2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)
March 29–31, 2018, Xiamen, China

Covering Rough Set-based Three-way Decision Feature Selection

Mengyuan Ren, Yanpeng Qu, and Ansheng Deng
Information Science and Technology College
Dalian Maritime University
Dalian, 116026, China
Email: {myren, yanpengqu, ashdeng}@dlmu.edu.cn

***Abstract*—Feature selection is the process of selecting a subset of features from the entire dataset such that the selected subset can be used on behalf of the entire features of dataset to reduce the complexity of further processing. Recently, many feature selection approaches have been developed effectively based on rough sets. To provide a reasonable semantic interpretation for the three regions of probability rough sets and decision theoretic rough sets, the three-way decision theory is proposed. By using of the covering rough set model, the positive region, boundary region and negative region of the three-way decision approach are established to conduct feature selection. In this paper, the covering rough set-based three-way decision feature selection method (*CTFS*) is proposed to extend the traditional rough set version. Furthermore, the condition entropy based on three-way decision regions can be used as an evaluation method to select features. The experimental results show that the proposed feature selection approach results in a smaller selected feature subsets compared to the state-of-the-art feature selection methods with higher classification accuracies.**

Keywords*—*Covering rough sets*; *Three-way decision*; *Feature selection

I. Introduction

As the dimensionality of the datasets has increased, the existence of the irrelevant or redundant features is more common to be observed in a given learning problem and classification task. Hence, the approach to extracting useful information effectively from a large-scale dataset has become a critical issue for the area such as information science, data mining. However, most of the learning techniques are confused and the learning performance is thus degenerated. Feature selection (*FS*) is an effective technique which can attempt at exploring a subset of selected features that contains most of the useful information. The purpose of *FS* is to find out a minimal feature subset from a problem domain [1], [2]. By removing the irrelevant and redundant features, the performance of the given learning problem and classification task can be improved significantly. In rough set, feature selection is also called attribute reduction.

In the research of intelligent information processing, how to effectively deal with the ubiquitous and imprecise information is an important problem to consider. Rough set theory [3], provides a kind of effective method to solve the problem of uncertainty in the data. It uses a pair of exact sets named the lower approximation and the upper approximation. In rough sets, the uncertainty of the target set is mainly caused by knowledge granulation or discernibility, calculated by difference between the upper and lower approximation. Rough sets are widely used in machine learning, artificial intelligence, data mining and many other fields and become an important intelligent information processing technology [4]. In addition, rough sets have many extensions, such as rough set based on similarity relation [5], rough set based on tolerance relation [6], decision-theoretic rough sets [7] and probabilistic rough sets [8]. Although the studied problems are formally different, the essence is essentially covering, only the ways of yielding covering are different. Therefore, it is necessary to study how to deal with the data using covering rough sets.

Covering rough sets are the extension of traditional rough sets [9], which are based on the relaxation of the partition to covering. Traditional rough sets are often too restrictive in classification problems, covering rough sets relax the definitions of the lower and upper approximation and covering rough sets are more suitable for feature selection and classification problem. Covering classes formed by the universe of coverings use to define the upper and lower approximations of any subset. Most of the studies of covering rough set model primarily are centered on the construction of the upper and lower approximations. Furthermore, covering-based approximation operators are discussed respect to the element, granule, and subsystem. A new method to select important coverings in covering-based decision systems by defining different coverings and it uses a discernibility matrix to calculate total features is proposed [10]. After that, covering rough sets studied in-depth and obtained a lot of meaningful research results.

The three-way decision theory extends the traditional two-way decision theory [11]. Two-way decisions only consider two choices: acceptance and rejection. However, in practice, it is often impossible to accept or reject information because of the uncertainty or incompleteness of information. When the information is uncertain, it is not enough to support the acceptance or the refusal, then the third choice is adopted, namely, the choice is not committed. The three-way decision theory is originally proposed to explain the three regions of rough sets. The main idea of the three-way decision making is to divide the whole into three independent parts. Meanwhile, adopts different processing methods for different parts. The main idea provides an effective strategy and method for complex problem

978-1-5386-4362-4/18/$31.00 © 2018 IEEE

solving. In the last few years, many researchers primarily are centered on model extensions and practical applications of three-way decision. For model extension, it primarily involves the extensions of rough set model, such as interval-valued fuzzy rough sets [12], neighborhood rough sets [13]. For practical application, it primarily involves text classification, information filtering, cluster analysis, *FS*, etc.

