Information Sciences 181 (2011) 2086–2110

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Positive approximation and converse approximation in interval-valued fuzzy rough sets

Yi Cheng [a,b,*], Duoqian Miao [b], Qinrong Feng [b]

[a] Department of Equipment and Engineering, Sichuan College of Architectural Technology, Deyang 618000, China
[b] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 13 June 2008
Received in revised form 25 January 2011
Accepted 26 January 2011
Available online 1 February 2011

Keywords:
Interval-valued fuzzy rough sets
Positive approximation
Converse approximation
Rule extraction
Granulation order

ABSTRACT

Methods of fuzzy rule extraction based on rough set theory are rarely reported in incomplete interval-valued fuzzy information systems. Thus, this paper deals with such systems. Instead of obtaining rules by attribute reduction, which may have a negative effect on inducting good rules, the objective of this paper is to extract rules without computing attribute reducts. The data completeness of missing attribute values is first presented. Positive and converse approximations in interval-valued fuzzy rough sets are then defined, and their important properties are discussed. Two algorithms based on positive and converse approximations, namely, mine rules based on the positive approximation (MRBPA) and mine rules based on the converse approximation (MRBCA), are proposed for rule extraction. The two algorithms are evaluated by several data sets from the UC Irvine Machine Learning Repository. The experimental results show that MRBPA and MRBCA achieve better classification performances than the method based on attribute reduction.

1. Introduction

1.1. Interval-valued fuzzy rough sets

A basic issue in a rule-based system is extracting rules for classification or inference. The rough set approach uses only internal knowledge, avoids external parameters, and does not rely on prior model assumptions such as probabilistic distribution in statistical methods and basic probability assignment in the Dempster–Shafer theory. Its basic idea is to search for an optimal attribute set to generate rules through an objective knowledge induction process.

The classical rough set theory developed by Pawlak [24,25] is used only to describe sets. We are interested in extending the rough set model of Pawlak in two ways. To describe crisp and fuzzy concepts, Dubois and Prade [5,6] extended the basic idea of rough sets to a new model called fuzzy rough sets. This new model has been proven a promising tool for pattern recognition, data mining, and knowledge discovery [1–7,9,13–17,22,23,26–29,32–40,43–47,49,51–53]. In addition, there are symbolic values, real values, or interval values in a practical database [34]. For example, data such as current, ID, temperature, time, and voltage are often described by interval values. The traditional fuzzy rough set theory effectively cannot deal with these kinds of data. Extending the rough set theory of Pawlak to a wider application is

* Corresponding author at: Department of Computer Science and Technology, Tongji University, Shanghai 201804, China. Tel.: +86 15183855604.
E-mail address: chengyimail@sohu.com (Y. Cheng).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.01.033

necessary. Thus, the model of interval-valued fuzzy rough (IVFR) sets was developed [7,38]. Here, we review two studies in this domain.

Sun et al. [38] defined IVFR sets and presented the attribute reduction method, which addresses interval-valued fuzzy information systems with both crisp condition and interval-valued fuzzy decision attributes. The reduction process has three steps: (1) computing the discernibility matrix of the information system; (2) searching the consistent set of the condition attribute set; and (3) obtaining a reduct by computing the minimum consistent set. However, the definition of the discernibility matrix is the same as that in the rough set theory by Pawlak, which means the discernibility matrix is effective only for nominal attributes. When the condition attributes are numerical or fuzzy interval values, the reduction method is ineffective because it cannot compute the discernibility matrix.

Gong et al. [7] proposed a knowledge discovery method for interval-valued fuzzy information systems. The method classifies each object in a decision class according to its maximal membership represented by a fuzzy interval. However, the method is designed for interval-valued fuzzy information systems with both crisp condition and interval-valued fuzzy decision attributes. If the condition attribute set includes *m* attributes, then the antecedent of the rule must include *m* conditions; overfull conditions may reduce the classification accuracy and the applicability of the rules. Two memberships represented by fuzzy intervals are incomparable when one interval is nested in the other; rules cannot be generated in this case.

Aside from [7,38], few studies on fuzzy rule extraction are based on rough sets in interval-valued fuzzy information systems. Establishing a more practical model for fuzzy rule extraction in interval-valued fuzzy information systems is necessary. The model should satisfy the following three requirements. First, it can be applied to three types of interval-valued fuzzy information systems, namely, (1) crisp condition and interval-valued fuzzy decision, (2) interval-valued fuzzy condition and crisp decision, and (3) interval-valued fuzzy condition and decision. Second, the computational complexity of the model should be relatively low. Third, rules can be generated when one interval is nested in the other.

1.2. Rule extraction in IVFR sets

Attribute reduction usually serves as a preparatory step before rule extraction [24], whose objective is to reduce attributes and thus reduce the complexity of the rule extraction process. Various attribute reduction methods have been proposed in rough sets and in fuzzy rough sets [2,4,12–16,23,37,39,40,43–45,52,53]. The IVFR set theory generalizes the traditional fuzzy rough set theory; thus, extracting rules based on attribute reduction is natural. However, this paper does not intend to extract rules based on attribute reduction due to the following reasons. Attribute reduction methods can be classified into three types: one based on the positive region [2,15,16,37], one based on the discernibility matrix [39,40,43,50], and another based on entropy [12–14]. For example, Shen and Jensen [16,37] conducted pioneering studies on attribute reduction based on a positive region and proposed an attribute reduction algorithm. However, an obvious limitation is the algorithm may not be convergent on many real data sets or the selected attributes are unreliable. Moreover, the computational complexity of the algorithm often increases exponentially with increasing samples and attributes [2]. Bhatt and Gopal [2] developed Shen's algorithm by improving the definition of the lower approximation on a compact computational domain. However, the degree of dependency of a selected reduct may be larger than that of the entire attribute set due to the computing method of the positive region [40]. This is unreasonable because more attributes will offer better approximations in a rough set framework [40]. Tsang et al. [39,40] proposed an algorithm using a discernibility matrix to compute all attribute reducts. However, the computational complexity is NP-hard [40]. Hu et al. [12,13] proposed an attribute reduction method based on information entropy. The attribute reduction concept is not constructed using existing fuzzy approximation operators [47], and studying the structure of attribute reduction is difficult [49]. Each attribute reduction method has its characteristics and flaws. Therefore, rule extraction based on attribute reduction may be faulty. This paper intends to avoid the attribute reduction process and establish the structure of the approximation by introducing granulation order, and then extracting rules based on it.

From the viewpoint of granular computing, a concept is described by the upper and lower approximations under static granulation in the IVFR set, as defined by Sun [38]. Provided the granulation is unchangeable, it is unacceptable when the granulation is too fine or too coarse. Excessively fine granulation may increase time and cost, while an excessively coarse one may not satisfy requirements. We consider describing a concept under dynamic granulation. This means a proper granulation family can be selected to describe a target concept according to the practical requirement.

Granulation order in sets was introduced by Qian and co-workers [20,30]. In our study, a granulation order is extended to fuzzy information systems. A positive granulation order is defined by adding one condition attribute at a time, which naturally defines a positive approximation space. Given a positive approximation space, a fuzzy concept can be described by the upper and lower approximations. Based on the positive approximation, a rule extraction algorithm called mine rules based on the positive approximation (MRBPA) is proposed. It is characterized by a gradually dwindling universe and a monotonously increasing approximation precision as the positive granulation order becomes longer. Thus, the computational complexity of the algorithm can be reduced effectively. Similarly, a converse granulation order involves deleting one condition attribute at a time, which defines a converse approximation space. Given a converse approximation space, a fuzzy concept can be described by the upper and lower approximations. As an application of the converse approximation, an algorithm called mine rules based on the converse approximation (MRBCA) is proposed for rule extraction. The main characteristic of MRBCA is that much simpler rules can be extracted by keeping the approximation precision invariant.

1.3. Outline

The rest of this paper is organized as follows. Section 2 briefly introduces related discussions about interval-valued fuzzy sets and IVFR sets. In Section 3, an algorithm called completeness of missing attribute values in interval-valued fuzzy information systems (CMAVIFIS) is presented for data completeness in interval-valued fuzzy information systems. In Section 4, the positive approximation is proposed, and important properties are obtained. A rule extraction algorithm called MRBPA based on the positive approximation is then designed; an example is illustrated. In Section 5, converse approximation is presented, and useful properties are deduced. The convergence degree of an interval-valued fuzzy set is defined and proven to increase in a converse granulation order. A new rule extraction algorithm called MRBCA based on the converse approximation is proposed and illustrated. In Section 6, the performances of CMAVIFIS, MRBPA, and MRBCA are evaluated by several data sets from the UC Irvine Machine Learning Repository (UCI). Section 7 concludes the paper.

2. Preliminaries

In this section, we briefly review the basic concepts of interval-valued fuzzy sets and IVFR sets.

2.1. Interval-valued fuzzy sets

As a generalization of Zadeh's fuzzy set, interval-valued fuzzy sets was put forward for the first time by Gorzalczany [8] and Turksen [41]. In a fuzzy set, interval-valued membership is easier to be determined than the single-valued one. The interval-valued fuzzy set theory has been applied to the fields of approximate inference, signal transmission, and so on. We first review some basic concepts.

Let I be a closed unit interval, i.e., $I = [0,1]$. Let $[I] = \{a = [a^-, a^+] | a^- \leqslant a^+, a^-, a^+ \in I\}$. For $\forall a \in I$, define $\bar{a} = [a, a]$, it is obvious that $a \in [I]$.

Definition 1. If $a_i \in [I]$, $i \in J$, $J = \{1, 2, \ldots, m\}$, define

$$(1)\ \bigvee_{i\in J}[a_i^-, a_i^+] = \left[\bigvee_{i\in J} a_i^-, \bigvee_{i\in J} a_i^+\right];\quad (2)\ \bigwedge_{i\in J}[a_i^-, a_i^+] = \left[\bigwedge_{i\in J} a_i^-, \bigwedge_{i\in J} a_i^+\right];\quad (3)\ [a_i^-, a_i^+]^c = [1 - a_i^+, 1 - a_i^-].$$

In particular, for $a_i \in [I]$, $i = 1, 2$, define

$$(4)\ a_1 = a_2 \iff a_1^- = a_2^-,\quad a_1^+ = a_2^+;\quad (5)\ a_1 \leqslant a_2 \iff a_1^- \leqslant a_2^-,\quad a_1^+ \leqslant a_2^+;\quad (6)\ a_1 < a_2 \iff a_1 \leqslant a_2, a_1 \neq a_2;$$

$$(7)\ a_1 \leqslant_w a_2 \iff a_1^- + a_1^+ \leqslant a_2^- + a_2^+,\quad a_1^+ - a_1^- \leqslant a_2^+ - a_2^-;\quad (8)\ a_1 <_w a_2 \iff a_1 \leqslant_w a_2, a_1 \neq a_2.$$

Definition 2. Let X be an ordinary non-empty set, and the mapping A: $X \to [I]$ is called an interval-valued fuzzy set on X. The set of interval-valued fuzzy sets on X is denoted by $F^I(X)$.

Similar to fuzzy sets, the operators $\subseteq$, $\cap$, $\cup$, and the completeness of interval-valued fuzzy sets are defined as follows. For A, $B \in F^I(X)$, $A \subseteq B$ means $A(x) \leqslant B(x)$ for $\forall x \in X$, $(A \cap B)(x) = \wedge\{A(x), B(x)\}$, $(A \cup B)(x) = \vee\{A(x), B(x)\}$, $(\sim A)(x) = 1 - A(x)$.

Definition 3. If $A \in F^I(X)$, let $A(x) = [A^-(x), A^+(x)]$, where $x \in X$, then two fuzzy sets A^-: $X \to I$ and A^+: $X \to I$ are called lower fuzzy set and upper fuzzy set about A, respectively.

2.2. IVFR sets

Due to the complementarity between interval-valued fuzzy sets and rough sets, Sun [38] proposed a model called IVFR sets. Let U be a non-empty finite universe. A binary interval-valued fuzzy subset R of $U \times U$ is called an interval-valued fuzzy relation on U.

Definition 4 [38]. Let U be a non-empty finite universe. For the interval-valued fuzzy relation $R(R \in F^I(U \times U))$ on the universe U,

(1) R is reflexive, if $R(x, y) = \bar{1}$, for any x, $y \in U$;
(2) R is symmetric, if $R(x,y) = R(y,x)$, for any x, $y \in U$;
(3) R is transitive, if $R(x,z) \geqslant R(x,y) \wedge R(y,z)$, for any x, y, $z \in U$.

If R is reflexive, symmetric, and transitive, then R is an interval-valued fuzzy equivalence relation. $[x]_R$ is the fuzzy block containing x. It is an interval-valued fuzzy set on U defined by $\mu_{[x]_R}(y) = \mu_R(x, y)$ for all $y \in U$. The collection of all fuzzy blocks is denoted as U/R.

Definition 5. [38] Let (U,R) be an interval-valued fuzzy information system, where R is an interval-valued fuzzy equivalence relation on U. For any interval-valued fuzzy set F, the lower and upper approximations of F in the interval-valued fuzzy information system (U,R) are defined as follows:

$$\mu_{\underline{apr}_R(F)}(x) = \inf_{y\in U} \max\{\bar{1} - \mu_R(x,y), \mu_F(y)\} = \left[\mu_{\underline{apr}_R(F^-)}(x), \mu_{\underline{apr}_R(F^+)}(x)\right]$$
$$= \left[\inf_{y\in U} \max\{1 - \mu_{R^+}(x,y), \mu_{F^-}(y)\}, \inf_{y\in U} \max\{1 - \mu_{R^-}(x,y), \mu_{F^+}(y)\}\right],$$

$$\mu_{\overline{apr}_R(F)}(x) = \sup_{y\in U} \min\{\mu_R(x,y), \mu_F(y)\} = \left[\mu_{\overline{apr}_R(F^-)}(x), \mu_{\overline{apr}_R(F^+)}(x)\right]$$
$$= \left[\sup_{y\in U} \min\{\mu_{R^-}(x,y), \mu_{F^-}(y)\}, \sup_{y\in U} \min\{\mu_{R^+}(x,y), \mu_{F^+}(y)\}\right].$$

In view of $\mu_R(x,y) = \mu_{[x]_R}(y)$, the lower and upper approximations can also be denoted as follows:

$$\mu_{\underline{apr}_R(F)}(x) = \inf_{y\in U} \max\{\bar{1} - \mu_{[x]_R}(y), \mu_F(y)\} = \left[\mu_{\underline{apr}_R(F^-)}(x), \mu_{\underline{apr}_R(F^+)}(x)\right]$$
$$= \left[\inf_{y\in U} \max\{1 - \mu_{[x]_{R^+}}(y), \mu_{F^-}(y)\}, \inf_{y\in U} \max\{1 - \mu_{[x]_{R^-}}(y), \mu_{F^+}(y)\}\right],$$

$$\mu_{\overline{apr}_R(F)}(x) = \sup_{y\in U} \min\{\mu_{[x]_R}(y), \mu_F(y)\} = \left[\mu_{\overline{apr}_R(F^-)}(x), \mu_{\overline{apr}_R(F^+)}(x)\right]$$
$$= \left[\sup_{y\in U} \min\{\mu_{[x]_{R^-}}(y), \mu_{F^-}(y)\}, \sup_{y\in U} \min\{\mu_{[x]_{R^+}}(y), \mu_{F^+}(y)\}\right].$$

If for $\forall x \in U, \mu_{\underline{apr}_R(F)}(x) = \mu_{\overline{apr}_R(F)}(x)$, then the interval-valued fuzzy set F is definable about (U,R). Otherwise, the interval-valued fuzzy set F is rough about (U,R), and F is called an IVFR set.

If F is an ordinary fuzzy set of universe U, R is a fuzzy similarity relation, then $F^- = F^+$. Therefore, the IVFR set degenerates into a classical fuzzy rough set.

3. Data completeness in interval-valued fuzzy information systems

Data completeness in interval-valued fuzzy information systems is the usual prerequisite for rule extraction. The process of converting an incomplete interval-valued fuzzy information system into a complete one, i.e., complementing the missing attribute values with specified values, is called the completeness of incomplete interval-valued fuzzy information systems.

Multiple completeness methods have been proposed [10,11,19,31,42,54]. A simple method is to either delete objects that miss their attribute values or replace their values with the most common one. The second method is based on probability statistics, e.g., Bayes method and multivariate linear regression analysis. Here, a Bayesian formula is used to determine the probability distribution of the missing value [19]. This method either chooses the most likely value or divides the object into fractional objects, each with one possible value weighted according to the probabilities. In application, determining the probability distribution is difficult due to the vast state-space of the data set. Thus, the traditional statistical technique may not be the best choice. The third method is based on a classical rough set theory such as the rough set theory-based incomplete data analysis approach (ROUSTIDA) [42,54] and the method based on value tolerance relation (VTR) [42]. For ROUSTIDA, the attribute differences of objects are reflected by a discernibility matrix, and missing values are replaced with those of indiscernible objects. However, ROUSTIDA may repeatedly compute the discernibility matrix and analyze completeness. The time complexity of computing one discernibility matrix is $O(|U|(|U|-1)|C|/2)$ (C is the attribute set); thus, the time complexity of ROUSTIDA is high. Moreover, many transient information systems are generated in the supplement process, which may distort data [31]. For each object with missing values, VTR computes its tolerance class and complements missing values with those of the object with maximal tolerance relation. The time complexity is $O(|C||U||MOS|)$, where MOS is the set of objects with missing values. Generally, $|MOS|$ is less than $|U|$. Thus, the time complexity of VTR is less than that of ROUSTIDA. However, ROUSTIDA and VTR can only be used in information systems with nominal attribute values. Data in interval-valued fuzzy information systems are numerical; thus, conventional methods should be extended to process the numerical attributes directly. Considering the time complexity, we present a new data completeness method based on value tolerance relation for interval-valued fuzzy information systems.

