The CSI Journal on
Computer Science and Engineering
Vol. 19, No. 2, 2025
Pages 45-51
doi.org/10.22034/jcse.2023.181962
Regular Paper

Using integer programming to rough set-based feature selection: An approach to find all reducts respectively

Seyed Majid Alavi[1] Narges Khazravi[1]

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

Abstract

Rough set theory (RST) is an important tool for finding feature subset selection. One of the most critical and challenging issues in RST is to find reducts and core. Since most applied sciences involve high-dimensional descriptions of input features, a large amount of research has been conducted on dimensional reduction. Feature Selection refers to the process of selecting the input features leading to the most predictable results. On the other hand, RST can be adopted to discover data dependencies and reduce the number of attributes in a data set using the data alone, requiring no extra information. Therefore, in this paper, we proposed a straightforward approach for feature subset selection through binary integer linear programming (BILP). Optimal solutions to the result of this problem in reducts that lead to feature subset selection. All reducts are obtained from the smallest cardinality to the largest cardinality, respectively. Also, to get the optimal solutions for BILP, we dealt with the Branch and Bound method and Genetic Algorithm. The steps of our approach are illustrated by an example.
Keywords: Rough Set Theory; Reduct; Core; Binary Integer linear programming; Feature Selection; Decision system

1. Introduction

One of the important issues in data mining is discovering the relationship between features of information systems. Some features are dependent on others ones. In many application problems, it is necessary to identify the main features and remove the dependent features. Construction of a partial order set is a way to find its minimal elements [1]. Moreover, one of the criteria for selecting features is that the accuracy of a predictive model
is not reduced [2]. Some researchers use feature selection to solve regression problems [3]. U. Stanczyk et al. study feature selection and reduction process for induction of decision rule with classical rough set [4]. Utkarsh et al. provided an overview of feature selection techniques and investigated the instability of the feature selection algorithm [5]. The significance of features in information systems is defined based on the upper and lower boundary of regions. The uncertainty of feature selection is affected by various parameters. Some researchers have used meta-heuristic algorithms to find the optimal parameters [6], [7], [8], [9], [10] . One of the tools to find reducts of initial attributes is to use rough set theory. This theory has attracted the attention of many researchers and is used in many fields. [11], [12], [13]. By reformulating the rough set reduction task in a propositional satisfiability (SAT) framework [14], solution techniques from SAT may be applied that should be able to discover such subsets, guaranteeing their minimalist. The algorithms based on the Davis-Logemann-Loveland algorithm have been emerging as representatives of the most effective to complete SAT solvers [1]. In many applied sciences related to data mining, the discernibility matrices are used for finding rules or reducts. After finding the set of all main concepts of the discernibility function, all the reducts of a system may be determined. Liyang Gao et al. propose the relevance assignation Feature Selection (RAFS) method based on the mutual information theory, which assigns the relevance evaluation according to the redundancy [15]. He Jiang et al. work on simultaneous feature selection and clustering based on square root optimization [16]. The resulting method is further extended to the continuous case by discernibility matrix. M. kelidari et al. use chaotic cuckoo optimization algorithm to find reduct[17]. In this paper, we represent a modified discernibility matrix and construct a system of linear inequality with binary variables.

We define a property partial order relation on the feasible solutions of this system. Minimal elements of this set result reducts and core. Reducts with minimum cardinality, can be found as binary integer linear programming (BILP). This property represents the selection or rejection of an opinion. We will apply special methods such as branch and bound algorithms with a divide-to-conquer technique or Genetic algorithm to obtain solutions. The continuation of this paper is structured as follows. Section 2 represents notation and basic definition in rough set theory. Section 3 introduces reduct and core in an information system. In section 4 we represent a concept of distance between equivalence classes. The rest of this paper focus on the novel method with an example and an easy algorithm to evaluate the proposed method.

2. Notations and basic definitions

2.1 partial order set

Partial order is a relation, $\preceq$, on a set, Ψ, so that $\preceq$ is reflexive (for all $x \in \Psi, x \preceq x$), anti-symmetric (for all elements x and y of Ψ , whenever $x \preceq y$ and $y \preceq x$ then $x = y$) and transitive (if $x \preceq y$ and $y \preceq z$ then $x \preceq z$). A set Ψ with partial order $\preceq$ is called partial order set (POS). Two elements x and y of Ψ are called comparable, if $x \preceq y$ or $y \preceq x$; otherwise, are called incomparable elements. An element $x_m \in \Psi$ is minimum of Ψ if for every $y \in \Psi, x_m \preceq y$. An element $x_0 \in \Psi$ is called minimal element if there is no element $y \in \Psi$, such that $y \preceq x_0$. Notice that a POS may has more than one minimal element but don't have more than one minimum element.

Proposition 1. Let $\Omega = \{(\chi_1, \chi_2, \dots, \chi_n) \mid \chi_i = 0 \text{ or } 1\}$ and Ψ be an arbitrary subset of Ω. Let X_1 and X_2 are two arbitrary elements of Ψ, we say that $X_1 \preceq X_2$ whenever $(X_1)_i \leq (X_2)_i$ for $i = 1,2, \dots, n$. It is clear that $(\Psi, \preceq)$ is partial order set (POS).

