condition attribute //A
  (c)  For ($j=0$; $j<m$; $j$++)// m: amount of equivalence classes of decision attribute$D$.

$$\text{If } \alpha = \frac{|A_i \cap D_J|}{|A_i|} > s \quad // \ 0.5 < s \neq 1 \text{ and chosen by user.}$$

  (d)  Select the examples with $A_i$, and build a decernibility matrix
  (e)  For the examples whose decision attribute values are not included in $D_j$ to use Boolean multiplication and Boolean sum to induce the factors that negate the rule $A_i \rightarrow D_j$ with $\alpha$. That is, use $f^x(B) = \prod_{y \in U} \{\Sigma \delta(x, y) : y \in U \text{ and }$

$\delta(x, y) \neq \varnothing\}$ to get the factors.

# 4 Hand-Written Chinese Characters Recognition

We have generated the feature vectors of 4 hand-written Chinese characters, where the values 1, 2, 3 and 4 of $Y$ denote the 4 different hand written Chinese characters (Figure 1), examples 1 and 2, 3 and 4, 5 and 6, 7 and 8 are the same hand-written Chinese characters respectively, after discretization we get Table 1.



**Fig. 1.** 4 hand-written Chinese characters

**Table 1.** The information table of 4 hand-written Chinese characters

| U | A | B | C | D | E | F | Y |
|---|---|---|---|---|---|---|---|
| 1 | 17 | 8 | 5 | 4 | 2 | 1 | 1 |
| 2 | 18 | 8 | 5 | 3 | 2 | 1 | 1 |
| 3 | 17 | 9 | 5 | 5 | 2 | 2 | 2 |
| 4 | 17 | 9 | 5 | 3 | 2 | 1 | 2 |
| 5 | 18 | 7 | 6 | 4 | 2 | 1 | 3 |
| 6 | 18 | 8 | 6 | 4 | 3 | 2 | 3 |
| 7 | 15 | 8 | 5 | 4 | 3 | 2 | 4 |
| 8 | 16 | 8 | 5 | 5 | 3 | 2 | 4 |

(1)  Using algorithm 1 to induce the factors

  (a)  The variable precision rules are:
    1)$A$=17→$Y$=2 with $\alpha$ =2/3; 2) $A$=18→$Y$=3 with $\alpha$ =2/3; 3) $E$=3 →$Y$=4 with $\alpha$ =2/3.

  (b)  After attribute reduction we get the following new condition attributes combinations: {$A, B, C$}, {$A, B, E$}, {$A, B, F$}, {$A, D$}, {$B, C, E$}, {$B, C, F$}.

  (c)  Select {$A, B, C$}, {$A, B, E$}, {$A, B, F$}, {$A, D$}, {$B, C, E$}, {$A, C, E, F$}, since the attribute sets {$B, C, F$} and {$B, C, D, F$} do not contain attribute $A$ and $E$ that are in the variable precision rules.

  (d)  For the selected attribute sets, the following rules are generated:
    4) $A$=17 ∧ $B$=8→$Y$=1; 5) $A$=18 ∧ $D$=3→$Y$=1; 6) $B$=9→$Y$=2; 7) $B$=7→$Y$=3; 8) $A$=18 ∧ $D$=4→$Y$=3; 9) $A$=15→$Y$=4; 10) $A$=16→$Y$=4; 11) $B$=8 ∧ $E$=2→$Y$=1; 12) $A$=18 ∧ $E$=3→$Y$=3; 13) $B$=8 ∧ $F$=1→$Y$=1; 14) $A$=18 ∧ $F$=2→$Y$=3; 15) $A$=17 ∧ $D$=4→$Y$=1; 16) $A$=17 ∧ $D$=5→$Y$=2; 17) $A$=17 ∧ $D$=3→$Y$=2; 18) $C$=6→$Y$=3; 19) $C$=5 ∧ $E$=3→$Y$=4; 20) $A$=18 ∧ $C$=5→$Y$=1; 21) $B$=8 ∧ $C$=5 ∧ $F$=2→$Y$=4.

