A New Method To Feature Selection In Rough Fuzzy Set Theory Based On Degree Of Separation

Narges Khazravi[a], Seyyed Majid Alavi[b*]

[a,b]Department of Mathematics, Islamic Azad University, Hamedan Branch, Iran

Correspondent: Seyyed Majid Alavi alavi_ma2004@yahoo.com

Article History: Received: 14 July 2020; Accepted: 2 January 2021; Published online: 5 February 2021

Abstract: Rough set theory (RST) is an important tool to find feature subset selection. One of the most important and challenging issues in RST is to find reducts and core. Most of the problems in many areas, including machine learning, involve high dimensional descriptions of input features. Therefore, it is not surprising to mention that several studies have been conducted on the dimensionality reduction. Feature selection refers to the problem of selecting those input features that are mostly predictive of a given result. RST can be used as a tool to discover data dependency and reduce the number of attributes contained in a data set via the data alone that require no extra information. There have been several studies in the area of finding reducts with minimal cardinal. In this paper, we have proposed the hybrid information system, in which their attributes consist of crisp and fuzzy variables. Fuzzy variables appear as linguistic variables. We first define the degree of separation between fuzzy numbers and then choose a threshold-level (γ) to clarify the objects based on attributes. Considering the threshold-level, we use discernibility matrices to find reducts and core. Experimental results show that the proposed algorithm can improve the feature selection.

Keywords: Rough Set Theory; Threshold-level; Feature Selection; Fuzzy numbers; Linguistic Variables

1. Introduction

Rough set theory is one of the most important tools in data analysis which can find hidden patterns in data. An important issue in data analysis is to discover the dependency between the attributes. For many application problems, it is often necessary to maintain a concise form of the information system. One way to implement this is to search for a minimal representation of the original data set[8]. For this purpose, the concept of a reduct is introduced and defined as a minimal subset of initial attribute. In other words, no attributes can be removed from the subset without affecting the dependency degree. A feature is mentioned to be relevant if it is predictive of decision feature(s), otherwise it is irrelevant. The main objective of feature selection (FS) is to determine a minimal feature subset from a problem domain while properly retaining the high accuracy in the representation of original features[6]. A detailed review of feature selection techniques devised for classification task can be found in (DashLiu, 1997). There are two main approaches to find reducts: one that considers the degree of dependency and on that concerns with the discernibility matrix. Heuristic methods such as [1, 3, 20, 10, 7], despite being useful and relatively quick in locating reducts, are not able to guarantee such minimal reductions. RST can be used as a tool to discover a minimal subset of initial attribute. It has become an extremely interesting topic to researchers and applied in many domains[14, 16]. It is an extension of the conventional set theory supporting approximations in decision-making, leading to application of stochastic-based approaches to this domain, such as Genetic Algorithms and extensions [4, 19], and Ant Colony Optimization[2]. By reformulating the rough set reduction task in a propositional satisfiability (SAT) framework [5], the solution techniques derived from SAT may be applied that should be able to discover such subsets and guarantee their minimality. Search algorithms based on the well-known Davis-Longman-Loveland algorithm (DPLL) have emerged as the representatives of the most efficient methods to complete SAT solvers[8]. Many applications of rough sets use

1

the discernibility matrices to find rules or reducts. By finding the set of all prime applicants of the discernibility function, all the minimal reducts of a system can be determined. Some of the researchers use clustering-based under sampling method [11]. The present article is structured as follows. First, the key concepts underpinning RST are reviewed, and the minimal reduct problem is formulated in the context of current solution methods. Next, the extension of rough set attribute reduction algorithm is proposed to optimally find discrete reducts. After that, the resulting method is extended to the continuous case by discernibility matrix. Finally the computational results are presented for the method on appropriate benchmark data [8]. Recently some authors have extended a rough set to fuzzy-rough and explained it [15]. In this paper, we proposed the hybrid information system, in which their attributes consist of crisp and fuzzy variables. Crisp variables appear as real numbers, and fuzzy variables appear as linguistic variables. Using syntactic and semantic rules, linguistic variables are generating and associating. We first defined the degree of separation between fuzzy numbers and then chose a threshold level. After that, we defined a binary corresponding to each attribute. Objects were classified based on these binary values. In this case, we use discernibility matrices to find reducts and core. In general form, this paper provides an efficient method, algorithm to find the minimal sets of data with the same knowledge as the original data. Section 2 represents the notation and basic definition in the rough set theory. Section 3 introduces the reducts and core in the information system. Section 4 represents a concept of degree of seperation between equivalence classes. The rest of this paper focuses on a novel method with an example and represents an easy algorithm.

Notations and basic definitions

In this section, we recall some basic definitions of fuzzy set and rough set theory.

2.1. **fuzzy theory.**

Definition 1. [9] A fuzzy number is a fuzzy set $u : \mathrm{R} \to I = [0,1]$ which satisfies (*i*) u is upper semiconscious.