Example 1. Let $\Psi = \{(x, y) \mid x, y \in \{0,1\}, x + y \neq 0\}$, define $\preceq$ on Ψ as $(x, y) \preceq (z, w)$ if and only if $x \leq z$ and $y \leq w$. In this case $\Psi = \{(1,0), (0,1), (1,1)\}$, moreover $(1,0) \preceq (1,1)$ and $(0,1) \preceq (1,1)$ minimal elements are $\{(0,1), (1,0)\}$. It is clear that $(0,1)$ and $(1,0)$ are not comparable elements.

2.2 Rough set theory

In this section, we recalled some basic definitions of the rough set theory. For example, an approximation space is a pair (U, R), in which U is a nonempty finite set called universe and R is an equivalence relation defined on U. For each $x \in U$ define $[x]_R$, the equivalence class of x, as follows:

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

Definition 1. [18] Suppose $S = (U, R)$ be an approximation space and X be a subset of U, the lower approximation of X by R in S is defined as $\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}$ and the upper approximation of X by R in S is defined as $\bar{R}X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}$.

A pair $IS = (U, A_{at})$ where A_{at} is a nonempty finite set of attributes that make an equivalence relation is called information system. For every $a \in A_{at}$ we have $a: U \to V_a$ where V_a is called domain of attribute a, if $X \subseteq U$ then $a(X) = \{a(x) \mid x \in X\}$. Note that a special kind of information system apears as $DS = (U, A_{at} = C_{at} \cup D_{at})$ where C_{at} contains condition attributes and D_{at} contains decision attributes is called decision system. Equivalence relations are a way to divide a set a set U into a union of disjoint subsets. Let R be an equivalence relation on U, If $IS = (U, A_{at})$ be an Information system, with any $B_{at} \subseteq A_{at}$ there is an associated equivalence relation $IND(B_{at}) = \{(x, y) \in U \times U \mid \forall b \in B_{at}, b(x) = b(y)\}$ is a equivalence relation on U that is called indiscernibility relation. One can see that $U/IND(B_{at}) = [x]_{IND(B_{at})}$ is partition of U generated by $IND(B_{at})$. Here for simplicity denote $U/IND(B_{at}), [x]_{IND(B_{at})}$ respectively by $U/B_{at}, [x]_{IND(B_{at})}$. A subset X of U is said to be R -definable in IS if and only if $\underline{R}X = \bar{R}X$. The boundary set is $\bar{R}(X) - \underline{R}(X)$ and denote it by $BN_R(X)$. It consists of objects that we can't decisively classify X to R. A subset X of U is called rough set if $BN_R X \neq \emptyset$; otherwise, the set X is crisp with respect to R.

Definition 2. [12] (Accuracy of Approximation) Let $IS = (U, A_{at})$ be an information system and X be a subset of universe set U and $B_{at} \subseteq A_{at}$ then accuracy of approximation X by B_{at} is defined as follow:

$$\alpha_{B_{at}}(X) = \frac{\left|\underline{B_{at}}(X)\right|}{\left|\overline{B_{at}}(X)\right|} \qquad (1)$$

where $|.|$ denotes the cardinality of a set. If $\alpha_{B_{at}}(X) = 1$, then X is crisp with respect to attributes in B_{at}, if $\alpha_{B_{at}}(X) < 1$ then X is rough with respect to attributes in B_{at}.

Example 2. Consider below decision table in which $C_{at} = \{a, b, c, d\}$ is

Table 1

U	a	b	c	d	*class*
1	a_2	b_1	c_1	d_1	*no*
2	a_2	b_1	c_2	d_2	*yes*
3	a_3	b_3	c_3	d_2	*no*
4	a_3	b_3	c_3	d_2	*yes*
5	a_2	b_1	c_2	d_2	*yes*
6	a_2	b_1	c_2	d_2	*yes*
7	a_1	b_1	c_2	d_2	*yes*
8	a_1	b_2	c_1	d_2	*no*
9	a_1	b_2	c_1	d_2	*no*
10	a_1	b_1	c_2	d_2	*yes*
11	a_2	b_1	c_1	d_1	*no*
12	a_1	b_1	c_2	d_2	*yes*
13	a_3	b_3	c_3	d_2	*yes*
14	a_3	b_3	c_3	d_2	*no*

a set of condition attributes, and the domain are $V_a = \{a_1, a_2, a_3\}, V_b = \{b_1, b_2, b_3\}, V_c = \{c_1, c_2, c_3\}$ also $V_d = \{d_1, d_2\}$ be condition attributes on U and D_{at} is decision attribute that its domain is $V_{\text{class}} =$ { yes, no } .

