



# A fuzzy similarity-based rough set approach for attribute selection in set-valued information systems

Shivani Singh<sup>1</sup> · Shivam Shreevastava<sup>2</sup> · Tanmoy Som<sup>2</sup> · Gaurav Somani<sup>2</sup>

Published online: 23 July 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

Databases obtained from different search engines, market data, patients' symptoms and behaviours, etc., are some common examples of set-valued data, in which a set of values are correlated with a single entity. In real-world data deluge, various irrelevant attributes lower the ability of experts both in speed and in predictive accuracy due to high dimension and insignificant information, respectively. Attribute selection is the concept of selecting those attributes that ideally are necessary as well as sufficient to better describe the target knowledge. Rough set-based approaches can handle uncertainty available in the real-valued information systems after the discretization process. In this paper, we introduce a novel approach for attribute selection in set-valued information system based on tolerance rough set theory. The fuzzy tolerance relation between two objects using a similarity threshold is defined. We find reducts based on the degree of dependency method for selecting best subsets of attributes in order to obtain higher knowledge from the information system. Analogous results of rough set theory are established in case of the proposed method for validation. Moreover, we present a greedy algorithm along with some illustrative examples to clearly demonstrate our approach without checking for each pair of attributes in set-valued decision systems. Examples for calculating reduct of an incomplete information system are also given by using the proposed approach. Comparisons are performed between the proposed approach and fuzzy rough-assisted attribute selection on a real benchmark dataset as well as with three existing approaches for attribute selection on six real benchmark datasets to show the supremacy of proposed work.

**Keywords** Set-valued data · Rough set · Fuzzy tolerance relation · Degree of dependency · Attribute selection

## 1 Introduction

Many real applications in the area of machine learning and data mining consist of set-valued data, i.e. the data where the attribute value of an object is not unique but a set of values, for example, in a venture investment company, the set of evaluation results given by expert (Qian et al. 2010b), the set of languages for each person from the foreign language ability test (Qian et al. 2009), and in a medical database, the set of patients symptoms and activities (He and Naughton 2009), etc. These kinds of information systems are called set-valued information systems, which are another important type of data tables and are generalized models of single-valued information systems. An incomplete information system (Kryszkiewicz 1998, 1999; Dai 2013; Dai and Xu 2012; Dai et al. 2013; Leung and Li 2003; Yang et al. 2011, 2012) can be converted into a set-valued information system by replacing all missing values with the set of all possible values of each

---

Communicated by V. Loia.

✉ Shivam Shreevastava  
shivam.rs.apm@itbhu.ac.in  
Shivani Singh  
shivanihthakur030@gmail.com  
Tanmoy Som  
tsom.apm@itbhu.ac.in  
Gaurav Somani  
gaurav.somani.mat14@itbhu.ac.in

<sup>1</sup> DST-Centre for Interdisciplinary Mathematical Sciences, Institute of Science, BHU, Varanasi 221005, India

<sup>2</sup> Department of Mathematical Sciences, IIT (BHU), Varanasi 221005, India

attribute. Set-valued information systems are always used to portray the inexact and lost information in a given dataset, in which the attribute set may vary with time as new information is added.

Dataset dimensionality is the main hurdle for the computational application in pattern recognition and other machine learning tasks. In many real-world applications, the generation and expansion of data occur continuously and thousands of attributes are stored in databases. Gathering useful information and mining-required knowledge from an information system is the most difficult task in the area of knowledge-based system. Not all attributes are relevant to the learning tasks as they reduce the real performance of proposed algorithms and increase the training and testing times. In order to enhance the classification accuracy and knowledge prediction, attribute subset selection (Hu et al. 2008; Yang and Li 2010; Qian et al. 2010a; Qian et al. 2011; Shi et al. 2011; Jensen and Shen 2009; Jensen et al. 2009) plays a key role via eliminating redundant and inconsistent attributes. Attribute selection is the process of selecting the most informative attributes of a given information system to reduce the classification time, complexity and overfitting.

Rough set approximations (proposed by Pawlak 1991 and Pawlak and Skowron 2007a, b, c) are the central point of approaches to knowledge discovery. Rough set theory (RST) uses only internal information and does not depend on prior model conventions, which can be used to extract and signify the hidden knowledge available in the information systems. It has many applications in the fields of decision support, document analysis, data mining, pattern recognition, knowledge discovery and so on. In rough sets, several discrete partitions are needed in order to tackle real-valued attributes and then dependency of decision attribute over conditional attributes is calculated. The intrinsic error due to this discretization process is the main issue while computing the degree of dependency of real-valued attributes.

Dubois and Prade (1992) combines a fuzzy set (Zadeh 1996) with rough set and proposed a fuzzy rough set to provide an important tool in reasoning with uncertainty for real-valued datasets. Fuzzy rough sets combine distinct concepts of indiscernibility (for rough sets) and vagueness (for fuzzy sets) available in the datasets and successfully applied to many fields. However, very few researchers are working in the area of set-valued information systems under the framework of rough set model in fuzzy environment.

In this paper, we introduce a novel approach for attribute selection in set-valued information system based on tolerance rough set theory. We define a fuzzy relation between two objects of a set-valued information system. A fuzzy tolerance relation is introduced by using a similarity

threshold to avoid misclassification and perturbation in order to tackle uncertainty in a much better way. Based on this relation, we calculate tolerance classes of each object to determine lower and upper approximations of any subset of the universe of discourse. Positive region of decision attribute over a subset of conditional attributes can be calculated using lower approximations. Degree of dependency of decision attribute over a subset of conditional attributes is the ratio of cardinality of positive region and cardinality of the universe of discourse. Analogous results of rough set theory are established in case of our proposed method for validation. Moreover, we present a greedy algorithm to clearly demonstrate our approach without calculating degree of dependencies for each pair of attributes. Illustrative example datasets are given for better understanding of our proposed approach. We compare the proposed approach with other existing approaches on real datasets and test the statistical significance of the obtained results.

The rest of the paper is structured as follows. Related works are given in Sect. 2. In Sect. 3, basic definitions related to incomplete and set-valued information systems are given. The proposed concept for set-valued datasets is presented and thoroughly studied in Sect. 4. In Sect. 5, analogous results of rough set theory are verified for the new proposed approach. An algorithm for attribute selection of set-valued information system is presented in Sect. 6. Illustrative examples with comparative analysis are given to demonstrate proposed model in Sect. 7. In Sect. 8, experimental analysis is performed on six real benchmark datasets. Section 9 concludes our work.

## 2 Related works

Nowadays, set-valued datasets are generated through many sources. Dimensionality reduction is a key issue for such type of datasets in order to reduce complexity, time and cost. Different criteria have been proposed by a few researchers to deal with set-valued datasets and to evaluate the best suitable attributes in the process of attribute selection. Lipski (1979, 1981) gave the idea of representing an incomplete information system as a set-valued information system and studied their basic properties. He also investigated the semantic and logical problems often occur in an incomplete information system. Concepts of internal and external interpretations are introduced in the paper. Internal interpretation is shown to lead towards the notion of topological Boolean algebra and a modal logic, whereas external interpretation is related to referring queries directly to reality leads towards Boolean algebra and classical logic. Orłowska and Pawlak (1984) established a method to deal with non-deterministic information system

which is considered as set-valued data. They defined a language in order to define non-deterministic information and introduced the concept of knowledge representation system.

A generalized decision logic, which is an extension of decision logic studied by Pawlak, in interval set-valued information system is presented by Yao and Liu (1999). They introduced two types of satisfiabilities of a formula, namely interval degree truth and interval-level truth. They also proposed generalized decision logic DGL and interpreted this concept based on two types of satisfiabilities. A detailed discussion on inference rules is also presented. Yao (2001) presented a concept of granulation for a universe of discourse in set-valued information systems and reviewed the corresponding approximation structure. The concept of ordered granulation and approximation structures are used in defining stratified rough set approximations. He first defined a nested sequence of granulations and then corresponding nested sequence of rough set approximations, which leads to a more general approximation structure.

Shoemaker and Ruiz (2003) introduced an extension of apriori algorithm that is able to mine association rules from a set-valued data. They introduced two different algorithms for mining association rules from set-valued data and compared their outcomes. They established a system based on one of these algorithms and applied it on some biological datasets for justification. Set-valued information systems were presented by Guan and Wang (2006). To classify the universe of discourse, they proposed a tolerance relation and used maximal tolerance classes. They introduced the concept of relative reduct of maximum tolerance classes and used Boolean reasoning technique for calculating relative reduct by defining a discernibility function. The concepts of E-lower, A-upper and A-lower relative reducts for set-valued decision systems are also discussed in details.

For the conjunctive/disjunctive type of set-valued ordered information systems, a dominance-based rough set approach was introduced by Qian et al. (2010b). This model is based on substitution of indiscernibility relation by a dominance relation. They also developed a new approach to sorting for objects in disjunctive set-valued ordered information systems. This approach is useful in simplifying a disjunctive set-valued ordered information system. Criterion reduction for a set-valued ordered information system is also discussed. Based on variable precision relation, Yang et al. (2010) generalized the notion of Qian et al. by defining an extended rough set model and propounded variable precision dominance relation for set-valued ordered information systems. They presented an attribute reduction method based on the discernibility matrix approach by using their proposed relation.

Zhang et al. (2012) proposed matrix approaches based on rough set theory with dynamic variation of attributes in set-valued information systems. In this paper, they defined the lower and upper approximations directly by using basic vector generated by the relation matrix in the set-valued information system. The concept of updation of the lower and upper approximations is also introduced by use of the variation of the relation matrix. Luo et al. (2013) investigated the updating mechanisms for computing lower and upper approximations with the variation of the object set. Authors proposed two incremental algorithms for the updation of the defined approximations in disjunctive/conjunctive set-valued information systems. After experiments on several datasets for checking the performance of the proposed algorithms, they showed that the incremental approaches are way better than the non-incremental approaches.

Wang et al. (2013) defined a new fuzzy preference relation and fuzzy rough set technique for disjunctive-type interval and set-valued information systems. They discussed the concept of relative significance measure of conditional attributes in interval and set-valued decision systems by using the degree of dependency approach. In this paper, authors mainly focused on semantic interpretation of disjunctive type only. They also presented an algorithm for calculating fuzzy positive region in interval and set-valued decision systems.

An incremental algorithm was designed to reduce the size of dynamic set-valued information systems by Lang et al. (2014). They presented three different relations and investigated their basic properties. Two types of discernibility matrices based on these relations for set-valued decision systems are also introduced. Furthermore, using the proposed relations and information system homomorphisms, a large-scale set-valued information system is compressed into a smaller information system. They addressed the compression updating via variations of feature set, immigration and emigration of objects and alterations of attribute values.

In set-valued ordered decision systems, Luo et al. (2014) worked on maintaining approximations dynamically and studied the approximations of decision classes by defining the dominant and dominated matrices via a dominance relation. The updating properties for dynamic maintenance of approximations were also introduced, when the evolution of the criteria values with time occurs in the set-valued decision system. Firstly, they constructed a matrix-based approach for computing lower and upper approximations of upward and downward unions of decision classes. Furthermore, incremental approaches for updating approximations are presented by modifying relevant matrices without retraining from the start on all accumulated training data.

Shu and Qian (2014) presented an attribute selection method for set-valued data based on mutual information of the unmarked objects. Mutual information-based feature selection methods use the concept of dependency among features. Unlike the traditional approaches, here the mutual information is calculated on the unmarked objects in the set-valued data. Furthermore, mutual information-based feature selection algorithm is developed and implemented on an universe of discourse to fasten the feature selection process. Due to the dynamic variation of criteria values in the set-valued information systems, Luo et al. (2015) presented the properties for dynamic maintenance of approximations. Two incremental algorithms for modernizing the approximations in disjunctive/conjunctive set-valued information system are presented corresponding to the addition and removal of criteria values, respectively.