(*ii*) $u(x) = 0$, outside some interval $[c,d]$. *iii*) There are real numbers $a,b : c \le a \le b \le d$ for which

1. $u(x)$ is monotonic increasing on $[c,a]$.
2. $u(x)$ is monotonic decreasing on $[b,d]$.
3. $u(x) = 1, a \le x \le b$

The set of all such fuzzy numbers is represented by E^1.

Definition 2. (Triangular fuzzy number) A fuzzy set A is called triangular fuzzy number with center$a$, left width $\alpha > 0$ and right width $\beta > 0$, if its membership function has the following form:

$$u(x) = \begin{cases} \frac{x-a}{\alpha} & if\ a-\alpha \le x < a \\ \frac{a-x}{\beta} & if\ a \le x < a+\beta, \\ 0 & othe\ wise \end{cases} \quad (2.1)$$

It often is denoted in brief as $u = (a, \alpha, \beta).$

Definition 3. (Trapezoidal fuzzy number) A fuzzy set A is called trapezoidal fuzzy number with tolerance interval $[a,b]$, left width α and right width β, if its membership function has the following form:

$$u(x) = \begin{cases} \frac{x-\alpha}{\alpha} & if\ a-\alpha \le x < a, \\ 1 & if\ a \le x < b, \\ \frac{b-x}{\beta} & if\ b \le x < b-\beta, \\ 0 & other\ wise. \end{cases} \quad (2.2)$$

Since the trapezoidal fuzzy numbers are completely characterized by four real numbers a,b,α,β, it is often denoted in brief as $u = (a,b,\alpha,\beta)$. If $a = b$ trapezoidal fuzzy number convert to triangular fuzzy number and is denoted by $u = (a,a,\alpha,\beta)$.

Definition 4. Following [13], we represent an arbitrary fuzzy number by an ordered pair of functions $(\underline{u}(r),\overline{u}(r))$; $0 \le r \le 1$, which satisfy the following requirements,

1): $\underline{u}(r)$ is a bounded left-continuous non- decreasing over [0,1].

2): $\overline{u}(r)$ is a bounded left-continuous non- increasing over [0,1]. **3):** $\underline{u}(r) \le \overline{u}(r)$, $0 \le r \le 1$.

Definition 5. The (crisp) set of elements that belong to the fuzzy number u which comprises all elements of < whose grade of membership in u is grater than or equal to r is called the r-level set and denoted by $[u]_r$ where $[u]_r = [\underline{u}(r),\overline{u}(r)]$.

Let $u = (a,\alpha,\beta)$ be a triangular fuzzy number then in parametric representation we have $u = (a - (1-r)\alpha, a + (1-r)\alpha)$ and $[u]_r = [(a - (1-r)\alpha, a + (1-r)\alpha)]$, if $u = (a,b,\alpha,\beta)$ be a trapezoidal fuzzy number then in parametric representation we have $u = (a-(1-r)\alpha, b+(1-r)\beta)$ and $[u]_r = [a - (1-r)\alpha, b + (1-r)\beta]$.

Definition 6. (support) Let u be a fuzzy number, the support of u, denoted $supp(u)$, is the crisp subset of < and is defined as below:

$$supp(u) = \{x \in < | u(x) > 0\}. \quad (2.3)$$

If u is a triangular(trapezoidal) fuzzy number then $supp(u) = (a - \alpha, a + \beta)(a - \alpha, b + \beta))$.

Definition 7. Let u and v are fuzzy numbers. We say that u j v if $u(x) \le v(x)$ for all $x \in <$.

Definition 8. Let $u = (\underline{u}(r),\overline{u}(r))$ and $v = (\underline{v}(r),\overline{v}(r))$ are two arbitrary fuzzy numbers, define $\delta(u,v)$ as below:

$$\delta(u,v) = \begin{cases} 1 - \max\{r \in [0,1] | [u]_r \cap [v]_r \neq \emptyset\} & Supp(u) \cap Supp(v) \neq \emptyset, \\ 1 & O.w, \end{cases}$$

that is called the degrees of separation between two fuzzy numbers. Due to definition7 if u, v are two arbitrary fuzzy number such that $u \subseteq v$ then $\delta(u,v) = 0$. Similarly if u_0 and v_0 are two arbitrary variables with crisp values(real, nominal, binary,...) then we define $\delta(u_0,v_0)$ as

$$\delta(u_0,v_0) = \begin{cases} 1 & u_0 \neq v_0, \\ 0 & O.w. \end{cases}$$

2.2. **Linguistic Variable.** One can use the fuzzy sets to represent linguistic variables. A linguistic variable can be considered as a variable whose value is a fuzzy set.