Let $S = (U, C \cup D)$ be an interval-valued fuzzy information system with decision attributes. We call S an interval-valued fuzzy decision table, where the condition attributes in C are represented by fuzzy interval numbers, and the decision attributes in D are crisp or fuzzy interval numbers. U and C are denoted as $U = \{x_1, x_2, \ldots, x_n\}$, $C = \{a_1, a_2, \ldots, a_m\}$, and $|U| = n$, $|C| = m$.

Definition 6. The set of objects with missing attribute values is defined as $MOS = \{x_i | a_k(x_i) = *,\ a_k \in C\}$, where $*$ denotes a missing attribute value.

Definition 7. The similarity degree of x_i and x_j with regard to a_k is defined as follows:

$$sim_{a_k}(x_i,x_j) = sim^-_{a_k}(x_i,x_j) \wedge sim^+_{a_k}(x_i,x_j) = \left(1 - \frac{|a^-_k(x_i) - a^-_k(x_j)|}{\max(a^-_k) - \min(a^-_k)}\right) \wedge \left(1 - \frac{|a^+_k(x_i) - a^+_k(x_j)|}{\max(a^+_k) - \min(a^+_k)}\right),$$

where $\max(a_k)$ and $\min(a_k)$ are the maximum and minimum values of a_k, respectively.

Definition 8. Based on the similarity degree, a tolerance relation with respect to condition attribute set C is defined as follows:

$$S_C = \{(x_i,x_j) \in U \times U | \forall a_k \in C, sim_{a_k}(x_i,x_j) \geqslant \xi \vee a_k(x_i) = * \vee a_k(x_j) = *\},$$

where $\xi \in [0,1]$ is a threshold.

Definition 9. The tolerance class of an object x_i with respect to condition attribute set C is defined as follows:

$$T_C(x_i) = \{x_j | (x_i,x_j) \in S_C, x_j \neq x_i\}.$$

Definition 10. For any two objects $x_i, x_j \in U$, if $a_k(x_j) \neq *$, then the probability of x_i is similar to x_j with respect to a_k is defined as follows:

$$P_{a_k}(x_i,x_j) = \frac{1}{2}\left(\frac{a^-_k(x_i)_{\max} - a^-_k(x_i)_{\min}}{\max(a^-_k) - \min(a^-_k)} + \frac{a^+_k(x_i)_{\max} - a^+_k(x_i)_{\min}}{\max(a^+_k) - \min(a^+_k)}\right),$$

where $a^-_k(x_i)_{\max}$ and $a^-_k(x_i)_{\min}$ are the upper and lower limits of $a^-_k(x_i)$, respectively, satisfying $sim^-_{a_k}(x_i,x_j) \geqslant \xi$ and $\min(a^-_k) \leqslant a^-_k(x_i) \leqslant \max(a^+_k)$; the meanings of $a^+_k(x_i)_{\max}$ and $a^+_k(x_i)_{\min}$ are similar.

Definition 11. The probability of x_i is similar to x_j with respect to condition attribute set C, defined as follows:

$$P(x_i,x_j) = \frac{1}{|CN|}\sum_{a_k \in CN} P_{a_k}(x_i,x_j),$$

where $CN = \{a_k \in C | a_k(x_j) \neq *\}$.

The coincidence degree of x_i and x_j is higher with larger $P(x_i,x_j)$.

A completeness algorithm for an incomplete interval-valued fuzzy information system is proposed as follows.

Algorithm of CMAVIFIS

Input: incomplete interval-valued fuzzy decision table $S = (U, C \cup D)$

Output: complete interval-valued fuzzy decision table

(1) Compute the set of objects with missing attribute values MOS;
(2) For $\forall x_i \notin MOS$, let $a_k(x_i) = a_k(x_i)$, $a_k \in C$;
(3) For $\forall x_i \in MOS$,
  3.1 Compute $T_C(x_i)$;
    3.1.1 If $|T_C(x_i)| = 1$, let $x_j \in T_C(x_i)$, if $a_k(x_j) = *$, then $a_k(x_i) = *$; otherwise $a_k(x_i) = a_k(x_j)$;
    3.1.2 If $|T_C(x_i)| \neq 1$, exists singleton $x_{j_0} \in T_C(x_i)$ and satisfies the condition $(a_k(x_{j_0}) \neq *)$, then $a_k(x_i) = a_k(x_{j_0})$;
    3.1.3 Otherwise, for $\forall x_j \in T_C(x_i)$ and $x_j \neq x_i$, compute $P(x_i,x_j)$. Let $x_{j\max} = \{x_j | P(x_i, x_{j\max}) = \max P(x_i,x_j)\}$, then $a_k(x_i) = a_k(x_{j\max})$;
(4) If data are still missing in the decision table, the combination completeness method is adopted for further processing;
(5) The end.

The main idea of CMAVIFIS is to complete the missing values using those of the object with the maximal tolerance relation. In Step (1), the time complexity for computing MOS is $O(|C||U|)$. In Step (3), the time complexity for computing $T_C(x_i)$ is $O(|C||U|)$, and the total time complexity of Step (3) is $O(|C||U||MOS|)$. Thus, the time complexity of CMAVIFIS is $O(|C||U||MOS|)$. Time complexity depends on the distribution and quantity of missing data. Usually, missing data comprise only a small portion of total data; thus, time complexity is relatively low.

We present an example to illustrate the operation of CMAVIFIS. Table 1 is an incomplete interval-valued fuzzy decision table, where $U = \{x_1, x_2, \ldots, x_{10}\}$ is a set of objects and C is a fuzzy condition attribute set that includes three attributes, a_1, a_2, a_3, each with corresponding linguistic terms, e.g., a_1 has terms a_{11}, a_{12} and a_{13}, and a_{11}, a_{12}, a_{13} are interval-valued fuzzy sets. The decision attribute d is also fuzzy and is separated into three linguistic terms, F_1, F_2, F_3, and F_1, F_2, F_3 are interval-valued fuzzy sets. Missing data are represented by $*$. In Table 1, each object $x \in U$ corresponds to a column, while each attribute a corresponds to a row.

Table 1
An incomplete interval-valued fuzzy decision table.

U			x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
C	a_1	a_{11}	[0.2,0.3]	[0.1,0.1]	*	[0.2,0.2]	[0.8,0.9]	[0.5,0.6]	*	[0.7,0.9]	[0.2,0.3]	[0.1,0.2]
		a_{12}	[0.5,0.6]	[0.2,0.6]	*	[0.55,0.7]	[0.3,0.4]	[0.4,0.5]	*	[0.3,0.4]	[0.6,0.7]	[0.2,0.3]
		a_{13}	[0.2,0.4]	[0.5,0.7]	*	[0.3,0.4]	[0.1,0.2]	[0.2,0.3]	*	[0.05,0.1]	[0.3,0.4]	[0.6,0.8]
	a_2	a_{21}	[0.6,0.8]	[0.2,0.4]	[0.7,0.9]	[0.2,0.4]	*	[0.8,0.8]	[0.3,0.3]	[0.6,0.8]	[0.7,0.8]	*
		a_{22}	[0.3,0.4]	[0.5,0.7]	[0.3,0.4]	[0.8,0.9]	*	[0.1,0.3]	[0.7,0.8]	[0.3,0.4]	[0.4,0.4]	*
	a_3	a_{31}	[0.1,0.1]	[0.85,0.9]	[0.05,0.1]	[0.1,0.2]	[0.05,0.1]	[0.1,0.3]	[0.7,0.8]	[0.1,0.3]	*	[0.6,0.7]
		a_{32}	[0.2,0.3]	[0.2,0.3]	[0.1,0.2]	[0.1,0.25]	[0.15,0.3]	[0.6,0.7]	[0.3,0.4]	[0.2,0.3]	*	[0.2,0.3]
		a_{33}	[0.7,0.8]	[0.1,0.2]	[0.7,0.75]	[0.6,0.6]	[0.3,0.4]	[0.2,0.3]	[0.1,0.2]	[0.3,0.4]	*	[0.1,0.2]
		a_{34}	[0.3,0.4]	[0.1,0.2]	[0.2,0.3]	[0.3,0.35]	[0.5,0.8]	[0.1,0.2]	[0.1,0.1]	[0.6,0.9]	*	[0.1,0.2]
D	d	F_1	[0.7,0.9]	[0.3,0.5]	[0.7,0.8]	[0.15,0.2]	[0.05,0.1]	[0.1,0.2]	[0.25,0.4]	[0.1,0.2]	[0.45,0.7]	[0.05,0.1]
		F_2	[0.15,0.2]	[0.5,0.7]	[0.3,0.4]	[0.5,0.8]	[0.2,0.3]	[0.35,0.5]	[1.0,1.0]	[0.25,0.4]	[0.25,0.3]	[0.8,0.9]
		F_3	[0.4,0.5]	[0.35,0.4]	[0.1,0.2]	[0.2,0.3]	[0.65,0.9]	[1.0,1.0]	[0.3,0.4]	[0.5,0.6]	[0.2,0.3]	[0.05,0.2]

Table 2
A complete interval-valued fuzzy decision table.

U			x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
C	a_1	a_{11}	[0.2,0.3]	[0.1,0.1]	[0.2,0.3]	[0.2,0.2]	[0.8,0.9]	[0.5,0.6]	[0.1,0.1]	[0.7,0.9]	[0.2,0.3]	[0.1,0.2]
		a_{12}	[0.5,0.6]	[0.2,0.6]	[0.6,0.7]	[0.55,0.7]	[0.3,0.4]	[0.4,0.5]	[0.2,0.6]	[0.3,0.4]	[0.6,0.7]	[0.2,0.3]
		a_{13}	[0.2,0.4]	[0.5,0.7]	[0.3,0.4]	[0.3,0.4]	[0.1,0.2]	[0.2,0.3]	[0.5,0.7]	[0.05,0.1]	[0.3,0.4]	[0.6,0.8]
	a_2	a_{21}	[0.6,0.8]	[0.2,0.4]	[0.7,0.9]	[0.2,0.4]	[0.6,0.8]	[0.8,0.8]	[0.3,0.3]	[0.6,0.8]	[0.7,0.8]	[0.3,0.3]
		a_{22}	[0.3,0.4]	[0.5,0.7]	[0.3,0.4]	[0.8,0.9]	[0.3,0.4]	[0.1,0.3]	[0.7,0.8]	[0.3,0.4]	[0.4,0.4]	[0.7,0.8]
	a_3	a_{31}	[0.1,0.1]	[0.85,0.9]	[0.05,0.1]	[0.1,0.2]	[0.05,0.1]	[0.1,0.3]	[0.7,0.8]	[0.1,0.3]	[0.1,0.1]	[0.6,0.7]
		a_{32}	[0.2,0.3]	[0.2,0.3]	[0.1,0.2]	[0.1,0.25]	[0.15,0.3]	[0.6,0.7]	[0.3,0.4]	[0.2,0.3]	[0.2,0.3]	[0.2,0.3]
		a_{33}	[0.7,0.8]	[0.1,0.2]	[0.7,0.75]	[0.6,0.6]	[0.3,0.4]	[0.2,0.3]	[0.1,0.2]	[0.3,0.4]	[0.7,0.8]	[0.1,0.2]
		a_{34}	[0.3,0.4]	[0.1,0.2]	[0.2,0.3]	[0.3,0.35]	[0.5,0.8]	[0.1,0.2]	[0.1,0.1]	[0.6,0.9]	[0.3,0.4]	[0.1,0.2]
D	d	F_1	[0.7,0.9]	[0.3,0.5]	[0.7,0.8]	[0.15,0.2]	[0.05,0.1]	[0.1,0.2]	[0.25,0.4]	[0.1,0.2]	[0.45,0.7]	[0.05,0.1]
		F_2	[0.15,0.2]	[0.5,0.7]	[0.3,0.4]	[0.5,0.8]	[0.2,0.3]	[0.35,0.5]	[1.0,1.0]	[0.25,0.4]	[0.25,0.3]	[0.8,0.9]
		F_3	[0.4,0.5]	[0.35,0.4]	[0.1,0.2]	[0.2,0.3]	[0.65,0.9]	[1.0,1.0]	[0.3,0.4]	[0.5,0.6]	[0.2,0.3]	[0.05,0.2]

According to CMAVIFIS, the set of objects with missing values is $MOS = \{x_3, x_5, x_7, x_9, x_{10}\}$. Let $\xi = 0.7$.

For object x_3, we have $T_C(x_3) = \{x_1, x_9\}$. Thus, $|T_C(x_3)| \neq 1$. Two elements in $T_C(x_3)$ have values that are not $*$. We need to compute $P(x_3, x_1)$ and $P(x_3, x_9)$. According to Definitions 10 and 11, we obtain $P(x_3, x_1) \approx 0.51$ and $P(x_3, x_9) \approx 0.53$. In view of $P(x_3, x_1) \leqslant P(x_3, x_9)$, we have $a_{11}(x_3) = a_{11}(x_9) = [0.2, 0.3]$; $a_{12}(x_3) = a_{12}(x_9) = [0.6, 0.7]$; $a_{13}(x_3) = a_{13}(x_9) = [0.3, 0.4]$.

For $x_5 \in MOS$, $T_C(x_5) = \{x_8\}$. $|T_C(x_i)| = 1$, we have $a_{21}(x_5) = a_{21}(x_8) = [0.6, 0.8]$; $a_{22}(x_5) = a_{22}(x_8) = [0.3, 0.4]$.

Analogously, the other missing values are completed as follows:

$$a_{11}(x_7) = a_{11}(x_2) = [0.1, 0.1];\quad a_{12}(x_7) = a_{12}(x_2) = [0.2, 0.6];\quad a_{13}(x_7) = a_{13}(x_2) = [0.5, 0.7],$$
$$a_{31}(x_9) = a_{31}(x_1) = [0.1, 0.1];\quad a_{32}(x_9) = a_{32}(x_1) = [0.2, 0.3];\quad a_{33}(x_9) = a_{33}(x_1) = [0.7, 0.8];\quad a_{34}(x_9) = a_{34}(x_1) = [0.3, 0.4],$$
$$a_{21}(x_{10}) = a_{21}(x_7) = [0.3, 0.3];\quad a_{22}(x_{10}) = a_{22}(x_7) = [0.7, 0.8].$$

The complete interval-valued fuzzy decision table is shown in Table 2.

The example shows that CMAVIFIS can easily process completeness analysis of an incomplete information system; thus, it can be adopted as a pretreatment method in data mining. The application of CMAVIFIS is in Section 6.

4. Positive approximation in IVFR sets

From the viewpoint of granular computing in IVFR defined by Sun [38], the concept is described under static granulation, i.e., a certain interval-valued fuzzy equivalence relation. However, we usually need to analyze and solve problems from multiviews and multilevels. Consider an extreme case. Suppose we select an interval-valued fuzzy equivalence relation R with the finest granulation; i.e., each fuzzy block contains only one object. An interval-valued fuzzy set F can be effectively expressed in the granulation space determined by R. However, such expression is only a simple enumeration of F, and the commonness of the objects constituting F cannot be fully mined. The extreme case encourages us to find a model that not only can effectively express F but also fully mines potential rules.

We review two existing methods to construct granulation. First, the universe can be divided according to the entire attribute set and obtain a granulation space. However, some potential community characteristics among objects may be lost due to excessively fine granulation. Second, instead of using the entire attribute set, a reduct set can divide the universe and obtain coarser granulation space. However, attribute reduction algorithms may have a negative effect on inducting good rules. Furthermore, even rules are extracted based on the reduct set, once the reduct set is selected, the corresponding granulation is determined. Solving problems in a uniform granulation inevitably makes granulation too fine or too coarse. Excessively fine granulation will lose potential community characteristics among objects and increase time and cost. If the granulation is too coarse, the target concept cannot be effectively expressed. Thus, we consider a model with a non-uniform granulation level, i.e., analyze a given object with a family of interval-valued fuzzy equivalence relations rather than a single one.

An interval-valued fuzzy equivalence relation R corresponds to a granulation space. A partial order sequence $R_1 \prec R_2 \prec \cdots \prec R_n$ corresponds to a sequence of spaces with granulations from fine to coarse. The changing process of granulation includes two cases: gradual refining and gradual coarsening. The former is used when a description of the object at a finer granulation level is required. For a family of interval-valued fuzzy equivalence relations with $R_1 \succ R_2 \succ \cdots \succ R_n$, the process of describing the object using gradual refining granulation is called positive approximation. The latter is used when excessively fine description leads to loss of community characteristics among objects. Accordingly, the process of describing the object by using gradual coarsening granulation is called converse approximation.