Most of the above approaches are based on classical and rough set techniques, which have their own limitations of discretization, which leads to information loss. Rough set in fuzzy environment-based methods deal with uncertainty as well as noise available in information system in a much better way as compared to classical and rough set-based approaches without requirement of any discretization process. Dai et al. (2013) defined a fuzzy relation between two objects and constructed a fuzzy rough set model for attribute reduction in set-valued information systems based on discernibility matrices. In this paper, the similarity of two objects in set-valued information system is taken up to a threshold value in order to avoid misclassification and perturbation and a tolerance rough set-based attribute selection is presented by using degree of dependency approach.

### 3 Preliminaries

In this section, we describe some basic concepts, symbolization and examples of set-valued information system.

**Definition 3.1** (Huang 1992) A quadruple  $IS = (U, AT, V, h)$  is called an Information System, where  $U = \{u_1, u_2, \dots, u_n\}$  is a non-empty finite set of objects, called the universe of discourse,  $AT = \{a_1, a_2, \dots, a_n\}$  is a non-empty finite set of attributes.  $V = \bigcup_{a \in AT} V_a$  where  $V_a$  is the set of attribute values associated with each attribute  $a \in AT$  and  $h : U \times AT \rightarrow V$  is an information function that assigns particular values to the objects against attribute set such that  $\forall a \in AT, \forall u \in U, h(u, a) \in V_a$ .

**Definition 3.2** (Huang 1992): In an information system, if each attribute has a single entity as attribute value, then it is called single-valued information system, otherwise it is known as set-valued information system. Set-valued

**Table 1** Set-valued information system

$U$	$c_1$	$c_2$	$c_3$	$c_4$
$u_1$	{1,2,3,4}	{0,1}	{1,2}	0.4
$u_2$	{2,3}	{2,3}	{1}	0.5
$u_3$	{1,2,3,4}	{1,2}	{1,2}	0.9
$u_4$	{2,3,4}	{0,1,2,3}	{0,1}	0.2
$u_5$	{2,4}	{0,1,2}	{0,1}	1

information system is a generalization of the single-valued information system, in which an object can have more than one attribute values. Table 1 illustrates a set-valued information system.

**Definition 3.3** (Guan and Wang 2006): An information system  $IS = (U, AT, V, h)$  is said to be a set-valued decision system if  $AT = C \cup D$  where  $C$  is a non-empty finite set of conditional attributes and  $D$  is a non-empty collection of decision attributes with  $C \cap D = \emptyset$ . Here  $V = V_C \cup V_D$  with  $V_C$  and  $V_D$  as the set of conditional attribute values and decision attribute values, respectively.  $h$  be a mapping from  $U \times C \cup D$  to  $V$  such that  $h : U \times C \rightarrow 2^{V_C}$  is a set-valued mapping and  $h : U \times D \rightarrow V_D$  is a single-valued mapping. Table 2 exemplifies a set-valued decision system.

To give a semantic interpretation of the set-valued data, many ways are given (Guan and Wang 2006), here we encapsulate them as two types. In *Type1*,  $h(u, a)$  is interpreted conjunctively, and in *Type2*,  $h(u, a)$  is interpreted disjunctively. For example, if  $a$  is an attribute, “speaking a language”, then  $h(u, a) = \{\text{Chinese, Spanish, English}\}$  can be inferred as:  $u$  speaks Chinese, Spanish and English in case of *Type1* and  $u$  speaks Chinese or Spanish or English, i.e.  $u$  can speak only one of them in case of *Type2*. Incomplete information systems with some unknown attribute values or partially known attribute values are of *Type2* set-valued information system.

In many real-world application problems, lots of missing data existed in the information system due to ambiguity and incompleteness. All missing values presented in

**Table 2** Set-valued decision system

$U$	$c_1$	$c_2$	$c_3$	$c_4$	$D$
$u_1$	{1,2,3,4}	{0,1}	{1,2}	0.4	1
$u_2$	{2,3}	{2,3}	{1}	0.5	1
$u_3$	{1,2,3,4}	{1,2}	{1,2}	0.9	2
$u_4$	{2,3,4}	{0,1,2,3}	{0,1}	0.2	1
$u_5$	{2,4}	{0,1,2}	{0,1}	1	2

**Table 3** Missing value dataset

$U$	$b_1$	$b_2$	$b_3$	$b_4$	$D$
$u_1$	0	1	1.0	10	1
$u_2$	1	2	*	15	1
$u_3$	*	*	*	20	2
$u_4$	0	3	2.0	15	2
$u_5$	*	*	1.5	25	1
$u_6$	1	1	2.5	20	1

**Table 4** Set-valued decision system obtained from Table 3

$U$	$b_1$	$b_2$	$b_3$	$b_4$	$D$
$u_1$	0	1	1.0	10	1
$u_2$	1	2	{1.0,1.5,2.0,2.5}	15	1
$u_3$	{0,1}	{1,2,3}	{1.0,1.5,2.0,2.5}	20	2
$u_4$	0	3	2.0	15	2
$u_5$	{0,1}	{1,2,3}	1.5	25	1
$u_6$	1	1	2.5	20	1

information system can be characterized by the set of all possible values of each attribute. This type of information systems can also be considered as a special case of set-valued information system. Table 3 is an example of an incomplete decision system, in which some attribute values of objects are missing. Table 4 illustrates the transformation of incomplete information system into a set-valued information system.

### 4 Proposed methodology

In this section, we define a new kind of fuzzy relation between two objects and supremacy over previously defined relation is shown using an example. Then, lower and upper approximations are defined using a threshold value on fuzzy similarity degree. Some basic properties of above-defined lower and upper approximations are also discussed.

#### 4.1 Fuzzy relation between objects

In this subsection, first we present the definition of tolerance relation available in the literature and then propose a new fuzzy relation between two objects. In continuation, we compared both the definition through an example.

**Definition 4.1** (Orlowska (1985), Yao (2001)) For a set-valued information system  $IS = (U, AT, V, h), \forall b \in AT$  and  $u_i, u_j \in U$ , tolerance relation is defined as

$$\widehat{T}_b = \{(u_i, u_j) | b(u_i) \cap b(u_j) \neq \phi\} \tag{1}$$

For  $B \subseteq AT$ , a tolerance relation is defined as

$$\widehat{T}_B = \{(u_i, u_j) | b(u_i) \cap b(u_j) \neq \phi, \forall b \in B\} = \bigcap_{b \in B} \widehat{T}_b \tag{2}$$

where  $(u_i, u_j) \in T_B$  implies that  $u_i$  and  $u_j$  are indiscernible (tolerant) with respect to a set of attributes  $B$ .

**Example 4.1** Let  $(U, AT, V, h)$  be a set-valued information system with  $b \in AT$  and  $u_1, u_2, u_3 \in U$  such that  $b(u_1) = \{v_1, v_2, v_3, v_2\}, b(u_2) = \{v_4, v_5, v_6, v_7\}$ . and  $b(u_3) = \{v_1, v_2, v_3\}$ . Then, by Definition 4.1, we say that both  $(u_1, u_2)$  and  $(u_1, u_3)$  belong to  $T_b$ , that is,  $u_1, u_2$  are indiscernible with respect to attribute  $b$  and  $u_1, u_3$  are indiscernible with respect to attribute  $b$  simultaneously.

It is obvious from above example that discernibility of  $u_1$  and  $u_3$  is more difficult than discernibility of  $u_1$  and  $u_2$ , but Definition 4.1 is not able to describe the extent to which two objects are related. To overcome this issue, we define a fuzzy relation for a set-valued dataset.

**Definition 4.2** Let  $SVIS = (U, AT, V, h), \forall b \in AT$  be a set-valued information system, then we define a fuzzy relation  $\widehat{R}_b$  as:

$$\mu_{\widehat{R}_b}(u_i, u_j) = \frac{2|b(u_i) \cap b(u_j)|}{|b(u_i)| + |b(u_j)|} \tag{3}$$

For a set of attributes  $B \subseteq A$ , a fuzzy relation  $\widehat{R}_B$  can be defined as

$$\mu_{\widehat{R}_B}(u_i, u_j) = \inf_{b \in B} \mu_{\widehat{R}_b}(u_i, u_j) \tag{4}$$

**Example 4.2** (Continued from Example 4.1). After calculating degree of similarity by using fuzzy relation as defined in Eq. (3), we get

$$\mu_{\widehat{R}_b}(u_1, u_2) = \frac{2|b(u_1) \cap b(u_2)|}{|b(u_1)| + |b(u_2)|} = \frac{2}{8} = 0.25$$

$$\mu_{\widehat{R}_b}(u_1, u_3) = \frac{2|b(u_1) \cap b(u_3)|}{|b(u_1)| + |b(u_3)|} = \frac{6}{7} = 0.86$$

Now, we can easily calculate the degree to which two objects are discernible. □

#### 4.2 Fuzzy tolerance relation-assisted rough approximations

In this subsection, a fuzzy tolerance relation using a threshold value is defined and lower and upper approximations of a set are presented

If we ignore some misclassification and perturbation by using a threshold value on fuzzy relation between two objects as given in Eq. (3), then the involvement of fuzzy

sets in the computation of fuzzy lower approximation will increase and fuzzy positive region enlarges. Thus, the knowledge representation ability becomes much stronger with respect to misclassification.

So we define a new kind of binary relation using a threshold value  $\alpha$  as follows:

$$\widehat{T}_b^\alpha = \left\{ (u_i, u_j) \mid \mu_{\widehat{R}_b}^-(u_i, u_j) \geq \alpha \right\} \tag{5}$$

where  $\alpha \in (0, 1)$  is a similarity threshold, which gives a level of similarity for insertion of objects within tolerance classes.

For a set of attributes  $B \subseteq AT$ , we define binary relation as:

$$\widehat{T}_B^\alpha = \left\{ (u_i, u_j) \mid \mu_{\widehat{R}_B}^-(u_i, u_j) \geq \alpha \right\} \tag{6}$$

where  $\mu_{\widehat{R}_B}^-(u_i, u_j)$  and  $\mu_{\widehat{R}_B}^+(u_i, u_j)$  are defined by Eqs. (3) and (4), respectively.