Definition 9. A linguistic variable is characterized by a quintuple $(x,T(x),U,G,M)$, in which x is the name of the variable, $T(x)$ is the term set of x, with each value being a fuzzy number defined on U, G is a syntactic rule for generating the names of values of x and M is a semantic rule for associating with each value its meaning

For example if $x = Age$ and $U = [7,65]$ then the term sets of linguistic variable *Age* i.e., $T(Age)$ may be as $T(Age) = \{baby, young, middle-age, old\}$, then we define the meaning of a fuzzy value such as *young* by a membership function in the form of trapezoidal or triangular fuzzy numbers as follow: $baby = (7,14,0,4)$, $young = (18,25,4,5)$, $middle-aged = (30,40,5,5)$ and $old = (45,60,5,0)$. Fig.1, represents the compatibility of *baby*, *young*, *middle aged* and *old* respectively. As we see in

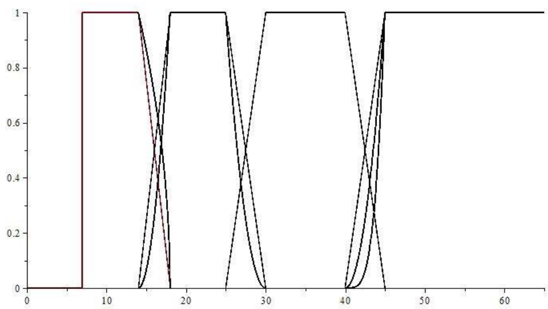

Figure 1. The components of Linguistic variable Age from left to right "baby, young, middle-aged, old"

Fig.1 for all $r > 0.5$, $[baby]_r \cap [young]_r = \emptyset$, $[young]_r \cap [middle-aged]_r = \emptyset$ and $[middle-aged]_r \cap [old]_r = \emptyset$. If A be a domain of a linguistic variable we can apply some words such as *very*(A), *more or less*(A), *very very*(A), etc. The membership function of theme, can be defined respectively

as $(A(x))^2$, $\sqrt{A(x)}$, $(A(x))^4$, etc. See Fig.2. Due to definition 8 we have $\delta(baby, middle-aged) = \delta(middle-aged, old) = 0.5$, $\delta(baby, more\ or\ less\ baby) = 0.6180$, $\delta(middle-aged, very\ old) = 0.3819$ and $\delta(middle-aged, very\ very\ old) = 0.2755$, $\delta(old, very\ old) = \delta((old, very\ very\ old)) = 0$, etc.

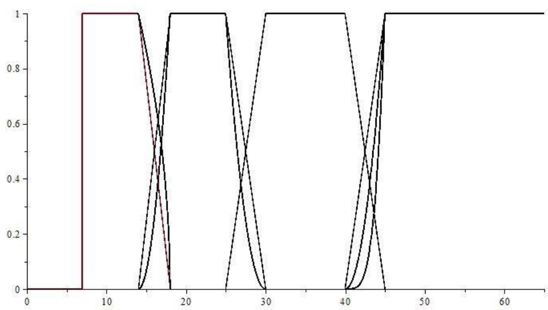

Figure 2. The components of Linguistic variable Age from left to right "baby, more or less baby, young, very young, middle-aged, old, very old, very very old"

2.3. **partial order set.** Let U be any set and Λ be some collections of its subsets, we say that the relationship R_b on Λ is a binary relation if for all $X, Y \in \Lambda$ that state $(X,Y) \in R_b$ is true or false.

Definition 10. A binary relation R_b on a set Λ is called

1- Reflexive, if and only if, for all $X \in \Lambda$, $(X,X) \in R_b$, 2- Antisymmetric, if and only, if for all elements X and Y of Λ, whenever $(X,Y) \in R_b$ and $(Y,X) \in R_b$ then $X = Y$,

3- Transitive, if $(X,Y) \in \Lambda$ and $(Y,Z) \in \Lambda$ then $(X,Z) \in \Lambda$.

Definition 11. A binary relation R_b on Λ is a partial order on Λ if and only if R_b is reflexive, antisymmetric and transitive. The set Λ together with a partial order R_b is called partially order set and denoted it by (Λ, R_b).

2.4. **Rough set theory.** An approximation space is a pair (U,R), where U is a nonempty finite set, that is called universe set and R is an equivalence relation defined on U. For each $x \in U$ define $[x]_R$, the equivalence class of x, as follows:

$[x]_R = \{y \in U | (x,y) \in R\}$

Definition 12. [12] Let $S = (U,R)$ be an approximation space and X be a subset of U, the lower approximation of X by R in S is defined as $\underline{R}X = \{x \in U | [x]_R \subseteq X\}$ and the upper approximation

—

of X by R in S is defined as $RX = \{x \in U | [x]_R \cap X =6 \emptyset\}$.

