Turkish Journal of Mathematics

http://journals.tubitak.gov.tr/math/

Research Article

Turk J Math
(2020) 44: 561 – 568
© TÜBİTAK
doi:10.3906/mat-1906-67

A new method for obtaining the inconsistent elements in a decision table based on dominance principle

Yasemen UÇAN*, **Ahmet TOPAL**, **Nilgün GÜLER BAYAZIT**
Department of Mathematical Engineering, Faculty of Chemical and Metallurgical Engineering,
Yıldız Technical University, İstanbul, Turkey

Received: 18.06.2019 • **Accepted/Published Online:** 21.02.2020 • **Final Version:** 17.03.2020

Abstract: In this study, using only the dominance relation, we propose a set defining inconsistent elements in the decision table. Then, we show the accuracy of our proposition with an example. We also express the computational complexity comparisons of the proposed method with general method in terms of the number of set intersection operations and the real number comparison operations.

Key words: Rough sets, dominance relation, inconsistent element, decision table

1. Introduction

As is known, the algebra of sets is the basis of mathematics. The classic definition of a set is ”a collection of objects”. The objects are elements of the set. That is, the membership of an element in a set is crisp. However, definitions of fuzzy, rough and soft sets are different because in these sets, an element could have partial membership in a set [7, 9, 13]. Fuzzy, rough, and soft sets become popular in engineering, economics, and many other fields as in mathematics [6, 10].

Rough set theory is an approximation which is based on mathematical foundations for expressing uncertainty and vagueness. The rough set theory, introduced for the first time by Pawlak [8] in the 1980s, has been extended to information systems. A special equivalence relation constitutes the basis of classical rough set approximation (CRSA). This equivalence relation is called the indiscernibility relation. With this relation, partitions of the universe set are formed. When any subset of the universe set is given, the characteristics of the subsets are determined by the approximation operator. By means of these operators, classification and decision rules can be extracted. Identifying inconsistent elements in decision tables is an extremely important issue [11]. CRSA cannot deal with inconsistency caused by the principle of dominance, whereas dominance-based rough set approximation (DRSA) can handle this problem. First introduced by Greco et al.[2, 3], DRSA is an extensions, and preserves the good properties of CRSA. The most fundamental difference between CRSA and DRSA is that DRSA is an approach based on the dominance principle instead of the indiscernibility relation in CRSA [5]. The goal of the current work is to obtain a set that consists of inconsistent elements by using dominance relation in decision table. There are other methods for finding inconsistent elements in literature [2, 4].

*Correspondence: ucan@yildiz.edu.tr
2010 *AMS Mathematics Subject Classification:* 03E75

However, in our proposed method, inconsistent elements can be obtained with less computational complexity by using only the dominance relation without the use of approximation operators.

In this study, the basic definitions of DRSA are given in the second section. In the third section, we present the proposition describing the inconsistent elements set in the decision table based on the DRSA. Subsequently, we proof the proposition and express the computational complexity comparisons of the proposed method with general method in terms of the number of set intersection operations and the real number comparison operations.

2. Dominance-based rough set approximation (rough sets Greco et al.)

In classification problems using CRSA, preference orders of variables and the values they have taken cannot be made. Since CRSA cannot eliminate the inconsistency resulting from the dominance principle, some methodological changes have been proposed. One of these is the Dominance-Based Rough Set Approximation (DRSA) proposed by Greco et al.[1]. This approximation is an expansion of classical rough set theory based on the dominance principle.

In DRSA, a decision table is defined as an information system $BS = \{E, A, V, f\}$ where E is a finite set of objects (universe), $A = C \cup D$ and the set C is a finite set of condition variables (attributes), D is a finite set of decision variables, V_a is the domain of the variables, $V = \bigcup_{a \in A} V_a$, $f : E \times A \to V$ is a information function. The preference-ordered attributes (variables) are called criteria. For the sake of simplicity in the following definitions, we employ a singleton as the set of decision variables, i.e. $D = \{d\}$. Thus, we denote the information system as $BS = \{E, A = C \cup \{d\}, V, f\}$.