  (e)  For variable precision rule 1) $A$=17→$Y$=2 with $\alpha$ =2/3, rule 4) $A$=17 ∧ $B$=8→$Y$=1 and rule 15) $A$=17 ∧ $D$=4→$Y$=1 should be selected, and after comparing rule 1) with rule 4) and rule 15) respectively, we can get that: $B$=8 and

$D$=4 are the factors that negate rule 1) $A$=17→$Y$=2 with $\alpha$ =2/3. For rule 2) $A$=18→$Y$=3 with $\alpha$ =2/3, rule 5) $A$=18 ∧ $D$=3→$Y$=1 and rule 20) $A$=18 ∧ $C$=5→$Y$=1 should be selected, and after comparing it with rule 5) rule 20) respectively, we can get that: $D$=3 and $C$=5 are the factors that negate the rule $A$=18→$Y$=3 with $\alpha$ =2/3. For rule 3) $E$=3 →$Y$=4 with $\alpha$ =2/3, rule 12) $A$=18 ∧ $E$=3→$Y$=3 should be selected, and after comparing rule 3) with rule12), we can get that: $A$=18 is the factor that negate $E$=3 →$Y$=4 with $\alpha$ =2/3.

(2)  Using algorithm 2 to induce the factors

   (a) The partitions of all the attributes can be gotten as follows:

$IND(A)$={$A_1,A_2,A_3,A_4$}={{1,3,4},{2,5,6},{7},{8}},
$IND(B)$={$B_1,B_2,B_3$}={{1,2,6,7,8},{3,4},{5}},
$IND(C)$={$C_1,C_2$ }={{1,2,3,4,7,8},{5,6}},
$IND(D)$={$D_1,D_2,D_3$}={{1,2,6,7,8},{3,4},{5}},
$IND(E)$={$E_1,E_2$ }={{1,2, 3,4, 5},{6,7,8}},
$IND(F)$={$F_1,F_2$ }={{1,2,4,5},{3,6,7,8}},
$IND(Y)$={ $Y_1$ , $Y_2$ , $Y_3$ ,$Y_4$ }={{1,2},{3,4},{5,6},{7,8}}.

Obviously,  only $\dfrac{|A_1 \cap Y_2|}{|A_1|}=\dfrac{2}{3}>0.5$ , $\dfrac{|A_2 \cap Y_3|}{|A_2|}=\dfrac{2}{3}>0.5$  and  $\dfrac{|E_2 \cap Y_4|}{|E_2|}=\dfrac{2}{3}>0.5$ can be held.

So for $A_1$={1,3,4}, we select examples 1, 3, and 4 to build the decernibility matrix, since the example 1 belongs to class 1, and examples 3, 4 belong to class 2; they should in different places in decernibility matrix, and for example 1 we using Boolean multiplication and Boolean sum to induce the factors that negate the variable rule.

From Table 2, since  ($B∨ D∨F$)∧ ($B∨ D$)= $B∨ D$, we can conclude that $B$=8 and $D$=4 are the factors that negate rule 1) $A$=17→$Y$=2 with $\alpha$ =2/3, or make $A$=17∧ $B$=8→$Y$=1 and $A$=17∧ $D$=4→$Y$=1. Namely we get the following knowledge:

      $A$=17→$Y$=2        with $\alpha$ =2/3
      $A$=17∧ $B$=8→$Y$=1   with $\alpha$ =1
      $A$=17∧ $D$=4→$Y$=1   with $\alpha$ =1
      $B$=8 and $D$=4 are the factors that negate the rule $A$=17→$Y$=2 with $\alpha$ =2/3

**Table 2.** Decernibility matrix for the examples in $A_1$

|   | 3 | 4 |
|---|---|---|
| 1 | BDF | BD |

For $A_2$={2,5,6}, we select examples 2, 5, and 6 to build the decernibility matrix.

**Table 3.** Decernibility matrix for the examples in $A_2$

|   | 5 | 6 |
|---|---|---|
| 2 | BC D | CDEF |

Since $(B \vee C \vee D) \wedge (C \vee D \vee E \vee F) = C \vee D$, we get that $D=3$ and $C=5$ are the factors that negate the rule $A=18 \rightarrow Y=3$ with $\alpha = 2/3$, or make $A=18 \wedge C=5 \rightarrow Y=1$ and $A=18 \wedge D=3 \rightarrow Y=1$.