A pair $S = (U,A)$ where $A = A_l \cup A_c$ in which A_l is a nonempty finite set of attributes that appear as fuzzy numbers(linguistic variable) and A_c is a finite set of attributes that appear as crisp value(real, nominal, ordinal,...). Making an equivalence relation is called *hybrid information system*. For every $a \in A$ we have $a : U \to V_a$ where V_a is called *domain of attribute a, if* $X \subseteq U$ *then* $a(X) = \{a(x)|x \in X\}$. *Through this paper we may assign fuzzy value to some attributes. Note that a special case of hybrid information system appears as* $S = (U, C \cup D)$ *where* $C = C_l \cup C_c$ *is the* hybrid condition attributes *and D is the* decision attributes *called* hybrid decision table. *Equivalence relations is a way to break up a set U into a union of disjoint subsets. Let A be an equivalence relation on U and* $S = (U,A)$ *be an hybrid information system, with any* $B \subseteq A$*, in which* $B = B_l \cup B_c$ *there is an associated equivalence relation*

$$IND(B) = \{(x,y) \in U \times U | \forall b \in B_l \cup B_c, \delta(b(x),b(y)) = 0\} \qquad (2.4)$$

is an equivalence relation on U called indiscernibility relation. *One can see that* $U/IND(B) = [x]_{IND(B)}$ *is partition of U generated by IND(B). Here for simplicity denote* $U/IND(B)$ *and* $[x]_{IND(B)}$ *respectively by* U/B *and* $[x]_B$.

—

A subset X of U is mentioned to be R-definable in S if and only if $\underline{R}X = RX$*. The boundary set*

—

is $R(X) - \underline{R}(X)$ *and denote it by* $BN_R(X)$*. It consist of those objects that we cannot certainly classify in to X in R. A subset X of U is called* rough set *if* BN_RX 6= $\emptyset$*, otherwise the set X is called crisp with respect to R.*

Definition 13. (Accuracy of Approximation:) Let $S = (U,A)$ be a hybrid information system where $A = A_l \cup A_c$ and X be a subset of universe set U and $B \subseteq A$ then accuracy of approximation X by B is defined as follow:

$$\gamma_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|}, \qquad (2.5)$$

where $|.|$ denotes the cardinality of a set.

If $\alpha_B(X) = 1$*, then X is crisp with respect to attributes in B, if* $\alpha_B(X) < 1$ *then X is rough with respect to attributes in B.*

Definition 14. Let C and D be a subset of A. It is mentioned that D depends on C in a degree $k (0 \leq k \leq 1)$, denoted by $C \to_k D$, if

$$k = \gamma(C,D) = \frac{|POS_C(D)|}{|U|},$$

where $POS_C(D) = {}^{S}{}_{X \in U/D} \underline{C}(X)$, is called *C-positive region* of D. Note that $k = 1$ means that D depends, totally, on C and $k < 1$, means that D depends partially(in a degree k) on C.

Example **1.** Consider below hybrid decision table in which $C = \{a,b,c\}$ is a set of hybrid condition attributes where $C_l = \{a,b\}$ are linguistic variables and $C_c = \{c\}$ is crisp attribute. Domain of theme are $V_a = \{a_1, a_2, a_3, very(a_1), more\ or\ less(a_3), very(a_3)\}$ in which $a_1 = (0,1,0,1)$, $a_2 = (2,1,1)$, $a_3 = (3,4,1,0)$ are triangular or trapezoidal fuzzy numbers.

$V_b = \{b_1,b_2,b_3,more\ or\ less(b_1),very(b_1),very(b_3)\}$ in which $b_1 = (2,3,0,1)$, $b_2 = (4,5,1,1)$,

Table 1

Segment	*a*	*b*	*c*	*e = class*	*frequency*
U_1	a_1	b_3	c_2	*yes*	30
U_2	$v(a_1)$	b_2	c_3	*no*	50
U_3	$ml(a_1)$	b_1	c_1	*yes*	70
U_4	a_1	$v(b_1)$	c_1	*no*	80
U_5	a_2	$v^2(b_1)$	c_2	*yes*	50
U_6	a_3	b_1	c_3	*no*	50
U_7	a_2	b_2	c_2	*yes*	50
U_8	$ml(a_2)$	$v^2(b_1)$	c_2	*yes*	35
U_9	$v^2(a_1)$	$v^2(b_3)$	c_2	*yes*	35
$U10$	$v^2(a_3)$	b_1	c_3	*no*	20

$b_3 = (6,8,1,0)$ are trapezoidal fuzzy numbers, also $V_c = \{c_1,c_2,c_3\}$ in which V_c is crisp set, are condition attributes on $U = \bigcup_{i=1}^{10} U_i$ where U_i is *ith* segment of U. For simplicity in decision table we replace *very*(a_i), *very very*(a_i) and *more or less*(a_i) respectively by $v(a_i)$, $v^2(a_i)$ and $ml(a_i)$, etc.