In order to demonstrate the validity of the proposed *FS* approach, comparative experimental studies are carried out on popular benchmark datasets. In particular, the work is compared with typical state-of-the-art *FS* and other dimensionality reduction techniques, including *CFS* [14], *FRFS* [15], *PCA* [16] and *MI* [17]. It is shown that the proposed work outperforms these algorithms, returning a high classification accuracy while leading to a reduced feature subset.

The remainder of this paper is structured as follows. In Section II, the theoretical background is given includes classic rough set, covering rough sets and the three-way decision theory. Section III introduces the *CTFS* method to select features in accordance with covering rough set-based three-way decision model in detail. Section IV provides an experimental analysis, and discusses the significance of the proposed *FS* methods. The paper is concluded in Section V, with an outline of proposed further work.

II. Theoretical Background

This section reviews some essential definitions associated with classic rough sets [18], the three-way decision [19], [20] and covering rough sets [21], [22].

A. Attribute Reduction and Rough Sets

Definition 1: Let $S = \{U, C \cup d\}$ be a decision system, where $U = \{x_1, x_2, x_3, \cdots, x_n\}$ is a nonempty finite set of objects called the universe. *C* is a finite nonempty set of condition attributes, *d* is a finite set of decision attribute. $C \cap d = \emptyset$, V_a represents values of a non-empty set of $a \in (C \cup d)$, and $f\colon (C \cup d) \to V_a$ represents an information function that maps an object in *U* to precisely one value in V_a.

Definition 2: A decision system is $S = (U, C \cup d)$, with any subset of attributes $P \subseteq C$, then an indiscernibility relation in the decision system is that:

$$\mathit{IND}\,(P) = \left\{(x,y) \in U^2 \mid \forall b \in P, f(x,b) = f(y,b)\right\}, \quad (1)$$

in which, the indiscernibility relation is reflexive, symmetric and transitive. The member of all equivalence classes of $\mathit{IND}\,(P)$ are denoted by $U/\mathit{IND}\,(P)$, U/P for short.

Definition 3: Let $X \subseteq U$ can be approximated using only the information contained in P by constructing the *P-lower* and *P-upper* approximations of *X*:

$$\underline{P}\,(X) = \{x \in U \mid [x]_P \subseteq X\} = \cup\{[x]_P \mid [x]_P \subseteq X\}, \quad (2)$$

$$\overline{P}\,(X) = \{x \in U \mid [x]_P \cap X \neq \emptyset\} = \cup\{[x]_P \mid [x]_P \cap X \neq \emptyset\}. \quad (3)$$

According to the lower and the upper approximation, it can obtain the positive region, boundary region and negative region:

$$POS_P(X) = \underline{P}(X), \quad (4)$$

$$BND_P(X) = \overline{P}(X) - \underline{P}(X), \quad (5)$$

$$NEG_P(X) = U - POS_P(X) \cup BND_P(X) = U - \overline{P}(X). \quad (6)$$

The $POS_P(X)$ represents the positive region that all objects in positive region definitely belong to the set *X*. The $NEG_P\,(X)$ represents the negative region that all objects in negative region definitely does not belong to the set *X*. The $BND_P\,(X)$ represents the boundary region that all object may belong to the set *X*. An *FS* model in rough set is a relative selection with respect to the decision attribute *d*, which is defined by necessitating that the positive region of the decision attribute *d* is invariable.

Definition 4: Attribute $a \in P \subseteq C$ is dispensable in *P*, if $\mathit{IND}\,(P) = \mathit{IND}\,(P - \{a\})$, otherwise a is indispensable in *P*.

Definition 5: A decision system $S = (U, C \cup d)$, and $U/d = \{d_1, d_2, \cdots, d_n\}$, an attribute set $P \subseteq C$ satisfies the following two conditions:

(1) $POS_P\,(d) = POS_C\,(d)$,

(2) $\forall a \in P$, $POS_{P-\{a\}}\,(d) \neq POS_C\,(d)$,

then the subset *P* called a reduction for *C*.

B. Covering Rough Set Model

This subsection reviews some essential definitions associated with covering-based rough sets, and introduced the positive region, boundary region and negative region of covering-based rough sets.

Definition 6: Let *U* be a universe, *C* is consist of a family of subsets of *U*, *C* is called a cover of *U*, if $\cup C = U$ and no subset in *C* is empty.

In covering rough sets, the concept of a covering is an extension of the concept of a partition. It is clear that a partition of *U* is certainly a covering of *U*.

Definition 7: Let $C = \{C_1, C_2, C_3, \cdots, C_m\}$ be a covering of *U*. For every $X \in U$, let $C_x = \cap\{C_j : C_j \in C, x \in C_j\}$, $Cov\,(C) = \{C_x : x \in U\}$ is also a covering of *U*, it is called the induced covering of *C*.