# 5   Micronutrient Data Set Experiments

(1) SARS Data Set and Discretization

The SARS data are the experimental results of micronutrients that are essential in minute amounts for the proper growth and metabolism of human beings. Among them, examples 31~60 are the results of SARS patients and 61~90 are the results of healthy human beings. Attributes "1", "2", "3", "4", "5", "6", "7" denote micronutrient Zn, Cu, Fe, Ca, Mg, K and Na respectively, and decision attribute "C" denotes the class "SARS" and "healthy". $V_C = \{0,1\}$, where "0" denotes "SARS", "1" denotes "healthy".

After discretization, some examples become a repeat. The amount of the total examples is reduced from 60 to 39. Table 4 describes the left 39 examples after discretization.

**Table 4.** Left examples after discretization

| U | 1 | 2 | 3 | 4 | 5 | 6 | 7 | C | U | 1 | 2 | 3 | 4 | 5 | 6 | 7 | C | U | 1 | 2 | 3 | 4 | 5 | 6 | 7 | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 31 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 54 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 69 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 |
| 32 | 2 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 55 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 70 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 |
| 34 | 1 | 1 | 1 | 0 | 1 | 2 | 1 | 0 | 56 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 71 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 |
| 39 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 59 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 72 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 |
| 41 | 0 | 1 | 1 | 0 | 0 | 2 | 2 | 0 | 60 | 1 | 2 | 1 | 0 | 1 | 1 | 1 | 0 | 73 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 |
| 42 | 2 | 1 | 1 | 0 | 1 | 2 | 1 | 0 | 61 | 2 | 1 | 1 | 2 | 2 | 0 | 1 | 1 | 74 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| 43 | 2 | 1 | 1 | 0 | 2 | 1 | 1 | 0 | 62 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 76 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 |
| 47 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 63 | 1 | 1 | 1 | 2 | 1 | 0 | 1 | 1 | 78 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| 48 | 2 | 1 | 0 | 2 | 1 | 1 | 1 | 0 | 64 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 79 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| 49 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 65 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 85 | 2 | 1 | 1 | 2 | 1 | 0 | 0 | 1 |
| 50 | 0 | 1 | 1 | 0 | 0 | 2 | 1 | 0 | 66 | 1 | 1 | 1 | 2 | 1 | 0 | 0 | 1 | 86 | 2 | 1 | 1 | 2 | 2 | 1 | 0 | 1 |
| 52 | 1 | 1 | 2 | 0 | 1 | 1 | 1 | 0 | 67 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 87 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 |
| 53 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 68 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 88 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |

(2) Induce the Negating Attribute Values

1) For examples 64, 70, 71, 73, 74, 78 and 88, we can conclude that "Fe=2 → C=1", whereas for example 52, we hold "Fe=2 → C=0". Namely "Fe=2 → C=1" with $\alpha = 7/8$, and "Fe=2 → C=0" with $\alpha = 1/8$ can be held. Which are the attribute value that negate the rule of "Fe=2 → C=1"? With the partial decernibility matrix we can find the factors.

Firstly, the examples with Fe=2 are selected. Secondly, generate the decernibility matrix with these examples. Thirdly, use the decernibility matrix to induce the rules that include the factors negating the rule "Fe=2 → C=1". Table 5 gives the results of the decernibility matrix.

**Table 5.** Decernibility matrix for the examples with Fe=2

|      | 64   | 70      | 71   | 73   | 74    | 78   | 88    |
|------|------|---------|------|------|-------|------|-------|
| 52   | 1,4  | 2,4,5,6 | 4    | 4,5  | 1,4,5 | 4,5  | 1,4,5 |

The attribute value that negates the rule "Fe=2 → C=1" can be induced as follows:

$(1 \lor 4) \land (2 \lor 4 \lor 5 \lor 6) \land 4 \land (4 \lor 5) \land (1 \lor 4 \lor 5) \land (4 \lor 5) \land (1 \lor 4 \lor 5)$=4 (Ca)
So for example 52, the following knowledge can be gotten:

<div style="margin-left:2em">

"Fe=2 → C=1"            with $\alpha$ =7/8;

"Fe=2 → C=0"            with $\alpha$ =1/8;

"Fe=2 ∧ Ca=0→ C=0"   with $\alpha$ =1;

"Ca=0"is the factor that negate the rule "Fe=2 → C=1"with $\alpha$ =7/8 (see Table 1).
</div>

2) For the examples 61, 63, 64, 65, 66, 67, 69, 70, 72, 74, 76, 78, 79, 85, 86, 87 and 88, we can conclude that "Ca=2 → C=1", while for example 48, we hold "Ca=2 → C=0". Similarly we get the following knowledge:

<div style="margin-left:2em">

"Ca=2 → C=1"                with $\alpha$ =17/18;

"Ca=2 → C=0"                with $\alpha$ =1/18;

"Ca=2 ∧ Fe=0 → C=0"       with $\alpha$ =1/18;

"Fe=0" is the factor that negate the rule "Ca=2 → C=1" with $\alpha$ =17/18.
</div>

3) For examples 49 and 53, we can induce that "Ca=1 → C=0 ", whereas for examples 62, 68, 71 and 73 we can get that "Ca=1 → C=1". Table 6 gives the results of the decernibility matrix.

**Table 6.** Decernibility matrix for the examples with Ca=1

|      | 62  | 68   | 71   | 73     |
|------|-----|------|------|--------|
| 49   | 7   | 7    | 3,7  | 3,5,7  |
| 53   | 6   | 6,7  | 3,6  | 3,5,6  |

With $7 \land 7 \land (3 \lor 7) \land (3 \lor 5 \lor 7)$ =7 (Na) for example 49 and with $6 \land (6 \lor 7) \land (3 \lor 6) \land (3 \lor 5 \lor 6)$=6 (K) for example 53, we can get the following knowledge:

<div style="margin-left:2em">

"Ca=1 → C=1"                with $\alpha$ =4/6;

"Ca=1 → C=0 "               with $\alpha$ =2/6;

"Ca=1 ∧ Na=2→ C=0 "       with $\alpha$ =1;

"Ca=1 ∧ K=0→ C=0 "        with $\alpha$ =1;

"Na=2" (see example 49) and "K=0" (example 53) are the factors that negate the rule "Ca=1 → C=1" with $\alpha$ =4/6.
</div>

## 6   Coronary Heart Disease Data Experiments

We have gotten 441 coronary heart disease cases from Beijing, and among all the 441 cases there are 161 ones who suffer from heart failure with 638 records in the course

of being in hospital. Among all the 638 records we get the following rules:

(1) Heart failure=true → Breath sounds in lungs = gruff (not sharp) with 192/638 and souffles in hearts =true with only 4/192;

(2) Heart failure=true → Breath sounds in lungs = decrease in soundness with 205/638 and souffles in hearts =true with 111/192;

## 7  Conclusions and Discussions

1) Three experimental results show that using the two algorithms can get the same results.

2) Since algorithm 1 contains the step of attribute reduction and the computational complexity of attribute reduction is NP hard, the computational complexity of algorithm 1 is NP hard too, whereas algorithm 2 does not need the step of attribute reduction, especially the amount of selected examples for building a decernibility matrix will be very small even there are only several ones. So the computational complexity of algorithm 2 is largely less than the one of algorithm 1.

3) The factors that negate a typical dependency embody the correlation between two rules, i.e., the factors negate a rule (dependency) whereas support another rule (dependency). This kind of knowledge differs from the exceptional rules, since the factors can give us the information of two rules, which is a kind of comparative knowledge, whereas the exceptional rule can only give us the information of one rule.

4) ANN, SVM, etc models can be viewed as "population based" as a single model is formed for the entire population (test data set), while the rough set approach follows an "individual (data object) based" paradigm. The "population based" tools determine features that are common to a population (training data set). The models (rules) created by rough set are explicit and easily understood. So for inducing easily understood knowledge, the rough set theory has an advantage over the black-box based machine learning methods such as ANN, SVM etc.

5) This kind of knowledge give us the knowledge that how a typical pattern change to the exceptional pattern.

6) The idea of decernibility matrix can be used to not only the whole data set but also the partial data. The current use of the decernibility matrix need whole data set, i.e., whether a part of data of the whole data set can be selected to be applied to the decernibility matrix for inducing particular knowledge has not been offered up to now.

## References

1. Pawlak Z.: Rough sets. Int. J. of Computer and Information Science. 1l (1982): 341-356
2. V.N. Vapnik: Statistical Learning Theory. John Wiley & Sons, (1998).
3. S.B.Cho."Pattern recognition with neural networks combined by genetic algorithm", Fuzzy sets and systems 103 (1999) 339-347,
4. Wong S K, Ziarko W.: On optimal decision rules in decision tables. Bulletin of Polish Academy of Sciences. 33(1985) 357-362