Due to 2.4

$U/C = \{W_1,W_2,W_3,W_4,W_5,W_6\}$,

in which $W_1 = \{U_1,U_9\}$, $W_2 = \{U_2\}$, $W_3 = \{U_3,U_4\}$, $W_4 = \{U_5,U_8\}$, $W_5 = \{U_6,U_{10}\}$, $W_6 = \{U_7\}$. Note that for all $x,y \in W_i$ we have $\delta(a(x),a(y)) = \delta(b(x),b(y)) = \delta(c(x),c(y)) = 0$. We observe that $D = \{e\}$ is decision attribute that it's domain is $V_D = \{yes,no\}$, then $U/D = \{D_1,D_2\}$ where $D_1 = \{U_1,U_3,U_5,U_7,U_8,U_9\}$ and $D_2 = \{U_2,U_4,U_6,U_{10}\}$ then $\underline{C}(D_1) = W_1 \cup W_4 \cup W_6 = \{U_1,U_5,U_7,U_8,U_9\}$ and $C(D_2) = W_2 \cup W_5 = \{U_2,U_6,U_{10}\}$. According to definition, $POS_C(D) =$

$\underline{C}(D_1) \cup \underline{C}(D_2)$, then $\overline{k} = \gamma(C,D) = \frac{|POS_C(D)|}{|U|} = \frac{320}{470} = 0.68$

Reducts and Core

Let $S = (U,A = C \cup D)$ *where* $C = C_l \cup C_c$ *be a hybrid decision table and* $\hat{c} \in C$ *then attribute* $\hat{c}$ *is called* dispensable *in S if* $POS_C(D) = POS_{C-\{\hat{c}\}}(D)$, *else* $\hat{c}$ *is called* indispensable, *in addition* $S = (U,A = C \cup D)$ *is independent if all* $\hat{c} \in C$ *are indispensable.*

Definition 15. Suppose $S = (U,A = C \cup D)$ where $C = C_l \cup C_c$ be a hybrid decision table. A subset R of C is called reduct of C if $S = (U,A = R \cup D)$ is independent and $POS_C(D) = POS_R(D)$.

Notice that a hybrid decision table may have many attribute reducts. The set of all reducts of C is denoted by $RED(C)$, *in other word,*

$$RED(C) = \{R \subseteq C | \gamma(R,D) = \gamma(C,D), \forall B \subset R, \gamma(C,D) 6= \gamma(C,B)\}. (3.1)$$

The intersection of all reducts of C is called the core of C, i.e. $Core(C) = {}^{T}Red(C)$.

δ-Matrix Let $S = (U,A = C \cup D)$ *where* $C = C_l \cup C_c$ *be a hybrid decision table and* $C_l =$

$\{a_1,a_2,...,a_{pl}\}$ *be a set of fuzzy condition attributes(Linguistic variables),* $C_c = \{a_{pl}+1,a_{pl}+2,...,a_p\}$ *is a set of crisp condition attributes and* $U/C = \{W_1,W_2,...,W_k\}$. *The δ-matrix of* $U = \{U_1,U_2,...,U_n\}$ *is a symmetric* $n \times n$ *matrix,* M_δ, *that their entries are given as:*

$(M_\delta)_{i,j} = (\delta(a_1(x),a_1(y)),\delta(a_2(x),a_2(y)),...,\delta(a_p(x),a_p(y)))$.

where $a_i \in C_l \cup C_c$ for i = 1,2,...,p. For convenience, we take $\delta_{i,j}^{a} = \{\delta(a(x),a(y)) | (x,y) \in$

$U_i \times U_j\}$ that are entries of δ-matrix, in other word

$(M_\delta)_{i,j} = (\delta_{i,j}^{a_1}, \delta_{i,j}^{a_2}, \ldots, \delta_{i,j}^{a_p})$.

notation in this matrix is used to show that U_i and U_j are in the same class. If the corresponding decisions of segment U_i and segment U_j are the same, then we do not assign $\delta_{i,j}$ in δ- Matrix, although, some attributes in objects U_i and U_j differ. In other word, if there exists a $D_s \in U/D$ such that U_i and U_j belong to D_s, the corresponding element in this matrix is removed. Due to example 1, δ-Matrix can be obtain as below:

Table 2

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
U_1	−	0,0.5,1	0,1,1	0,1,1		1,1,1				1,1,1
U_2	*	−	0,0.5,1	0,0.62,1	0,0.72,1		0.62,0,1	1,0.72,1	0,0.72,1	
U_3	*	*	−		0.32,0,1	1,0,1	0.32,0.5,1	1,0,1	0,1,1	1,0,1
U_4	*	*	*	−	0.5,0,1	1,0,0	0.5,0.62,1	1,0,1	0,1,1	1,0,1
U_5	*	*	*	*	−	0.5,0,1				0.72,0,1
U_6	*	*	*	*	*	−	0.5,0.5,1	0,0,1	1,1,1	
U_7	*	*	*	*	*	*	−			0.72,0.5,1
U_8	*	*	*	*	*	*	*	−		0,0,1
U_9	*	*	*	*	*	*	*	*	−	1,1,1
U_{10}	*	*	*	*	*	−	*	*	*	−