**Definition 4.3** A fuzzy binary relation  $\widetilde{R}_b(u_i, u_j)$  between objects  $u_i, u_j \in U$  is said to be a fuzzy tolerance relation if it is reflexive (i.e.  $\widetilde{R}_b(u_i, u_i) = 1, \forall u_i \in U$ ) and symmetric (i.e.  $\widetilde{R}_b(u_i, u_j) = \widetilde{R}_b(u_j, u_i), \forall u_i, u_j \in U$ )

**Lemma 4.1**  $\widehat{T}_b^\alpha$  is a tolerance relation.

*Proof*

(i) Reflexive:

$$\begin{aligned} \therefore \mu_{\widehat{R}_b}^-(u_i, u_j) &= \frac{2|b(u_i) \cap b(u_j)|}{|b(u_i)| + |b(u_j)|} \Rightarrow \mu_{\widehat{R}_b}^-(u_i, u_i) \\ &= \frac{2|b(u_i) \cap b(u_i)|}{|b(u_i)| + |b(u_i)|} = \frac{|b(u_i)|}{|b(u_i)|} = 1 \geq \alpha \\ &\Rightarrow (u_i, u_i) \in \widehat{T}_b^\alpha \end{aligned}$$

(ii) Symmetric:

Let  $(u_i, u_j) \in \widehat{T}_b^\alpha$   
 therefore,  $\mu_{\widehat{R}_b}^-(u_i, u_j) = \frac{2|b(u_i) \cap b(u_j)|}{|b(u_i)| + |b(u_j)|} \geq \alpha$   
 Now,  $\mu_{\widehat{R}_b}^-(u_j, u_i) = \frac{2|b(u_j) \cap b(u_i)|}{|b(u_j)| + |b(u_i)|} = \mu_{\widehat{R}_b}^-(u_i, u_j) \geq \alpha$   
 $\Rightarrow (u_j, u_i) \in \widehat{T}_b^\alpha$

Therefore,  $\widehat{T}_b^\alpha$  is a fuzzy tolerance relation. □

**Example 4.3** If we take  $\alpha = 0.3$  in Eq. (5) and apply it on Example 4.1, then we can see that only  $(u_1, u_3)$  belongs to  $\widehat{T}_b$ , i.e. only  $u_1$  and  $u_3$  are indiscernible with respect to

attribute  $b$ . So, our proposed definition gives more precise tolerance relation than previous ones.

Now, we define tolerance classes for an object  $u_i$  with respect to  $b \in B$  as follows:

$$[\widehat{T}_b]^\alpha(u_i) = \left\{ u_j \in U \mid u_i \widehat{T}_b^\alpha u_j \right\} \tag{7}$$

For a set of attributes  $B \subseteq AT$ ,

$$[\widehat{T}_B]^\alpha(u_i) = \left\{ u_j \in U \mid u_i \widehat{T}_B^\alpha u_j, \forall b \in B \right\} \tag{8}$$

Then we propose, lower and upper approximations of any object set  $X \subseteq U$  as:

$$\widehat{T}_b^\alpha \downarrow X = \left\{ u_i \in U \mid [\widehat{T}_b]^\alpha(u_i) \subseteq X \right\} \tag{9}$$

$$\widehat{T}_b^\alpha \uparrow X = \left\{ u_i \in U \mid [\widehat{T}_b]^\alpha(u_i) \cap X \neq \phi \right\} \tag{10}$$

The tuple  $\langle \widehat{T}_b^\alpha \downarrow X, \widehat{T}_b^\alpha \uparrow X \rangle$  is called a tolerance rough set.

### 4.3 Properties of lower and upper approximations

In this subsection, we will examine the results on lower and upper approximations equivalent to Dubois and Prade (1992) for our proposed approach

Let  $(U, C \cup D, V, h)$  be a set-valued decision system. Let  $B \subseteq C$  and  $X \subseteq U, \alpha \in (0, 1)$

**Theorem 4.1**  $\widehat{T}_B^\alpha \downarrow X \subseteq X \subseteq \widehat{T}_B^\alpha \uparrow X$

*Proof* Let  $x \in \widehat{T}_B^\alpha \downarrow X \Rightarrow [\widehat{T}_B]^\alpha(x) \subseteq X$

Since  $x \in [\widehat{T}_B]^\alpha(x) \Rightarrow x \in X$ .

Therefore,  $\widehat{T}_B^\alpha \downarrow X \subseteq X$ .

Now, let  $x \in X$ , since  $x \in [\widehat{T}_B]^\alpha(x) \Rightarrow [\widehat{T}_B]^\alpha(x) \cap X \neq$

$\phi \Rightarrow x \in \widehat{T}_B^\alpha \uparrow X$

Therefore,  $X \subseteq \widehat{T}_B^\alpha \uparrow X$

Hence,  $\widehat{T}_B^\alpha \downarrow X \subseteq X \subseteq \widehat{T}_B^\alpha \uparrow X$  □

**Theorem 4.2** Let  $B_1 \subseteq B_2 \subseteq C$ , then

(i)  $\widehat{T}_{B_1}^\alpha \downarrow X \subseteq \widehat{T}_{B_2}^\alpha \downarrow X$

(ii)  $\widehat{T}_{B_2}^\alpha \uparrow X \subseteq \widehat{T}_{B_1}^\alpha \uparrow X$

**Proof**

(i) Let  $x \in \widehat{T}_{B_1}^\alpha \downarrow X$  then  $[\widehat{T}_{B_1}^\alpha](x) \subseteq X$ ,  
 $\because B_1 \subseteq B_2 \Rightarrow [\widehat{T}_{B_2}^\alpha](x) \subseteq [\widehat{T}_{B_1}^\alpha](x)$   
 Thus,  $[\widehat{T}_{B_2}^\alpha](x) \subseteq X \Rightarrow x \in \widehat{T}_{B_2}^\alpha \downarrow X$ .  
 Hence,  $\widehat{T}_{B_1}^\alpha \downarrow X \subseteq \widehat{T}_{B_2}^\alpha \downarrow X$

$$\left\{ \begin{array}{l} \because B_1 \subseteq B_2 \\ [\widehat{T}_{B_1}^\alpha](x) = \{y \in U \mid \mu_{R_b}^\alpha(x, y) \geq \alpha, \forall b \in B_1\} \\ [\widehat{T}_{B_2}^\alpha](x) = \{y \in U \mid \mu_{R_b}^\alpha(x, y) \geq \alpha, \forall b \in B_2\} \\ \Rightarrow [\widehat{T}_{B_2}^\alpha](x) \subseteq [\widehat{T}_{B_1}^\alpha](x) \end{array} \right.$$

(ii) Let  $x \in \widehat{T}_{B_2}^\alpha \uparrow X$ , then  $[\widehat{T}_{B_2}^\alpha](x) \cap X \neq \varphi$ ,  
 Since,  $B_1 \subseteq B_2 \Rightarrow [\widehat{T}_{B_2}^\alpha](x) \subseteq [\widehat{T}_{B_1}^\alpha](x)$   
 Thus,  $[\widehat{T}_{B_1}^\alpha](x) \cap X \neq \varphi \Rightarrow x \in \widehat{T}_{B_1}^\alpha \uparrow X$ .  
 Hence,  $\widehat{T}_{B_2}^\alpha \uparrow X \subseteq \widehat{T}_{B_1}^\alpha \uparrow X$

**Theorem 4.3** Let  $\alpha_1 \leq \alpha_2$ , then

- (i)  $\widehat{T}_B^{\alpha_1} \downarrow X \subseteq \widehat{T}_B^{\alpha_2} \downarrow X$
- (ii)  $\widehat{T}_B^{\alpha_2} \uparrow X \subseteq \widehat{T}_B^{\alpha_1} \uparrow X$

**Proof**

(i) Let  $x \in \widehat{T}_B^{\alpha_1} \downarrow X \Rightarrow [\widehat{T}_B^{\alpha_1}](x) \subseteq X$   
 since  $[\widehat{T}_B^{\alpha_1}](x) = \{y \in U \mid \mu_{R_b}^\alpha(x, y) > \alpha_1, \forall b \in B\}$   
 and  $[\widehat{T}_B^{\alpha_2}](x) = \{y \in U \mid \mu_{R_b}^\alpha(x, y) > \alpha_2, \forall b \in B\}$   
 $\Rightarrow [\widehat{T}_B^{\alpha_2}](x) \subseteq [\widehat{T}_B^{\alpha_1}](x)$  (Since,  $\alpha_2 \geq \alpha_1$ )  
 $\Rightarrow [\widehat{T}_B^{\alpha_2}](x) \subseteq X$  (Since,  $[\widehat{T}_B^{\alpha_1}](x) \subseteq X$ ) (11)  
 $\Rightarrow x \in \widehat{T}_B^{\alpha_2} \downarrow X$

Hence,  $\widehat{T}_B^{\alpha_1} \downarrow X \subseteq \widehat{T}_B^{\alpha_2} \downarrow X$

(ii) Let  $y \in \widehat{T}_B^{\alpha_2} \uparrow X$ , then  $[\widehat{T}_B^{\alpha_2}](y) \cap X \neq \varphi$ . Since,  
 $[\widehat{T}_B^{\alpha_2}](y) \subseteq [\widehat{T}_B^{\alpha_1}](y)$  {by (11)}  
 $\Rightarrow [\widehat{T}_B^{\alpha_1}](y) \cap X \neq \varphi \Rightarrow y \in \widehat{T}_B^{\alpha_1} \uparrow X$ ,  
 Hence,  $\widehat{T}_B^{\alpha_2} \uparrow X \subseteq \widehat{T}_B^{\alpha_1} \uparrow X$

**Theorem 4.4**  $\widehat{T}_B^\alpha \downarrow (X^C) = (\widehat{T}_B^\alpha \uparrow (X))^C$ , where  $X^C$  denotes complement of set  $X$ .

**Proof**  $y \in \widehat{T}_B^\alpha \downarrow (X^C) \Leftrightarrow [\widehat{T}_B^\alpha](y) \subseteq X^C \Leftrightarrow [\widehat{T}_B^\alpha](y) \cap X = \varphi \Leftrightarrow y \notin \widehat{T}_B^\alpha \uparrow (X) \Leftrightarrow y \in (\widehat{T}_B^\alpha \uparrow (X))^C$ . Hence,  $\widehat{T}_B^\alpha \downarrow (X^C) = (\widehat{T}_B^\alpha \uparrow (X))^C$   $\square$

**Theorem 4.5** Let  $Y \subseteq U$  be another set of objects, then following properties hold.

- (i)  $\widehat{T}_B^\alpha \downarrow (X \cap Y) = \widehat{T}_B^\alpha \downarrow (X) \cap \widehat{T}_B^\alpha \downarrow (Y)$
- (ii)  $\widehat{T}_B^\alpha \uparrow (X \cup Y) = \widehat{T}_B^\alpha \uparrow (X) \cup \widehat{T}_B^\alpha \uparrow (Y)$

**Proof**

(i)  $z \in \widehat{T}_B^\alpha \downarrow (X \cap Y) \Leftrightarrow [\widehat{T}_B^\alpha](z) \subseteq X \cap Y \Leftrightarrow [\widehat{T}_B^\alpha](z) \subseteq X$  and  $[\widehat{T}_B^\alpha](z) \subseteq Y \Leftrightarrow z \in \widehat{T}_B^\alpha \downarrow (X)$  and  $z \in \widehat{T}_B^\alpha \downarrow (Y) \Leftrightarrow z \in \widehat{T}_B^\alpha \downarrow (X) \cap \widehat{T}_B^\alpha \downarrow (Y)$ .  
 Hence,  $\widehat{T}_B^\alpha \downarrow (X \cap Y) = \widehat{T}_B^\alpha \downarrow (X) \cap \widehat{T}_B^\alpha \downarrow (Y)$

(ii)  $z \in \widehat{T}_B^\alpha \uparrow (X \cup Y) \Leftrightarrow [\widehat{T}_B^\alpha](z) \cap (X \cup Y) \neq \varphi \Leftrightarrow ([\widehat{T}_B^\alpha](z) \cap X) \cup ([\widehat{T}_B^\alpha](z) \cap Y) \neq \varphi$   
 Either  $[\widehat{T}_B^\alpha](z) \cap X \neq \varphi$  or  $[\widehat{T}_B^\alpha](z) \cap Y \neq \varphi$ ,  
 either  $z \in \widehat{T}_B^\alpha \uparrow (X)$  or  $z \in \widehat{T}_B^\alpha \uparrow (Y)$ .  
 Therefore,  $z \in \widehat{T}_B^\alpha \uparrow (X) \cup \widehat{T}_B^\alpha \uparrow (Y)$   
 Hence,  $\widehat{T}_B^\alpha \uparrow (X \cup Y) = \widehat{T}_B^\alpha \uparrow (X) \cup \widehat{T}_B^\alpha \uparrow (Y)$   $\square$