Let $X = \{x \in U \mid \text{class}(x) = no\} = \{1,3,8,9,11,14\}$ and $B_{at} = \{b, c\} \subset C_{at}$. We have

$$U/B_{at} = \{\{1,11\}, \{2,5,6,7,10,12\}, \{3,4,13,14\}, \{8,9\}\}$$

$\underline{IS}_{B_{at}}(X) = \{\{1,11\}, \{8,9\}\}$ and $\overline{IS}_{B_{at}}(X) = \{\{1,11\}, \{3,4,13,14\}, \{8,9\}\}$ then $\underline{B_{at}}(X) = \{1,11,8,9\}, \overline{B_{at}}(X) = \{1,11,8,9,3,4,13,14\}$. Due to definition $BN_{B_{at}}(X) = \{3,4,13,14\}$ hence $\alpha_{B_{at}}(X) = \frac{1}{2}$.

Definition 3. [12] Let C_{at} and D_{at} be subset of A_{at}. It said that D_{at} depend on C_{at} in a degree $k(0 \le k \le 1)$, denoted by $C_{at} \rightarrow_k D_{at}$, if

$$\kappa = \gamma(C_{at}, D_{at}) = \frac{\left|POS_{C_{at}}(D_{at})\right|}{|U|}$$

where POS $_{C_{at}}(D_{at}) = \bigcup_{X \in U/Dat} \frac{C_{at}}{D}(X)$, is called C_{at} - positive region of D_{at}. Note that $\kappa = 1$ means that D_{at} depends, totally, on C_{at} and $\kappa < 1$ means that D_{at} depends partially (in a degree κ) on C_{at}.

Due to 2, attribute conditions are $C_{at} = \{a, b, c, d\}$ and $D_{at} =$ { class } is decision attribute then

$$U/C_{at} = \{\{1,11\}, \{2,5,6\}, \{3,4,13,14\}, \{7,10,12\}, \{8,9\}\}$$
$$U/D_{at} = \{D_1, D_2\} = \{\{1,3,8,9,11,14\}, \{2,4,5,6,7,10,12,13\}\}$$

then

$$\underline{C_{at}}(D_1) = \{1,11,8,9\}, \underline{C_{at}}(D_2) = \{2,5,6,7,10,12\}$$

It results that

$$POS_{C_{at}}(D_{at}) = \underline{C_{at}}(D_1) \cup \underline{C}(D_2) = \{1,11,8,9,2,5,6,7,10,12\}.$$

Then

$$\kappa = \gamma(C_{at}, D_{at}) = \frac{\left|POS_{C_{at}}(D_{at})\right|}{|U|} = \frac{5}{7}$$

Now let $C_{at} = \{b, c\}$ then $U/C_{at} = \{\{1,11\}, \{2,5,6,7,10,12\}, \{3,4,13,14\}, \{8,9\}\}$. One can see that $\kappa = \gamma(C_{at}, D_{at}) = \frac{|POS_{C_{at}}(D_{at})|}{|U|} = \frac{4}{14}$.

3. Reducts and Core

Let $DS = (U, A_{at} = C_{at} \cup D_{at})$ be a decision system and $c \in C_{at}$ then attribute c is called dispensable in DS if $POS_{C_{at}}(D_{at}) = POS_{C_{at}-\{c\}}(D_{at})$ else c is called indispensable, in addition $DS = (U, A_{at} = C_{at} \cup D_{at})$ is independent if all $c \in C_{at}$ are indispensable.

Definition 4. [12] Suppose $DS = (U, A_{at} = C_{at} \cup D_{at})$ be a decision system. A subset R of C_{at} is called reduct of C_{at} if $DS = (U, A_{at} = R \cup D_{at})$ is independent and $POS_{C_{at}}(D) = POS_R(D_{at})$.

It should be noted that a decision system may have many reducts. A set of all reducts of C_{at} is denoted by RED(C_{at}), in other word

$$RED(C_{at}) = \{R \subseteq C_{at} \mid \gamma(R, D_{at}) = \gamma(C_{at}, D_{at}), \forall B_{at} \subset R, \gamma(C_{at}, D_{at}) \neq \gamma(C_{at}, B_{at})\}. \quad (2)$$

The intersection of all reducts of C_{at} is also called its core, i.e. $\text{Core}(C_{at}) = \cap \text{Red}(C_{at})$. Every rough set can also be described by rough membership function. The rough membership function determines the degree of relative overlap between the set X and the equivalence class $[x]_R$. It is defined as follows:

$$\mu_X^R: U \rightarrow [0,1], \mu_X^R(x) = \frac{|X \cap [x]_R|}{|[x]_R|} \quad (3)$$

Discernibility Matrix: Let DS $= (U, A_{at} = C_{at} \cup D_{at})$ be a decision system where $C_{at} = \{a_1, a_2, \dots, a_p\}$ be a set of condition attributes and $U/C_{at} = \{C_1, C_2, \dots, C_k\}$. The discernibility matrix of U/C_{at} is a symmetric $k \times k$ matrix with entries given as follows:

$$b_{i,j} = f(C_i, C_j) = \{a \in C_{at} \mid a(x) \neq a(y) \text{ where } (x, y) \in C_i \times C_j\}$$

where $f: U/C_{at} \times U/C_{at} \rightarrow C_{at}$ is set value function. Note that we can rewrite $b_{i,j}$ as $b_{i,j} = f(C_i, C_j) = \{a \in C_{at} \mid a(C_i) \neq a(C_j)\}$. It means that each $b_{i,j}$ consists of the set of attributes upon which classes C_i and C_j differ. We say that two classes C_i and C_j are mergeable, if there exists $D_s \in U/D_{at}$, such that C_i and C_j belong to D_s. If this is the case, we can remove $b_{i,j}$ from the discernibility matrix. Since $b_{i,i} = \emptyset$ and $b_{i,j} = b_{j,i}$ this matrix is represented only as an upper triangular matrix. The concept of indiscernibility relation helps us to determine redundant attributes. Due to Example 1, discernibility matrix can be obtained as below:

Table 2

	C_1	C_2	C_3	C_4	C_5
C_1	—	$\emptyset$	*a, b, c, d*	*a, c, d*	$\emptyset$
C_2	*	—	$\emptyset$	*a*	*a, b, c*
C_3	*	*	—	$\emptyset$	$\emptyset$
C_4	*	*	*	—	*b, c*
C_5	*	*	*	*	—

4. Distance in equivalence classes

Let $C_{At} = \{a_1, a_2, \dots, a_p\}$ is a nonempty finite set of attributes and $U/C_{At} = \{C_1, C_2, \dots, C_k\}$, define a mapping F from U to $\Re^k$ by

$$F: U \rightarrow \Re^p$$
$$F(x) = (a_1(x), a_2(x), \dots, a_p(x))$$

Let $\hat{C} \in U/C_{At}$. It is obviously that for all $x, y \in \hat{C}, F(x) = F(y) = V_{\hat{C}}$, it means that corresponding to each $\hat{C} \in U/C_{At}$ define a constant vector $V_{\hat{C}} \in \Re^p$ as

$$V_{\hat{C}} = (\hat{a_1}, \hat{a_2}, \dots, \hat{a_p}) \quad (4)$$

Here, we presented obtain different vectors, including $V_{C_1}, V_{C_2}, \dots, V_{C_N}$, where V_{C_i} is a constant vector function from C_i to $\Re^k$. Notice that if V is a vector function from U to $\Re^p$, such that $V(x) = V_{\hat{C}}$ then $x \in \hat{C}$

Definition 5. let $DS = (U, C_{at} \cup D_{at})$ be a decision system in which $C_{at} = \{a_1, a_2, \dots, a_p\}$ is a nonempty finite set of condition attributes. Based on the classical set theory, a subset R_{At} of C_{at} is defined by its characteristic function $\chi_{R_{At}}$s below:

$$\chi_{R_{At}}: C \longrightarrow \{0,1\}$$
$$\chi_{R_{At}}(a) = \begin{cases} 1 & a \in R_{At}, \\ 0 & O.W \end{cases}$$

Furthermore, let $P(C_{at}) = \{R_{At} \mid R_{At} \subseteq C_{at}\}$ be power set of C_{at} based on definition 5, each $R_{At} \in P(C_{at})$ can be obtained by

$$R_{At} = \left\{\left(a_1, \chi_{R_{At}}(a_1)\right), \left(a_2, \chi_{R_{At}}(a_2)\right), \dots, \left(a_p, \chi_{R_{At}}(a_p)\right)\right\}$$

in which $\chi_{R_{At}}(a_i) \in \{0,1\}$. To simplify the computation, we defined a one-to-one correspondence function ((bijection function)) as

$$\phi: P(C_{at}) \to \{0,1\}^p$$
$$\phi(R_{At}) = (\chi_1, \chi_2, \dots, \chi_p), \quad (5)$$

Where for each $i = 1,2, \dots, p, \chi_i \in \{0,1\}$. In addition, $R_{At} \in P(C_{at})$ is mapped to exactly an element of the $\{0,1\}^p$ and each $(\chi_1, \chi_2, \dots, \chi_p)$ is mapped on to exactly a R_{At} of $P(C_{at})$ denote $\chi_{R_{At}}(a_i)$ by χ_i. Therefore $R_{At} \subseteq C_{at}$ can be rewritten as $R_{At} = \{(a_1, \chi_1), (a_2, \chi_2), \dots, (a_p, \chi_p)\}$ in which for each $i = 1,2, \dots, p, \chi_i \in \{0,1\}$. In other word $\phi: P(C_{at}) \to \{0,1\}^p$ is invertible.

Definition 6. (Discernibility vector:) Let $DS = (U, C_{at} \cup D_{at})$ be a decision system, where C_{at} is condition attributes, i.e, $C_{at} = \{a_1, a_2, \dots, a_p\}$. Due to 4 assume $W = \{V_X \mid X \in U/C\}$, define a mapping H from $W \times W$ to $\{0,1\}^p$ as below

$$H: W \times W \longrightarrow \{0,1\}^p$$
$$H(X,Y) = (\delta_1, \delta_2, \dots, \delta_p) \quad (6)$$

where

$$\delta_i = \begin{cases} 1 & (V_X)_i \neq (V_Y)_i, \\ 0 & O.w \end{cases}$$

H is called discernibility vector.

Take $\Psi_1 = \{H(X,Y) \mid X, Y \in U/C_{at}\}$, according to proposition 1, $(\Psi_1, \preceq)$ is a partial order set.