In the process of the positive approximation, coarser granulation is first selected, and then objects whose decision class can be determined in the corresponding granulation space are deleted. Thus, only the rest needs to be considered, namely, objects whose decision class cannot be determined for the moment. In the finer granulation space with more detailed information, objects whose decision class can be determined are deleted. Objects that require further investigation in the universe are considered the next research objectives each time. A sequence of expressions with different granulation levels can then be generated. In family of interval-valued fuzzy equivalence relations, the positive approximation can effectively express the given interval-valued fuzzy set by the least knowledge granules.

4.1. The concept of positive approximation

Granulation order in sets was introduced by Qian and co-workers [20,30]. To obtain granulation order in interval-valued fuzzy information systems, we first extend the partial relation $\prec$ on 2^A. Let $S = (U,A)$ be an interval-valued fuzzy information system, where U is a non-empty set of finite objects (the universe), attributes in A are represented by fuzzy interval numbers, and I, Q are subsets of A. Define a partial relation $\prec$ as follows: $I \prec Q(Q \succ I)$ if and only if, for every $I_k \in U/I$, there exists $Q_j \in U/Q$ such that $\forall x \in U,\ \mu_{I_k^-}(x) \leqslant \mu_{Q_j^-}(x)$ and $\mu_{I_k^+}(x) \leqslant \mu_{Q_j^+}(x)$, where $U/I = \{I_1 = [I_1^-, I_1^+],\ I_2 = [I_2^-, I_2^+], \ldots, I_m = [I_m^-, I_m^+]\}$ and $U/Q = \{Q_1 = [Q_1^-, Q_1^+],\ Q_2 = [Q_2^-, Q_2^+], \ldots, Q_n = [Q_n^-, Q_n^+]\}$ are fuzzy blocks induced by I and Q, respectively.

A fuzzy partition induced by an interval-valued fuzzy equivalence relation provides granulation space to describe a target concept. Let R_k ($k = 1,2,\ldots,n$) be a family of interval-valued fuzzy equivalence relations with, $R_1 \succ R_2 \succ \cdots \succ R_n$. The sequence of granulation spaces from coarse to fine, as determined by R_k ($k = 1,2,\ldots,n$), is called positive granulation order. The upper and lower approximations of positive approximation in a positive granulation order are defined as follows.

Definition 12. Let $S = (U,A)$ be an interval-valued fuzzy information system, F be an interval-valued fuzzy set of U, and $P = \{R_1, R_2, \ldots, R_n\}$ be a family of interval-valued fuzzy attribute sets with $R_1 \succ R_2 \succ \cdots \succ R_n$, where R_k ($k = 1,2,\ldots,n$) is a subset of A. P-upper approximation $\overline{apr}_P(F)$ and P-lower approximation $\underline{apr}_P(F)$ of positive approximation of F are defined as follows:

$$\mu_{\overline{apr}_P(F)}(x) = [\mu_{\overline{apr}_P(F^-)}, \mu_{\overline{apr}_P(F^+)}] = \left[\sup_{y\in U} \min\{\mu_{[x]_{R_n^-}}(y), \mu_{F^-}(y)\}, \sup_{y\in U} \min\{\mu_{[x]_{R_n^+}}(y), \mu_{F^+}(y)\}\right],$$

$$\mu_{\underline{apr}_P(F)}(x) = [\mu_{\underline{apr}_P(F^-)}, \mu_{\underline{apr}_P(F^+)}] = \begin{cases} \left[\inf_{y\in U} \max\left\{1-\mu_{[x]_{R_1^+}}(y), \mu_{F^-}(y)\right\}, \inf_{y\in U} \max\left\{1-\mu_{[x]_{R_1^-}}(y), \mu_{F^+}(y)\right\}\right], x \in W_1, \\ \cdots \\ \left[\inf_{y\in U} \max\left\{1-\mu_{[x]_{R_n^+}}(y), \mu_{F^-}(y)\right\}, \inf_{y\in U} \max\left\{1-\mu_{[x]_{R_n^-}}(y), \mu_{F^+}(y)\right\}\right], x \in W_n, \\ \left[\inf_{y\in U} \max\left\{1-\mu_{[x]_{R_{n+1}^+}}(y), \mu_{F^-}(y)\right\}, \inf_{y\in U} \max\left\{1-\mu_{[x]_{R_{n+1}^-}}(y), \mu_{F^+}(y)\right\}\right], x \in U_{n+1}, \end{cases}$$

where $\mu_{[x]_{R_k}}(y) = \mu_{R_k}(x,y),\ (k = 1,2,\ldots,n), [x]_{R_k} \in U_k/R_k\ (k = 1,2,\ldots,n)$, especially $[x]_{R_{n+1}} \in U_{n+1}/R_n$ and $U_1 = U,\ U_i = U_{i-1} - W_{i-1}\ (i = 2,3,\ldots,n+1),\ W_{i-1} = \{x|\mu_{\underline{apr}_{R_{i-1}}(F)}(x) = [\inf_{y\in U} \max\{1-\mu_{[x]_{R_{i-1}^+}}(y), \mu_{F^-}(y)\}, \inf_{y\in U} \max\{1-\mu_{[x]_{R_{i-1}^-}}(y), \mu_{F^+}(y)\}] \geqslant [\eta^-, \eta^+]\} = \{x|\inf_{y\in U} \max\{1-\mu_{[x]_{R_{i-1}^+}}(y), \mu_{F^-}(y)\} \geqslant \eta^-, \inf_{y\in U} \max\{1-\mu_{[x]_{R_{i-1}^-}}(y), \mu_{F^+}(y)\} \geqslant \eta^+\}$ $\eta^-, \eta^+ \in [0.5,1]$ and $\eta = [\eta^-, \eta^+] \in [I]$ is a suitable threshold.

The boundary $BN_P(F)$ of F is defined as follows:

$$\mu_{BN_P(F)}(x) = \left[\left(\sup_{y\in U} \min\left\{\mu_{[x]_{R_n^-}}(y), \mu_{F^-}(y)\right\}\right) \wedge \left(1 - \inf_{y\in U} \max\left\{1-\mu_{[x]_{R_n^-}}(y), \mu_{F^+}(y)\right\}\right), \left(\sup_{y\in U} \min\left\{\mu_{[x]_{R_n^+}}(y), \mu_{F^+}(y)\right\}\right) \wedge \left(1 - \inf_{y\in U} \max\left\{1-\mu_{[x]_{R_n^+}}(y), \mu_{F^-}(y)\right\}\right)\right].$$

The differentiation index of F in (U,A) is defined as follows:

$$diff_{R_k}(F) = \frac{1}{2}\left(\frac{\sum_{x\in U}\mu_{F^-}(x)}{\sum_{x\in U}\mu_{F^-}(x) + \sum_{x\in U}\mu_{BN_{R_k}(F^-)}(x)} + \frac{\sum_{x\in U}\mu_{F^+}(x)}{\sum_{x\in U}\mu_{F^+}(x) + \sum_{x\in U}\mu_{BN_{R_k}(F^+)}(x)}\right).$$

Remark 1. The main idea of Definition 12 is that in the coarsest granulation space decided by R_1, objects whose decision class can be determined are deleted to obtain an updated universe, $U_1 - W_1$. In the coarser granulation space determined by adding a condition attribute, for the updated universe $U_1 - W_1$, objects whose decision class can be determined are deleted, and the universe is updated again. This process is repeated until the updated universe becomes an empty set or no condition attribute can be added. Although approximation operators are equivalent to the ones in Definition 5, the structure of the approximation operators reflects the granulation spaces changing from coarse to fine. Definition 12 shows that the universe dwindles as the granulation space becomes fine. This helps reduce computational complexity.

Remark 2. Upper and lower approximations are not symmetrical. In many applications, computing the upper approximation is not always necessary. For simplicity, the upper approximation operator is not represented in a structural form.

Remark 3. If $S = (U,A)$ is an information system with both crisp condition and interval-valued fuzzy decision attributes, and Definition 12 is usable. Here, $P = \{R_1, R_2, \ldots, R_n\}$ is a family of equivalence relations with $R_1 \succ R_2 \succ \cdots \succ R_n$, and $\mu_{[x]_{R_j^-}}(y) = \mu_{[x]_{R_j^+}}(y) = \mu_{[x]_{R_j}}(y) = \begin{cases} 1, & y \in [x]_{R_j} \\ 0, & y \notin [x]_{R_j} \end{cases}$, $j = 1, 2, \ldots, n+1$. Analogously, if $S = (U,A)$ is an information system with both interval-valued fuzzy condition and crisp decision attributes, Definition 12 is still effective. Here, F is a set of U, and

$$\mu_{F^-}(y) = \mu_{F^+}(y) = \mu_F(y) = \begin{cases} 1, & y \in F, \\ 0, & y \notin F. \end{cases}$$

Remark 4. The differentiation index provides a quantitative depiction of the object. Clearly, $0 \leqslant diff_{R_k}(F) \leqslant 1$. If $diff_{R_k}(F) = 1$, then we have $BN_{R_k}(F) = \varnothing$.

Definition 12 shows that a target fuzzy concept is approached by the upper approximation $\overline{apr}_P(F)$ and variable lower approximation $\underline{apr}_P(F)$. In the process of approximate classification, the approximate classification result and the decision class is usually incompatible due to the unavoidable roughness of the problem description. The closer the lower approximation $\underline{apr}_P(F)$ is to F, the higher the compatibility between the approximate classification result and the decision class. The result of the positive approximation is that the universe is decomposed into a union of several subsets, i.e., $U = W_1 \cup W_2 \cup \cdots W_n \cup U_{n+1}$. Each subset is located in different granulation levels; the maximal subset satisfies the given threshold in the corresponding granulation. The number of "$\cup$" quantitatively reflects the compatible extent between the approximate classification result and the decision class.

Theorem 1. *Let $S = (U,A)$ be an interval-valued fuzzy information system, F be an interval-valued fuzzy set of U, and $P = \{R_1, R_2, \ldots, R_n\}$ be a family of interval-valued fuzzy attribute sets with $R_1 \succ R_2 \succ \cdots \succ R_n$, where R_k $(k = 1, 2, \ldots, n)$ is a subset of A. Let $P_i = \{R_1, R_2, \ldots, R_i\}$. Then for $\forall P_i$, $(i = 1, 2, \ldots, n)$, the following properties hold:*

$$\underline{apr}_P(F) \subseteq F \subseteq \overline{apr}_P(F), \tag{1}$$

$$\underline{apr}_{P_1}(F) \subseteq \underline{apr}_{P_2}(F) \subseteq \cdots \subseteq \underline{apr}_{P_n}(F), \tag{2}$$

$$BN_{P_1}(F) \supseteq BN_{P_2}(F) \supseteq \cdots \supseteq BN_{P_n}(F), \tag{3}$$

$$diff_{P_1}(F) \leqslant diff_{P_2}(F) \leqslant \cdots \leqslant diff_{P_n}(F). \tag{4}$$

Proof. $\mu_{apr_P(F)}(x) = \mu_{apr_{R_n}(F)}(x)$. Then, $\mu_{\underline{apr}_P(F)}(x) = \mu_{\underline{apr}_{R_n}(F)}(x) \leqslant \mu_F(x) \leqslant \mu_{\overline{apr}_{R_n}(F)}(x) = \mu_{\overline{apr}_P(F)}(x)$, for $\forall x \in U$. That is, $\underline{apr}_P(F) \subseteq F \subseteq \overline{apr}_P(F)$.

To prove (2), we prove $\underline{apr}_{P_1}(F) \subseteq \underline{apr}_{P_2}(F)$ first.

$$\mu_{\underline{apr}_{P_1}(F)}(x) = \begin{cases} \left[\inf\limits_{y\in U} \max\left\{1 - \mu_{[x]_{R_1^+}}(y), \mu_{F^-}(y)\right\}, \inf\limits_{y\in U} \max\left\{1 - \mu_{[x]_{R_1^-}}(y), \mu_{F^+}(y)\right\}\right], & x \in W_1, \\ \left[\inf\limits_{y\in U} \max\left\{1 - \mu_{[x]_{R_{1+1}^+}}(y), \mu_{F^-}(y)\right\}, \inf\limits_{y\in U} \max\left\{1 - \mu_{[x]_{R_{1+1}^-}}(y)\}, \mu_{F^+}(y)\right\}\right], & x \in U_2 = U_1 - W_1, \end{cases}$$

$$\mu_{\underline{apr}_{P_2}(F)}(x)=\begin{cases}\left[\inf\limits_{y\in U}\max\left\{1-\mu_{[x]_{R_1^+}}(y),\mu_{F^-}(y)\right\},\inf\limits_{y\in U}\max\left\{1-\mu_{[x]_{R_1^-}}(y),\mu_{F^+}(y)\right\}\right], & x\in W_1,\\ \left[\inf\limits_{y\in U}\max\left\{1-\mu_{[x]_{R_2^+}}(y),\mu_{F^-}(y)\right\},\inf\limits_{y\in U}\max\left\{1-\mu_{[x]_{R_2^-}}(y),\mu_{F^+}(y)\right\}\right], & x\in W_2\subseteq U_2=U_1-W_1,\\ [\inf\limits_{y\in U}\max\{1-\mu_{[x]_{R_{2+1}^+}}(y),\mu_{F^-}(y)\},\inf\limits_{y\in U}\max\{1-\mu_{[x]_{R_{2+1}^-}}(y),\mu_{F^+}(y)\}], & x\in U_3=U_2-W_2=U_1-W_1-W_2.\end{cases}$$

When $x\in W_1$, $\mu_{\underline{apr}_{P_1}(F)}(x)=\mu_{\underline{apr}_{P_2}(F)}(x)$; otherwise, $\mu_{\underline{apr}_{P_1}(F)}(x)\leqslant\mu_{\underline{apr}_{P_2}(F)}(x)$. That is, for $\forall x\in U$, $\mu_{\underline{apr}_{P_1}(F)}(x)\leqslant\mu_{\underline{apr}_{P_2}(F)}(x)$. Thus, $\underline{apr}_{P_1}(F)\subseteq\underline{apr}_{P_2}(F)$. Similarly, we can show other inequalities. Therefore, $\underline{apr}_{P_1}(F)\subseteq\underline{apr}_{P_2}(F)\subseteq\cdots\subseteq\underline{apr}_{P_n}(F)$.

To prove (3), we prove $BN_{P_1}(F)\supseteq BN_{P_2}(F)$ first.

$$\mu_{BN_{P_1}(F)}(x)=\left[\left(\sup_{y\in U}\min\left\{\mu_{[x]_{P_1^-}}(y),\mu_{F^-}(y)\right\}\right)\wedge\left(1-\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_1^-}}(y),\mu_{F^+}(y)\right\}\right),\left(\sup_{y\in U}\min\left\{\mu_{[x]_{P_1^+}}(y),\mu_{F^+}(y)\right\}\right)\wedge\left(1-\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_1^+}}(y),\mu_{F^-}(y)\right\}\right)\right],$$

$$\mu_{BN_{P_2}(F)}(x)=\left[\left(\sup_{y\in U}\min\left\{\mu_{[x]_{P_2^-}}(y),\mu_{F^-}(y)\right\}\right)\wedge\left(1-\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_2^-}}(y),\mu_{F^+}(y)\right\}\right),\left(\sup_{y\in U}\min\left\{\mu_{[x]_{P_2^+}}(y),\mu_{F^+}(y)\right\}\right)\wedge\left(1-\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_2^+}}(y),\mu_{F^-}(y)\right\}\right)\right].$$

Clearly, $[x]_{P_1}\supseteq[x]_{P_2}$, then $\sup_{y\in U}\min\left\{\mu_{[x]_{P_1^-}}(y),\ \mu_{F^-}(y)\right\}\geqslant\sup_{y\in U}\min\left\{\mu_{[x]_{P_2^-}}(y),\ \mu_{F^-}(y)\right\}$, $\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_1^-}}(y),\ \mu_{F^+}(y)\right\}\leqslant\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_2^-}}(y),\ \mu_{F^+}(y)\right\}$. Therefore, $\left(\sup_{y\in U}\min\left\{\mu_{[x]_{P_1^-}}(y),\ \mu_{F^-}(y)\right\}\right)\wedge\left(1-\inf_{y\in U}\max\{1-\mu_{[x]_{P_1^-}}(y),\ \mu_{F^+}(y)\}\right)\geqslant\left(\sup_{y\in U}\min\left\{\mu_{[x]_{P_2^-}}(y),\ \mu_{F^-}(y)\right\}\right)\wedge\left(1-\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_2^-}}(y),\ \mu_{F^+}(y)\right\}\right)$. Similarly, we can show that $\left(\sup_{y\in U}\min\left\{\mu_{[x]_{P_1^+}}(y),\mu_{F^+}(y)\right\}\right)\wedge\left(1-\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_1^+}}(y),\mu_{F^-}(y)\right\}\right)\geqslant\left(\sup_{y\in U}\min\left\{\mu_{[x]_{P_2^+}}(y),\mu_{F^+}(y)\right\}\right)\wedge\left(1-\inf_{y\in U}\max\left\{1-\mu_{[x]_{P_2^+}}(y),\mu_{F^-}(y)\right\}\right)$. Thus, $\mu_{BN_{P_1}(F)}(x)\geqslant\mu_{BN_{P_2}(F)}(x)$, i.e., $BN_{P_1}(F)\supseteq BN_{P_2}(F)$. Analogously, the others can be proven. □