**Theorem 4.6**

$$\widehat{T}_B^\alpha \downarrow (U) = U = \widehat{T}_B^\alpha \uparrow (U) \text{ and } \widehat{T}_B^\alpha \downarrow (\varphi) = \varphi = \widehat{T}_B^\alpha \uparrow (\varphi)$$

**Proof** Easy to check.  $\square$

**Theorem 4.7**

$$\widehat{T}_B^\alpha \downarrow ([\widehat{T}_B^\alpha](x)) = [\widehat{T}_B^\alpha](x) = \widehat{T}_B^\alpha \uparrow ([\widehat{T}_B^\alpha](x))$$

**Proof**

Since  $\widehat{T}_B^\alpha \downarrow (X) \subseteq X \subseteq \widehat{T}_B^\alpha \uparrow (X)$ ,  
 hence  $\widehat{T}_B^\alpha \downarrow \left( \left[ \widehat{T}_B^\alpha \right] (x) \right) \subseteq \left[ \widehat{T}_B^\alpha \right] (x) \subseteq \widehat{T}_B^\alpha \uparrow \left( \left[ \widehat{T}_B^\alpha \right] (x) \right)$  (12)

Now, we have to show that,

$$\left[ \widehat{T}_B^\alpha \right] (x) \subseteq \widehat{T}_B^\alpha \downarrow \left( \left[ \widehat{T}_B^\alpha \right] (x) \right) \text{ and } \left[ \widehat{T}_B^\alpha \right] (x) \supseteq \widehat{T}_B^\alpha \uparrow \left( \left[ \widehat{T}_B^\alpha \right] (x) \right).$$

If  $z \in \left[ \widehat{T}_B^\alpha \right] (x)$ , then  $\mu_{\widehat{R}_b}(x, z) \geq \alpha, \forall b \in B$  (13)

If  $y \in \left[ \widehat{T}_B^\alpha \right] (z)$ , then  $\mu_{\widehat{R}_b}(z, y) \geq \alpha, \forall b \in B$  (14)

If  $\widehat{T}_B^\alpha$  is an equivalence relation,

then,  $\min \left\{ \left[ \widehat{T}_B^\alpha \right] (x, z), \left[ \widehat{T}_B^\alpha \right] (z, y) \right\} \subseteq \left[ \widehat{T}_B^\alpha \right] (x, y) \forall x, y, z \in U$ , (by transitivity of  $\widehat{T}_B^\alpha$ ) (15)

From (13), (14) and (15), we can conclude that  $\mu_{\widehat{R}_b}(x, y) \geq \alpha, \forall b \in B$  then,  $y \in \left[ \widehat{T}_B^\alpha \right] (x)$ , hence,  $\left[ \widehat{T}_B^\alpha \right] (z) \subseteq \left[ \widehat{T}_B^\alpha \right] (x)$ , then  $z \in \widehat{T}_B^\alpha \downarrow \left( \left[ \widehat{T}_B^\alpha \right] (x) \right)$

hence,  $\left[ \widehat{T}_B^\alpha \right] (x) \subseteq \widehat{T}_B^\alpha \downarrow \left( \left[ \widehat{T}_B^\alpha \right] (x) \right)$ . (16)

Now, if  $z \in \widehat{T}_B^\alpha \uparrow \left( \left[ \widehat{T}_B^\alpha \right] (x) \right)$ , then  $\left[ \widehat{T}_B^\alpha \right] (z) \cap \left[ \widehat{T}_B^\alpha \right] (x) \neq \emptyset$  then  $\exists y \in U$  such that  $y \in \left[ \widehat{T}_B^\alpha \right] (z) \cap \left[ \widehat{T}_B^\alpha \right] (x)$ , then  $y \in \left[ \widehat{T}_B^\alpha \right] (z)$  and  $y \in \left[ \widehat{T}_B^\alpha \right] (x)$  then  $\mu_{\widehat{R}_b}(y, z) \geq \alpha$  and  $\mu_{\widehat{R}_b}(y, x) \geq \alpha, \forall b \in B$ .

therefore,  $\widehat{T}_B^\alpha \uparrow \left( \left[ \widehat{T}_B^\alpha \right] (x) \right) \subseteq \left[ \widehat{T}_B^\alpha \right] (x)$ . (17)

Hence, from (12), (16) and (17), we get the required result.  $\square$

**Theorem 4.8** (i)  $\widehat{T}_B^\alpha \downarrow \left( \{x\}^C \right) = \left( \left[ \widehat{T}_B^\alpha \right] (x) \right)^C$  (ii)  $\widehat{T}_B^\alpha \uparrow \left( \{x\} \right) = \left[ \widehat{T}_B^\alpha \right] (x)$

**Proof** (i)  $z \in \widehat{T}_B^\alpha \downarrow \left( \{x\}^C \right)$  iff  $\left[ \widehat{T}_B^\alpha \right] (z) \subseteq \{x\}^C \Leftrightarrow \left[ \widehat{T}_B^\alpha \right] (z) \not\subseteq \{x\} \Leftrightarrow z \notin \left[ \widehat{T}_B^\alpha \right] (x) \Leftrightarrow z \in \left( \left[ \widehat{T}_B^\alpha \right] (x) \right)^C$ .  
 (ii) if  $z \in \widehat{T}_B^\alpha \uparrow \left( \{x\} \right)$ , then,  $\left[ \widehat{T}_B^\alpha \right] (z) \cap \{x\} \neq \emptyset$ , then,  $x \in \left[ \widehat{T}_B^\alpha \right] (z)$ , then  $z \in \left[ \widehat{T}_B^\alpha \right] (x)$   $\square$

**Theorem 4.9**

- (i)  $\widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \downarrow (X) \right) = \widehat{T}_B^\alpha \downarrow (X)$
- (ii)  $\widehat{T}_B^\alpha \uparrow \left( \widehat{T}_B^\alpha \uparrow (X) \right) = \widehat{T}_B^\alpha \uparrow (X)$

**Proof** (i) Since,  $\widehat{T}_B^\alpha \downarrow (X) \subseteq X$ , now, replacing  $X$  by  $\widehat{T}_B^\alpha \downarrow (X)$ , we get

$$\widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \downarrow (X) \right) \subseteq \widehat{T}_B^\alpha \downarrow (X) \tag{18}$$

Now, let  $y \in \widehat{T}_B^\alpha \downarrow (X)$ , we have to show that  $y \in \widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \downarrow (X) \right)$

If  $y \in \widehat{T}_B^\alpha \downarrow (X)$ , then  $\left[ \widehat{T}_B^\alpha \right] (y) \subseteq X$  (19)

Let  $z \in \left[ \widehat{T}_B^\alpha \right] (y)$ , then  $\mu_{\widehat{R}_b}(z, y) \geq \alpha, \forall b \in B$  (20)

If  $u \in \left[ \widehat{T}_B^\alpha \right] (z)$ , this implies that  $\mu_{\widehat{R}_b}(u, z) \geq \alpha, \forall b \in B$  (21)

If  $\widehat{T}_B^\alpha$  is an equivalence relation, then from (20), (21) and transitivity of  $\widehat{T}_B^\alpha$ , we get  $\mu_{\widehat{R}_b}(u, y) \geq \alpha, \forall b \in B$ , it implies that  $u \in \left[ \widehat{T}_B^\alpha \right] (y)$ . From (19), we get  $u \in X$ . Since  $u \in \left[ \widehat{T}_B^\alpha \right] (z)$  and  $u \in X$ , hence,  $\left[ \widehat{T}_B^\alpha \right] (z) \subseteq X$ , then  $z \in \widehat{T}_B^\alpha \downarrow (X)$ .

Since  $z \in \left[ \widehat{T}_B^\alpha \right] (y)$  and  $z \in \widehat{T}_B^\alpha \downarrow (X)$ . This gives that  $\left[ \widehat{T}_B^\alpha \right] (y) \subseteq \widehat{T}_B^\alpha \downarrow (X)$ , this implies that  $y \in \widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \downarrow (X) \right)$ ,

hence,  $\widehat{T}_B^\alpha \downarrow (X) \subseteq \widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \downarrow (X) \right)$  (22)

Hence, from (18) and (22), we get the required result

(ii) Since,  $X \subseteq \widehat{T}_B^\alpha \uparrow (X)$ , then, replacing  $X$  by  $\widehat{T}_B^\alpha \uparrow (X)$ , we get

$$\widehat{T}_B^\alpha \uparrow (X) \subseteq \widehat{T}_B^\alpha \uparrow \left( \widehat{T}_B^\alpha \uparrow (X) \right) \tag{23}$$

Now, let  $y \in \widehat{T}_B^\alpha \uparrow \left( \widehat{T}_B^\alpha \uparrow (X) \right)$ , this implies that  $\left[ \widehat{T}_B^\alpha \right] (y) \cap \widehat{T}_B^\alpha \uparrow (X) \neq \emptyset$ . Let us consider  $z \in \left[ \widehat{T}_B^\alpha \right] (y) \cap \widehat{T}_B^\alpha \uparrow (X)$ , then  $z \in \left[ \widehat{T}_B^\alpha \right] (y)$  and  $z \in \widehat{T}_B^\alpha \uparrow (X)$  this gives that  $\mu_{\widehat{R}_b}(z, y) \geq \alpha, \forall b \in B$  and  $\left[ \widehat{T}_B^\alpha \right] (z) \cap X \neq \emptyset$ , this implies that  $\exists u$ , such that  $u \in \left[ \widehat{T}_B^\alpha \right] (z)$  and  $u \in X \Rightarrow \mu_{\widehat{R}_b}(u, z) \geq \alpha, \forall b \in B$  and  $u \in X$ . Since,  $\mu_{\widehat{R}_b}(z, y) \geq \alpha, \forall b \in B$  and  $\mu_{\widehat{R}_b}$



$(u, z) \geq \alpha, \forall b \in B$ , hence, using transitivity of  $\widehat{T}_B^\alpha$ , we can conclude that  $\mu_{R_b}^\alpha(u, y) \geq \alpha, \forall b \in B \Rightarrow u \in [\widehat{T}_B^\alpha](y)$ .