Definition 8: Let $\Delta = \{C_i : i = 1, \cdots, n\}$ be a family of coverings of *U*. For any $x \in U$, let $\Delta_x = \cap\{C_{ix} : C_{ix} \in Cov\,(C_i)\}$, then $Cov\,(\Delta) = \{\Delta_x : x \in U\}$ is also a covering of *U*, it is called the induced covering of Δ.

The Δ_x represents the intersection of all the elements in every C_i including *x*. For any $x \in U$, Δ_x is the minimal set in $Cov\,(\Delta)$ including *x*. If the elements in the Δ represent attribute, then every covering represents an attribute.

For any $X \subset U$, the lower and the upper approximation of *X* respect to Δ are defined by:

$$\underline{\Delta}\,(X) = \cup\{\Delta_x : \Delta_x \subseteq X\}, \quad (7)$$

$$\overline{\Delta}\,(X) = \cup\{\Delta_x : \Delta_x \cap X \neq \emptyset\}. \quad (8)$$

Obviously, the positive region, boundary region and negative region of *X* for Δ is that:

$$POS_\Delta\,(X) = \underline{\Delta}\,(X),$$

$$BND_\Delta\,(X) = \overline{\Delta}(X) - \underline{\Delta}(X),$$

$$NEG_\Delta\,(X) = U - \overline{\Delta}(X).$$

C. Three-way Decision Theory

The three-way decision theory is introduced in this subsection. The three-way decision theory has only two states and three actions, three actions respectively are accept, non-commitment and reject. The two states set $\Omega = (X, X^c)$, in that X^c represents denotes the complement of *X*. The three actions set $X = \{m_P, m_B, m_N\}$ represents three actions of deciding that an object is in the set $POS(X)$, $BND(X)$ or $NEG(X)$, and if an element belongs to *X*, let $\lambda_{PP}, \lambda_{BP}, \lambda_{NP}$ represents the costs that doing the actions m_P, m_B, m_N. If an element belongs to X^c, let $\lambda_{PN}, \lambda_{BN}, \lambda_{NN}$ represents the costs that doing the same three actions. So the loss functions show as follows:

	m_P	m_B	m_N
X	λ_{PP}	λ_{BP}	λ_{NP}
X^c	λ_{PN}	λ_{BN}	λ_{NN}

The equivalence class of *x* respect to *IND* (P) is the $[x]_P$, and the probabilities that belong to *X* or X^c are denoted by:

$$Pr(X \mid [x]_P) = \frac{|X \cap [x]_P|}{|[x]_P|},$$

$$Pr(X^c \mid [x]_P) = 1 - Pr(X \mid [x]_P).$$

Then the expected loss connected with doing different actions can be denoted by:

$$v(m_P \mid [x]) = \lambda_{PP} Pr(X \mid [x]_P) + \lambda_{PN} Pr(X^c \mid [x]_P),$$

$$v(m_B \mid [x]) = \lambda_{BP} Pr(X \mid [x]_P) + \lambda_{BN} Pr(X^c \mid [x]_P),$$

$$v(m_N \mid [x]) = \lambda_{BP} Pr(X \mid [x]_P) + \lambda_{BN} Pr(X^c \mid [x]_P).$$

The bayesian decision procedure illustrates the following minimum-cost decision rules:

(i) If $v(m_P \mid [x]_P) \leq v(m_B \mid [x]_P)$ and $v(m_P \mid [x]_P) \leq v(m_N \mid [x]_P)$, then $x \in POS_P(X)$,

(ii) If $v(m_B \mid [x]_P) \leq v(m_P \mid [x]_P)$ and $v(m_B \mid [x]_P) \leq v(m_N \mid [x]_P)$, then $x \in BND_P(X)$,

(iii) If $v(m_N \mid [x]_P) \leq v(m_P \mid [x]_P)$ and $v(m_N \mid [x]_P) \leq v(m_B \mid [x]_P)$, then $x \in NEG_P(X)$.

It is easy to get $\lambda_{PP} \leq \lambda_{BP} \leq \lambda_{NP}$, $\lambda_{NN} \leq \lambda_{BN} \leq \lambda_{PN}$, so the formula above can be written as:

(I) If $Pr(X \mid [x]_P) \geq \alpha$, and $Pr(X \mid [x]_P) \geq \gamma$, then $x \in POS_P(X)$,

(II) If $Pr(X \mid [x]_P) \leq \alpha$, and $Pr(X \mid [x]_P) \geq \beta$, then $x \in BND_P(X)$,

(III) If $Pr(X \mid [x]_P) \leq \beta$, and $Pr(X \mid [x]_P) \leq \gamma$, then $x \in NEG_P(X)$,

where,

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})},$$

$$\beta = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})},$$

$$\gamma = \frac{\lambda_{PN} - \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}.$$