Definition 2.1 *The notion of outranking relation (preference relation) $\succcurlyeq_a$ is a pre-order relation in the domain of variables $a \in A$. For any $x_i, x_j \in E$, $x_i \succcurlyeq_a x_j$ implies that "x_i is at least as good as x_j", with respect to the variable $a \in A = C \cup \{d\}$.*

If the domain of variables a is a subset of real numbers $(V_a \subseteq \mathbb{R})$, the outranking relation is a simple ordered relation in real numbers. In this case, the following relation is provided for "the more, the better" type criteria:

$$x_i \succcurlyeq_a x_j \iff f(x_i, a) \geq f(x_j, a)$$

Definition 2.2 *Let $M = \{1, ..., r\}$. The decision criterion d containing r criteria values partitions E into r classes $Cl = \{Cl_m, m \in M\}$, where $Cl_m = \{x_i \in E : f(x_i, d) = m\}$. In other words, each element $x_i \in E$ belongs to one and only one class Cl_m.*

Definition 2.3 *Let d be a decision criterion and $\succcurlyeq_d$ be an outranking relation on E. The classes are called preference-ordered if $x_i \succcurlyeq_d x_j$ for all $t > s$, $x_i \in Cl_t$ and $x_j \in Cl_s$.*

Definition 2.4 *Upward and downward unions of classes are described by unions of decision class by means of the concepts 'at least' and 'at most', respectively as follows:*

$$Cl_m^{\geq} = \bigcup_{n \geq m} Cl_n, \quad Cl_m^{\leq} = \bigcup_{n \leq m} Cl_n$$

Definition 2.5 *Let x_i and x_j be two elements of E. It is said that x_i p-dominates x_j if "x_i is at least as good as x_j" for all $c \in P \subseteq C$, is denoted by $x_i D_p x_j$.*

Proposition 2.6 *The dominance relation D_p is a partial preorder relation.*

Definition 2.7 *Given an element $x_i \in E$ and $P \subseteq C$. The P-dominating set and P-dominated set are defined respectively, as follows:*

$$D_p^+(x_i) = \{x_j \in E | x_j D_P x_i\},$$

$$D_p^-(x_i) = \{x_j \in E | x_i D_P x_j\}.$$

Definition 2.8 *The P-lower approximations of $Cl_m^{\geq}$ and $Cl_m^{\leq}$ are defined follows, respectively.*

$$\underline{P}\left(Cl_m^{\geq}\right) = \left\{x_i \in E | D_p^+(x_i) \subseteq Cl_m^{\geq}\right\} \;\; for \;\; m = 2, ..., r,$$

$$\underline{P}\left(Cl_m^{\leq}\right) = \left\{x_i \in E | D_p^-(x_i) \subseteq Cl_m^{\leq}\right\} \;\; for \;\; m = 1, 2, ..., r-1.$$

Definition 2.9 *The P-upper approximations of $Cl_m^{\geq}$ and $Cl_m^{\leq}$ are defined follows, respectively.*

$$\overline{P}\left(Cl_m^{\geq}\right) = \left\{x_i \in E | D_p^-(x_i) \bigcap Cl_m^{\geq} \neq \varnothing\right\} \;\; for \;\; m = 2, ..., r,$$

$$\overline{P}\left(Cl_m^{\leq}\right) = \left\{x_i \in E | D_p^+(x_i) \bigcap Cl_m^{\leq} \neq \varnothing\right\} \;\; for \;\; m = 1, 2, ..., r-1.$$