Since,  $u \in [\widehat{T}_B^\alpha](y)$  and  $u \in X$ , this provides that  $[\widehat{T}_B^\alpha](y) \cap X \neq \emptyset$ , this implies that

$$y \in \widehat{T}_B^\alpha \uparrow (X). \text{ Hence, } \widehat{T}_B^\alpha \uparrow \left( \widehat{T}_B^\alpha \uparrow (X) \right) \subseteq \widehat{T}_B^\alpha \uparrow (X) \quad (24)$$

From (23) and (24), we get the result.  $\square$

**Theorem 4.10**

$$(i) \widehat{T}_B^\alpha \downarrow (X) \subseteq \widehat{T}_B^\alpha \uparrow \left( \widehat{T}_B^\alpha \downarrow (X) \right) \subseteq \widehat{T}_B^\alpha \uparrow (X)$$

$$(ii) \widehat{T}_B^\alpha \downarrow (X) \subseteq \widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \uparrow (X) \right) \subseteq \widehat{T}_B^\alpha \uparrow (X)$$

**Proof**

(i) Since,  $X \subseteq \widehat{T}_B^\alpha \uparrow (X)$ , then, replacing  $X$  by  $\widehat{T}_B^\alpha \downarrow (X)$ , we get

$$\widehat{T}_B^\alpha \downarrow (X) \subseteq \widehat{T}_B^\alpha \uparrow \left( \widehat{T}_B^\alpha \downarrow (X) \right).$$

Now, let  $y \in \widehat{T}_B^\alpha \uparrow \left( \widehat{T}_B^\alpha \downarrow (X) \right)$ , it results in,

$$[\widehat{T}_B^\alpha](y) \cap \widehat{T}_B^\alpha \downarrow (X) \neq \emptyset.$$

It implies that  $\exists z \in U$ , such that

$$z \in [\widehat{T}_B^\alpha](y) \text{ and } z \in \widehat{T}_B^\alpha \downarrow (X), \text{ then, } z \in [\widehat{T}_B^\alpha]$$

$$(y) \text{ and } [\widehat{T}_B^\alpha](z) \subseteq X,$$

this implies that  $z \in [\widehat{T}_B^\alpha](y)$  and  $z \in X$ , this provides that

$$[\widehat{T}_B^\alpha](y) \cap X \neq \emptyset, \text{ then } y \in \widehat{T}_B^\alpha \uparrow (X). \text{ Hence,}$$

$$\widehat{T}_B^\alpha \uparrow \left( \widehat{T}_B^\alpha \downarrow (X) \right) \subseteq \widehat{T}_B^\alpha \uparrow (X).$$

(ii) Since,  $\widehat{T}_B^\alpha \downarrow (X) \subseteq X$ , then replacing  $X$  by  $\widehat{T}_B^\alpha \uparrow (X)$ , we can conclude that

$$\widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \uparrow (X) \right) \subseteq \widehat{T}_B^\alpha \uparrow (X).$$

Now, let  $y \in \widehat{T}_B^\alpha \downarrow (X)$ , this implies that  $[\widehat{T}_B^\alpha](y) \subseteq X$ .

Since,  $X \subseteq \widehat{T}_B^\alpha \uparrow (X)$ , hence,  $[\widehat{T}_B^\alpha](y) \subseteq \widehat{T}_B^\alpha \uparrow (X)$ , this results in,  $y \in \widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \uparrow (X) \right)$ .

$$\text{Hence, } \widehat{T}_B^\alpha \downarrow (X) \subseteq \widehat{T}_B^\alpha \downarrow \left( \widehat{T}_B^\alpha \uparrow (X) \right). \quad \square$$

In next section, we propose an attribute selection method for set-valued information system based on degree

of dependency approach using above-defined lower approximation.

**5 Degree of dependency-based attribute selection**

Based on rough set as defined in Sect. 4, positive region of set of decision attributes  $D$  over set of conditional attributes  $B$  is defined as:

$$POS_B^\alpha(D) = \bigcup_{X \in U/D} (\widehat{T}_B^\alpha \downarrow X) \quad (25)$$

where  $U/D$  = collection of classes having objects with same decision values.

**Theorem 5.1** Let  $(U, C \cup D, V, h)$  be a set-valued decision system and  $X \subseteq U, \alpha \in (0, 1)$ . If  $B_1 \subseteq B_2 \subseteq C$ , then  $POS_{B_1}^\alpha(D) \subseteq POS_{B_2}^\alpha(D)$ .

**Proof** If  $B_1 \subseteq B_2$ , we have  $\widehat{T}_{B_1}^\alpha \downarrow X \subseteq \widehat{T}_{B_2}^\alpha \downarrow X$ , as proved in Theorem 4.3(i), so that  $\bigcup_{X \in U/D} (\widehat{T}_{B_1}^\alpha \downarrow X) \subseteq \bigcup_{X \in U/D} (\widehat{T}_{B_2}^\alpha \downarrow X)$ . Therefore,  $POS_{B_1}^\alpha(D) \subseteq POS_{B_2}^\alpha(D)$ .  $\square$

**Theorem 5.2** Let  $(U, C \cup D, V, h)$  be a set-valued decision system and  $X \subseteq U, \alpha \in (0, 1)$ . If  $\alpha_1 \leq \alpha_2$ , then  $POS_B^{\alpha_1}(D) \subseteq POS_B^{\alpha_2}(D)$ .

**Proof** If  $\alpha_1 \leq \alpha_2$ , we have  $\widehat{T}_B^{\alpha_1} \downarrow X \subseteq \widehat{T}_B^{\alpha_2} \downarrow X$ , as proved in Theorem 4.2(i), so that  $\bigcup_{X \in U/D} (\widehat{T}_B^{\alpha_1} \downarrow X) \subseteq \bigcup_{X \in U/D} (\widehat{T}_B^{\alpha_2} \downarrow X)$ . Hence,  $POS_B^{\alpha_1}(D) \subseteq POS_B^{\alpha_2}(D)$ .  $\square$

Now, using the definition of positive region, we compute degree of dependency of decision attribute  $D$  over set of conditional attributes  $B$  as:

$$\Gamma_B(D) = \frac{|POS_B^\alpha(D)|}{|U|} \quad (26)$$

where  $|\cdot|$  = cardinality of a set and  $\Gamma_B(D) \in [0, 1]$

**Theorem 5.3** (Monotonicity of  $\Gamma_B(D)$ ) Suppose that  $B \subseteq AT, \{c\}$  be an arbitrary conditional attribute that belong to the dataset and  $D$  be the set of decision attributes and  $\alpha \in (0, 1)$ , then  $\widehat{\Gamma}_{B \cup \{c\}}^\alpha(D) \geq \widehat{\Gamma}_B(D)$

**Proof** Since,  $\widehat{T}_B^\alpha = \left\{ (u_i, u_j) \mid \mu_{R_B}^\alpha(u_i, u_j) \geq \alpha \right\}, \widehat{T}_{B \cup \{c\}}^\alpha = \left\{ (u_i, u_j) \mid \mu_{R_{B \cup \{c\}}}^\alpha(u_i, u_j) \geq \alpha \right\} \Rightarrow \widehat{T}_{B \cup \{c\}}^\alpha \subseteq \widehat{T}_B^\alpha$ . Therefore,  $[\widehat{T}_{B \cup \{c\}}^\alpha](u_i) \subseteq [\widehat{T}_B^\alpha](u_i) \Rightarrow \widehat{T}_{B \cup \{c\}}^\alpha \downarrow X \supseteq \widehat{T}_B^\alpha \downarrow X$ . Since,  $POS_B(D) = \bigcup_{X \in U/D} (\widehat{T}_B^\alpha \downarrow X)$   $\square$

By Eq. (2),  $POS_{B \cup \{c\}}(D) \supseteq POS_B(D)$

Now, since  $\Gamma_B(D) = \frac{|POS_B(D)|}{|U|}$ , this implies that

$$\Gamma_{B \cup \{c\}}(D) \geq \Gamma_B(D).$$

**Input:** C, the set of all conditional attributes; D, the set of decision attributes;  $\Gamma_C(D)$ : degree of dependency of D over C.

**Output:** R, the attribute subset (reduct set)

```

R ← {};  $\Gamma_{new} = 0$ ;  $\Gamma_{old} = 0$ ;
do
{
    S ← R
     $\Gamma_{new} = \Gamma_{old}$ 
    foreach u ∈ (C - R)
        if  $\Gamma_{R \cup \{u\}}(D) > \Gamma_S(D)$ 
            S ← R ∪ {u}
             $\Gamma_{old} = \Gamma_S(D)$ 
            R ← S
    }
while  $\Gamma_{new} \neq \Gamma_{old}$ 
return R
    
```

It selects those conditional attributes, which provide a maximum increase in the degree of dependency of decision attribute. The proposed algorithm is given as follows:

A subset B of the conditional attribute set C is said to be a reduct of SVDS if

$$\Gamma_B(D) = \Gamma_C(D) \tag{27}$$

$$\Gamma_{B - \{b_i\}}(D) < \Gamma_B(D), \forall b_i \in B$$

The selection of attributes in reduct set is achieved by comparing the degree of dependencies of decision attribute over sets of conditional attributes. Attributes are selected one by one until the reduct set provides the same quality of classifications as the original set.

### 6 An algorithm for tolerance rough set-based attribute selection of set-valued data

In this section, a quick reduct algorithm for attribute selection of set-valued information system is presented by using degree of dependency method based on tolerance relation. Initially, the proposed algorithm starts with an empty set and add attributes one by one to calculate degree of dependencies of decision attribute over a set of condi-

The main advantage of proposed algorithm is that it produces a close-to-minimal reduct set of a decision system without thoroughly checking all possible subsets of conditional attributes.

Now, we apply above proposed algorithm on some example datasets to demonstrate our approach.

## 7 Illustrative examples

**Example 7.1** Consider a set-valued decision system as given in Table 2. A fuzzy tolerance relation between objects  $u_i, u_j \in U$ , calculated by using Eq. (3), is given in Table 5.

We calculate the degree of dependency of decision attribute d over conditional attribute  $c_1$  as follows, taking  $\alpha = 0.70$ ,

Tolerance Classes:

$$[\hat{T}_{c_1}^\alpha](u_1) = \{u_1, u_3, u_4\}, [\hat{T}_{c_1}^\alpha](u_2) = \{u_2\},$$

$$[\hat{T}_{c_1}^\alpha](u_3) = \{u_1, u_3, u_4\},$$

$$[\hat{T}_{c_1}^\alpha](u_4) = \{u_1, u_3, u_4\}, [\hat{T}_{c_1}^\alpha](u_5) = \{u_5\}$$

**Table 5** Fuzzy tolerance relation

$\mu_{\tilde{R}_{c_1}}(u_i, u_j)$		$\mu_{\tilde{R}_{c_2}}(u_i, u_j)$		$\mu_{\tilde{R}_{c_3}}(u_i, u_j)$		$\mu_{\tilde{R}_{c_4}}(u_i, u_j)$													
1	0.67	1	0.86	0.67	1	0	0.5	0.67	0.8	1	0.67	1	0.5	0.5	1	0.5	0.9	0.8	0.4
0.67	1	0.67	0.8	0.5	0	1	0.5	0.67	0.4	0.67	1	0.67	0.67	0.67	0.9	1	0.6	0.7	0.5
1	0.67	1	0.86	0.67	0.5	0.5	1	0.67	0.8	1	0.67	1	0.5	0.5	0.5	0.6	1	0.3	0.9
0.86	0.8	0.86	1	0.8	0.67	0.67	0.67	1	0.86	0.5	0.67	0.5	1	1	0.8	0.7	0.3	1	0.2
0.67	0.5	0.67	0.8	1	0.8	0.4	0.8	0.87	1	0.5	0.67	0.5	1	1	0.4	0.5	0.9	0.2	1

$U/d = \{d_1, d_2\}$

$d_1 = \{u_1, u_2, u_4\}, d_2 = \{u_3, u_5\}$

Lower approximation of  $U/d$  is calculated as:

$$\tilde{T}_{c_1}^\alpha \downarrow d_1 = \{u_2\}, \tilde{T}_{c_1}^\alpha \downarrow d_2 = \{u_5\}$$

So, positive region of  $d$  over  $c_1$  is calculated as:

$$\begin{aligned} POS_{c_1}(d) &= \bigcup_{X \in U/d} (\tilde{T}_b^\alpha \downarrow X) = (\tilde{T}_b^\alpha \downarrow d_1) \cup (\tilde{T}_b^\alpha \downarrow d_2) \\ &= \{u_2\} \cup \{u_5\} = \{u_2, u_5\} \end{aligned}$$