It can be easily found the thresholds α, β can replace γ, so the formula above simplified as:

$$POS_P^{(\alpha,\beta)}(X) = \{x \mid \alpha < Pr(X \mid [x]_P) \leq 1\}, \quad (9)$$

$$BND_P^{(\alpha,\beta)}(X) = \{x \mid \beta < Pr(X \mid [x]_P) \leq \alpha\}, \quad (10)$$

$$NEG_P^{(\alpha,\beta)}(X) = \{x \mid 0 < Pr(X \mid [x]_P) \leq \beta\}. \quad (11)$$

III. Covering Rough Set-based Three-way Decision Feature Selection

The limitation of *FS* process of traditional rough sets is that, it can only operate effectively with data sets whose attributes can induce partitions. Covering rough sets relax the partition to a cover, so that more information is used to select features and the result is more effective.

A. A Covering Rough Set-based Three-way Decision Model

Generally, a covering-based three-way decision system can be represented *CTS* $= (U, \Delta \cup d, \alpha, \beta)$, in which $U = \{x_1, x_2, \cdots, x_n\}$ is a nonempty finite set of objects called the universe. $\Delta = \{C_1, C_2, \cdots, C_m\}$ is a family of coverings of *U*, the elements of Δ represent the condition attribute, *d* is a decision attribute set, α, β $(0 \leq \beta < \alpha \leq 1)$ are thresholds for the three regions pair-wise disjoint.

Definition 9: Let *CTS* $= (U, \Delta \cup d, \alpha, \beta)$ be a covering-based three-way decision system, $U/d = \{d_1, d_2, \cdots, d_n\}$ be an equivalence class set constituted by decision attribute *d* on the universe *U*. For the condition attribute set Δ, the positive, boundary and negative regions of *d* in the three-way decision system are defined by:

$$POS_\Delta^{(\alpha,\beta)}(d) = \cup_{i=1}^n \{x \mid x \in d_i, \alpha < Pr(d_i \mid \Delta_x) \leq 1\}, \quad (12)$$

$$BND_\Delta^{(\alpha,\beta)}(d) = \cup_{i=1}^n \{x \mid x \in d_i, \beta < Pr(d_i \mid \Delta_x) \leq \alpha\}, \quad (13)$$

$$NEG_\Delta^{(\alpha,\beta)}(d) = \cup_{i=1}^n \{x \mid x \in d_i, 0 < Pr(d_i \mid \Delta_x) \leq \beta\}, \quad (14)$$

where $P(d_i \mid \Delta_x) = \frac{|\Delta_x \cap d_i|}{|\Delta_x|}$.

The *FS* of decision regions in the three-way decision model are those decision regions with respect to the decision partition U/d remained unchanged. The region *FS* of a covering-based three-way decision system is defined in the following:

(1) *P* is a positive region *FS* result of *CTS* if it satisfies:
   a) $POS_P^{(\alpha,\beta)}(d) = POS_\Delta^{(\alpha,\beta)}(d)$,
   b) $\forall R \subset P, POS_R^{(\alpha,\beta)}(d) \neq POS_P^{(\alpha,\beta)}(d)$.

(2) *P* is a boundary region *FS* result of *CTS* if it satisfies:
   a) $BND_P^{(\alpha,\beta)}(d) = BND_\Delta^{(\alpha,\beta)}(d)$,
   b) $\forall R \subset P, BND_R^{(\alpha,\beta)}(d) \neq BND_P^{(\alpha,\beta)}(d)$.

(3) *P* is a negative region *FS* result of *CTS* if it satisfies:
   a) $NEG_P^{(\alpha,\beta)}(d) = NEG_\Delta^{(\alpha,\beta)}(d)$,
   b) $\forall R \subset P, NEG_R^{(\alpha,\beta)}(d) \neq NEG_P^{(\alpha,\beta)}(d)$.

In above definition, condition (a) for the positive, boundary and negative region is called the region *FS* conditions, respectively. Correspondingly, condition (b) is called the independency conditions.