Based on $BN_{P_1}(F)\supseteq BN_{P_2}(F)\supseteq\cdots\supseteq BN_{P_n}(F)$, for $\forall x\in U$, we obtain $\mu_{BN_{P_1}(F^-)}(x)\geqslant\mu_{BN_{P_2}(F^-)}(x)\geqslant\cdots\geqslant\mu_{BN_{P_n}(F^-)}(x)$, then $\sum_{x\in U}\mu_{BN_{P_1}(F^-)}(x)\geqslant\sum_{x\in U}\mu_{BN_{P_2}(F^-)}(x)\geqslant\cdots\geqslant\sum_{x\in U}\mu_{BN_{P_n}(F^-)}(x)$; therefore, $\frac{\sum_{x\in U}\mu_{F^-}(x)}{\sum_{x\in U}\mu_{F^-}(x)+\sum_{x\in U}\mu_{BN_{P_1}(F^-)}(x)}\leqslant\frac{\sum_{x\in U}\mu_{F^-}(x)}{\sum_{x\in U}\mu_{F^-}(x)+\sum_{x\in U}\mu_{BN_{P_2}(F^-)}(x)}\leqslant\cdots\leqslant\frac{\sum_{x\in U}\mu_{F^-}(x)}{\sum_{x\in U}\mu_{F^-}(x)+\sum_{x\in U}\mu_{BN_{P_n}(F^-)}(x)}$. Analogously, $\frac{\sum_{x\in U}\mu_{F^+}(x)}{\sum_{x\in U}\mu_{F^+}(x)+\sum_{x\in U}\mu_{BN_{P_1}(F^+)}(x)}\leqslant\frac{\sum_{x\in U}\mu_{F^+}(x)}{\sum_{x\in U}\mu_{F^+}(x)+\sum_{x\in U}\mu_{BN_{P_2}(F^+)}(x)}\leqslant\cdots\leqslant\frac{\sum_{x\in U}\mu_{F^+}(x)}{\sum_{x\in U}\mu_{F^+}(x)+\sum_{x\in U}\mu_{BN_{P_n}(F^+)}(x)}$, i.e., $diff_{P_1}(F)\leqslant diff_{P_2}(F)\leqslant\cdots\leqslant diff_{P_n}(F)$.

Theorem 1 states that the lower approximation enlarges as the positive granulation order becomes longer by adding interval-valued fuzzy equivalence relations. At the same time, the boundary set dwindles and the differentiation index increases.

To describe the uncertainty of concepts in a positive granulation order, the approximation precision is defined as follows.

Definition 13. Let $S=(U,A)$ be an interval-valued fuzzy information system, F be an interval-valued fuzzy set of U, and $P=\{R_1,R_2,\ldots,R_n\}$ be a family of interval-valued fuzzy attribute sets with $R_1\succ R_2\succ\cdots\succ R_n$, where $R_k\ (k=1,2,\ldots,n)$ is a subset of A. The approximation precision $\alpha_P(F)$ is defined as follows:

$$\alpha_P(F)=\alpha_P(F^-)\wedge\alpha_P(F^+)=\frac{\sum_{x\in U}\mu_{\underline{apr}_P(F^-)}(x)}{\sum_{x\in U}\mu_{\overline{apr}_P(F^-)}(x)}\wedge\frac{\sum_{x\in U}\mu_{\underline{apr}_P(F^+)}(x)}{\sum_{x\in U}\mu_{\overline{apr}_P(F^+)}(x)}.$$

Theorem 2. *Let $S=(U,A)$ be an interval-valued fuzzy information system, F be an interval-valued fuzzy set of U, and $P=\{R_1,R_2,\ldots,R_n\}$ be a family of interval-valued fuzzy attribute sets with $R_1\succ R_2\succ\cdots\succ R_n$, where $R_k\ (k=1,2,\ldots,n)$ is a subset of A. Let $P_i=\{R_1,R_2,\ldots,R_i\}$, then for $\forall P_i$, $(i=1,2,\ldots,n)$, we have:*

$$\alpha_{P_1}(F)\leqslant\alpha_{P_2}(F)\leqslant\cdots\leqslant\alpha_{P_n}(F).$$

Proof. For $\forall x \in U, \mu_{\overline{apr}_{P_1}(F^-)}(x) = \sup_{y\in U}\min\left\{\mu_{[x]_{R_1^-}}(y), \mu_{F^-}(y)\right\} \geqslant \sup_{y\in U}\min\left\{\mu_{[x]_{R_2^-}}(y), \mu_{F^-}(y)\right\} = \mu_{\overline{apr}_{P_2}(F^-)}(x) \geqslant \cdots \geqslant \mu_{\overline{apr}_{P_n}(F^-)}(x)$. Moreover, Theorem 1 shows that $\mu_{\underline{apr}_{P_1}(F^-)}(x) \leqslant \mu_{\underline{apr}_{P_2}(F^-)}(x) \leqslant \cdots \leqslant \mu_{\underline{apr}_{P_n}(F^-)}(x)$; thus, according to Definition 13, we can easily obtain $\alpha_{P_1}(F^-) \leqslant \alpha_{P_2}(F^-) \leqslant \cdots \leqslant \alpha_{P_n}(F^-)$. Similarly, $\alpha_{P_1}(F^+) \leqslant \alpha_{P_2}(F^+) \leqslant \cdots \leqslant \alpha_{P_n}(F^+)$. Therefore, $\alpha_{P_1}(F) \leqslant \alpha_{P_2}(F) \leqslant \cdots \leqslant \alpha_{P_n}(F)$. □

Theorem 2 states that $\alpha_P(F)$ increases as the positive granulation order becomes longer.

4.2. Fuzzy rule extraction algorithm based on positive approximation

In the fuzzy rough set theory, rule extraction is usually performed under uniform granulation; thus, the dynamic property is deficient in the process of rule extraction. However, extract rules need to be dynamically extracted according to user requirements. In a decision table, granulation is mainly reflected as the hierarchy relation between condition and decision attribute set. The positive approximation approaches a target concept by the change in granulation, which can fully embody this hierarchy relation. Based on the positive approximation, a rule extraction algorithm called MRBPA is proposed.

Definition 14. Let $S = (U, C \cup D)$ be an interval-valued fuzzy decision table, where the condition attributes in C are represented by fuzzy interval numbers, and the decision attributes in D are fuzzy interval numbers or crisp numbers. The positive region of D with regard to C is defined as follows:

$$\mu_{pos_C(D)}(x) = \sup_{F\in U/D} \mu_{\underline{apr}_C(F)}(x) = \left[\sup_{F\in U/D} \mu_{\underline{apr}_C(F^-)}(x), \sup_{F\in U/D} \mu_{\underline{apr}_C(F^+)}(x)\right].$$

The dependency degree $\gamma_C(D)$ of C with regard to D is defined as follows:

$$\gamma_C(D) = \frac{\sum_{x\in U} \sup_{F\in U/D} \mu_{\underline{apr}_C(F^-)}(x) + \sum_{x\in U} \sup_{F\in U/D} \mu_{\underline{apr}_C(F^+)}(x)}{2|U|}.$$

Algorithm: MRBPA

Input: decision table with interval-valued fuzzy decision attributes $S = (U, C \cup D)$
Output: decision rules

(1) For $\forall c \in C$, compute the dependency degree $\gamma_{\{c\}}(D)$, let $\gamma_{\{c_1\}}(D) = \max\{\gamma_{\{c\}}(D) | c \in C\}$ and $P_1 = c_1$;
(2) $U/D = \{F_1, F_2, \ldots, F_d\}$, where F_k $(k = 1, 2, \ldots, d)$ is an interval-valued fuzzy set;
(3) Let $P = \{P_1\}$, $U_1 = U$, $Rule' = Rule = \emptyset$, $i = 1$;
(4) Let $W_i = \bigcup_{k=1}^{d}\{x | \mu_{\underline{apr}_P(F_k^-)}(x) = \inf_{y\in U}\max\left\{1 - \mu_{[x]_{P_i^+}}(y), \mu_{F_k^-}(y)\right\} \geqslant \eta^-,\ \mu_{\underline{apr}_P(F_k^+)}(x) = \inf_{y\in U}\max\left\{1 - \mu_{[x]_{P_i^-}}(y),\ \mu_{F_k^+}(y)\right\} \geqslant \eta^+, x \in U_i\}$. If $W_i \neq \emptyset$, then for$\forall x \in W_i$, put $des_P(x) \to des_{F_k}(x)$ $(k = 1, 2, \ldots, d)$ into $Rule'$. Let $Rule = Rule \cup Rule'$ and $U_{i+1} = U_i - W_i$;
(5) If $C - P = \emptyset$ and $U_{i+1} \neq \emptyset$, then for $\forall x \in U_{i+1}$, let $T = \left\{x | \mu_{\underline{apr}_P(F)}(x) \geqslant_w \eta\right\} = \bigcup_{k=1}^{d}\left\{x | \mu_{\underline{apr}_P(F_k^-)}(x) + \mu_{\underline{apr}_P(F_k^+)}(x) \geqslant \eta^- + \eta^+,\ \mu_{\underline{apr}_P(F_k^+)}(x) - \mu_{\underline{apr}_P(F_k^-)}(x) \geqslant \eta^+ - \eta^-\right\}$. For $\forall x \in T$, put $des_P(x) \to des_{F_k}(x)$ $(k = 1, 2, \ldots, d)$ into $Rule$, go to (8);
(6) If $U_{i+1} = \emptyset$, go to (8);
(7) For $\forall c \in C - P$, compute $\gamma_{P\cup\{c\}}(D)$, let $\gamma_{P\cup\{c_2\}}(D) = \max\{\gamma_{P\cup\{c\}}(D) | c \in C - P\}$. Let $P_{i+1} = P_i \cup \{c_2\}$, $P = P \cup P_{i+1}$, $i = i + 1$, go to (4);
(8) Output $Rule$.

Remark 5. In Step (4), $\eta = [\eta^-, \eta^+]$ is a threshold, and η^-, $\eta^+ \in [0.5, 1]$. Generally, more conditions must be satisfied in the rules, and the applicability of the rules decreases with increasing η. That is, η determines the granulation of the rules to some extent. The selection of η is determined by the actual requirement provided by the user.

Remark 6. In Step (4), $des_P(x)$ is the antecedent of the rule, and $des_{F_k}(x)$ is the consequent. For example, "If a_3 is a_{31}, then d is F_2", where "a_3 is a_{31}" is $des_P(x)$, "d is F_2" is $des_{F_k}(x)$.

According to MRBPA, a family of rules can be extracted with granulation changing from coarse to fine. The dynamic classification results can approximate the decision classification closely. MRBPA not only fully considers the potential community characteristics among objects but also possesses high efficiency. The time complexity to extract rules is polynomial.

In Step (1), for $\forall c \in C$, the time complexity for computing $\gamma_c(D)$ is $O(|C||U|^2)$.
In Step (2), the time complexity for computing U/D is $O(|U|^2)$.

In Step (4), the time complexity for computing W_i is $O(|P_i||U_i|^2)$.
In Step (5), the time complexity for computing T is $O(|C||U_{i+1}|^2)$.

In Step (7), the time complexity for computing $\gamma_{P\cup\{c\}}(D)$ is $O(|C-P_i|(|P_i|+1)|U_{i+1}|^2)$. From Steps (4) to (7), $|C|$ is the maximum value of the circle times. Therefore, the time complexity is:

$$\sum_{i=1}^{|C|}(O(|P_i||U_i|^2)+O(|C||U_{i+1}|^2)+O(|C-P_i|(|P_i|+1)|U_{i+1}|^2)) \qquad (*)$$

Evidently, $|P_i|\leqslant|C|$, $|U_i|\leqslant|U|$, $|U_{i+1}|<|U|$; thus, the time complexity of $(*)$ is smaller than $O(|C|^3|U|^2)$. Other steps will not be considered because their time complexities are constant. Hence, the time complexity of the algorithm MRBPA is:

$$O(|C||U|^2)+O(|U|^2)+\sum_{i=1}^{|C|}(O(|P_i||U_i|^2)+O(|C||U_{i+1}|^2)+O(|C-P_i|(|P_i|+1)|U_{i+1}|^2))\leqslant O(|C|^3|U|^2).$$

Generally, the time complexity of the rule extraction algorithm based on attribute reduction (RIA) is $O(|C|^3|U|^2)$; MRBPA is much smaller because the universe dwindles gradually.

Remark 7. The main differences between MRBPA and the method based on attribute reduction include two aspects. First, rule extraction is based on a granulation order; thus, the adverse effects of attribute reduction are excluded as much as possible. Second, the time complexity of the model is effectively reduced because of the dwindling universe.

4.3. An example

An example illustrates the MRBPA operation. The decision table is in Table 2, where $U=\{x_1,x_2,\ldots,x_{10}\}$ is a set of objects; and C is a fuzzy condition attribute set that includes three attributes a_1, a_2, a_3, each with corresponding linguistic terms, for example, a_1 has terms a_{11}, a_{12} and, a_{13}; and a_{11}, a_{12}, a_{13} are interval-valued fuzzy sets. The decision attribute d is fuzzy and is separated into three linguistic terms, F_1, F_2, F_3, and F_1, F_2, F_3 are interval-valued fuzzy sets.

According to MRBPA, the dependency degrees of a_1, a_2, a_3 can be computed with regard to d. We obtain $\gamma_{\{a_1\}}(d)=\frac{99}{200}$, $\gamma_{\{a_2\}}(d)=\frac{74}{200}$, $\gamma_{\{a_3\}}(d)=\frac{103}{200}$.

Hence, $P_1=\{a_3\}$, $P=\{P_1\}$ and $U_1/a_3=\{a_{31},a_{32},a_{33},a_{34}\}$. For $x\in a_{31}$, we have $\mu_{\underline{apr}_P(F_1^-)}(x)=0.25$, $\mu_{\underline{apr}_P(F_1^+)}(x)=0.4$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.5$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.7$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.3$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.4$; for $x\in a_{32}$, we have $\mu_{\underline{apr}_P(F_1^-)}(x)=0.3$, $\mu_{\underline{apr}_P(F_1^+)}(x)=0.4$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.35$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.5$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.6$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.7$; for $x\in a_{33}$, we obtain $\mu_{\underline{apr}_P(F_1^-)}(x)=0.4$, $\mu_{\underline{apr}_P(F_1^+)}(x)=0.4$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.2$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.3$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.2$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.3$; for $x\in a_{34}$, we obtain $\mu_{\underline{apr}_P(F_1^-)}(x)=0.1$, $\mu_{\underline{apr}_P(F_1^+)}(x)=0.4$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.2$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.4$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.5$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.6$.

Let $\eta=[\eta^-,\eta^+]=[0.5,0.6]$. For $x\in a_{31}$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.5\geqslant\eta^-$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.7\geqslant\eta^+$; $x\in a_{32}$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.6\geqslant\eta^-$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.7\geqslant\eta^+$; $x\in a_{34}$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.5\geqslant\eta^-$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.6\geqslant\eta^+$, then we have $W_1=\{x_2,x_5,x_6,x_7,x_8,x_{10}\}$, *Rule* = {$r_1$: If a_3 is a_{31} then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x)\geqslant[0.5,0.7]$; r_2: If a_3 is a_{32} then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x)\geqslant[0.6,0.7]$; r_3 : If a_3 is a_{34}, then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x)\geqslant[0.5,0.6]$} and $U_2=U_1-W_1=\{x_1,x_3,x_4,x_9\}$.

For $C-P\neq\emptyset$ and $U_2\neq\emptyset$, continue to compute the dependency degrees of the rest of the attributes a_1, a_2 with respect to d. The computation of the dependency degree is based on the updated U_2. We obtain: $\gamma_{\{a_1,a_3\}}(d)=\frac{340}{800}$, $\gamma_{\{a_2,a_3\}}(d)=\frac{455}{800}$. Choose a_2 as c_2, then $P_2=\{a_2,a_3\}$, $P=\{a_2,a_3\}$ and $U_2/P=\{a_{21}\cap a_{33}, a_{22}\cap a_{33}\}$. For $x\in a_{21}\cap a_{33}$, we have $\mu_{\underline{apr}_P(F_1^-)}(x)=0.45$, $\mu_{\underline{apr}_P(F_1^+)}(x)=0.7$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.2$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.3$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.2$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.3$; for $x\in a_{22}\cap a_{33}$, we have $\mu_{\underline{apr}_P(F_1^-)}(x)=0.4$, $\mu_{\underline{apr}_P(F_1^+)}(x)=0.4$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.5$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.6$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.4$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.4$. For $x\in a_{22}\cap a_{33}$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.5\geqslant\eta^-$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.6\geqslant\eta^+$, then $W_2=\{x_4\}$ and *Rule* = {r_1: If a_3 is a_{31}, then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x)\geqslant[0.5,0.7]$. r_2: If a_3 is a_{32}, then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x)\geqslant[0.6,0.7]$. r_3: If a_3 is a_{34}, then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x)\geqslant[0.5,0.6]$; r_4: If a_3 is a_{33} and a_2 is a_{22}, then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x)\geqslant[0.5,0.6]$} and $U_3=U_2-W_2=\{x_1,x_3,x_9\}$.