Now, degree of dependency of  $d$  over  $c_1$  is calculated as:

$$\Gamma_{\{c_1\}}(d) = \frac{|POS_{c_1}(d)|}{|U|} = \frac{2}{5} = 0.4$$

Similarly, we can calculate degree of dependency of decision attribute with respect to other conditional attributes,

$$\begin{aligned} \Gamma_{\{c_2\}}(d) &= \frac{3}{5} = 0.6, \quad \Gamma_{\{c_3\}}(d) = \frac{1}{5} = 0.2, \\ \Gamma_{\{c_4\}}(d) &= \frac{3}{5} = 0.6, \end{aligned}$$

since,  $\Gamma_{\{c_3\}}(d) < \Gamma_{\{c_1\}}(d) < \Gamma_{\{c_2\}}(d) = \Gamma_{\{c_4\}}(d)$

Either  $c_2$  or  $c_4$  will be the member of reduct set. Suppose,  $c_2$  is first reduct member. We will add other attributes to  $c_2$  one by one and calculate corresponding degree of dependencies by using Eq. (8), we get

$$\begin{aligned} \tilde{T}_{\{c_1, c_2\}}^\alpha(u_1) &= \{u_1\}, \tilde{T}_{\{c_1, c_2\}}^\alpha(u_2) = \{u_2\}, \\ \tilde{T}_{\{c_1, c_2\}}^\alpha(u_3) &= \{u_3, u_4\}, \\ \tilde{T}_{\{c_1, c_2\}}^\alpha(u_4) &= \{u_3, u_4\}, \tilde{T}_{\{c_1, c_2\}}^\alpha(u_5) = \{u_5\} \\ \tilde{T}_{\{c_1, c_2\}}^\alpha \downarrow d_1 &= \{u_1, u_2\}, \tilde{T}_{\{c_1, c_2\}}^\alpha \downarrow d_2 = \{u_5\} \\ POS_{\{c_1, c_2\}}(d) &= \tilde{T}_{\{c_1, c_2\}}^\alpha \downarrow d_1 \cup \tilde{T}_{\{c_1, c_2\}}^\alpha \downarrow d_2 = \{u_1, u_2, u_5\} \\ \Gamma_{\{c_1, c_2\}}(d) &= \frac{3}{5} = 0.6 \end{aligned}$$

Similarly,  $\Gamma_{\{c_2, c_3\}}(d) = \frac{3}{5} = 0.6, \quad \Gamma_{\{c_2, c_4\}}(d) = \frac{5}{5} = 1$

Since degree of dependency cannot exceed 1,  $\{c_2, c_4\}$  will be the reduct set of set-valued decision system as given in Table 2.

Applying the method of Dai et al. (2013) on this example set-valued dataset, we get the same reduct set  $\{c_2, c_4\}$ , but, when we change the value of parameter  $\alpha$  from 0.7 to 0.9, our approach gives  $\{c_2\}$  as reduct set. Therefore, proposed approach gives the facility to get the best minimal reduct for a set-valued decision system.

**Example 7.2** After converting an incomplete decision system as given in Table 3 into a set-valued decision system by replacing missing attribute values to set of all possible attribute values for any object, we get Table 4. Again, similar to Example 7.1, the fuzzy tolerance relation between objects  $u_i, u_j \in U$  is given in Table 6.

Taking,  $\alpha = 0.4$ ,

$d_1 = \{u_1, u_2, u_5, u_6\}, d_2 = \{u_3, u_4\}$

Calculating degree of dependency of decision attribute  $D$  over conditional attribute  $c_1$ ,

$$\begin{aligned} [\tilde{T}_{c_1}^\alpha](u_1) &= \{u_1, u_3, u_4, u_5\}, [\tilde{T}_{c_1}^\alpha](u_2) = \{u_2, u_3, u_5, u_6\}, \\ [\tilde{T}_{c_1}^\alpha](u_3) &= \{u_1, u_2, u_3, u_4, u_5, u_6\}, \\ [\tilde{T}_{c_1}^\alpha](u_4) &= \{u_1, u_3, u_4, u_5\}, [\tilde{T}_{c_1}^\alpha](u_5) = \{u_1, u_2, u_3, u_4, u_5, u_6\}, \\ [\tilde{T}_{c_1}^\alpha](u_6) &= \{u_2, u_3, u_5, u_6\} \end{aligned}$$

Now,  $\tilde{T}_{c_1}^\alpha \downarrow d_1 = \phi, \tilde{T}_{c_1}^\alpha \downarrow d_2 = \phi$

$POS_{c_1}(d) = \phi$  and  $\Gamma_{\{c_1\}}(d) = 0$

Similarly, for other conditional attributes,

$$\Gamma_{\{c_2\}}(d) = 0, \Gamma_{\{c_3\}}(d) = 0, \Gamma_{\{c_4\}}(d) = 0.67,$$

Since the degree of dependency is the highest for  $c_4$ ,  $c_4$  will be the first member of the reduct set. Similar to Example 7.1, on adding other attributes to  $c_4$ , we can calculate corresponding degree of dependencies as follows:

$$\Gamma_{\{c_1, c_4\}}(d) = 1, \Gamma_{\{c_2, c_4\}}(d) = 1, \Gamma_{\{c_3, c_4\}}(d) = 0.67$$

**Table 6** Fuzzy tolerance relation

$\mu_{\tilde{R}_{b_1}}(u_i, u_j)$		$\mu_{\tilde{R}_{b_2}}(u_i, u_j)$		$\mu_{\tilde{R}_{c_3}}(u_i, u_j)$		$\mu_{\tilde{R}_{b_4}}(u_i, u_j)$																	
1	0	0.67	1	0.67	0	1	0	0.5	0	0.5	1	1	0.4	0.4	0	0	0	1	0	0	0	0	1
0	1	0.67	0	0.67	1	0	1	0.5	0	0.5	0	0.4	1	1	0.4	0.4	0.4	0	1	0	1	0	0
0.67	0.67	1	0.67	1	0.67	0.5	0.5	1	0.5	1	0.5	0.4	1	1	0.4	0.4	0.4	0	0	1	0	0	0
1	0	0.67	1	0.67	0	0	0	0.5	1	0.5	0	0	0.4	0.4	1	0	0	0	1	0	1	0	0
0.67	0.67	1	0.67	1	0.67	0.5	0.5	1	0.5	1	0.5	0	0.4	0.4	0	1	0	0	0	0	0	1	0
0	1	0.67	0	0.67	1	1	0	0.5	0	0.5	1	0	0.4	0.4	0	0	1	1	0	0	0	0	1

**Table 7** Effect of  $\alpha$  on reduct set

Values of $\alpha$	Reducts
$\alpha = 0.4$	$\{c_1, c_4\}$ or $\{c_2, c_4\}$
$\alpha = 0.5$	$\{c_3, c_4\}$
$\alpha = 0.6$	$\{c_2, c_4\}$ or $\{c_3, c_4\}$
$\alpha = 0.7$	$\{c_2, c_3\}$ or $\{c_2, c_4\}$

Hence,  $\{c_1, c_4\}$  or  $\{c_2, c_4\}$  will be the reduct set of incomplete decision system as given in Table 3.

Here,  $\alpha$  is user-oriented, so we can find the best minimal reduct by changing the value of  $\alpha$  as follows (Table 7):

So, expert can decide the value of  $\alpha$  according to domain in order to find the best suitable reduct set of a decision system with missing values.

**Example 7.3** Let us consider a practical situation from the foreign language ability test in Shanxi University, China. Results can be inferred as a conjunctive set-valued information system. We have classified the whole test into four factors: Audition, Spoken language, Reading and Writing. The test results are given in Table 8, which can be downloaded from (<http://www.yuhuaqian.com>), where  $U = \{u_1, u_2, u_3, \dots, u_{49}, u_{50}\}$ . For convenience purpose, we have used abbreviations for Audition, Spoken language, Reading and Writing as A, S, R and W, respectively.

We have applied the same process as described in Example 7.1 and calculate reduct for this dataset by taking  $\alpha = 0.67$  as follows:

$$\Gamma_{\{c_1\}}(D) = 0, \Gamma_{\{c_2\}}(D) = \frac{2}{50}, \Gamma_{\{c_3\}}(D) = \frac{9}{50},$$

$$\Gamma_{\{c_4\}}(D) = \frac{16}{50}$$

$$\Gamma_{\{c_1, c_4\}}(D) = \frac{32}{50}, \Gamma_{\{c_2, c_4\}}(D) = \frac{43}{50}, \Gamma_{\{c_3, c_4\}}(D) = \frac{31}{50},$$

$$\Gamma_{\{c_1, c_2, c_4\}}(D) = \frac{48}{50}, \Gamma_{\{c_2, c_3, c_4\}}(D) = \frac{48}{50},$$

$$\Gamma_{\{c_1, c_2, c_3, c_4\}}(D) = \frac{48}{50}.$$

So, reduct of the decision system is either  $\{c_2, c_3, c_4\}$  or  $\{c_1, c_2, c_4\}$ .

So far, we have performed the experimental analysis for attribute selection of set-valued information system by applying proposed rough set-based approach in Example 7.1. We have compared proposed approach with an existing approach to find the close-to-minimal reduct set by changing the value of the parameter  $\alpha$ . In Example 7.2, we have dealt the problem of attribute selection in an incomplete information system through conversion into set-valued information system by replacing missing attribute values for an object with the set of all possible attribute values. Also the effect of parameter  $\alpha$  has been shown to find a minimal reduct set, which depend on users' choice. In Example 7.3, we have successfully applied our approach in a practical situation obtained from foreign language ability test in Shanxi University, China.

### 8 Experimental results and analysis

To check the efficiency of the proposed attribute reduction algorithm for set-valued information systems, we perform some experiments on a PC with specifications given in Table 9. We conduct our experiments on six real datasets taken from the University of California, Irvine (UCI) Machine Learning Repository in (Blake 1998). All six real datasets are incomplete decision systems (special case of set-valued decision system) given in Table 10. For the experimental work, we use the WEKA tool with ten fold cross-validation technique (Hall et al. 2009). In the experiments, we select three attribute reduction algorithms for comparisons. There are two statistical approaches Relief-F (Robnik-Šikonja and Kononenko 2003) and correlation-based feature selection (Hall 1999), and one fuzzy rough set model (Dai and Tian 2013). For convenience, we denote them as Relief-F, CFS and FRSM, respectively. For calculation of classification accuracies, we use two classifiers, namely PART and J48. Finally, a paired  $t$  test is performed to ensure the significance of experimental

**Table 8** A set-valued decision table obtained from foreign language ability test in Shanxi University

Students	Audition	Spoken language	Reading	Writing	Evaluation
$x_1$	{E}	{E}	{F,G}	{F,G}	Poor
$x_2$	{E,F,G}	{E,F,G}	{F,G}	{E,F,G}	Good
$x_3$	{F,G}	{F}	{F,G}	{F,G}	Good
$x_4$	{E,F}	{E,G}	{F,G}	{F}	Poor
$x_5$	{F,G}	{F,G}	{F,G}	{F}	Poor
$x_6$	{F}	{F}	{E,F}	{E,F}	Poor
$x_7$	{E,F,G}	{E,F,G}	{E,G}	{E,F,G}	Good
$x_8$	{F,G}	{F}	{F,G}	{F,G}	Good
$x_9$	{E,G}	{G}	{F,G}	{F,G}	Poor
$x_{10}$	{E,F}	{E,G}	{F,G}	{E,F}	Good
$x_{11}$	{F}	{E,F}	{F}	{G}	Good
$x_{12}$	{E,F,G}	{E,G}	{E,F,G}	{E,G}	Poor
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_{46}$	{E,G}	{E,G}	{E,F}	{F,G}	Good
$x_{47}$	{E,F,G}	{E,F,G}	{E,F,G}	{E,G}	Poor
$x_{48}$	{E,F}	{E,G}	{F,G}	{F}	Poor
$x_{49}$	{E,F,G}	{F,G}	{E,F,G}	{E,G}	Poor
$x_{50}$	{F,G}	{F}	{F,G}	{F,G}	Good

**Table 9** The description of experiment environment

No.	Names	Model	Parameters
1.	CPU	Intel(R) Core(TM) i5-4210U	1.70 GHz, 2 Cores
2.	Memory	DDR3 SDRAM	8 GB 2401 MHz
3.	Hard disk	ST1000LM024	1 TB
4.	System	Windows 10	64-bit
5.	Platform	Python 2.7	Anaconda distribution for windows

**Table 10** The description of datasets

No.	Datasets	Abbreviation	Objects	Features	Classes
1.	Audiology_Standardized	Audiology	226	69	24
2.	Soyabean_Large	Soyabean	307	35	19
3.	Dermatology	Dermatology	366	34	6
4.	Hepatitis	Hepatitis	115	19	2
5.	Zoo	Zoo	101	17	7
6.	Processed_Cleveland	Cleveland	303	14	5

results obtained from the proposed approach, where the significance level is specified to be 0.05.