Given a covering-based three-way decision system, *CTS* $= (U, \Delta \cup d, \alpha, \beta)$, in that $U/d = \{d_1, d_2, \cdots, d_n\}$ is an set of equivalence class. For a subset $P \subseteq \Delta$, the positive dependency, boundary dependency and negative dependency of *d* with respect to *P* in a three-way decision system are defined by:

$$\mu_{POS_P^{(\alpha,\beta)}} = \frac{\left|POS_P^{(\alpha,\beta)}(d)\right|}{|U|}, \tag{15}$$

$$\mu_{BND_P^{(\alpha,\beta)}} = \frac{\left|BND_P^{(\alpha,\beta)}(d)\right|}{|U|}, \tag{16}$$

$$\mu_{NEG_P^{(\alpha,\beta)}} = \frac{\left|NEG_P^{(\alpha,\beta)}(d)\right|}{|U|}, \tag{17}$$

these equations can illustrate the significance of regions in the universe *U*. It is easy to get $\mu_{POS_P^{(\alpha,\beta)}} + \mu_{BND_P^{(\alpha,\beta)}} + \mu_{NEG_P^{(\alpha,\beta)}} = 1$.

B. Information Measures for Covering Three-way Decision System

In this section, some basic concepts of *FS* in a covering-based three-way decision system are introduced. In rough set model, the *FS* is monotonic, but in a covering-based three-way decision system, the monotonicity of each region *FS* does not hold. In order to solve this problem, three new coverings and three monotonic information measures are constructed by using variants of the condition entropy.

$$U/R_{POS_P^{(\alpha,\beta)}} = \left\{POS_P^{(\alpha,\beta)}(d_1), \cdots, POS_P^{(\alpha,\beta)}(d_n), U\right\}, \tag{18}$$

$$U/R_{BND_P^{(\alpha,\beta)}} = \left\{BND_P^{(\alpha,\beta)}(d_1), \cdots, BND_P^{(\alpha,\beta)}(d_n), U\right\}, \tag{19}$$

$$U/R_{NEG_P^{(\alpha,\beta)}} = \left\{NEG_P^{(\alpha,\beta)}(d_1), \cdots, NEG_P^{(\alpha,\beta)}(d_n), U\right\}. \tag{20}$$

Here, three new covering classes generated by the three coverings $U/R_{POS_P^{(\alpha,\beta)}}$, $U/R_{BND_P^{(\alpha,\beta)}}$, $U/R_{NEG_P^{(\alpha,\beta)}}$ can be defined by:

$$\Delta_{R_{POS_P^{(\alpha,\beta)}}}(x_i) = \begin{cases} \left\{ POS_P^{(\alpha,\beta)}(d_j) \middle| x_i \in d_j \ and \ x_i \in POS_P^{(\alpha,\beta)}(d_j) \right\}, \\ \left\{ U \middle| x_i \in d_j \ and \ x_i \notin POS_P^{(\alpha,\beta)}(d_j) \right\} \end{cases} \tag{21}$$

$$\Delta_{R_{BND_P^{(\alpha,\beta)}}}(x_i) = \begin{cases} \left\{ BND_P^{(\alpha,\beta)}(d_j) \middle| x_i \in d_j \ and \ x_i \in BND_P^{(\alpha,\beta)}(d_j) \right\}, \\ \left\{ U \middle| x_i \in d_j \ and \ x_i \notin BND_P^{(\alpha,\beta)}(d_j) \right\} \end{cases} \tag{22}$$

$$\Delta_{R_{NEG_P^{(\alpha,\beta)}}}(x_i) = \begin{cases} \left\{ NEG_P^{(\alpha,\beta)}(d_j) \middle| x_i \in d_j \ and \ x_i \in NEG_P^{(\alpha,\beta)}(d_j) \right\} \\ \left\{ U \middle| x_i \in d_j \ and \ x_i \notin NEG_P^{(\alpha,\beta)}(d_j) \right\}. \end{cases} \tag{23}$$

For covering class, if *x* belongs to *d* and positive region, then the covering class of *x* is positive region, if *x* belongs to *d* and does not belong to positive region, then the covering class of *x* is *U*. According to the formula above, the condition entropy of $\Delta_{R_{POS_P^{(\alpha,\beta)}}}(x_i)$, $\Delta_{R_{BND_P^{(\alpha,\beta)}}}(x_i)$, $\Delta_{R_{NEG_P^{(\alpha,\beta)}}}(x_i)$ respect to *P* is defined by:

$$H\left(R_{POS_\Delta^{(\alpha,\beta)}} \mid P\right) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{|P_{x_i} \cap \Delta_{R_{POS_\Delta^{(\alpha,\beta)}}}(x_i)|}{|P_{x_i}|}, \tag{24}$$

$$H\left(R_{BND_\Delta^{(\alpha,\beta)}} \mid P\right) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{|P_{x_i} \cap \Delta_{R_{BND_\Delta^{(\alpha,\beta)}}}(x_i)|}{|P_{x_i}|}, \tag{25}$$

$$H\left(R_{NEG_\Delta^{(\alpha,\beta)}} \mid P\right) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{|P_{x_i} \cap \Delta_{R_{NEG_\Delta^{(\alpha,\beta)}}}(x_i)|}{|P_{x_i}|}. \tag{26}$$