For $C-P\neq\emptyset$ and $U_3\neq\emptyset$, the last attribute a_1 is added to P, i.e., $P=\{a_1,a_2,a_3\}$. Then, $U_3/P=\{a_{12}\cap a_{21}\cap a_{33}\}$.

When $x\in a_{12}\cap a_{21}\cap a_{33}$, we obtain $\mu_{\underline{apr}_P(F_1^-)}(x)=0.45$, $\mu_{\underline{apr}_P(F_1^+)}(x)=0.7$, $\mu_{\underline{apr}_P(F_2^-)}(x)=0.3$, $\mu_{\underline{apr}_P(F_2^+)}(x)=0.4$, $\mu_{\underline{apr}_P(F_3^-)}(x)=0.3$, $\mu_{\underline{apr}_P(F_3^+)}(x)=0.4$.

Clearly, $W_3=\emptyset$ and $U_4=U_3$. $C-P=\emptyset$ and $U_4\neq\emptyset$; then for $x\in U_4=\{x_1,x_3,x_9\}$, we obtain $\mu_{\underline{apr}_P(F_1^-)}(x)+\mu_{\underline{apr}_P(F_1^+)}(x)=0.45+0.7\geqslant\eta^-+\eta^+=0.5+0.6$, $\mu_{\underline{apr}_P(F_1^+)}(x)-\mu_{\underline{apr}_P(F_1^-)}(x)=0.7-0.45\geqslant\eta^+-\eta^-=0.6-0.5$, i.e., $\mu_{\underline{apr}_P(F_1)}(x)\geqslant_w\eta$. Thus, $T=\{x_1,x_3,x_9\}$ and r_5: If a_3 is a_{33}, a_2 is a_{21} and a_1 is a_{12}, then d is F_1 and $\mu_{\underline{apr}_P(F_1)}(x)\geqslant[0.45,0.7]$ is added to *Rule*. The algorithm is stopped, and the rules are obtained as follows:

Rule = {r_1: If a_3 is a_{31} Then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x) \geqslant [0.5, 0.7]$;
r_2: If a_3 is a_{32} Then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x) \geqslant [0.6, 0.7]$;
r_3: If a_3 is a_{34} Then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x) \geqslant [0.5, 0.6]$;
r_4: If a_3 is a_{33} and a_2 is a_{22} Then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x) \geqslant [0.5, 0.6]$;
r_5: If a_3 is a_{33}, a_2 is a_{21} and a_1 is a_{12} Then d is F_1 and $\mu_{\underline{apr}_P(F_1)}(x) \geqslant [0.45, 0.7]$}.

5. Converse approximation in IVFR sets

The positive approximation approaches a target concept by the change in granulation. Due to the positive approximation, the approximation precision $\alpha_P(F)$ increases as the positive granulation order becomes longer, and a family of fuzzy rules with granulation changing from coarse to fine can be obtained. However, in some applications, the approximation precision is restricted by the decision requirements or preference of decision makers [30]. An obvious problem is extracting simpler rules based on keeping the approximation precision invariant. The positive approximation appears unsuitable for this purpose. Therefore, the converse approximation in IVFR sets is proposed.

In the process of the converse approximation, the objects that require further investigation in the universe are considered the next research objectives. A sequence of expressions with different granulation levels can then be generated. In the family of interval-valued fuzzy equivalence relations, the converse approximation can not only effectively reduce the knowledge granules describing the interval-valued fuzzy set, it also fully mines potential community characteristics among objects based on keeping the approximation precision invariant.

5.1. The concept of converse approximation

Let R_k ($k = 1,2,\ldots,n$) be a family of interval-valued fuzzy equivalence relations with $R_1 \prec R_2 \prec \cdots \prec R_n$. The sequence of granulation spaces from fine to coarse determined by R_k ($k = 1,2,\ldots,n$) is called converse granulation order. The upper and lower approximations of converse approximation are defined as follows.

Definition 15. Let $S = (U,A)$ be an interval-valued fuzzy information system, F be an interval-valued fuzzy set of U, and $P = \{R_1,R_2,\ldots,R_n\}$ be a family of interval-valued fuzzy attribute sets with$R_1 \prec R_2 \prec \cdots \prec R_n$, where R_k ($k = 1,2,\ldots,n$) is a subset of A. Let $P_i = \{R_1,R_2,\ldots,R_i\}$. P_i-upper approximation $\overline{apr}_{P_i}(F)$ and P_i-lower approximation $\underline{apr}_{P_i}(F)$ of P_i-converse approximation of F are defined as follows:

$$\mu_{\overline{apr}_{P_i}(F)}(x) = [\mu_{\overline{apr}_{P_i}(F^-)}(x), \mu_{\overline{apr}_{P_i}(F^+)}(x)] = \left[\sup_{y\in U} \min\{\mu_{[x]_{R_1^-}}(y), \mu_{F^-}(y)\}, \sup_{y\in U} \min\{\mu_{[x]_{R_1^+}}(y), \mu_{F^+}(y)\}\right],$$

$$\mu_{\underline{apr}_{P_i}(F)}(x) = [\mu_{\underline{apr}_{P_i}(F^-)}(x), \mu_{\underline{apr}_{P_i}(F^+)}(x)] = \begin{cases} \left[\bigvee_{h=1}^{j} \inf_{y\in U} \max\{1-\mu_{[x]_{R_h^+}}(y), \mu_{F^-}(y)\}, \bigvee_{h=1}^{j} \inf_{y\in U} \max\{1-\mu_{[x]_{R_h^-}}(y), \mu_{F^+}(y)\}\right], & \text{if } \exists j, \\ \left[\inf_{y\in U} \max\{1-\mu_{[x]_{R_1^+}}(y), \mu_{F^-}(y)\}, \inf_{y\in U} \max\{1-\mu_{[x]_{R_1^-}}(y), \mu_{F^+}(y)\}\right], & \text{otherwise}, \end{cases}$$

where $j = \max\left\{t | \mu_{\underline{apr}_{R_t}(F^-)}(x) = \inf_{y\in U} \max\left\{1-\mu_{[x]_{R_t^+}}(y), \mu_{F^-}(y)\right\} \geqslant \zeta^-, \mu_{\underline{apr}_{R_t}(F^+)}(x) = \inf_{y\in U} \max\left\{1-\mu_{[x]_{R_t^-}}(y),\ \mu_{F^+}(y)\right\} \geqslant \zeta^+, 1 \leqslant t \leqslant i\right\}$, $\mu_{[x]_{R_h}}(y) = \mu_{R_h}(x,y), \zeta^-, \zeta^+ \in [0.5, 1]$ and $\zeta = [\zeta^-,\zeta^+] \in [I]$ is a suitable threshold.

The boundary $BN_P(F)$ of F is defined as follows:

$$\mu_{BN_{P_i}(F)}(x) = \left[\left(\sup_{y\in U} \min\left\{\mu_{[x]_{(\cup_{R_i\in P_i} R_i)^-}}(y), \mu_{F^-}(y)\right\}\right) \wedge \left(1 - \inf_{y\in U} \max\{1-\mu_{[x]_{(\cup_{R_i\in P_i} R_i)^-}}(y), \mu_{F^+}(y)\}\right),\right.$$
$$\left.\left(\sup_{y\in U} \min\{\mu_{[x]_{(\cup_{R_i\in P_i} R_i)^+}}(y), \mu_{F^+}(y)\}\right) \wedge \left(1 - \inf_{y\in U} \max\left\{1-\mu_{[x]_{(\cup_{R_i\in P_i} R_i)^+}}(y), \mu_{F^-}(y)\right\}\right)\right].$$

The differentiation index of F in (U,A) is defined as follows:

$$diff_{P_i}(F) = \frac{1}{2}\left(\frac{\sum_{x\in U}\mu_{F^-}(x)}{\sum_{x\in U}\mu_{F^-}(x) + \sum_{x\in U}\mu_{BN_{P_i}(F^-)}(x)} + \frac{\sum_{x\in U}\mu_{F^+}(x)}{\sum_{x\in U}\mu_{F^+}(x) + \sum_{x\in U}\mu_{BN_{P_i}(F^+)}(x)}\right).$$

Remark 8. The main idea of the converse approximation is that the number of fuzzy blocks used to describe the target concept is reduced as the converse granulation order becomes longer. That is, new fuzzy blocks under different granulations are induced by combining known fuzzy blocks. Although the approximation operators are equivalent to the ones in Definition 5, the converse approximation emphasizes the change in the construction of the target concept. The structure of the approximation operators reflects the granulation spaces changing from fine to coarse.

Remark 9. The upper and lower approximations are not symmetrical. In many applications, computing the upper approximation is not always necessary. For simplicity, the upper approximation operator does not denote its structural form.

Remark 10. Definition 15 is still usable for information systems with both crisp condition and interval-valued fuzzy decision attributes, or for those with both interval-valued fuzzy condition and crisp decision attributes.

Remark 11. $\zeta = [\zeta^-, \zeta^+]$ is a threshold, and $\zeta^-, \zeta^+ \in [0.5, 1]$. Generally, more conditions must be satisfied in the rules, and the applicability of the rules decreases with increasing ζ. That is, ζ determines the granulation of the fuzzy rules to some extent. The selection of ζ is determined by the actual requirement provided by the user.

Remark 12. The differentiation index provides a quantitative depiction of the object. Clearly, $0 \leqslant diff_{R_k}(F) \leqslant 1$. If $diff_{R_k}(F) = 1$, then we have $BN_{R_k}(F) = \varnothing$.

Definition 15 shows that a target concept can be approached by the upper approximation $\overline{apr}_{P_i}(F)$ and the variable lower approximation $\underline{apr}_{P_i}(F)$. In particular, when $i = n$, we denote $\overline{apr}_{P_n}(F)$ as $\overline{apr}_P(F)$ and $\underline{apr}_{P_n}(F)$ as $\underline{apr}_P(F)$. $\overline{apr}_P(F)$ and $\underline{apr}_P(F)$ are called P-upper approximation and P-lower approximation of P-converse approximation of F, respectively.

Theorem 3. *Let $S = (U, A)$ be an interval-valued fuzzy information system, F be an interval-valued fuzzy set of U, and $P = \{R_1, R_2, \ldots, R_n\}$ be a family of interval-valued fuzzy attribute sets with $R_1 \prec R_2 \prec \cdots \prec R_n$, where R_k $(k = 1, 2, \ldots, n)$ is a subset of A. Let $P_i = \{R_1, R_2, \ldots, R_i\}$, then for $\forall P_i$, $(i = 1, 2, \ldots, n)$, the following properties hold:*

$$\underline{apr}_{P_1}(F) = \underline{apr}_{P_2}(F) = \cdots = \underline{apr}_{P_n}(F), \tag{5}$$

$$\underline{apr}_{P_i}(F) \subseteq F \subseteq \overline{apr}_{P_i}(F), \tag{6}$$

$$BN_{P_1}(F) = BN_{P_2}(F) = \cdots = BN_{P_n}(F), \tag{7}$$

$$diff_{P_1}(F) = diff_{P_2}(F) = \cdots = diff_{P_n}(F). \tag{8}$$

Proof. For $\forall P_i$, if j exists, then $\mu_{\underline{apr}_{P_i}(F)}(x) = [\overset{j}{\underset{h=1}{\vee}} \inf_{y\in U} \max\{1 - \mu_{[x]_{R_h^+}}(y), \mu_{F^-}(y)\}, \overset{j}{\underset{h=1}{\vee}} \inf_{y\in U} \max\{1 - \mu_{[x]_{R_h^-}}(y), \mu_{F^+}(y)\}] = [\inf_{y\in U} \max\{1 - \mu_{[x]_{R_1^+}}(y), \mu_{F^-}(y)\}, \inf_{y\in U} \max\{1 - \mu_{[x]_{R_1^-}}(y), \mu_{F^+}(y)\}] = \mu_{\underline{apr}_{R_1}(F)}(x)$; if j does not exist, then $\mu_{\underline{apr}_{P_i}(F)}(x) = [\inf_{y\in U} \max\{1 - \mu_{[x]_{R_1^+}}(y), \mu_{F^-}(y)\}, \inf_{y\in U} \max\{1 - \mu_{[x]_{R_1^-}}(y), \mu_{F^+}(y)\}] = \mu_{\underline{apr}_{R_1}(F)}(x)$. Therefore, $\underline{apr}_{P_1}(F) = \underline{apr}_{P_2}(F) = \cdots = \underline{apr}_{P_n}(F) = \underline{apr}_{R_1}(F)$ can be obtained, i.e., 5(5).

Moreover, $\mu_{\underline{apr}_{P_i}(F)}(x) = \mu_{\underline{apr}_{R_1}(F)}(x) \leqslant \mu_F(x) \leqslant \mu_{\overline{apr}_{R_1}(F)}(x) = \mu_{\overline{apr}_{P_i}(F)}(x)$. Thus, $\underline{apr}_{P_i}(F) \subseteq F \subseteq \overline{apr}_{P_i}(F)$.

For $\forall i$, $\mu_{BN_{P_i}(F)}(x) = \left[\left(\sup_{y\in U} \min\left\{\mu_{[x]_{(\cup_{R_i\in P_i} R_i)^-}}(y), \mu_{F^-}(y)\right\}\right) \wedge \left(1 - \inf_{y\in U} \max\left\{1 - \mu_{[x]_{(\cup_{R_i\in P_i} R_i)^-}}(y), \mu_{F^+}(y)\right\}\right),\right.$ $\left(\sup_{y\in U} \min\left\{\mu_{[x]_{(\cup_{R_i\in P_i} R_i)^+}}(y), \mu_{F^+}(y)\right\}\right) \wedge \left(1 - \inf_{y\in U} \max\left\{1 - \mu_{[x]_{(\cup_{R_i\in P_i} R_i)^+}}(y), \mu_{F^-}(y)\right\}\right)] = \left[\left(\sup_{y\in U} \min\left\{\mu_{[x]_{R_1^-}}(y), \mu_{F^-}(y)\right\}\right)\right.$ $\wedge\left(1 - \inf_{y\in U} \max\left\{1 - \mu_{[x]_{R_1^-}}(y), \mu_{F^+}(y)\right\}\right), \left(\sup_{y\in U} \min\left\{\mu_{[x]_{R_1^+}}(y), \mu_{F^+}(y)\right\}\right) \wedge \left(1 - \inf_{y\in U} \max\left\{1 - \mu_{[x]_{R_1^+}}(y), \mu_{F^-}(y)\right\}\right)] = \mu_{BN_{R_1}(F)}(x)$. Then (7) is proven.

Based on $BN_{P_1}(F) = BN_{P_2}(F) = \cdots = BN_{P_n}(F)$, for $\forall x \in U$, we obtain $\mu_{BN_{P_1}(F^-)}(x) = \mu_{BN_{P_2}(F^-)}(x) = \cdots = \mu_{BN_{P_n}(F^-)}(x)$, then $\sum_{x\in U} \mu_{BN_{P_1}(F^-)}(x) = \sum_{x\in U} \mu_{BN_{P_2}(F^-)}(x) = \cdots = \sum_{x\in U} \mu_{BN_{P_n}(F^-)}(x)$. Therefore, $\frac{\sum_{x\in U} \mu_{F^-}(x)}{\sum_{x\in U} \mu_{F^-}(x) + \sum_{x\in U} \mu_{BN_{P_1}(F^-)}(x)} = \frac{\sum_{x\in U} \mu_{F^-}(x)}{\sum_{x\in U} \mu_{F^-}(x) + \sum_{x\in U} \mu_{BN_{P_2}(F^-)}(x)} = \cdots = \frac{\sum_{x\in U} \mu_{F^-}(x)}{\sum_{x\in U} \mu_{F^-}(x) + \sum_{x\in U} \mu_{BN_{P_n}(F^-)}(x)}$. Analogously, $\frac{\sum_{x\in U} \mu_{F^+}(x)}{\sum_{x\in U} \mu_{F^+}(x) + \sum_{x\in U} \mu_{BN_{P_1}(F^+)}(x)} = \frac{\sum_{x\in U} \mu_{F^+}(x)}{\sum_{x\in U} \mu_{F^+}(x) + \sum_{x\in U} \mu_{BN_{P_2}(F^+)}(x)} = \cdots = \frac{\sum_{x\in U} \mu_{F^+}(x)}{\sum_{x\in U} \mu_{F^+}(x) + \sum_{x\in U} \mu_{BN_{P_n}(F^+)}(x)}$, i.e., $diff_{P_1}(F) = diff_{P_2}(F) = \cdots = diff_{P_n}(F)$.