Definition 2.10 *The P-boundaries of $Cl_m^{\geq}$ and $Cl_m^{\leq}$, denoted as BP($Cl_m^{\geq}$) and BP($Cl_m^{\leq}$), respectively, are defined in terms of P-lower and P-upper approximations of $Cl_m^{\geq}$ and $Cl_m^{\leq}$ as follows:*

$$BP\left(Cl_m^{\geq}\right) = \overline{P}\left(Cl_m^{\geq}\right) - \underline{P}\left(Cl_m^{\geq}\right),$$

$$BP\left(Cl_m^{\leq}\right) = \overline{P}\left(Cl_m^{\leq}\right) - \underline{P}\left(Cl_m^{\leq}\right).$$

Theorem 2.11 *Let BS be an information system in which P$\subseteq$ C. Then some properties of the lower and upper approximations are as follows [12]:*

i. $\underline{P}\left(Cl_m^{\geq}\right) \subseteq Cl_m^{\geq} \subseteq \overline{P}\left(Cl_m^{\geq}\right),$

ii. $\underline{P}\left(Cl_m^{\leq}\right) \subseteq Cl_m^{\leq} \subseteq \overline{P}\left(Cl_m^{\leq}\right),$

iii. $\underline{P}\left(Cl_m^{\geq}\right) = E - \overline{P}\left(Cl_{m-1}^{\leq}\right)(m = 2, ..., r),$

iv. $\underline{P}\left(Cl_m^{\leq}\right) = E - \overline{P}\left(Cl_{m+1}^{\geq}\right)(m = 1, ..., r-1),$

v. $\overline{P}\left(Cl_m^{\geq}\right) = E - \underline{P}\left(Cl_{m-1}^{\leq}\right)(m = 2, ..., r),$

vi. $\overline{P}\left(Cl_m^{\leq}\right) = E - \underline{P}\left(Cl_{m+1}^{\geq}\right)(m = 1, ..., r-1),$

vii. $BP\left(Cl_m^{\geq}\right) = BP\left(Cl_{m-1}^{\leq}\right) \quad (m = 2, ..., r),$

viii. $BP\left(Cl_m^{\leq}\right) = BP\left(Cl_{m+1}^{\geq}\right) \quad (m = 1, ..., r-1).$

Definition 2.12 *Let $c \in P \subseteq C$ and d be condition criteria and decision criterion, respectively. Then*
i) The decision table is called p-consistent if $x_i \succcurlyeq_c x_j$ then $x_i \succcurlyeq_d x_j$, for all $c \in P$.
ii) The decision table is called p-inconsistent if $x_i \succcurlyeq_c x_j$ then $x_j \succcurlyeq_d x_i$, for all $c \in P$.

3. On inconsistency in decision tables

The information corresponding to distinct elements is stored in the rows of data tables where the columns represent different variables(attributes). Values taken by the elements for each related variable are in the intersections of the rows and columns [6].

Proposition 3.1 *Let $BS = (E, C \cup \{d\}, V, f)$ be a decision table and $Cl = \{Cl_m, m \in M\}$, where $M = \{1, ..., r\}$ and $Cl_m = \{x_i \in E : f(x_i, d) = m\}$ is a partition of E with respect to a decision variable.*

Thus, the set $\bigcup_{i,j \in I}\left(D_p^+(x_j) \bigcap D_p^-(x_i)\right)$ consists of inconsistent elements of the decision table, where $x_i \in Cl_s$, $x_j \in Cl_t$ and $s < t$.

Proof As is known, the set of inconsistent elements is associated with $\bigcup_{t=2}^{r} BP\left(Cl_t^{\geq}\right)$ *and* $\bigcup_{s=1}^{r-1} BP\left(Cl_s^{\leq}\right)$ in [11]. Thus, it must be shown that the elements of the set defined in the proposition belong to union of boundary sets.

Let us assume that $x_k \in \bigcup_{i \neq j}\left(D_p^+(x_j) \bigcap D_p^-(x_i)\right)$.

By considering set theory, it is clear that x_k are elements of both $D_p^+(x_{j_0})$ and $D_p^-(x_{i_0})$, that is, $x_k \in D_p^+(x_{j_0})$ and $x_k \in D_p^-(x_{i_0})$ $\quad$ *for* $\exists\, i_0, j_0 \in I$).