According to the formula above, the attribute significance of $\{C_i\}$ respected to positive region, boundary region and negative region in Δ respect to *P* can be defined by the follows:

$$\begin{aligned} SIG(\{C_i\}, P, R_{POS_\Delta^{(\alpha,\beta)}(d)}) = \\ H(R_{POS_\Delta^{(\alpha,\beta)}(d)} | P) - H(R_{POS_\Delta^{(\alpha,\beta)}(d)} | P \cup \{C_i\}), \end{aligned} \tag{27}$$

$$\begin{aligned} SIG(\{C_i\}, P, R_{BND_\Delta^{(\alpha,\beta)}(d)}) = \\ H(R_{BND_\Delta^{(\alpha,\beta)}(d)} | P) - H(R_{BND_\Delta^{(\alpha,\beta)}(d)} | P \cup \{C_i\}), \end{aligned} \tag{28}$$

$$\begin{aligned} SIG(\{C_i\}, P, R_{NEG_\Delta^{(\alpha,\beta)}(d)}) = \\ H(R_{NEG_\Delta^{(\alpha,\beta)}(d)} | P) - H(R_{NEG_\Delta^{(\alpha,\beta)}(d)} | P \cup \{C_i\}). \end{aligned} \tag{29}$$

The three attribute significances can be used to value the significance of attribute $\{C_i\}$ in a covering-based three-way decision system. Therefore, the algorithms of *FS* can be depicted as follows.

The algorithm 1 explains the process of the positive region *CTFS* method that proposed in this paper. This algorithm mainly calculates the significance of attributes. The attribute significances are completely determined by the condition entropy, while the condition entropy need calculate covering class. The *Step* 2 in this algorithm, condition entropy for each attribute can be calculated. The steps of boundary and negative regions *FS* are similar to the positive region *FS* algorithm. The boundary region *FS* algorithm only use $H\left(R_{BND_\Delta^{(\alpha,\beta)}(d)} \mid C_i\right)$ and $SIG(\{C_i\}, P, R_{BND_\Delta^{(\alpha,\beta)}(d)})$ to replace $H\left(R_{POS_\Delta^{(\alpha,\beta)}(d)} \mid C_i\right)$ and $SIG(\{C_i\}, P, R_{POS_\Delta^{(\alpha,\beta)}(d)})$, finally cessation condition $BND_P^{(\alpha,\beta)}(d) = BND_\Delta^{(\alpha,\beta)}(d)$. The negative region *FS* algorithm only use $H\left(R_{NEG_\Delta^{(\alpha,\beta)}(d)} \mid C_i\right)$ and $SIG(\{C_i\}, P, R_{NEG_\Delta^{(\alpha,\beta)}(d)})$ to replace $H\left(R_{POS_\Delta^{(\alpha,\beta)}(d)} \mid C_i\right)$ and $SIG(\{C_i\}, P, R_{POS_\Delta^{(\alpha,\beta)}(d)})$, finally cessation condition $NEG_P^{(\alpha,\beta)}(d) = NEG_\Delta^{(\alpha,\beta)}(d)$.

Algorithm 1 Covering rough set-based three-way decision systems positive region *FS*

Input: A covering rough set-based three-way decision system $CTS = (U, \Delta \cup d, \alpha, \beta)$

Output: A positive region *FS* result *P*

1: Let $P = \emptyset$, initialize the parameters α, β;
2: For every attribute covering $\{C_i\} \in \Delta$, compute the condition entropy of attribute C_i, $H\left(R_{POS_{\Delta}^{(\alpha,\beta)}(d)} \mid C_i\right)$;
3: Calculate the $H\left(R_{POS_{\Delta}^{(\alpha,\beta)}(d)} \mid C_i\right)$, and select the attribute with the maximum value, record it as $\{b_{max}\}$;
4: $P \leftarrow P \cup \{b_{max}\}$;
5: For every attribute $\{C_i\} \in \Delta - P$,compute the significance of condition attribute $SIG(\{C_i\}, P, R_{POS_{\Delta}^{(\alpha,\beta)}(d)})$;
6: Calculate the $SIG(\{C_i\}, P, R_{POS_{\Delta}^{(\alpha,\beta)}(d)})$, and select the attribute with the maximum value, record it as $\{a_{max}\}$;
7: $P \leftarrow P \cup \{a_{max}\}$;
8: If $POS_{P}^{(\alpha,\beta)}(d) \neq POS_{\Delta}^{(\alpha,\beta)}(d)$, then go back Step 5, else go to Step 9;
9: The set *P* is the selected positive region feature.