This completes the proof. □

Theorem 3 states that the lower and upper approximations of P-converse approximation, the boundary set, and the differentiation index remain invariant as the converse granulation order becomes longer.

Theorem 4. *Let $S = (U,A)$ be an interval-valued fuzzy information system, F be an interval-valued fuzzy set of U, and $P = \{R_1, R_2, \ldots, R_n\}$ be a family of interval-valued fuzzy attribute sets with $R_1 \prec R_2 \prec \cdots \prec R_n$, where R_k $(k = 1,2,\ldots,n)$ is a subset of A. Let $P_i = \{R_1, R_2, \ldots, R_i\}$, then for $\forall P_i$, $(i = 1,2,\ldots,n)$, the following properties hold:*

$$\alpha_{P_1}(F) = \alpha_{P_2}(F) = \cdots = \alpha_{P_n}(F),$$

where $\alpha_{P_i}(F) = \alpha_{P_i}(F^-) \wedge \alpha_{P_i}(F^+) = \frac{\sum_{x\in U}\mu_{\underline{apr}_{P_i}(F^-)}(x)}{\sum_{x\in U}\mu_{\overline{apr}_{P_i}(F^-)}(x)} \wedge \frac{\sum_{x\in U}\mu_{\underline{apr}_{P_i}(F^+)}(x)}{\sum_{x\in U}\mu_{\overline{apr}_{P_i}(F^+)}(x)}$ *is approximation precision.*

Proof. It follows from Definition 15 that $\overline{apr}_{P_1}(F) = \overline{apr}_{P_2}(F) = \cdots = \overline{apr}_{P_n}(F) = \overline{apr}_{R_1}(F)$. Then for $\forall x \in U$, $\mu_{\overline{apr}_{P_1}(F^-)}(x) = \mu_{\overline{apr}_{P_2}(F^-)}(x) = \cdots = \mu_{\overline{apr}_{P_n}(F^-)}(x)$, $\mu_{\overline{apr}_{P_1}(F^+)}(x) = \mu_{\overline{apr}_{P_2}(F^+)}(x) = \cdots = \mu_{\overline{apr}_{P_n}(F^+)}(x)$. From Theorem 3, we obtain $\underline{apr}_{P_1}(F) = \underline{apr}_{P_2}(F) = \cdots = \underline{apr}_{P_n}(F)$, then for $\forall x \in U$, $\mu_{\underline{apr}_{P_1}(F^-)}(x) = \mu_{\underline{apr}_{P_2}(F^-)}(x) = \cdots = \mu_{\underline{apr}_{P_n}(F^-)}(x)$, $\mu_{\underline{apr}_{P_1}(F^+)}(x) = \mu_{\underline{apr}_{P_2}(F^+)}(x) = \cdots = \mu_{\underline{apr}_{P_n}(F^+)}(x)$. Therefore, $\alpha_{P_1}(F) = \alpha_{P_2}(F) = \cdots = \alpha_{P_n}(F)$.

This completes the proof. □

Definition 16. Let $S = (U,A)$ be an interval-valued fuzzy information system, R be a subset of A, and any attribute in R be an interval-valued fuzzy equivalence relation, $\Gamma = \{F_1, F_2, \ldots, F_m\}$ be a fuzzy partition of U and F_k $(k = 1,2,\ldots,m)$ be an interval-valued fuzzy set. Lower and upper approximations of Γ with respect to R are defined as follows:

$$\underline{apr}_R\Gamma = \{\underline{apr}_R(F_1), \underline{apr}_R(F_2), \cdots, \underline{apr}_R(F_m)\},$$
$$\overline{apr}_R\Gamma = \{\overline{apr}_R(F_1), \overline{apr}_R(F_2), \cdots, \overline{apr}_R(F_m)\}.$$

The target concept is described by fuzzy blocks. For the given universe U, the fuzzy blocks are determined by R. Thus, the lower and upper approximations of Γ have a close relationship with R. We need to define a new measure to evaluate the convergence of Γ with respect to R, which is helpful in understanding the construction of the lower approximation.

Definition 17. Let $S = (U,A)$ be an interval-valued fuzzy information system, R be a subset of A, any attribute in R be an interval-valued fuzzy equivalence relation, and $\Gamma = \{F_1, F_2, \ldots, F_m\}$ be a fuzzy partition of U, where F_k $(k = 1,2,\ldots,m)$ is an interval-valued fuzzy set. The convergence degree of Γ with respect to R is defined as follows:

$$C(R,\Gamma) = C(R,\Gamma^-) \wedge C(R,\Gamma^+) = \left(\sum_{k=1}^{m} \frac{|F_k^-|}{|U|} \sum_{j=1}^{s_k} p^2\left(F_k^{-j}\right)\right) \wedge \left(\sum_{k=1}^{m} \frac{|F_k^+|}{|U|} \sum_{j=1}^{s_k} p^2\left(F_k^{+j}\right)\right),$$

where $|F_k^-| = \sum_{x\in U}\mu_{F_k^-}(x)$, $|F_k^+| = \sum_{x\in U}\mu_{F_k^+}(x)$, $p(F_k^{-j}) = \frac{|F_k^{-j}|}{|F_k^-|}$, $p(F_k^{+j}) = \frac{|F_k^{+j}|}{|F_k^+|}$. s_k is the number of blocks that satisfy $\mu_{\underline{apr}_R(F_k)}(x) = \left[\inf_{y\in U}\max\left\{1-\mu_{[x]_{R^+}}(y), \mu_{F_k^-}(y)\right\}, \inf_{y\in U}\max\left\{1-\mu_{[x]_{R^-}}(y), \mu_{F_k^+}(y)\right\}\right] \geqslant [\zeta^-, \zeta^+] = \zeta$. Let $M_{Rk} = \left\{[x]_R \mid \mu_{\underline{apr}_R(F_k)}(x) = \left[\inf_{y\in U}\max\left\{1-\mu_{[x]_{R^+}}(y), \mu_{F_k^-}(y)\right\}, \inf_{y\in U}\max\left\{1-\mu_{[x]_{R^-}}(y), \mu_{F_k^+}(y)\right\}\right] \geqslant [\zeta^-, \zeta^+] = \zeta, x \in U\right\}$, where $[x]_R$ is determined by $\mu_{[x]_R}(y) = \mu_R(x,y)$, then $s_k = |M_{Rk}|$.

Remark 13. Without losing generality, $M_{Rk} = \{A_1, A_2, \ldots, A_{s_k}\}$, can be assumed, where $A_j = [x_{t_j}]_R$, $j = 1,2,\ldots,s_k$,, $t_j \in \{1,2,\ldots,|U|\}$. The convergence degree is then denoted as follows:

$$C(R,\Gamma) = C(R,\Gamma^-) \wedge C(R,\Gamma^+) = \left(\sum_{k=1}^{m} \frac{\sum_{x\in U}\mu_{F_k^-}(x)}{|U|} \sum_{j=1}^{s_k}\left(\frac{\sum_{x\in A_j}\mu_{F_k^-}(x)}{\sum_{x\in U}\mu_{F_k^-}(x)}\right)^2\right) \wedge \left(\sum_{k=1}^{m} \frac{\sum_{x\in U}\mu_{F_k^+}(x)}{|U|} \sum_{j=1}^{s_k}\left(\frac{\sum_{x\in A_j}\mu_{F_k^+}(x)}{\sum_{x\in U}\mu_{F_k^+}(x)}\right)^2\right).$$

Remark 14. If $\Gamma = \{F\}$, then $C(R,\Gamma) = C(R,\Gamma^-) \wedge C(R,\Gamma^+) = \left(\sum_{j=1}^{s} p^2(F^{-j})\right) \wedge \left(\sum_{j=1}^{s} p^2(F^{+j})\right) = \left(\frac{1}{|F^-|^2}\sum_{j=1}^{s}\left(\sum_{x\in[x_{t_j}]_R}\mu_{F^-}(x)\right)^2\right) \wedge \left(\frac{1}{|F^+|^2}\sum_{j=1}^{s}\left(\sum_{x\in[x_{t_j}]_R}\mu_{F^+}(x)\right)^2\right)$.

Definition 18. Let $S = (U,A)$ be an interval-valued fuzzy information system, $\Gamma = \{F_1, F_2, \ldots, F_m\}$ be a fuzzy partition of U, and $P = \{R_1, R_2, \ldots, R_n\}$ be a family of interval-valued fuzzy attribute sets with $R_1 \prec R_2 \prec \cdots \prec R_n$, where F_k $(k = 1,2,\ldots,m)$ is an interval-valued fuzzy set, and R_k $(k = 1,2,\ldots,n)$ is a subset of A. Let $P_i = \{R_1, R_2, \ldots, R_i\}$. The convergence degree of Γ with respect to P is defined as follows:

$$C(P,\Gamma) = C(P,\Gamma^-) \wedge C(P,\Gamma^+) = \left(\sum_{k=1}^{m} \frac{|F_k^-|}{|U|} \sum_{j=1}^{s_k} p^2\left(F_k^{-j}\right)\right) \wedge \left(\sum_{k=1}^{m} \frac{|F_k^+|}{|U|} \sum_{j=1}^{s_k} p^2\left(F_k^{+j}\right)\right),$$

where $|F_k^-| = \sum_{x\in U}\mu_{F_k^-}(x),\ |F_k^+| = \sum_{x\in U}\mu_{F_k^+}(x),\ p\left(F_k^{-j}\right) = \frac{|F_k^{-j}|}{|F_k^-|},\ p\left(F_k^{+j}\right) = \frac{|F_k^{+j}|}{|F_k^+|},\ s_k = |M_{Pk}|, M_{Pk} = \left\{\bigcup_{h=1}^{j_x}[x]_{R_h} | j_x = \max\left\{t | \mu_{\underline{apr}_{R_t}(F_k)}(x) = \left[\inf_{y\in U}\max\left\{1-\mu_{[x]_{R_t^+}}(y), \mu_{F_k^-}(y)\right\},\ \inf_{y\in U}\max\left\{1-\mu_{[x]_{R_t^-}}(y), \mu_{F_k^+}(y)\right\}\right] \geqslant [\zeta^-, \zeta^+]\right\}\right\}$.

Remark 15. Without losing generality, $M_{Pk} = \{A_1, A_2, \ldots, A_{s_k}\}$ can be assumed, where $A_j = [x_{t_j}]_{R_{I_j}},\ j = 1, 2, \ldots, s_k, I_j \in \{1, 2, \ldots, n\}, t_j \in \{1, 2, \ldots, |U|\}$. The convergence degree is then represented as follows:

$$C(P,\Gamma) = C(P,\Gamma^-) \wedge C(P,\Gamma^+) = \left(\sum_{k=1}^{m}\frac{\sum_{x\in U}\mu_{F_k^-}(x)}{|U|}\sum_{j=1}^{s_k}\left(\frac{\sum_{x\in A_j}\mu_{F_k^-}(x)}{\sum_{x\in U}\mu_{F_k^-}(x)}\right)^2\right) \wedge \left(\sum_{k=1}^{m}\frac{\sum_{x\in U}\mu_{F_k^+}(x)}{|U|}\sum_{j=1}^{s_k}\left(\frac{\sum_{x\in A_j}\mu_{F_k^+}(x)}{\sum_{x\in U}\mu_{F_k^+}(x)}\right)^2\right).$$

Remark 16. If $\Gamma = \{F\}$, then:

$$C(P,\Gamma) = C(P,\Gamma^-) \wedge C(P,\Gamma^+) = \left(\sum_{j=1}^{s} p^2(F^{-j})\right) \wedge \left(\sum_{j=1}^{s} p^2(F^{+j})\right)$$
$$= \left(\frac{1}{|F^-|^2}\sum_{j=1}^{s}(\sum_{x\in A_j}\mu_{F^-}(x))^2\right) \wedge \left(\frac{1}{|F^+|^2}\sum_{j=1}^{s}\left(\sum_{x\in A_j}\mu_{F^+}(x)\right)^2\right),$$

where $A_j = [x_{t_j}]_{R_{I_j}}$.

Theorem 5. *Let S = (U,A) be an interval-valued fuzzy information system, F be an interval-valued fuzzy set of U, and P = $\{R_1, R_2, \ldots, R_n\}$ be a family of interval-valued fuzzy attribute sets with $R_1 \prec R_2 \prec \cdots \prec R_n$, where R_k (k = 1,2,...,n) is a subset of A. Let $P_i = \{R_1, R_2, \ldots, R_i\}$, then for $\forall P_i$ (i = 1,2,...,n), the following property holds:*

$$C(P_1, F) \leqslant C(P_2, F) \leqslant \cdots \leqslant C(P_n, F).$$

Proof. Suppose $1 \leqslant a < b \leqslant n$, $M_{P_ak} = \{A_1, A_2, \cdots, A_m\}$, and $M_{P_bk} = \{B_1, B_2, \cdots, B_n\}$. Clearly, $M_{P_ak} \subseteq M_{P_bk}$ and $m > n$. Then, a partition $\{C_1, C_2, \ldots, C_n\}$ of $\{1, 2, \ldots, m\}$ may exist, such that $B_t = \bigcup_{l\in C_t}A_l,\ t = 1, 2, \ldots, n$. Therefore, the following can be obtained:

$$C(P_b, F) = C(P_b, F^-) \wedge C(P_b, F^+) = \left(\frac{1}{|F^-|^2}\sum_{t=1}^{s_b}\left(\sum_{x\in B_t}\mu_{F^-}(x)\right)^2\right) \wedge \left(\frac{1}{|F^+|^2}\sum_{t=1}^{s_b}\left(\sum_{x\in B_t}\mu_{F^+}(x)\right)^2\right)$$
$$= \left(\frac{1}{|F^-|^2}\sum_{t=1}^{s_b}\left(\sum_{x\in\bigcup_{l\in C_t}A_l}\mu_{F^-}(x)\right)^2\right) \wedge \left(\frac{1}{|F^+|^2}\sum_{t=1}^{s_b}\left(\sum_{x\in\bigcup_{l\in C_t}A_l}\mu_{F^+}(x)\right)^2\right)$$
$$= \left(\frac{1}{|F^-|^2}\sum_{t=1}^{s_b}\left(\sum_{l\in C_t}\sum_{x\in A_l}\mu_{F^-}(x)\right)^2\right)$$
$$\wedge \left(\frac{1}{|F^+|^2}\sum_{t=1}^{s_b}\left(\sum_{l\in C_t}\sum_{x\in A_l}\mu_{F^+}(x)\right)^2\right) \geqslant \left(\frac{1}{|F^-|^2}\sum_{t=1}^{s_b}\sum_{l\in C_t}\left(\sum_{x\in A_l}\mu_{F^-}(x)\right)^2\right) \wedge \left(\frac{1}{|F^+|^2}\sum_{t=1}^{s_b}\sum_{l\in C_t}\left(\sum_{x\in A_l}\mu_{F^+}(x)\right)^2\right)$$
$$= \left(\frac{1}{|F^-|^2}\sum_{l=1}^{s_a}\left(\sum_{x\in A_l}\mu_{F^-}(x)\right)^2\right) \wedge \left(\frac{1}{|F^+|^2}\sum_{l=1}^{s_a}\left(\sum_{x\in A_l}\mu_{F^+}(x)\right)^2\right) = C(P_a, F^-) \wedge C(P_a, F^+) = C(P_a, F).$$

Thus, $C(P_1,F) \leqslant C(P_2,F) \leqslant \cdots \leqslant C(P_n,F)$. □

Theorem 6. *Let S = (U,A) be an interval-valued fuzzy information system, $\Gamma = \{F_1, F_2, \ldots, F_m\}$ be a fuzzy partition of U, where F_k (k = 1,2,..., m) is an interval-valued fuzzy set, and P = $\{R_1, R_2, \ldots, R_n\}$ is a family of interval-valued fuzzy attribute sets with $R_1 \prec R_2 \prec \cdots \prec R_n$, where R_k (k = 1,2,..., n) is a subset of A, which is an interval-valued fuzzy equivalence relation. Let $P_i = \{R_1, R_2, \ldots, R_i\}$, then for $\forall P_i$, (i = 1,2,...,n), the following property holds:*

$$C(P_1,\Gamma) \leqslant C(P_2,\Gamma) \leqslant \cdots \leqslant C(P_n,\Gamma).$$

Proof. It follows from Theorem 5 that $C(P_1,F_k) \leqslant C(P_2,F_k) \leqslant \cdots \leqslant C(P_n,F_k)$ for $\forall F_k(k \leqslant m)$. Suppose $1 \leqslant a < b \leqslant n$, then $C(P_a,F_k) \leqslant C(P_b,F_k)$. Therefore, we obtain $C(P_a,\Gamma) = \left(\sum_{k=1}^{m}\frac{|F_k^-|}{|U|}\sum_{j=1}^{s_k}p^2\left(F_k^{-j}\right)\right) \wedge (\sum_{k=1}^{m}\frac{|F_k^+|}{|U|}\sum_{j=1}^{s_k}p^2\left(F_k^{+j}\right)) = \left(\sum_{k=1}^{m}\frac{|F_k^-|}{|U|}\cdot C\left(P_a,F_k^-\right)\right) \wedge\left(\sum_{k=1}^{m}\frac{|F_k^+|}{|U|}\cdot C\left(P_a,F_k^+\right)\right) \leqslant \left(\sum_{k=1}^{m}\frac{|F_k^-|}{|U|}\cdot C\left(P_b,F_k^-\right)\right) \wedge \left(\sum_{k=1}^{m}\frac{|F_k^+|}{|U|}\cdot C\left(P_b,F_k^+\right)\right) = C(P_b,\Gamma)$. Thus, $C(P_1,\Gamma) \leqslant C(P_2,\Gamma) \leqslant \cdots \leqslant C(P_n,\Gamma)$. □

Theorem 6 shows that the convergence degree of Γ with respect to P_i increases as the converse granulation order becomes longer. New blocks under different granulations are induced by combining known blocks; thus, the number of blocks to describe the target concept is reduced. This suggests a new idea to describe a target concept with as few blocks as possible based on keeping the approximation precision invariant. This may have potential applications in the IVFR set theory, such as the description of multi-target concepts, approximation classification, and rule extraction.