Definition 16. Let $S = (U,A = C \cup D)$ where $C = C_l \cup C_c$ be a hybrid decision table and $C_l =$

$\{a_1,a_2,...,a_{pl}\}$ be a set of fuzzy condition attributes(Linguistic variables), $C_c = \{a_{pl}+1,a_{pl}+2,...,a_p\}$ is a set of crisp condition attributes. Assuming $U/C = \{W_1,W_2,...,W_k\}$, then for each $a \in C_l \cup C_c$ define

$$\Delta^a(W_i,W_j) = \min\{\delta^a(U_i,U_j)|(U_i,U_j) \in (W_i \times W_j)\}, \qquad (3.2)$$

where $\delta^a(U_i,U_j) = \delta^a(a(x),a(y),(a(x),a(y)) \in (U_i,U_j)$.

Using this relation we define the Δ-matrix of $U/C = \{W_1,W_2,...,W_k\}$ that is a symmetric $k \times k$ matrix, Δ, that their entries are given as:

$\Delta i,j = (\Delta a1(Wi,Wj),\Delta a2(Wi,Wj),...,\Delta ap(Wi,Wj))$

For instance in previous decision table we have, $W_1 = \{U_1,U_9\}$, $W_2 = \{U_2\}$ *then* $\Delta^a(W_1,W_2) = \min\{\delta^a(U_1,U_2),\delta^a(U_2,U_9)\} = \min\{0,0\} = 0$, $\Delta^b(W_1,W_2) = \min\{\delta^b(U_1,U_2),\delta^b(U_2,U_9)\} = \min\{0.5,0.75\} = 0.5$ *and* $\Delta^c(W_1,W_2) = \min\{\delta^c(U_1,U_2),\delta^c(U_2,U_9)\} = \min\{1,1\} = 1$, *then* $\Delta_{1,2} = \{0,0.5,1\}$. *As this process continues, the following table is obtained*

Threshold-level Matrix We say that two class W_i and W_j, approximately differ in attribute $a \in A$ in γ− level, whenever $\Delta^a_{i,j} \geq \gamma,\ \gamma \in [0.5,1]$ *is the threshold level. First corresponding to each attribute $a \in A$, define a binary variable as follow:*

$$b_\gamma(\Delta^a_{i,j}) = \begin{cases} 1 & \Delta a i,j \geq \gamma, \\ 0 & O.w. \end{cases} \qquad (3.3)$$

Table 3

	W_1	W_2	W_3	W_4	W_5	W_6
W_1	−	0,0.5,1	0,1,1		1,1,1	
W_2	*	−	0,0.5,1	1,0.72,1		0.62,0,1
W_3	*	*	−	0.32,0,1	1,0,0	0.32,0.5,1
W_4	*	*	*	−	0,0,1	

W_5	*	*	*	*	−	0.5,0.5,1
W_6	*	*	*	*	*	−

By considering all the features, we make the threshold-level matrix that is obtained from ∆−matrix. The threshold-level matrix of $U/C = \{W_1,W_2,...,W_k\}$ *is a symmetric* $k \times k$ *matrix,* T^γ*, that their entries are given as:*

$$T^{\gamma}_{i,j} = (b_\gamma(\Delta^{a_1}_{i,j}), b_\gamma(\Delta^{a_2}_{i,j}), \ldots, b_\gamma(\Delta^{a_p}_{i,j})). \quad (3.4)$$

For each γ*, threshold-level matrix become a discernibility matrix, due to the example 1, thresholdlevel matrix with* ($\gamma = 0.6$) *and* ($\gamma = 0.5$) *respectively can be obtain as below:*

Table 4. 0.6-discernibility matrix

	W_1	W_2	W_3	W_4	W_5	W_6
W_1	−	0,0,1	0,1,1		1,1,1	
W_2	*	−	0,0,1	1,1,1		1,0,1
W_3	*	*	−	0,0,1	1,0,0	0,0,1
W_4	*	*	*	−	0,0,1	
W_5	*	*	*	*	−	0,0,1
W_6	*	*	*	*	*	−

Table 5. 0.5-discernibility matrix

	W_1	W_2	W_3	W_4	W_5	W_6
W_1	−	0,1,1	0,1,1		1,1,1	
W_2	*	−	0,1,1	1,1,1		1,0,1
W_3	*	*	−	0,0,1	1,0,0	0,1,1
W_4	*	*	*	−	0,0,1	
W_5	*	*	*	*	−	1,1,1
W_6	*	*	*	*	*	−