TABLE I. Benchmark Data and Unreduced Feature Number

Dataset	*Objects*	*Features*
bupa	345	6
diabetes	768	8
fertility	100	9
ilpd	578	9
plrx	182	12
sonar	208	60

IV. Experimentation

This part includes the experimental steps and experimental results and presents an experimental evaluation of the selection methods in terms of classification accuracy. In this article, positive region and boundary region *FS* method result of covering-based three-way decision *FS* method are compared with other methods, in that, *CTFSP* represents positive region *FS* method, *CTFSB* represents boundary region *FS* method. Table I summarises the datasets used to conduct this experimentation. The experiments are run on datasets taken from *UCI* machine learning Repository databases [23].

The experimental results in terms of reduced data set size gained by *CTFSP* and *CTFSB* are compared to those of the state-of-the-art, such as *CFS*, *FRFS*, *PCA* and *MI*. Designed for operational simplicity, the threshold values (α, β) for the different data sets are same and valued $\alpha = 0.7$, $\beta = 0.3$. The selected feature subset sizes for six approaches are shown in Table II.

Table III, Table IV and Table V experimentally compare the classification accuracy for the reduced data sets gained by the use of *CTFSP*, *CTFSB*, *CFS*, *FRFS*, *PCA* and *MI*, respectively. For completeness, the classification methods use *SMO* [24], *Adaboost* [25] and *PART* [26]. Stratified 10×10-fold cross validation (10-*FCV*) is based on all original and reduced datasets is performed, it used to generate the classification results throughout the experimentation.

TABLE II. Reduction Size Results

Dataset	*CTFSP*	*CTFSB*	*CFS*	*FRFS*	*PCA*	*MI*
bupa	3	3	1	6	3	3
diabetes	5	5	4	8	4	4
fertility	5	5	1	7	4	4
ilpd	3	5	5	9	4	4
plrx	2	2	1	5	6	6
sonar	4	15	19	5	30	30

TABLE III. Classification Accuracy by *SMO*

Dataset	*CTFSP*	*CTFSB*	*CFS*	*FRFS*	*PCA*	*MI*
bupa	57.98	57.98	57.98	57.98	57.98	57.98
diabetes	77.04	76.74	76.97	76.80	65.11	75.43
fertility	88.00	88.00	88.00	88.00	88.00	88.00
ilpd	71.50	71.50	71.50	71.50	71.50	71.50
plrx	71.46	71.46	71.46	71.46	71.46	71.46
sonar	75.51	71.85	77.75	73.18	61.87	78.71

TABLE IV. Classification Accuracy by *AdaBoost*

Dataset	*CTFSP*	*CTFSB*	*CFS*	*FRFS*	*PCA*	*MI*
bupa	62.88	60.51	61.71	65.96	62.98	62.98
diabetes	74.89	74.83	74.84	74.92	67.75	73.99
fertility	87.10	87.10	88.00	85.9	86.40	86.40
ilpd	72.33	71.23	70.50	70.47	71.47	71.50
plrx	71.46	71.46	71.46	71.46	70.42	70.42
sonar	72.82	72.06	76.27	70.84	69.36	76.79

TABLE V. Classification Accuracy by *PART*

Dataset	*CTFSP*	*CTFSB*	*CFS*	*FRFS*	*PCA*	*MI*
bupa	62.32	60.32	62.18	65.25	60.15	60.15
diabetes	73.88	73.47	73.48	73.45	64.06	74.55
fertility	85.10	86.30	88.00	84.20	84.40	84.40
ilpd	70.93	70.47	70.11	69.13	71.24	71.45
plrx	71.46	71.46	71.46	71.46	71.46	71.46
sonar	71.00	70.78	76.21	69.35	65.80	75.74

In Table III, the classification accuracy of *CTFS* outperforms other *FS* methods in this paper. In Table IV, the classification accuracy of *CTFS* is superior to most of the *FS* methods and the subset of attributes is smaller than subsets of some other methods. In Table V, the classification accuracy of *CTFS* is suboptimal for these selection methods in this paper. The classification accuracy of *CTFS* is similar as other *FS* methods in this paper for a certain dataset. In general, these results illustrate that the proposed approach has a similar overall performance in terms of classification accuracy and smaller feature subset size.