Size and the classification accuracy of the reduced feature subset obtained by FSRS are statistically compared with those acquired by FRSM, Relief-F and CFS by using the paired  $t$  test. The obtained results are listed in the last columns of Tables 11, 12 and 13. In tables, “w” denotes the number of win, “\*” denotes the number of tie and “l” denotes the number of loss achieved by the proposed FSRS approach, which is also written next to the values in Table 11, 12 and 13. Here, win means that the cardinality (or accuracy) of the feature subset obtained by FSRS is

significantly fewer (or higher) than that of FRSM, Relief-F or CFS; tie means that the results obtained by FSRS have no statistical difference with that of FRSM, Relief-F or CFS; and loss means that the proposed approach is statistically poor than other approaches.

### 8.1 Reduct size

After comparing proposed approach with other three approaches on chosen datasets, reduced average (avg.) feature subset size is given in Table 11. Effect of parameter  $\alpha$  on reduct size is also shown for the proposed approach.

**Table 11** Comparison of feature subset size (Avg. subset size)

Datasets	FSRS				FRSM	Relief-F	CFS	Paired <i>t</i> test (w/t/l)
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$				
Audiology	11.6	12.3	9.6	9.6	17.0 w	17.0 w	18.0 w	(3/0/0)
Soyabean	0	0	8.9	11.5	16.0 w	17.0 w	17.0 w	(3/0/0)
Dermatology	6.0	6.0	6.0	6.0	11.7 w	16.0 w	15.0 w	(3/0/0)
Hepatitis	2.9	2.9	2.9	2.9	6.3 w	5.0 w	9.6 w	(3/0/0)
Zoo	6.2	6.2	6.2	6.2	7.7 *	9.0 w	10.0 w	(2/1/0)
Cleveland	3.0	3.0	3.0	3.0	4.8 w	3.5*	3.0 *	(1/2/0)

**Table 12** Comparison of classification accuracies (rules-PART)

Datasets	Original	FSRS				FRSM	Relief-F	CFS	Paired <i>t</i> test (w/*/l)
		$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$				
Audiology	78.31	77.87	78.31	75.66	75.66	76.12*	80.08*	77.43*	(0/3/0)
Soyabean	91.94	0	0	83.74	88.43	87.70*	87.99 1	85.06*	(0/2/1)
Dermatology	94.53	91.26	91.26	91.26	91.26	90.25w	91.53w	90.71w	(3/0/0)
Hepatitis	67.74	83.22	83.22	83.22	83.22	76.13w	84.51*	81.93*	(1/2/0)
Zoo	92.07	95.04	95.04	95.04	95.04	85.14w	95.04*	95.04*	(1/2/0)
Cleveland	53.79	54.12	54.12	54.12	54.12	49.17w	57.75 1	54.78*	(1/1/1)

**Table 13** Comparison of classification accuracies (trees-J48)

Datasets	Original	FSRS				FRSM	Relief-F	CFS	Paired <i>t</i> test(w/*/l)
		$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$				
Audiology	77.87	78.31	78.76	77.87	77.87	76.12*	78.76*	77.87*	(0/3/0)
Soyabean	91.50	0	0	86.23	87.99	87.99*	87.84 1	85.36*	(0/2/1)
Dermatology	93.98	92.62	92.62	92.62	92.62	88.28*	90.44*	87.70*	(0/3/0)
Hepatitis	58.06	83.22	83.22	83.22	83.22	78.06w	84.51*	81.29*	(1/2/0)
Zoo	92.07	94.05	94.05	94.05	94.05	89.10w	95.04*	95.04*	(1/2/0)
Cleveland	52.14	54.12	54.12	54.12	54.12	54.12*	56.10*	54.12*	(0/3/0)

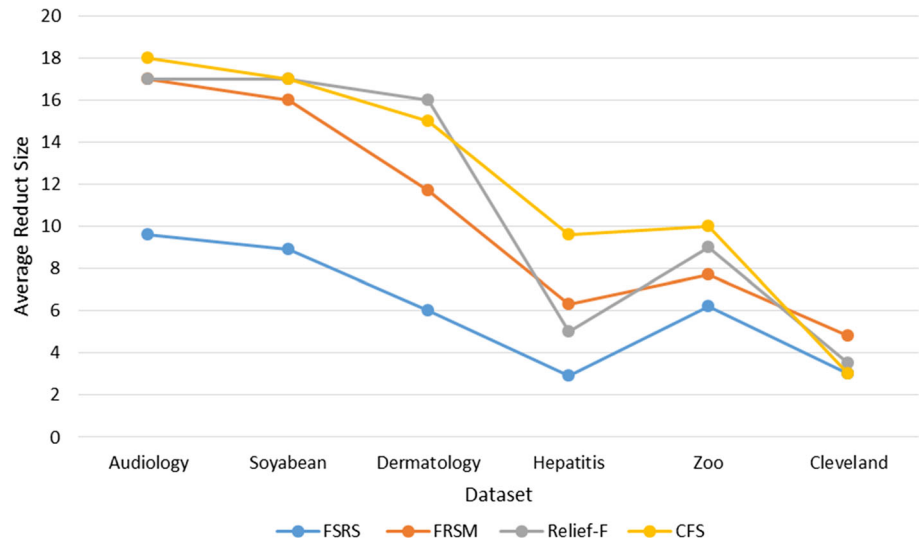
Results obtained from Table 11 indicate that all four feature selection algorithms exclude most of the features available in unreduced datasets. But it can be observed that the proposed approach provides more reduced or equal reduct size as compared to other three approaches. As for the hepatitis dataset having 19 attributes, proposed approach (FSRS) selects 3 (nearest integer is taken) attributes while FRSM, Relief-F and CFS select 6, 5 and 10 attributes, respectively. It shows that FSRS has a redundancy-removing capacity, while other algorithms do not completely eradicate the redundant features from the selected feature subset.

**Effect of parameter  $\alpha$**  The parameter  $\alpha$  need to be set individually according to different datasets because of their different correlation strengths. In Table 11, selected feature subset for soyabean dataset is varying with the change in parameter  $\alpha$ . At  $\alpha = 0$  and  $\alpha = 0.3$ , FSRS does not provide

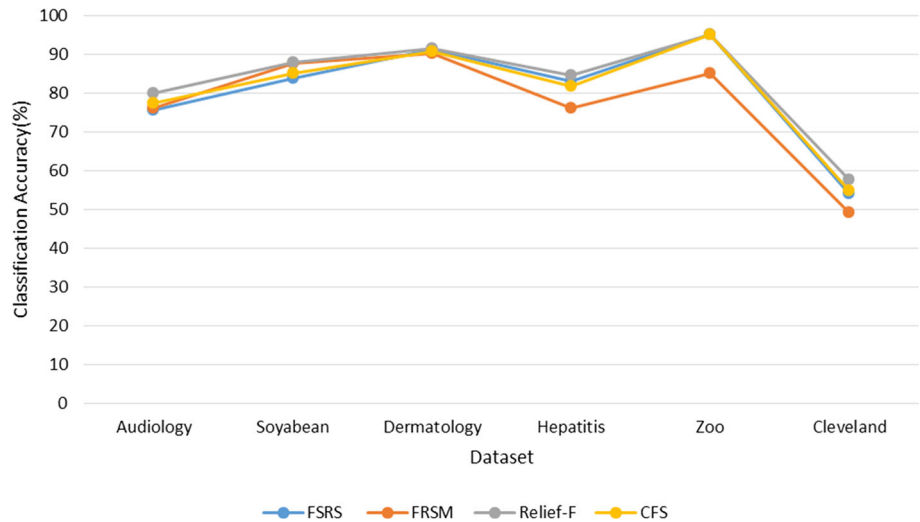
any reduct elements, but as we increase the threshold parameter  $\alpha$ , FSRS outperforms other approaches. For example, at  $\alpha = 0.5$  and  $\alpha = 0.7$ , FSRS select 9 and 12 attributes, respectively, while FRSM, Relief-F and CFS select 16, 17 and 17 attributes, respectively. Also, for audiology data, subset size is affected by parametric value  $\alpha$  as at  $\alpha = 0.1$ ,  $\alpha = 0.3$  and  $\alpha = 0.5$ , FSRS gives reduct sizes 12, 12 and 10, respectively, but at  $\alpha = 0.7$ , reduct size is same as at  $\alpha = 0.5$ . Remaining 4 datasets show no variation for the value of parameter  $\alpha$  as they provide the same number of selected attributes for all chosen values of  $\alpha$ .

**Statistical analysis** It can be seen clearly from the results of *t* test which is applied between reduct sizes of FSRS (at  $\alpha = 0.5$ ), FRSM, Relief-F and CFS algorithms (presented in Table 11) that, for almost all the datasets, FSRS outperforms the other three reduction algorithms in terms of

**Fig. 1** Variation of reduced feature subset sizes with four algorithms



**Fig. 2** Variation of classification accuracies with classifier PART for four algorithms



cardinality of the feature subset. In Table 11, FSRS achieves significantly fewer features for all the datasets except the dataset Cleveland and zoo. For Cleveland dataset, FSRS is significantly equivalent to Relief-F and CFS algorithms in terms of subset size. In summary, out of total 18 paired *t* test performance results it gets 15 wins, 3 ties and 0 loss.