Definition 17. let $S = (U,C \cup D)$ be a decision table where C is a nonempty finite set of condition attributes ,i.e, $C = \{a_1,a_2,...,a_k\}$, based on classical set theory, a subset R of C is defined by its characteristic function χ_R, that is a mapping from the elements of C to the set $\{0,1\}$ by

$$\chi_R : C \longrightarrow \{0,1\}$$

$$\chi_R(a) = \begin{cases} 1 & a \in R, \\ 0 & O.W \end{cases}$$

Let $P(C) = \{R | R \subseteq C\}$ *be the power set of* C*, based on definition17, each* $R \in P(C)$ *can be obtained by* $R = \{(a_1,\chi_R(a_1)),(a_2,\chi_R(a_2)),...,(a_p,\chi_R(a_k))\}$ *in which* $\chi_R(a_i) \in \{0,1\}$*. To facilitate the computation, a one to one correspondence function*(bijection function) *is defined as*

$$\varphi : P(C) \rightarrow \{0,1\}^k$$

$$\varphi(R) = (\chi_1,\chi_2,...,\chi_k), \quad (3.5)$$

in which for each $i = 1,2,...,k;$ $\chi_i = \chi_R(a_i) \in \{0,1\}$*. Then, each* $R \in P(C)$ *is mapped to exactly one element of the* $\{0,1\}^p$ *and each* $(\chi_1,\chi_2,...,\chi_p)$ *is mapped to exactly one* R *of* $P(C)$*. In other word,* $\varphi : P(C) \rightarrow \{0,1\}^p$ *is invertible and*

$$R = \varphi^{-1}(\chi_1,\chi_2,...,\chi_p) = \{(a_1,\chi_1),(a_2,\chi_2),...,(a_k,\chi_k)\} \quad (3.6)$$

is an equivalence relation such that $R \subseteq C$. Due to table 3, one can see that according to $\gamma = 0.6$, C_1 and C_2 approximately differ in $\varphi^{-1}(\{b_\gamma(\Delta^a), b_\gamma(\Delta^b), b_\gamma(\Delta^c)\}) = \{(a,0),(b,0),(c,1)\} = \{c\}$, but according to $\gamma = 0.5$, C_1 and C_2 approximately differ in

$\varphi^{-1}(\{b_\gamma(\Delta^a), b_\gamma(\Delta^b), b_\gamma(\Delta^c)\}) = \{(a,0),(b,1),(c,1)\} = \{b,c\}$.

Each component points to all the attributes that make object i and object j different. In order to find the reduct we can apply the decision-relative discernibility. Accordingly, A. Skowron et al [17], *introduce the discernibility function. According to their definition, a discernibility function f_D is a boolean function of m boolean variables $b_\gamma(\Delta^{a1}), b_\gamma(\Delta^{a2}), ..., b_\gamma(\Delta^{am})$(corresponding to attributes $a_1, a_2, ..., a_m$) defined as below:*

$$f_D(b_\gamma(\Delta^{a1}), b_\gamma(\Delta^{a2}), ..., b_\gamma(\Delta^{am})) = \wedge\{\vee T_{i,j}^\gamma \mid 1 \leq j \leq i \leq k, T_{i,j}^\gamma 6= 0\}, \quad (3.7)$$

in which $\vee T_{i,j}^\gamma = (b_\gamma(\Delta_{i,j}^{a_1}) \vee b_\gamma(\Delta_{i,j}^{a_2}) \vee \ldots \vee b_\gamma(\Delta_{i,j}^{a_p}))$. Note that 3.7 represent all reducts of C based on γ. Now we define $S_{i,j}$ as

$$S_{i,j}^\gamma = \phi^{-1}(b_\gamma(\Delta_{i,j}^{a_1}), b_\gamma(\Delta_{i,j}^{a_2}), \ldots, b_\gamma(\Delta_{i,j}^{a_k})). \quad (3.8)$$

Theorem 3.1. *Let $S = (U, A = C \cup D)$ where $C = C_l \cup C_c$ be a hybrid decision table such that $C_l =$*

$\{a_1, a_2, ..., a_{pl}\}$ is a set of fuzzy condition attributes(Linguistic variables), $C_c = \{a_{pl}+1, a_{pl}+2, ..., a_p\}$ is a set of crisp condition attributes and $U/C = \{W_1, W_2, ..., W_k\}$. Assuming $\gamma_1 \leq \gamma_2$ and R be a reduct of C based on threshold-levels γ_2 then R is a a reduct of C based on threshold-levels γ_1.

Proof. Assume $\gamma_1 \leq \gamma_2$ and R is a reduct of C based on threshold-levels γ_2. Let $a \in R$ then for some $i,j \in \{1,2,...,k\}$, $b_{\gamma 2}(\Delta^a_{i,j}) = 1$, due to 3.3 we have $\Delta^a_{i,j} \leq \gamma_1$ then $\Delta_{i,j}{}^a \geq \gamma_2$ therefor

$b_{\gamma_1}(\Delta^a_{i,j}) = 1$. Hence R is a redut reduct of C based on threshold-levels γ_1.