V. CONCLUSION

This paper has presented a *CTFS* method to handle data obtained from real-world application. In proposed method, in order to solve non-monotonicity of covering-based three-way decision systems, new coverings of the positive region, negative region and boundary region are defined. Moreover, each feature is represented by a covering, information entropy of coverings is employed to guide feature selection process. An optimal feature subset that contains sufficient information is obtained. Also, the usage of *FS* algorithm of negative region is similar to *FS* algorithm of positive region and boundary region. The proposed *FS* algorithm has been implemented and tested against the state-of-the-art *FS* methods on benchmark datasets. The experimental results have shown that the proposed *FS* approach can often identify feature subsets of much smaller in size than those competing existing methods. The proposed *FS* algorithm can lead to similar classification accuracy for other methods.

Topics for further investigation include a more comprehensive study of how the covering rough set-based three-way decision model could be used for other tasks such as classification. An investigation into how the proposed method may be extended to covering rough set-based multi-way decision model. Another further extension to this work would enhance the classification accuracy of the task of feature selection.

ACKNOWLEDGMENT

This work is jointly supported by the National Natural Science Foundation of China (No. 61502068), the China Postdoctoral Science Foundation (No. 2013M541213 and 2015T80239), the Royal Society International Exchanges Cost Share Award with NSFC (No. IE160875).

REFERENCES

[1] A. Stevanovic, B. Xue, and M. Zhang, "Feature selection based on pso and decision-theoretic rough set model," pp. 2840–2847, 2013.

[2] L. Cervante, B. Xue, L. Shang, and M. Zhang, "A multi-objective feature selection approach based on binary pso and rough set theory," in *European Conference on Evolutionary Computation in Combinatorial Optimization*, pp. 25–36, 2013.

[3] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[4] L. Cervante, B. Xue, L. Shang, and M. Zhang, "A dimension reduction approach to classification based on particle swarm optimisation and rough set theory," in *Australasian Joint Conference on Artificial Intelligence*, pp. 313–325, 2012.

[5] Y. Qu, Q. Shen, Parthal, N. M. In, C. Shang, and W. Wu, "Fuzzy similarity-based nearest-neighbour classification as alternatives to their fuzzy-rough parallels," *International Journal of Approximate Reasoning*, vol. 54, no. 1, pp. 184–195, 2013.

[6] G. Cattaneo, *Abstract Approximation Spaces for Rough Theories.* 1998.

[7] A. Stevanovic, B. Xue, and M. Zhang, "Feature selection based on pso and decision-theoretic rough set model," pp. 2840–2847, 2013.

[8] L. Cervante, B. Xue, L. Shang, and M. Zhang, "Binary particle swarm optimisation and rough set theory for dimension reduction in classification," in *Evolutionary Computation*, pp. 2428–2435, 2013.

[9] G. Lin, J. Liang, and Y. Qian, "Multigranulation rough sets: From partition to covering," *Information Sciences*, vol. 241, no. 12, pp. 101–118, 2013.

[10] D. Chen, C. Wang, and Q. Hu, "A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets," *Information Sciences*, vol. 177, no. 17, pp. 3500–3518, 2007.

[11] Y. Yao, *An Outline of a Theory of Three-Way Decisions.* Springer Berlin Heidelberg, 2012.

[12] B. Sun, Z. Gong, and D. Chen, "Rough set theory for the interval-valued fuzzy information systems," *Computer Engineering and Applications*, vol. 178, no. 13, pp. 2794–2815, 2011.

[13] Y. Chen, Z. Zeng, Q. Zhu, and C. Tang, "Three-way decision reduction in neighborhood systems," *Applied Soft Computing*, vol. 38, no. C, pp. 942–954, 2016.

[14] M. A. Hall, "Correlation-based feature selection for machine learning," 1999.

[15] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 824–838, 2009.

[16] L. I. Smith, "A tutorial on principal components analysis," *Information Fusion*, vol. 51, no. 3, p. 52, 2002.

[17] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[18] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 38, no. 11, pp. 88–95, 1995.

[19] Y. Y. Yao, "Three-way decisions with probabilistic rough sets," *Information Sciences*, vol. 180, no. 3, pp. 341–353, 2010.

[20] B. Xue, L. Cervante, L. Shang, W. N. Browne, and M. Zhang, "Binary pso and rough set theory for feature selection: A multi-objective filter based approach," *International Journal of Computational Intelligence and Applications*, vol. 13, no. 02, pp. 1450009–, 2014.

[21] Y. Yao and B. Yao, "Covering based rough set approximations," *Information Sciences*, vol. 200, no. 1, pp. 91–107, 2012.

[22] W. Zhu and F. Y. Wang, "On three types of covering-based rough sets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 8, pp. 1131–1144, 2007.

[23] C. Blake, "Uci repository of machine learning databases," 1998.

[24] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[25] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Thirteenth International Conference on International Conference on Machine Learning*, pp. 148–156, 1996.

[26] E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," 1998.