By the reflexivity of dominance principle, we have

$$x_k \in D_p^+(x_{j_0}) \quad \Rightarrow \quad x_{j_0} \in D_p^-(x_k), \tag{3.1}$$

$$x_k \in D_p^-(x_{i_0}) \quad \Rightarrow \quad x_{i_0} \in D_p^+(x_k). \tag{3.2}$$

By definition of the upward and downward unions of classes, we have

$$x_{i_0} \in Cl_s^{\leq} \quad \text{and} \quad x_{i_0} \notin Cl_t^{\geq}, \tag{3.3}$$

$$x_{j_0} \notin Cl_s^{\leq} \quad \text{and} \quad x_{j_0} \in Cl_t^{\geq}. \tag{3.4}$$

As a consequence of (3.1), (3.2), (3.3) and (3.4), it can be seen that $x_k \notin \underline{P}\left(Cl_s^{\leq}\right)$ and $x_k \notin \underline{P}\left(Cl_t^{\geq}\right)$ from the definition of the P-lower approximation of $Cl_s^{\leq}$ and $Cl_t^{\geq}$.

Additionally, from (3.1), (3.2), (3.3) and (3.4), $D_p^+(x_k)\bigcap Cl_s^{\leq}$ and $D_p^-(x_k)\bigcap Cl_t^{\geq}$ contain at least one element each, called x_{i_0} and x_{j_0}, respectively. It means that these sets are not empty. Thus, $x_k \in \overline{P}\left(Cl_s^{\leq}\right)$ and $x_k \in \overline{P}\left(Cl_t^{\geq}\right)$ are valid for the definition of the P-upper approximation of $Cl_s^{\leq}$ and $Cl_t^{\geq}$:

$$\begin{aligned} x_k \notin \underline{P}\left(Cl_s^{\leq}\right) \quad &\text{and} \quad x_k \in \overline{P}\left(Cl_s^{\leq}\right) \Rightarrow x_k \in BP\left(Cl_s^{\leq}\right), \\ x_k \notin \underline{P}\left(Cl_t^{\geq}\right) \quad &\text{and} \quad x_k \in \overline{P}\left(Cl_t^{\geq}\right) \Rightarrow x_k \in BP\left(Cl_t^{\geq}\right), \end{aligned}$$

by using the definition of the P-boundaries of $Cl_s^{\leq}$ and $Cl_t^{\geq}$.

Hence, we get $x_k \in \bigcup_{i,j\in I}\left(D_p^+(x_j)\bigcap D_p^-(x_i)\right) \Rightarrow x_k \in \bigcup_{t=2}^{r} BP\left(Cl_t^{\geq}\right)$ and $x_k \in \bigcup_{s=1}^{r-1} BP\left(Cl_s^{\leq}\right)$.

On the other hand, assume that $x_k \in BP\left(Cl_t^{\geq}\right)$ and $x_k \in BP\left(Cl_s^{\leq}\right)$ for $s < t$.

The following statements are covered by the definition of the P-boundaries of $Cl_s^{\leq}$ and $Cl_t^{\geq}$.

$$x_k \in \overline{P}\left(Cl_t^{\geq}\right) \text{ and } x_k \in \overline{P}\left(Cl_s^{\leq}\right)$$

By using the P-upper approximation definition

$$D_p^-(x_k)\bigcap Cl_t^{\geq} \neq \varnothing \; and \; D_p^+(x_k)\bigcap Cl_s^{\leq} \neq \varnothing. \tag{3.5}$$

By considering (3.5), we can say

$$\begin{gathered} \left(x_{j_0} \in Cl_t^{\geq}\right)\left(x_{j_0} \in D_p^-(x_k)\right) \text{ for } \exists\, x_{j_0}, \\ \left(x_{i_0} \in Cl_s^{\leq}\right)\left(x_{i_0} \in D_p^+(x_k)\right) \text{ for } \exists\, x_{i_0}. \end{gathered}$$

Applying the reflexivity principle, we have

$$x_k \in D_p^+(x_{j_0}) \; and \; x_k \in D_p^-(x_{i_0}) \; for \; \exists\, x_{i_0}, x_{j_0}.$$

□

The above expression implies that $x_k \in \bigcup_{i,j\in I}\left(D_p^+(x_j)\bigcap D_p^-(x_i)\right)$.