Example **2.** Due to table 3 and table 3, reducts based on $\gamma = 0.6$ and $\gamma = 0.5$ are obtained respectively as below: $f_D(b_{0.6}(\Delta^a), b_{0.6}(\Delta^b), b_{0.6}(\Delta^c)) = \{0\vee0\vee1\}\wedge\{0\vee0\vee1\}\wedge\{0\vee1\vee1\}\wedge\{1\vee1\vee1\}\wedge\{1\vee0\vee1\} = \{1,0,1\}$.

Therefore reduct is

$R = \varphi^{-1}(1,1,0) = \{(a,1),(b,0),(c,1)\} = \{a,c\}$

Similarly

$f_D(\Delta^a(0.5), \Delta^b(0.5), \Delta^c(0.5)) = \{0\vee1\vee1\}\wedge\{1\vee1\vee1\}\wedge\{1\vee0\vee1\}\wedge\{0\vee0\vee1\}\wedge\{1\vee0\vee0\} = \{1,0,1\}$.

Thus reduct is

$R = \varphi^{-1}(1,0,1) = \{(a,1),(b,0),(c,1)\} = \{a,c\}$

By finding the set of all prime implicants of the discernibility function, all the minimal reducts of a system may be determined.

conclusion

This paper represents a simple method for feature selection with hybrid attribute(Fuzzy and Crisp). Each reduct is corresponds to a minimal element of feasible solutions. Due to the properties of rough set theory. The reducts and core are obtained by binary operations. The proposed method is an extremely useful tool in handling the rough set theory. In the future work, the proposed algorithm can be extended under the neighborhood concept. The efficiency of the proposed method is illustrated by some examples.

References

1. H. Chen, T. Li, X. Fan, C. Luo, Feature selection for imbalanced data based on neighborhood rough sets, Information sciences, 483(2019) 1-20.
2. Y. Chen, D. Miao, R. Wang, A rough set approach to feature selection based on ant colony optimization, Pattern Recognition Letters 31 (3) (2010) 226-232.
3. Chouchoulas, Q. Shen, Rough set-aided keyword reduction for text categorisation, Applied Artificial Intelligence 15 (9) (2001)843-873.
4. K. Das, S. Sengupta, S. Bahttacharyya, A group incremental feature selection for classification using rough set theory based genetic algorithm, Applied Soft Computing 65(2018) 400-411.

5. M. Davis, G. Logemann, D. Loveland, A machine program for theorem proving, Communications of the ACM 5 (1962) 394-397.
6. R. Jensen, Q. Shen, Rough set-based feature selection: A review,All content following this page was uploaded by Richard Jensen on 29 May 2014.
7. R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: rough and fuzzy-rough based approaches, IEEE Transactions on Knowledge and Data Engineering 16 (12) (2004) 1457-1471.
8. R. Jensen, A. Tuson, Q. Shen, Finding rough and fuzzy-rough set reducts with SAT, Information Sciences 255 (2014) 100-120.
9. Klir GJ, Clair US, Yuan B. Fuzzy set theory: foundations and applications. Prentice-Hall Inc.; 1997
10. T. Y. Lin, P. Yin, Heuristically fast finding of the shortest reducts, rough sets and current trends in computing, Lecture Notes in Computer Science 3066(2004) 465-470.
11. W.C. Lin, C.F. Tsai, Y.H. Hu, J.S. Jhang, Clustering-based under sampling in class-imbalanced data, Inf. Sci. (Ny) 409?410 (2017) 17-26, doi: 10.1016/j.ins. 2017.05.008
12. Z. Lu, Z. Qin,Y. Zhang, J. Fang, A fast feature selection approach based on rough set boundary regions,Pattern Recognition Letters 36 (2014) 81-88.
13. Ma M, Friedman M, Kandel A. A new fuzzy arithmetic. Fuzzy Sets Syst 1999;108:83–90.
14. N. Mac Parthal?in, R. Jensen, Unsupervised fuzzy-rough set-based dimensionality reduction, Information Sciences 229 (2013) 106-121.
15. Moayedikia, K.L. Ong, Y.L. Boo, W.G. Yeoh, R. Jensen, Feature selection for high dimensional imbalanced class data using harmony search, Eng. Appl. Artif. Intell. 57 (2017) 38-49, doi: 10.1016/j.engappai.2016.10.008
16. .
17. Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishing, Dordrecht, 1991.
18. Skowron and C. Rauszer. The discernibility matrices and functions in information systems. In [340], pp. 331?362. 1992.
19. Starzyk, J.A., Nelson, D. E. Sturtz, K. A Mathematical Foundation for Improved Reduct Generation in Information systems. Journal of Knowledge and Information Systems, 2,(2) (2000) pp.131-146
20. J. Wroblewski, Finding minimal reducts using genetic algorithms, in: Proceedings of the 2nd Annual Joint Conference on Information Sciences, 1995, pp. 186-189.
21. N. Zhong, J. Dong, S. Ohsuga, Using rough sets with heuristics for feature selection, Journal of Intelligent Information Systems 16 (3) (2001) 199-214.