**8.2 Classification accuracy**

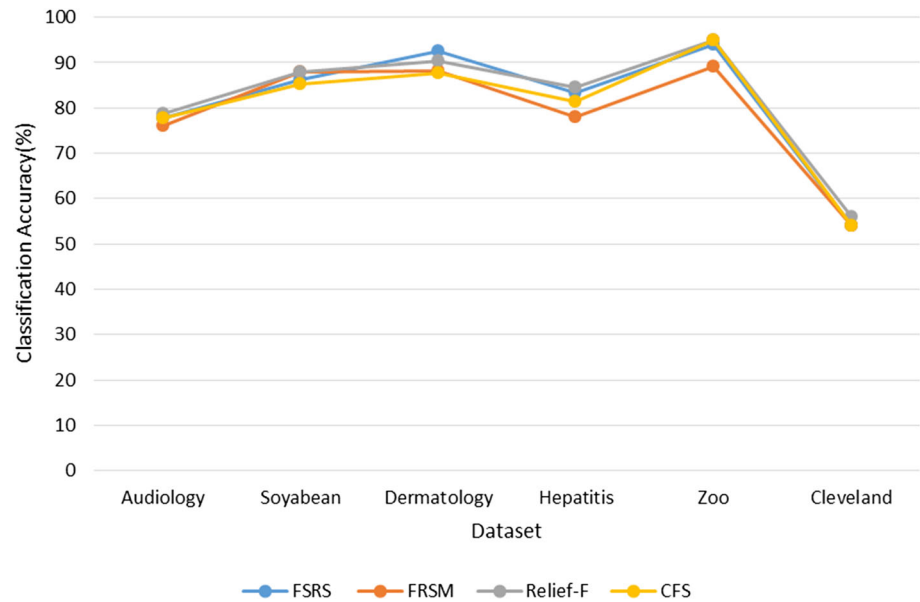
Comparison of classification accuracies for classifiers PART and J48 is presented in Tables 12 and 13, respectively. The classification accuracies are presented in percentage. In Table 12, for soyabean, dermatology and zoo datasets, classification accuracies evaluated by FSRS algorithm are higher or equal as compared to rest three algorithms while for other datasets, it shows mixed behaviour. For example, classification accuracy for Cleveland

dataset is 54.12 in case of FSRS approach and 49.17, 57.75 and 54.78 in case of FRSM, Relief-F and CFS approaches, respectively. Similarly, for hepatitis dataset, proposed algorithm provides better classification accuracy as FRSM and CFS, but less than the Relief-F algorithm. In Table 13, we can find similar kind of results as in Table 12.

*Effect of parameter  $\alpha$*  As we change the value of  $\alpha$ , corresponding classification accuracies are also changing for datasets soyabean and audiology but not for other datasets. In Table 12, for  $\alpha = 0$  and  $\alpha = 0.3$ , classification accuracies in case of soyabean dataset are 0 but for  $\alpha = 0.5$  and  $\alpha = 0.7$ , classification accuracies are 83.74 and 88.43, respectively. So, by changing the value of parameter  $\alpha$ , we can get better classification accuracies than rest three approaches. There is no effect of parameter on rest four datasets.

*Statistical analysis* Paired *t* test is applied between classification accuracies of FSRS (at  $\alpha = 0.5$ ), FRSM, Relief-F and CFS approaches, and results are shown in the last column

**Fig. 3** Variation of classification accuracies with classifier J48 for four algorithms



of Tables 12 and 13. In Table 12, FSRS achieves significantly higher or equivalent accuracy for all datasets except the datasets soyabean and Cleveland where it loses to Relief-F approach. In summary, out of total 18 paired  $t$  test performance results FRSR approach gets 6 wins, 10 ties and 2 losses. Similarly, in Table 13, out of total 18 paired  $t$  test performance results it gets 2 wins, 16 ties and 1 loss. Therefore, the proposed FSRS approach is effective and way better than other approaches in terms of both acquiring few features and achieving high classification accuracy.

More detailed change trendline of each approach on the six datasets is displayed in Figs. 1, 2 and 3. Figure 1 represents a comparison of reduced average feature subset size for all four algorithms on six datasets. It can be observed that the proposed FSRS algorithm selects least number of features as a member of reduct set. Figures 2 and 3 display more detailed change trend of the algorithms in classification accuracy with the number of selected attributes on all chosen dataset. It is obvious from figures that the proposed approach provides either higher or nearly equal classification accuracy for all six datasets.

After summarizing the comparison tables and graphs above, we can finally conclude that the proposed FSRS algorithm is an acceptable choice to select the best feature subsets in set-valued decision systems.

## 9 Conclusion and future work

In this paper, we have defined a tolerance relation for set-valued decision systems and given a novel approach for attribute selection based on the rough set concept using a similarity threshold. Lower and upper approximations have been defined by using fuzzy tolerance relation and

presented a method to calculate degree of dependency of decision attribute over a subset of conditional attributes. Some important results on lower and upper approximations, positive regions and the degree of dependencies have been validated using our approach. Moreover, we have presented an algorithm along with some illustrative examples for better understanding of the proposed approach. In Example 7.2, we have applied our method to an incomplete information system, in which some attribute values were missing. Effect of parameter  $\alpha$  on reduct set of set-valued decision systems has been shown. We have compared the proposed approach with three existing approaches on a six real benchmark datasets and observed that our model is able to find the minimal reduct with higher accuracy. We have also ensured that the proposed approach is statistically more significant in comparison with the other approaches by using paired  $t$  test technique.

In the future, we will investigate some robust models for set-valued information system to avoid misclassification and noise. Set-valued information systems with missing decision values will be taken into consideration from the viewpoint of updating the process of knowledge discovery. We intend to find some generalizations of fuzzy rough set-based attribute selection for set-valued decision systems.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Research involving human participants and/or animals** This article does not contain any studies with human participants or animals performed by any of the authors.



## References

- Blake CL (1998) UCI Repository of machine learning databases, Irvine, University of California. <http://www.ics.uci.edu/~mllearn/MLRepository.html>. Accessed 1 Feb 2019
- Dai J (2013) Rough set approach to incomplete numerical data. *Inf Sci* 241:43–57
- Dai J, Tian H (2013) Fuzzy rough set model for set-valued data. *Fuzzy Sets Syst* 229:54–68
- Dai J, Xu Q (2012) Approximations and uncertainty measures in incomplete information systems. *Inf Sci* 198:62–80
- Dai J, Wang W, Tian H, Liu L (2013) Attribute selection based on a new conditional entropy for incomplete decision systems. *Knowl-Based Syst* 39:207–213
- Dubois D, Prade H (1992) Putting rough sets and fuzzy sets together. In: Słowiński R (ed) *Intelligent decision support*. Springer, Dordrecht, pp 203–232
- Guan YY, Wang HK (2006) Set-valued information systems. *Inf Sci* 176(17):2507–2525
- Hall M (1999) Correlation-based feature selection for machine learning. PhD Thesis, Department of Computer Science, Waikato University, New Zealand
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *ACM SIGKDD Explor Newsl* 11(1):10–18
- He Y, Naughton JF (2009) Anonymization of set-valued data via top-down, local generalization. *Proc VLDB Endow* 2(1):934–945
- Hu Q, Yu D, Liu J, Wu C (2008) Neighborhood rough set based heterogeneous feature subset selection. *Inf Sci* 178(18):3577–3594
- Huang SY (ed) (1992) *Intelligent decision support: handbook of applications and advances of the rough sets theory*, vol 11. Springer, Berlin
- Jensen R, Shen Q (2009) New approaches to fuzzy-rough feature selection. *IEEE Trans Fuzzy Syst* 17(4):824–838
- Jensen R, Cornelis C, Shen Q (2009) Hybrid fuzzy-rough rule induction and feature selection. In: *FUZZ-IEEE 2009, IEEE international conference on fuzzy systems, 2009*. IEEE, pp. 1151–1156
- Kryszkiewicz M (1998) Rough set approach to incomplete information systems. *Inf Sci* 112(1–4):39–49
- Kryszkiewicz M (1999) Rules in incomplete information systems. *Inf Sci* 113(3–4):271–292
- Lang G, Li Q, Yang T (2014) An incremental approach to attribute reduction of dynamic set-valued information systems. *Int J Mach Learn Cybern* 5(5):775–788
- Leung Y, Li D (2003) Maximal consistent block technique for rule acquisition in incomplete information systems. *Inf Sci* 153:85–106
- Lipski W Jr (1979) On semantic issues connected with incomplete information databases. *ACM Trans Database Syst (TODS)* 4(3):262–296
- Lipski W Jr (1981) On databases with incomplete information. *J ACM (JACM)* 28(1):41–70
- Luo C, Li T, Chen H, Liu D (2013) Incremental approaches for updating approximations in set-valued ordered information systems. *Knowl-Based Syst* 50:218–233
- Luo C, Li T, Chen H (2014) Dynamic maintenance of approximations in set-valued ordered decision systems under the attribute generalization. *Inf Sci* 257:210–228
- Luo C, Li T, Chen H, Lu L (2015) Fast algorithms for computing rough approximations in set-valued decision systems while updating criteria values. *Inf Sci* 299:221–242
- Orłowska E (1985) Logic of nondeterministic information. *Stud Logica* 44(1):91–100
- Orłowska E, Pawlak Z (1984) Representation of nondeterministic information. *Theor Comput Sci* 29(1–2):27–39
- Pawlak Z (1991) *Rough Sets: theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Dordrecht
- Pawlak Z, Skowron A (2007a) Rough sets and Boolean reasoning. *Inf Sci* 177(1):41–73
- Pawlak Z, Skowron A (2007b) Rough sets: some extensions. *Inf Sci* 177(1):28–40
- Pawlak Z, Skowron A (2007c) Rudiments of rough sets. *Inf Sci* 177(1):3–27
- Qian Y, Dang C, Liang J, Tang D (2009) Set-valued ordered information systems. *Inf Sci* 179(16):2809–2832
- Qian Y, Liang J, Pedrycz W, Dang C (2010a) Positive approximation: an accelerator for attribute reduction in rough set theory. *Artif Intell* 174(9–10):597–618
- Qian YH, Liang JY, Song P, Dang CY (2010b) On dominance relations in disjunctive set-valued ordered information systems. *Int J Inf Technol Decis Mak* 9(01):9–33
- Qian J, Miao DQ, Zhang ZH, Li W (2011) Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation. *Int J Approx Reason* 52(2):212–230
- Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn* 53(1–2):23–69
- Shi Y, Yao L, Xu J (2011) A probability maximization model based on rough approximation and its application to the inventory problem. *Int J Approx Reason* 52(2):261–280
- Shoemaker CA, Ruiz C (2003) Association rule mining algorithms for set-valued data. In: *International conference on intelligent data engineering and automated learning*, Springer, Berlin, pp. 669–676
- Shu W, Qian W (2014) Mutual information-based feature selection from set-valued data. In: *26th IEEE international conference on tools with artificial intelligence (ICTAI), 2014*, IEEE, pp. 733–739
- Wang H, Yue HB, Chen XE (2013) Attribute reduction in interval and set-valued decision information systems. *Appl. Math.* 4(11):1512
- Data sets in articles. <http://www.yuhuaqian.com>
- Yang T, Li Q (2010) Reduction about approximation spaces of covering generalized rough sets. *Int J Approx Reason* 51(3):335–345
- Yang QS, Wang GY, Zhang QH, MA XA (2010) Disjunctive set-valued ordered information systems based on variable precision dominance relation. *J. Guangxi Normal Univ Nat Sci Ed* 3:84–88
- Yang X, Zhang M, Dou H, Yang J (2011) Neighborhood systems-based rough sets in incomplete information system. *Knowl Based Syst* 24(6):858–867
- Yang X, Song X, Chen Z, Yang J (2012) On multigranulation rough sets in incomplete information system. *Int J Mach Learn Cybern* 3(3):223–232
- Yao YY (2001) Information granulation and rough set approximation. *Int J Intell Syst* 16(1):87–104
- Yao YY, Liu Q (1999) A generalized decision logic in interval-set-valued information tables. In: *International workshop on rough sets, fuzzy sets, data mining, and granular-soft computing*, Springer, Berlin, pp. 285–293
- Zadeh LA (1996) Fuzzy sets. In: *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*, pp. 394–432
- Zhang J, Li T, Ruan D, Liu D (2012) Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems. *Int J Approx Reason* 53(4):620–635