5.2. Fuzzy rule extraction algorithm based on converse approximation

In this section, we apply the converse approximation to rule extraction. Converse approximation is based on dynamic granulation. Thus, the decision classifications induced by decision attributes can be regarded as target concepts, and the condition attribute sets can be used to construct a converse granulation order. The converse approximation approaches a target concept by the change in granulation, which can fully embody the hierarchy relation between condition and decision attribute sets. Based on the converse approximation, a rule extraction algorithm called MRBCA is designed.

Given a decision table $S = (U, C \cup D)$, for $\forall c \in C$, the significance of c with respect to D is defined by $sig_C^D(c) = \gamma_C(D) - \gamma_{C-\{c\}}(D)$, where $\gamma_C(D) = \frac{\sum_{x \in U} \sup_{F \in U/D} \mu_{\underline{apr}_C(F^-)}(x) + \sum_{x \in U} \sup_{F \in U/D} \mu_{\underline{apr}_C(F^+)}(x)}{2|U|}$.

Algorithm: MRBCA
Input: decision table $S = (U, C \cup D)$
Output: decision rules

(1) Compute decision classes $U/D = \{F_1, F_2, \ldots, F_d\}$, where F_k $(k = 1, 2, \ldots, d)$ is an interval-valued fuzzy set;
(2) Let $Rule = \emptyset$, $P_1 = \{\{C_1\}\}$, $j = 1$, $C_1 = C$;
(3) For $\forall c \in C_j$, compute the significance $sig_{C_j}^D(c)$. Let $B = \left\{c_0 | sig_{C_j}^D(c_0) = \min\left\{sig_{C_j}^D(c), c \in C_j\right\}\right\}$. If $|B| \neq 1$, then let $\gamma_{\{c_0\}}(D) = \min\{\gamma_{\{c\}}(D), c \in B\}$;
(4) Let $C_{j+1} = C_j - \{c_0\}$, $P_{j+1} = P_j \cup \{\{C_{j+1}\}\}$;
(5) $j = j + 1$. If $j < |C|$, go to (3); otherwise, go to (6);
(6) Let $k = 1$;
(7) Let $P = P_j$. Compute $\underline{apr}_P(F_k)$, M_{Pk};
(8) Put every decision rule $des([x]) \to des_{F_k}(x)$ into *Rule*, where $[x] \in M_{Pk}$;
(9) $k = k + 1$. If $k \leqslant d$, go to (7); otherwise, go to (10);
(10) Let $T = \bigcup_{k=1}^d \{x | \mu_{\underline{apr}_P(F_k)}(x) \geqslant_w \zeta, [x]_P \notin M_{Pk}\} = \bigcup_{k=1}^d \{x | \mu_{\underline{apr}_P(F_k^-)}(x) + \mu_{\underline{apr}_P(F_k^+)}(x) \geqslant \zeta^- + \zeta^+, \mu_{\underline{apr}_P(F_k^+)}(x) - \mu_{\underline{apr}_P(F_k^-)}(x) \geqslant \zeta^+ - \zeta^-, [x]_P \notin M_{Pk}\}$. For $\forall x \in T$, put $des_P(x) \to des_{F_k}(x) (k = \overline{1, 2, \ldots, d})$ into *Rule*.
(11) Output *Rule*.

Remark 17. In Step (8), $des([x])$ is the antecedent of the rule, and $des_{F_k}(x)$ is the consequent. "If a_3 is a_{31}, then d is F_2," where "a_3 is a_{31}" is $des([x])$, "d is F_2" is $des_{F_k}(x)$.

The time complexity to extract rules is a polynomial.

In Step (1), the time complexity to compute a decision partition is $O(|U|^2)$.

In Step (3), the time complexity to compute significance is $O(|C_j||U|^2)$, and the time complexity to compute $sig_{C_j}^D(c)$ for $\forall c \in C_j$ is $O(|C_j|^2|U|^2)$. The time complexity to choose the minimum of significance is $O(|C_j|)$. In Steps (3)–(5), because $|C| - 1$ is the maximum value for the circle times, the time complexity to construct P_j is

$$\begin{aligned}\sum_{j=1}^{|C|-1}(O(|C_j|^2|U|^2) + O(|C_j|)) &= \sum_{j=1}^{|C|-1} O(|C_j|^2|U|^2) + \sum_{j=1}^{|C|-1} O(|C_j|) \\ &= O(|C|^2|U|^2 + (|C|-1)^2|U|^2 + (|C|-2)^2|U|^2 \cdots + 2^2|U|^2) + O(|C| + (|C|-1) + \cdots + 2) \\ &= O\left(\left(\frac{1}{6}(2|C|^3 + 3|C|^2 + |C|) - 1\right)|U|^2\right) + O(\frac{1}{2}(|C|^2 + |C| - 2) = O(|C|^3|U|^2).\end{aligned}$$

In Step (7), the time complexity for computing $\underline{apr}_P(F_k)$,M_{Pk} is $O(|C||U|^2)$.

In Step (8), the time complexity for putting each decision rule into the rule base is $O(|M_{Pk}|)$.

In Step (10), the time complexity for computing T is $O(|C||U|^2)$.

In Step (11), the time complexity is $O(|U|)$.

In Steps (7)–(9), d is the circle times. Therefore, the time complexity of the algorithm MRBCA is $O(|U|^2) + O(|C|^3|U|^2) + \sum_{k=1}^d (O(|C||U|^2) + O(|M_{Pk}|)) + O(|C||U|^2) + O(|U|) = O(|C|^3|U|^2)$.

The time complexity of this algorithm can be reduced to $O(|C|^3|U| \log_2 |U|)$ of a classification is computed using the ranking technique.

Remark 18. The differences between MRBCA and the method based on attribute reduction include two aspects. First, rule extraction is based on a converse granulation order instead of attribute reduction. Second, MRBCA can extract much simpler rules based on keeping the approximation precision invariant.

5.3. An example

An example using the data set in Table 2 is presented to discuss the application of MRBCA.

According to MRBCA, a converse granulation order is constructed first. The significances of a_1, a_2, a_3 with regard to d are computed. For $C_1 = C = \{a_1,a_2,a_3\}$, we obtain $sig^d_{C_1}(a_1) = \frac{10}{2000}$, $sig^d_{C_1}(a_2) = \frac{115}{2000}$, $sig^d_{C_1}(a_3) = \frac{70}{2000}$, thus, $c_0 = a_1$, $C_2 = C_1 - a_1 = \{a_2,a_3\}$, $P_2 = \{C_1,C_2\}$.

For C_2, we obtain $sig^d_{C_2}(a_2) = \frac{145}{2000}$, $sig^d_{C_2}(a_3) = \frac{435}{2000}$. As $sig^d_{C_2}(a_2) < sig^d_{C_2}(a_3)$, $C_3 = C_2 - a_2 = \{a_3\}$, $P_3 = \{C_1, C_2, C_3\} = \{\{a_1, a_2, a_3\}, \{a_2, a_3\}, \{a_3\}\}$, and $P = P_3$.

Let $\zeta = [0.5,0.6]$. According to the definition of the converse approximation, we obtain:

$$\mu_{\underline{apr_P}(F_1)}(x_1) = [0.45, 0.7], \quad \mu_{\underline{apr_P}(F_1)}(x_2) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_1)}(x_3) = [0.45, 0.7], \quad \mu_{\underline{apr_P}(F_1)}(x_4) = [0.4, 0.45],$$
$$\mu_{\underline{apr_P}(F_1)}(x_5) = [0.2, 0.4],$$

$$\mu_{\underline{apr_P}(F_1)}(x_6) = [0.4, 0.5], \quad \mu_{\underline{apr_P}(F_1)}(x_7) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_1)}(x_8) = [0.2, 0.4], \quad \mu_{\underline{apr_P}(F_1)}(x_9) = [0.45, 0.7],$$
$$\mu_{\underline{apr_P}(F_1)}(x_{10}) = [0.3, 0.4],$$

$$M_{PF_1} = \emptyset, \quad Rule = \emptyset,$$

$$\mu_{\underline{apr_P}(F_2)}(x_1) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_2)}(x_2) = [0.5 \vee 0.5 \vee 0.5, 0.7 \vee 0.7 \vee 0.7],$$

$$\mu_{\underline{apr_P}(F_2)}(x_3) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_2)}(x_4) = [0.5 \vee 0.5, 0.6 \vee 0.6],$$

$$\mu_{\underline{apr_P}(F_2)}(x_5) = [0.2, 0.5], \quad \mu_{\underline{apr_P}(F_2)}(x_6) = [0.4, 0.5], \quad \mu_{\underline{apr_P}(F_2)}(x_7) = [0.5 \vee 0.5 \vee 0.5, 0.7 \vee 0.7 \vee 0.7],$$
$$\mu_{\underline{apr_P}(F_2)}(x_8) = [0.2, 0.5],$$

$$\mu_{\underline{apr_P}(F_2)}(x_9) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_2)}(x_{10}) = [0.5 \vee 0.5 \vee 0.5, 0.7 \vee 0.7 \vee 0.7];$$

$M_{PF_2} = \{(a_{13} \cap a_{22} \cap a_{31}) \cup (a_{22} \cap a_{31}) \cup a_{31}, (a_{12} \cap a_{22} \cap a_{33}) \cup (a_{22} \cap a_{33})\} = \{a_{31}, a_{22} \cap a_{33}\}$, and

Rule = {r'_1: If a_3 is a_{31} Then d is F_2 and $\mu_{\underline{apr_P}(F_2)}(x) \geqslant [0.5, 0.7]$;

r'_2: If a_2 is a_{22} and a_3 is a_{33}, then d is F_2 and $\mu_{\underline{apr_P}(F_2)}(x) \geqslant [0.5, 0.6]$}.

$$\mu_{\underline{apr_P}(F_3)}(x_1) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_3)}(x_2) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_3)}(x_3) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_3)}(x_4) = [0.4, 0.45],$$

$$\mu_{\underline{apr_P}(F_3)}(x_5) = [0.5 \vee 0.5 \vee 0.5, 0.6 \vee 0.6 \vee 0.6], \quad \mu_{\underline{apr_P}(F_3)}(x_6) = [0.7 \vee 0.7 \vee 0.6, 0.8 \vee 0.7 \vee 0.7], \quad \mu_{\underline{apr_P}(F_3)}(x_7) = [0.3, 0.4],$$

$$\mu_{\underline{apr_P}(F_3)}(x_8) = [0.5 \vee 0.5 \vee 0.5, 0.6 \vee 0.6 \vee 0.6], \quad \mu_{\underline{apr_P}(F_3)}(x_9) = [0.3, 0.4], \quad \mu_{\underline{apr_P}(F_3)}(x_{10}) = [0.3, 0.4];$$

$M_{PF_3} = \{(a_{11} \cap a_{21} \cap a_{34}) \cup (a_{21} \cap a_{34}) \cup a_{34}, (a_{11} \cap a_{21} \cap a_{32}) \cup (a_{21} \cap a_{32}) \cup a_{32}\} = \{a_{34}, a_{32}\}$, and

Rule = {r'_1: If a_3 is a_{31} then d is F_2 and $\mu_{\underline{apr_P}(F_2)}(x) \geqslant [0.5, 0.7]$;
r'_2: If a_2 is a_{22} and a_3 is a_{33}, then d is F_2 and $\mu_{\underline{apr_P}(F_2)}(x) \geqslant [0.5, 0.6]$;

r'_3:If a_3 is a_{34}, then d is F_3 and $\mu_{\underline{apr_P}(F_3)}(x) \geqslant [0.5, 0.6]$;
r'_4: If a_3 is a_{32}, then d is F_3 and $\mu_{\underline{apr_P}(F_3)}(x) \geqslant [0.6, 0.7]$}.

Moreover, for $x \in a_{12} \cap a_{21} \cap a_{33}$, we obtain $\mu_{\underline{apr_P}(F_1^-)}(x) + \mu_{\underline{apr_P}(F_1^+)}(x) = 0.45 + 0.7 \geqslant \zeta^- + \zeta^+ = 0.5 + 0.6$, $\mu_{\underline{apr_P}(F_1^+)}(x) - \mu_{\underline{apr_P}(F_1^-)}(x) = 0.7 - 0.45 \geqslant \zeta^+ - \zeta^- = 0.6 - 0.5$ and $a_{12} \cap a_{21} \cap a_{33} \notin M_{PF_k}$, k = 1, 2, 3. Therefore, $T = \{x_1,x_3,x_9\}$ and r'_5: If a_1 is a_{12}, a_2 is a_{21} and a_3 is a_{33} then d is F_1 and $\mu_{\underline{apr_P}(F_1)}(x) \geqslant [0.45, 0.7]$ is added to*Rule*. Rules can be obtained as follows:

Rule = {r'_1: If a_3 is a_{31}, then d is F_2 and $\mu_{\underline{apr_P}(F_2)}(x) \geqslant [0.5, 0.7]$;
r'_2: If a_2 is a_{22} and a_3 is a_{33}, then d is F_2 and $\mu_{\underline{apr_P}(F_2)}(x) \geqslant [0.5, 0.6]$;
r'_3: If a_3 is a_{34}, then d is F_3 and $\mu_{\underline{apr_P}(F_3)}(x) \geqslant [0.5, 0.6]$;
r'_4: If a_3 is a_{32}, then d is F_3 and $\mu_{\underline{apr_P}(F_3)}(x) \geqslant [0.6, 0.7]$;
r'_5: If a_1 is a_{12},a_2 is a_{21} and a_3 is a_{33}, then d is F_1 and $\mu_{\underline{apr_P}(F_1)}(x) \geqslant [0.45, 0.7]$}

Table 3
Data description.

	Data set	Abbreviation	Samples	Attributes	Classes
1	Parkinsons	parkinsons	197	23	2
2	Housing	housing	506	13	Continuous
3	Concrete compressive strength	concrete	1030	8	Continuous
4	Image segmentation	image	2310	19	7
5	Page blocks	page	5473	10	5
6	Waveform database generator	wave	5000	21	3
7	Magic gamma telescope	magic	19020	11	2

Comparing r'_1, r'_2, r'_3, r'_4, and r'_5 with r_1, r_4, r_3, r_2, and r_5, respectively, in Section 4.3 shows that the rules extracted from MRBPA and MRBCA are the same.

The previously discussed examples reveal three merits of MRBPA and MRBCA. First, the computational complexity of MRBPA is less than that of the knowledge discovery method [7] due to the gradual dwindling of the universe. Second, for the algorithm MRBCA, much simpler rules can be extracted based on keeping the approximation precision invariant. Third, rules can be generated when an interval is nested in the other due to Step (5) in MRBPA and Step (10) in MRBCA.

6. Experimental analysis

In this section, we first evaluate the performance of CMAVIFIS. We then compare the computing time and classification accuracy of MRBPA, MRBCA, and the fuzzy rule induction algorithm (RIA) on different data sets.

We download several data sets from the UCI Machine Learning database [55] to test our proposed methods. The data sets are outlined in Table 3. In the seven sets, two have a continuous class attribute, while the others have a categorical class attribute. Furthermore, the number of samples is between 197 and 19,020.

The experiment is performed on a 400 MHz Pentium Server with 512 MB of memory running on Windows XP. Algorithms are coded in Matlab 7.1. Considering that CMAVIFIS, MRBPA, and MRBCA mainly deal with interval-valued data, some

Table 4
Accuracy of CMAVIFIS.