Definition 7. According to definition 6, we metricized the set of U/C_{at} by

$$\text{dist}(X,Y) = \|H(X,Y)\| = \sum_{i=1}^{p} \delta_i$$

that is called distance between members of U/C_{at}. Assuming $R_{At} = \{(a_1, \chi_{R_{At}}(a_1), (a_2, \chi_{R_{At}}(a_2), \dots, (a_p, \chi_{R_{At}} \quad C_{at}$. Considering R_{At} , distance between members of U/C_{at} is defined by

$$\text{dist}_{R_{At}}(X,Y) = \sum_{i=1}^{p} \chi_i \delta_i,$$

where $\chi_i = \chi_{R_{At}}(a_i)$. It is clear that $\text{dist}_{C_{at}}(X,Y) = \text{dist}(X,Y)$

Additionally, let $DS = (U, C_{at} \cup D_{at})$ be a decision system, where C_{at} is condition attributes, If $R_{At} \subseteq C_{at}$ make an equivalence relation on U and $X, Y \in U/C_{at}$ then $0 \leq \text{dist}_{R_{At}}(X,Y) \leq |R_{At}|$. In addition, if $X \cap Y = \emptyset$ then $\text{dist}_{R_{At}}(X,Y) \neq 0$ in other words, $1 \leq \text{dist}_{R_{At}}(X,Y) \leq |R_{At}|$.

5. The proposed method

Lemma 1. Let $DS = (U, C_{at} \cup D_{at})$ be a decision system, where C_{at} is condition attributes. Also let us consider that R_{At}, R'_{At} are two subsets of C_{at} so that $R'_{\text{At}} \subseteq R_{\text{At}}$ then partition generated by $R_{At}(U/R_{At})$ is finer than the partition generated by $R'_{At}(U/R_{At})$.

Proof. Consider that $U/R_{At} = \{C_1, C_2, \dots, C_k\}$ and $U/R'_{At} = \{C'_1, C'_2, \dots, C'_{k'}\}$. If $X \in U/R_{At}$ and $Y \in U/R'_{At}$ so that $X \cap Y \neq \emptyset$ there would exist a $\hat{x} \in X \cap Y$ so that $[\hat{x}]_{R_{At}} = X$ and $[\hat{x}]_{R'_{At}} = Y$. Now let us suppose that $x \in X$ then $a(x) = a(\hat{x})$ for all $a \in R_{At}$. Moreover, since $R'_{At} \subseteq R_{At}$ we would have $a(x) = a(\hat{x})$ for all $a \in R'_{At}$ then $x \in [\hat{x}]_{R'_{At}}$ it results that $x \in Y$

Proposition 2. Considering the discernibility vector, the discernibility matrix can be represented as table 3, in which, all entries are the discernibility vectors, it is means that the discernibility matrix has become symmetric $|W| \times |W|$ matrix that is defined by

$$\overrightarrow{b_{i,j}} = \begin{cases} H(V_i, V_j) & \text{if } C_i \text{ and } C_j \text{ are not mergeable}, \\ \emptyset & O.w \end{cases}$$

For finding reducts, the discernibility matrix is more interest. Note that $H(V_i, V_i) = \vec{0}$, so diagonal entries of the discernibility matrix are zero.

As a result of this, discernibility matrix of example 2 can be changed to table 4 as below:

Table 3

Dat	V_1	V_2	V_3	. . .	$Vk-1$	V_k
V_1	→0	→$b1$, 2	→$b1,3$	. . .	→$b1$, $k-1$	→$b1$, k
V_2		→0	→$b2,3$	. . .	→$b2$, $N-1$	→$b2$, k
V_3			→0	. . .	→$b3$, $k-1$	→$b3$, k
.				. . .	.	.
$VN-1$					→0	→$bk-1,k$
V_N						→0

Table 4

	V_1	V_2	V_3	V_4	V_5
V_1	→0	∅	(0, 1, 1, 0)	(1, 0, 1, 1)	∅
V_2	*	→0	∅	(0, 0, 1, 1)	(1, 1, 1, 0)
V_3	*	*	→0	∅	∅
V_4	*	*	*	→0	(1, 1, 1, 1)
V_5	*	*	*	*	→0

Theorem 1. Let $DS = (U, C_{at} \cup D_{at})$ be a decision system, where C_{at} is condition attributes and D_{at} is decision attributes if $U/C_{at} = \{C_1, C_2, \dots, C_k\}$, using the definition of inner product on vectors, suppose $\chi = (\chi_1, \chi_2, \dots, \chi_p) \in \{0,1\}^p$ be a solution to linear inequality system

$$< H\left(V_{C_i}, V_{C_j}\right), \chi >\geq 1, \text{ for } 1 \leq i < j \leq k \text{ and } \vec{b}_{i,j} \neq \emptyset. \quad (7)$$

then $\text{POS}_{C_{at}}(D_{at}) = \text{POS}_{R_{At}}(D_{at})$, where $R_{At} = \phi^{-1}(\chi) = \{(a_1, \chi_1), (a_2, \chi_2), \dots, (a_p, \chi_p)\}$