Example: In Table, conditional variables and decision variable are pre-ordered and the candidates are classified as A (Accept) or R (Reject) with respect to the conditional variables which are defined as follows based on pre-order relations.

a_1 : Level of Piano (↑)

a_2 : Level of Violin (↑)

a_3 : Level of Trumpet (↑)

a_4 : Level of Guitar (↑)

(The symbol (↑) means that the preference order increases with the value taken by the variable.)

$Cl_1 = Cl_R = \{x_4, x_7\}$ and $Cl_2 = Cl_A = \{x_1, x_2, x_3, x_5, x_6\}$ are partitions of E= $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$

Table . Decision table for candidate selection to music high school [14].

objects	a_1	a_2	a_3	a_4	d
x_1	4	4	3	4	A
x_2	5	5	2	4	A
x_3	4	4	2	4	A
x_4	4	4	2	4	R
x_5	5	5	2	4	A
x_6	4	4	2	3	A
x_7	4	3	2	3	R

P-Dominating or P-Dominated sets of some candidates(elements) are identified, where P is the set of all conditional variables.

$D_p^+(x_1) = \{x_1\}$

$D_p^+(x_2) = \{x_2, x_5\}$

$D_p^+(x_3) = \{x_1, x_2, x_3, x_4, x_5\}$

$D_p^+(x_5) = \{x_2, x_5\}$

$D_p^+(x_6) = \{x_1, x_2, x_3, x_4, x_5, x_6\}$

$D_p^-(x_4) = \{x_3, x_4, x_6, x_7\}$

$D_p^-(x_7) = \{x_7\}$

The inconsistent elements set is determined as follows:

$D_p^+(x_1) \bigcap D_p^-(x_4) = \varnothing$

$D_p^+(x_1) \bigcap D_p^-(x_7) = \varnothing$

$D_p^+(x_2) \bigcap D_p^-(x_4) = \varnothing$

$D_p^+(x_2) \bigcap D_p^-(x_7) = \varnothing$

$D_p^+(x_3) \bigcap D_p^-(x_4) = \{x_3, x_4\}$

$D_p^+(x_3) \bigcap D_p^-(x_7) = \varnothing$

$D_p^+(x_5) \bigcap D_p^-(x_4) = \varnothing$

$D_p^+(x_5) \bigcap D_p^-(x_7) = \varnothing$

$D_p^+(x_6) \bigcap D_p^-(x_4) = \{x_3, x_4, x_6\}$

$D_p^+(x_6) \bigcap D_p^-(x_7) = \varnothing$

The union of all the sets written above consists of inconsistent elements, i.e. $\{x_3, x_4, x_6\})$ is the set of inconsistent elements in the music high school decision table. Indeed, this result has been obtained by means of calculating the lower and upper approximations in [14].

3.1. Computational complexity analysis

Let there be k classes and let P be the set of attributes. To simplify the analysis, let us further assume that each class has $\frac{n}{k}$ elements for a total of n elements.

We express the computational complexity comparisons of the proposed method with general method in terms of the number of set intersection operations and the real number comparison operations.

1. Since the $D_p^+(x)$ is not computed for Cl_1 and $D_p^-(x)$ is not computed for Cl_k in the proposed method, the savings in number of real number comparisons may be expressed as

$$\Delta_{comp}(D_p^+) + \Delta_{comp}(D_p^-),$$

where

$$\Delta_{comp}(D_p^+) = \Delta_{comp}(D_p^-) = \frac{n}{k}(n-1)|P|,$$

and $|.|$ is the cardinality of set.