Proof. Proof by contradiction, let $POS_{C_{at}}(D_{at}) \neq POS_{R_{At}}(D_{at})$, based on the Lemma 1, since $R_{At} \subseteq C_{at}$ then partition generated by C_{at} is finer than the partition generated by R_{At}, in other word if $U/C_{at} = \{C_1, C_2, \dots, C_k\}$ and $U/R_{At} =$

$\{C'_1, C'_2, \dots, C'_{k'}\}$ then $k' \le k$ and each element of U/C_{at} is the union of one or more elements of U/R_{At}. Since $POS_{C_{at}}(D_{at}) \neq POS_{R_{At}}(D_{at})$ then there exist $M'_{i_0} \in U/R_{At}, C_{j_0} \in U/C_{at} - \text{Pos}_{C_{at}}(D_{at})$ and $W \subseteq \{1,2,\dots,k\}$ such that $C'_{i_0} = C_{j_0} \cup_{j\in W} C_j$, therefor $\text{dist}_{R_{At}}(C_{j_0}, C_j) = 0$ for $j \in W$ then $< H(V_{j_0}, V_j), \chi >= 0$ it is contradiction.

For simplicity we take $h = \{(i,j) \mid 1 \le i < j \le k$ and $\vec{b}_{i,j} \neq \emptyset\}$ then we can rewrite relation 7 as $\tilde{A}\chi \ge b$, in which i_{th} row of $\tilde{A}$ is extracted from table 5, $\tilde{A}$ is a matrix of order $|h| \times p$

Proposition 3. If $\chi = (\chi_1, \chi_2, \dots, \chi_p)$ satisfies $< H\left(V_{C_i}, V_{C_j}\right), \chi >\ge 1$ and $H\left(V_{C_i}, V_{C_j}\right) \le H\left(V_{C_{i'}}, V_{C_{j'}}\right)$, then χ satisfies $< H\left(V_{C_{i'}}, V_{C_{j'}}\right), \chi >\ge 1$. Then inequality linear system $< H\left(V_{C_{i'}}, V_{C_{j'}}\right), \chi >\ge 1$ is redundant and we remove $b_{i',j'}$ from discernibility matrix and denote this component by $\emptyset$. By continuing this manner, we remove all redundant inequalities from the system of linear inequality 7. Finally, the modified coefficient matrix is obtained. Denote it by ($\tilde{A}_m$) and replace to $\tilde{A}$ in $\tilde{A}\chi \ge b$ then for finding reducts it is enough to search all feasible solutions of $\tilde{A}_m\chi \ge b$. Therefore, to find reducts of decision system $DS = (U, C_{at} \cup D_{at})$ we need to find all feasible solutions of

$$\begin{cases} \tilde{A}_m\chi \ge b \\ \chi_i \in \{0,1\} \end{cases} \quad (8)$$

The set of all feasible solutions of this modified system is denoted by Λ. The relation $\preceq$ over Λ is binary relation and $(\Lambda, \preceq)$ is a POS. We present an easy approach to find all the reducts from the minimum cardinality, to the maximum cardinality. Assuming $\chi^k = (\chi_1^k, \chi_2^k, \dots, \chi_n^k) \in \Lambda$, take $I_{\chi^k} = \{i \mid \chi_i^k = 1\}$, denote $\prod_{\chi^k}(\chi) = \prod_{\chi^k}(\chi_1, \chi_2, \dots, \chi_n) = \prod_{j\in I_{\chi^k}} \chi_j$ and $\sum_{\chi^k} = \sum_{j\in I_{\chi^k}} \chi_j$.

Proposition 4. Based on the above description if χ^0 is an optimal solution for the problem

$$\begin{cases} \min \sum_{i=1}^{n} \chi_i \\ s.t \\ \tilde{A}_m\chi \ge b \\ \chi_i \in \{0,1\} \end{cases} \quad (9)$$

then χ^0 is a reduct with minimum cardinality. Lemma 2. Let χ^0 is an optimal solution of 9, take $\Lambda_0 = \{\chi \in \Lambda \mid \prod_{\chi^0}(\chi) \neq 0\}$, then we have $\Lambda_0 = \{\chi \in \Lambda \mid \chi^0 \preceq \chi\}$
Proof. χ^0 is an optimal solution of 9, if χ is an element of Λ_0 then $\chi_i = 1$ for $i \in I_{\chi^0}$ and $\chi_i = 0$ for $i \notin I_{\chi_0}$, subsequently $\chi_i^0 \le \chi_i$ for $1 \le i \le n$.

Proposition 5. Consider that χ^0 is an optimal solution for the 9 and χ_1 is an optimal solution to

$$\begin{cases} \min \sum_{i=1}^{n} \chi_i + M \prod_{\chi^0}(\chi) \\ s.t \\ \tilde{A}_m\chi \ge b \\ \chi_i \in \{0,1\} \end{cases} \quad (10)$$

where M is a large positive penalty constant. then $\chi^1 \neq \chi^0$ is a minimal of Λ and then is a reduct of the decision system $DS = (U, C_{at} \cup D_{at})$.
Finally, to find the nearest reduct to $R_0 = \phi^{-1}(\chi^0)$ with respect to cardinality we only need to solve BIP10. To find other reducts, in the theorem below, all the reducts are sequentially presented.

Theorem 2. Assuming that M is a large positive penalty constant and $\Omega_{k-1} = \{\chi^0, \chi^1, \dots, \chi^{k-1}\}$ is a set of minimal elements of Λ (for each $i = 1,2,\dots,k-1, \phi^{-1}(\chi^i)$ is a reduct) so that $|\phi^{-1}(\chi^0)| \le |\phi^{-1}(\chi^1)| \le \dots \le |\phi^{-1}(\chi^{k-1})|$. Let χ^k be an optimal solution of

$$\begin{cases} \min \sum_{i=1}^{n} \chi_i + M\left(\prod_{\chi^0}(\chi) + \prod_{\chi^1}(\chi) + \dots + \prod_{\chi^{k-1}}(\chi)\right) \\ s.t \\ \tilde{A}_m\chi \ge b \\ \chi_i \in \{0,1\} \end{cases} \quad (11)$$

then χ^k is a minimal element of Λ so that $\chi^k \notin \Omega_{k-1}$ and $\sum \chi^k \ge \sum \chi^{k-1}$, i.e., $|\phi^{-1}(\chi^k)| \ge |\phi^{-1}(\chi^{k-1})|$.
Proof. Proposition 5 and Lemma 2, lead to proof.

To find the optimal solutions for 9, 10 or generally 11, we can apply either one of the metaheuristic algorithms such as Genetic Algorithm or standard methods such as the Branch and Bound method. Genetic algorithm is also a powerful method. Running the algorithm to system 9, it gives us only one of the global optima corresponding to one of the reducts, with minimum cardinality, with minimum cardinality. Then, if we apply genetic algorithm to 11, we obtain another reduts satisfying the assumptions of last theorem. On the other hand, the branch and bound algorithm is useful for discrete, hybrid and mathematical optimization problems. It consists of a systematic enumeration of candidate solutions utilizing, state space search. This set of candidate solutions creates a rooted tree with a complete set at the root. The branch and bound algorithm search this tree's branches representing subsets of the solution set. Before enumerating the candidate solutions of a branch, the branch is checked against upper and lower estimated bounds on the optimal solution, If it cannot find a better solution than the one found by the branch and bound algorithm, it is discarded. Therefore, the goal of the branch-and-bound algorithm is to find the optimal solution for systems 9, 10 or 11. As a result, to find all reducts and core, we apply one of the mentioned approaches.

Example 3. Consider example 2, table 5 determine discernibility matrix and gives a non-equality linear system $\tilde{A}\chi \ge b$ as below:

$$\begin{cases} \chi_2 + \chi_3 \ge 1 \\ \chi_1 + \chi_3 + \chi_4 \ge 1 \\ \chi_3 + \chi_4 \ge 1 \\ \chi_1 + \chi_2 + \chi_3 \ge 1 \\ \chi_1 + \chi_2 + \chi_3 + \chi_4 \ge 1 \\ \chi_i \in \{0,1\}, i = 1,2,3,4 \end{cases} \quad (12)$$

After removing redundant component of table 4, the modified discernibility matrix is obtained as table 5. Using discernibility matrix, inequality linear

Table 5

	V_1	V_2	V_3	V_4	V_5
V_1	→0	∅	(0, 1, 1, 0)	(1, 0, 1, 1)	∅
V_2	*	→0	∅	(0, 0, 1, 1)	∅
V_3	*	*	→0	∅	∅
V_4	*	*	*	→0	∅
V_5	*	*	*	*	→0

system 12, lead to

$$\begin{cases}\chi_2+\chi_3\geq 1\\ \chi_1+\chi_3+\chi_4\geq 1\\ \chi_3+\chi_4\geq 1\\ \chi_i\in\{0,1\}, i=1,2,3,4\end{cases} \quad (13)$$

To find reducts with minimum cardinality, we solve BIP below:

$$\begin{cases}\min\chi_1+\chi_2+\chi_3+\chi_4\\ S.t\\ \chi_2+\chi_3\geq 1\\ \chi_1+\chi_3+\chi_4\geq 1\\ \chi_3+\chi_4\geq 1\\ \chi_i\in\{0,1\}, i=1,2,3,4\end{cases} \quad (14)$$

The optimal solution is $\chi^0=(0,0,1,0)$ then $(R_{at})_0=\phi^{-1}(\chi_0)=\{(a,0),(b,0),(c,1),(d,0)\}=\{c\}$. To find another reduct we solve

$$\begin{cases}\min\chi_1+\chi_2+\chi_3+\chi_4+M\chi_3\\ S.t\\ \chi_2+\chi_3\geq 1\\ \chi_1+\chi_3+\chi_4\geq 1\\ \chi_3+\chi_4\geq 1\\ \chi_i\in\{0,1\}, i=1,2,3,4\end{cases} \quad (15)$$

The optimal solutions are $\chi_1=(0,1,0,1)$ and $\chi_2=(1,1,0,0)$ then $(R_{at})_1=\phi^{-1}(\chi_1)=\{(a,0),(b,1),(c,0),(d,1)\}=\{b,d\}$ and $(R_{at})_2=\phi^{-1}(\chi_2)=\{(a,1),(b,1),(c,0),(d,0)\}=\{a,b\}$.