Additionally, there is a complexity savings of

$$\Delta_{AND}(D_p^+) + \Delta_{AND}(D_p^-) = \frac{n}{k}(n-1),$$

in the number of logical AND's. The savings advantage increases (decreases) if more (fewer) than average number of elements reside in Cl_1 and Cl_k, but is always greater than zero.

2. Set intersection operations: We first note that the complexities of set intersection and subset check operations are equivalent and are $O(n)$ when efficiently performed by a hash table lookup. Indeed, the subset check operations in the P-lower approximations may be performed by a set intersection. Hence, the complexity savings in the set intersection operations due to employing the proposed method in place of the general method may be expressed as

$$\Delta_{\cap} = 4nk - \frac{k(k-1)}{2}\left[\frac{n^2}{k^2}\right].$$

We note that the proposed method requires fewer set intersection operations than the general method for small $\frac{n}{k}$. For instance, for $k=2$ and $n=16$, $\Delta_{\cap}=64$ whereas for $k=2$ and $n=32$, $\Delta_{\cap}=0$. On the other hand, as k grows with fixed $\frac{n}{k}$, the complexity savings of the proposed method increases, e.g. for $k=4$ and $n=32$, $\Delta_{\cap}=128$ and for $k=8$ and $n=64$, $\Delta_{\cap}=256$.
The complexity advantage of the proposed method in the number of comparisons is more significant since a comparison typicaly takes substantially more CPU time (cycles/instruction) than a table look up operation for performing set intersection or subset check operations.

4. Conclusion

The problem on finding inconsistent elements in a decision tables plays an important role on information systems. In this paper, we present a new proposition on finding inconsistent elements in a decision table on dominance principle. Then, we prove that our proposition is simpler and more understandable than previous methods.

References

[1] Greco S, Matarazzo B, Slowinski R. Multicriteria classification by dominance-based rough set approach. Università degli studi, 2000.

[2] Greco S, Matarazzo B, Slowinski R. Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 2001; 129: 1-47.

[3] Greco S, Matarazzo B, Slowinski R. Rough approximation by dominance relations. International Journals of Intelligent Systems 2002; 17 (2): 153-171.

[4] Janos D, Thi VD, Giang NL. On finding all reducts of consistent decision tables. Cybernetics and Information Technologies 2014; 14 (4): 3-10.

[5] Kacprzyk J, Pedrycz W. (Eds.). Springer Handbook of Computational Intelligence. New York, USA: Springer, 2015.

[6] Liou, JJ, Tzeng GH. A dominance-based rough set approach to customer behavior in the airline market. Information Sciences 2010; 180 (11): 2230-2238.

[7] Molodtsov D. Soft set theory-first results. Computers & Mathematics with Applications 1999; 37 (4-5): 19-31.

[8] Pawlak Z. Rough set. International Journal of Computer & Information Sciences 1982; 11 (5): 341-356.

[9] Pawlak Z, Skowron A. Rough membership functions: a tool for reasoning with uncertainty. Algebraic Methods in Logic and in Computer Science 1993; Banach Center Publications, 28, institute of mathematics polish academy of sciences warszawa, 135-150.

[10] Salama, AS. Some topological properties of rough sets with tools for data mining. International Journal of Computer Science Issues (IJCSI) 2011; 8(3): 588-594.

[11] Słowiński R, Stefanowski J, Greco S, Matarazzo B. Rough set based processing of inconsistent information in decision analysis. Control and Cybernetics 2000; 29 (1): 379-404.

[12] Yang X, Yang J. Incomplete information system and rough set theory. Models and attribute reductions. Science Press: Springer, 2012.

[13] Zadeh LA. Fuzzy sets. Information and Control 1965; 8 (3): 338-353.

[14] Zaras K. The dominance-based rough set approach (DRSA) applied to bankruptcy prediction modeling for small and medium businesses. Multiple Criteria Decision Making 2011; 287-295.