What was stated to feature subsets selection, can be summarized in the following algorithm.

Algorithm

-Input $DS=(U,C_{at}\cup D_{at})$ where $C_{at}=\{a_1,a_2,\dots,a_p\}, D_{at}=\{d_1,d_2,\dots,d_l\},, b$ is a vector of order $p\times 1$ as $b=(1,1,\dots,1)^T$.

- Output Reduccts and core

1. construct $U/C_{at}=\{C_1, C_2, \dots, C_k\}$, $U/D_{at}=\{D_1, D_2, \dots, D_q\}$,
2. Due to 4 correspondening to each C_i consider vector V_Ci for i = 1, 2, . . ., k
3. Construct modified discernibility matrix.
4. correspondent to modified discernibility matrix create system of linear inequalities A˜mχ ≥ b, in which each row of A˜m is one of the nonempty components of this matrix.
5. Using the mentioned methods, we could solve binary linear integer pro- grammings 9 and 11. All reducts are obtained from the smallest to the largest cardinality, sequently.
6. Intersection of all the reducts is considered core.

6. Conclusion

This paper presents a new method for feature selection using binary linear integer programming systems. By solving these systems, all redducts from minimum to maximum cardinality are found. The intersection of all this reducts would be core.Optimal solutions can be found using standard meth- ods such as branch and bound or meta-heuristic algorithms such as genetic algorithm. The introduced method is suggested as a valuable tool to handle the rough set theory. In future work, the proposed algorithm can be extended to fuzzy or hybrid decision systems.

References

[1] R. Jensen, A. Tuson, and Q. Shen, "Finding rough and fuzzy-rough set reducts with SAT," *Inf. Sci.*, vol. 255, pp. 100–120, 2014.

[2] R. Jensen and Q. Shen, "Rough set-based feature selection: A review," 2014. [Online]. Available: uploaded by Richard Jensen on May 29, 2014.

[3] F. Amini and G. Hu, "A two-layer feature selection method using genetic algorithm and elastic net," *Expert Syst. Appl.*, vol. 166, p. 114072, 2021.

[4] U. Stanczyk and B. Zielosko, "Heuristic-based feature selection for rough set approach," *Int. J. Approx. Reason.*, vol. 125, pp. 187–202, 2020.

[5] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *In press, Corrected Proof*, 2020.

[6] A. Chouchoulas and Q. Shen, "Rough set-aided keyword reduction for text categorisation," *Appl. Artif. Intell.*, vol. 15, no. 9, pp. 843–873, 2001.

[7] R. Jensen and Q. Shen, "Semantics-preserving dimensionality reduction: Rough and fuzzy-rough based approaches," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1457–1471, 2004.

[8] T. Y. Lin and P. Yin, "Heuristically fast finding of the shortest reducts," in *Rough Sets and Current Trends in Computing*, *Lect. Notes Comput. Sci.*, vol. 3066, pp. 465–470, 2004.

[9] N. Zhong, J. Dong, and S. Ohsuga, "Using rough sets with heuristics for feature selection," *J. Intell. Inf. Syst.*, vol. 16, no. 3, pp. 199–214, 2001.

[10] Y. Chen, D. Miao, and R. Wang, "A rough set approach to feature selection based on ant colony optimization," *Pattern Recognit. Lett.*, vol. 31, no. 3, pp. 226–232, 2010.

[11] N. Mac Parthalin and R. Jensen, "Unsupervised fuzzy-rough set-based dimensionality reduction," *Inf. Sci.*, vol. 229, pp. 106–121, 2013.

[12] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*, Dordrecht, Netherlands: Kluwer Academic Publishers, 1991.

[13] G. Lang, J. Luo, and Y. Yao, "Three-way conflict analysis: A unification of models based on rough sets and formal concept analysis," *Knowl.-Based Syst.*, vol. 194, p. 105556, 2020.

[14] M. Davis, G. Logemann, and D. Loveland, "A machine program for theorem proving," *Commun. ACM*, vol. 5, pp. 394–397, 1962.

[15] L. Gao and W. Wu, "Relevance assignation feature selection method based on mutual information for machine learning," *Knowl.-Based Syst.*, vol. 209, p. 106439, 2020.

[16] H. Jiang, S. Luo, and Y. Dong, "Simultaneous feature selection and clustering based on square root optimization," *Eur. J. Oper. Res.*, vol. 289, no. 1, pp. 214–231, 2021.

[17] M. Kelidari and J. Hamidzadeh, “Feature selection by using chaotic cuckoo optimization algorithm with levy flight, opposition-based learning and disruption operator,” *Soft Comput.*, vol. 25, no. 4, pp. 2911–2933, 2021.
[18] Z. Lu, Z. Qin, Y. Zhang, and J. Fang, “A fast feature selection approach based on rough set boundary regions,” *Pattern Recognit. Lett.*, vol. 36, pp. 81–88, 2014.