Missing ratio	Test times and accuracy (%)					Average accuracy (%)
	1	2	3	4	5	
0.02	0.948	0.952	0.950	0.945	0.942	0.9474
0.05	0.918	0.923	0.920	0.921	0.916	0.9196
0.08	0.908	0.917	0.912	0.913	0.906	0.9112

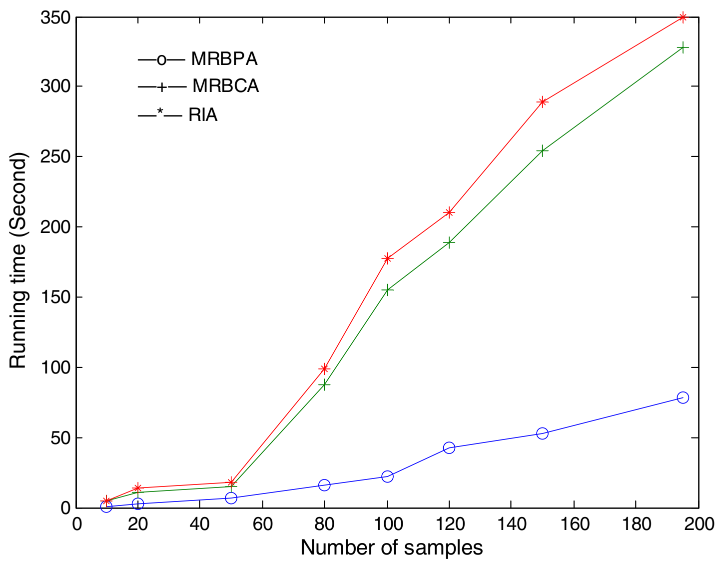

Fig. 1. Comparison of running time (parkinsons).

pretreatments, including the fuzzification of the data set and the conversion of a fuzzy set to an interval-valued fuzzy set, are necessary.

A simple algorithm [48] is used to generate a triangular membership function defined as follows:

$$T_1(x) = \begin{cases} 1, & x \leqslant m_1, \\ (m_2 - x)/(m_2 - m_1), & m_1 < x < m_2, \\ 0, & m_2 \leqslant x, \end{cases}$$

$$T_k(x) = \begin{cases} 1, & x \geqslant m_k, \\ (x - m_{k-1})/(m_k - m_{k-1}), & m_{k-1} < x < m_k, \\ 0, & x \leqslant m_{k-1}, \end{cases}$$

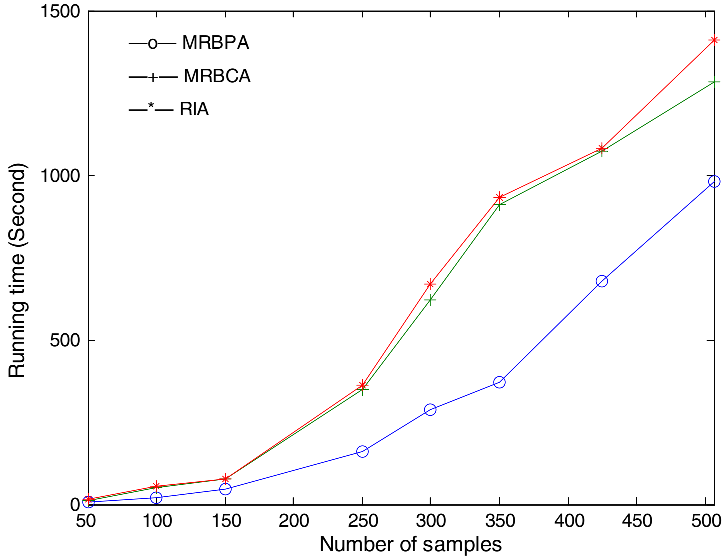

Fig. 2. Comparison of running time (housing).

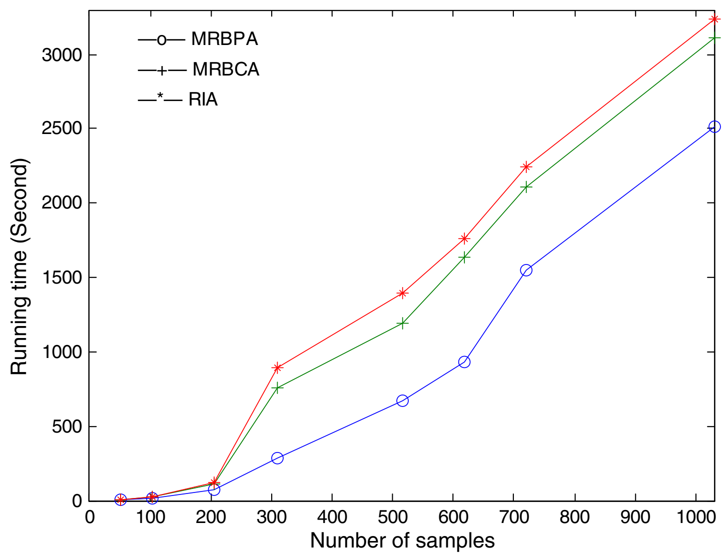

Fig. 3. Comparison of running time (concrete).

$$T_i(x) = \begin{cases} 0, & x \geqslant m_{i+1}, \\ (m_{i+1} - x)/(m_{i+1} - m_i), & m_i \leqslant x \leqslant m_{i+1}, \\ (x - m_{i-1})/(m_i - m_{i-1}), & m_{i-1} < x < m_i, \\ 0, x \leqslant m_{i-1}. \end{cases}$$

The slopes of the triangular membership functions are selected such that adjacent membership functions cross at the membership value 0.5. In this case, the only parameter to be determined is $M = \{m_i, i = 1,2,\ldots,k\}$. The center, m_i, can be calculated using the feature-maps algorithm by Kohonen [18].

After the fuzzification of data, each attribute has three linguistic terms, each of which is a fuzzy set on the data set. A construction theorem is used to construct an interval-valued fuzzy set from a fuzzy set [21]. Seven interval-valued fuzzy information systems can be obtained. The following experiment consists of two parts: the validity test of CMAVIFIS and the rule extraction based on MRBPA and MRBCA.

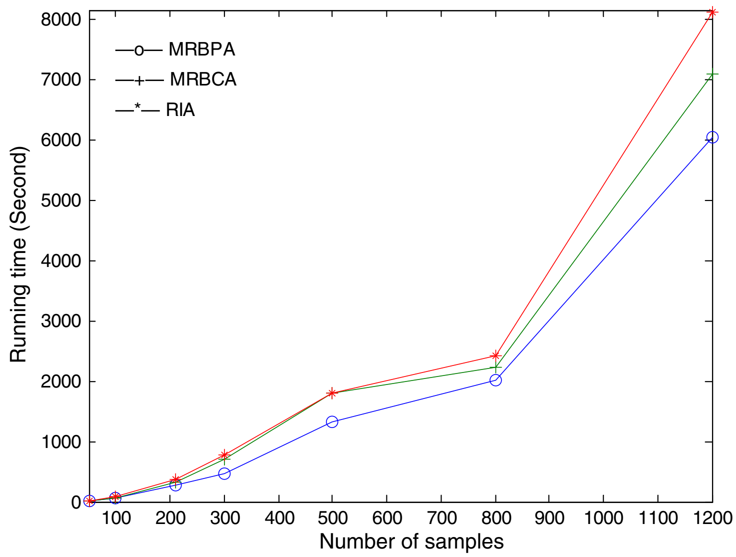

Fig. 4. Comparison of running time (image).

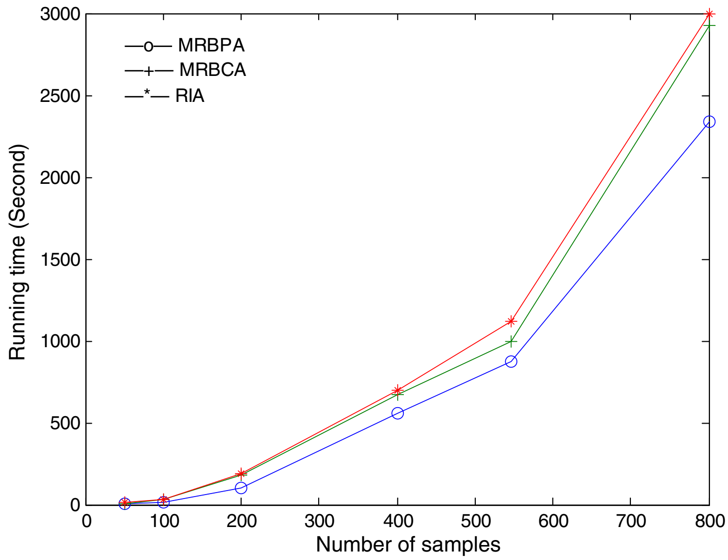

Fig. 5. Comparison of running time (page).

6.1. Validity test of CMAVIFIS

The data set ''housing'' is used to evaluate the performance of CMAVIFIS. In the corresponding interval-valued fuzzy information system, there is no missing datum. Missing values with a certain ratio are randomly generated to obtain an incomplete interval-valued fuzzy information system. The ratios of the missing values are 2, 5, and 8%, respectively. The test is carried out five times. The results are summarized in Table 4 ($\xi = 0.65$). Table 4 also shows that the overall accuracy decreases as the ratio of missing values increases. This is reasonable because more useful information is lost when the amount of missing data increases. The average accuracy of CMAVIFIS is over 91%, which shows that CMAVIFIS sufficiently utilizes the hidden commonness in the data set. Hardly any existing completeness method of incomplete interval-valued fuzzy information systems exists; thus, we cannot directly compare the complete information system obtained from CMAVIFIS with others. Comparing the accuracies with those of the completeness methods for symbolic-valued information systems [10,11,54], the experiment results are acceptable.

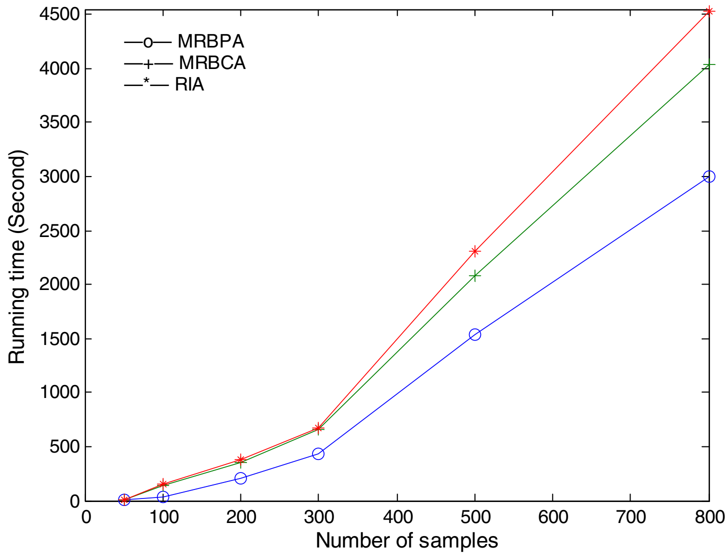

Fig. 6. Comparison of running time (wave).

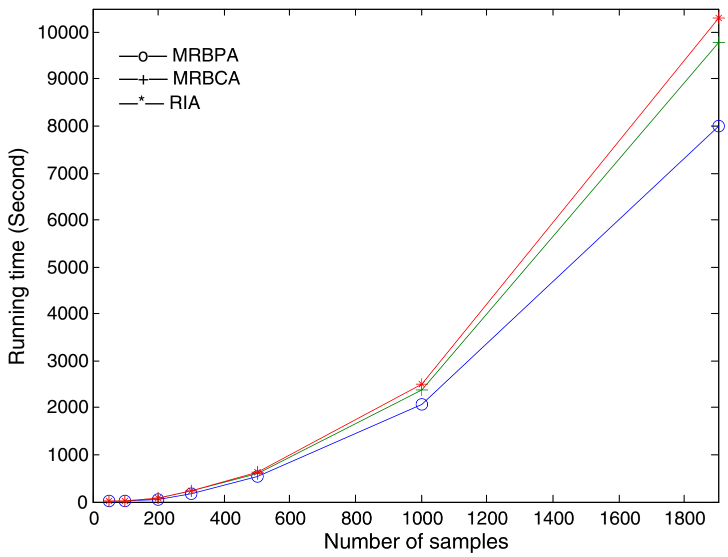

Fig. 7. Comparison of running time (magic).

6.2. Rule extraction based on MRBPA and MRBCA

In this section, rules are extracted using MRBPA and MRBCA. There is almost no existing rule extraction method for interval-valued fuzzy information systems with both interval-valued fuzzy condition and decision attributes; thus, we cannot directly compare the performance of MRBPA and MRBCA with that of others before generalizing the rule extraction method in Ref. [37].

6.2.1. Rule extraction based on attribute reduction

Ref. [37] is a representative work on fuzzy rule extraction based on fuzzy rough sets. The rule extraction method consists of two steps: (1) extension relative reduction to fuzzy rough sets and development of an algorithm to compute a reduct, and

Table 5
Classification accuracies of MRBPA, MRBCA and RIA (%).

	Data set	MRBPA	MRBCA	RIA
1	parkinsons	0.9744 ± 0.0161	0.9568 ± 0.0230	0.7846 ± 0.0822
3	concrete	0.6867 ± 0.1034	0.6599 ± 0.1123	0.5276 ± 0.0823
4	image	0.8345 ± 0.0373	0.8157 ± 0.0409	0.7311 ± 0.0698
5	page	0.5513 ± 0.0312	0.5507 ± 0.0398	0.2608 ± 0.1034
6	wave	0.6012 ± 0.0978	0.5747 ± 0.0830	0.5462 ± 0.1208
7	magic	0.6179 ± 0.0723	0.6135 ± 0.0676	0.5840 ± 0.0857

Table 6
Classification accuracies of MRBPA, MRBCA and RIA.

Number of samples in training set	Number of samples in test set	Method	Training accuracy (%)	Testing accuracy (%)
506	506	MRBPA($\eta^- = 0.5, \eta^+ = 0.55$)	0.65217	0.65217
		MRBPA($\eta^- = 0.5, \eta^+ = 0.52$)	0.73715	0.73715
		MRBPA($\eta^- = 0.5, \eta^+ = 0.5$)	0.79249	0.79249
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.55$)	0.66333	0.66333
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.52$)	0.74877	0.74877
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.5$)	0.80049	0.80049
		RIA($\alpha = 0.85$)	0.65115	0.65115
		RIA($\alpha = 0.8$)	0.54941	0.54941
425	81	MRBPA($\eta^- = 0.5, \eta^+ = 0.55$)	0.66824	0.53086
		MRBPA($\eta^- = 0.5, \eta^+ = 0.52$)	0.78353	0.60494
		MRBPA($\eta^- = 0.5, \eta^+ = 0.5$)	0.84235	0.72840
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.55$)	0.67489	0.54870
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.52$)	0.77890	0.61852
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.5$)	0.85647	0.72141
		RIA($\alpha = 0.85$)	0.65412	0.51852
		RIA($\alpha = 0.8$)	0.54321	0.50529
350	156	MRBPA($\eta^- = 0.5, \eta^+ = 0.55$)	0.71429	0.55128
		MRBPA($\eta^- = 0.5, \eta^+ = 0.52$)	0.83714	0.68590
		MRBPA($\eta^- = 0.5, \eta^+ = 0.5$)	0.85429	0.69872
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.55$)	0.71633	0.56989
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.52$)	0.83429	0.69282
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.5$)	0.84286	0.70154
		RIA($\alpha = 0.85$)	0.57688	0.56154
		RIA($\alpha = 0.8$)	0.55143	0.55333
250	256	MRBPA($\eta^- = 0.5, \eta^+ = 0.55$)	0.74000	0.53516
		MRBPA($\eta^- = 0.5, \eta^+ = 0.52$)	0.82000	0.69531
		MRBPA($\eta^- = 0.5, \eta^+ = 0.5$)	0.84800	0.71875
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.55$)	0.73300	0.54000
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.52$)	0.82400	0.67688
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.5$)	0.83600	0.70900
		RIA($\alpha = 0.85$)	0.58000	0.50984
		RIA($\alpha = 0.8$)	0.50400	0.48828
150	356	MRBPA($\eta^- = 0.5, \eta^+ = 0.55$)	0.84667	0.59270
		MRBPA($\eta^- = 0.5, \eta^+ = 0.52$)	0.89333	0.65169
		MRBPA($\eta^- = 0.5, \eta^+ = 0.5$)	0.97333	0.75281
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.55$)	0.83112	0.59860
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.52$)	0.90667	0.66966
		MRBCA($\zeta^- = 0.5, \zeta^+ = 0.5$)	0.95903	0.75225
		RIA($\alpha = 0.85$)	0.59333	0.54775
		RIA($\alpha = 0.8$)	0.52000	0.54494

(2) extraction of rules based on an existing fuzzy RIA. To extend the method to interval-valued fuzzy information systems with both interval-valued fuzzy condition and decision attributes, the dependency degree and the fuzzy subsethood should be redefined. The definition of dependency degree is the same as in Definition 14. The fuzzy subsethood is redefined as follows:

$$S(A,B)=\frac{M(A\cap B)}{M(A)}=\frac{\sum_{u\in U}\min\{\mu_{A^-}(u),\mu_{B^-}(u)\}}{2\sum_{u\in U}\mu_{A^-}(u)}+\frac{\sum_{u\in U}\min\{\mu_{A^+}(u),\mu_{B^+}(u)\}}{2\sum_{u\in U}\mu_{A^+}(u)},$$

where A and B are interval-valued fuzzy sets.

The subsethood values indicate the relationship between condition and decision attributes. A suitable threshold, $\alpha\in[0,1]$, must be chosen to determine whether the terms are close enough.

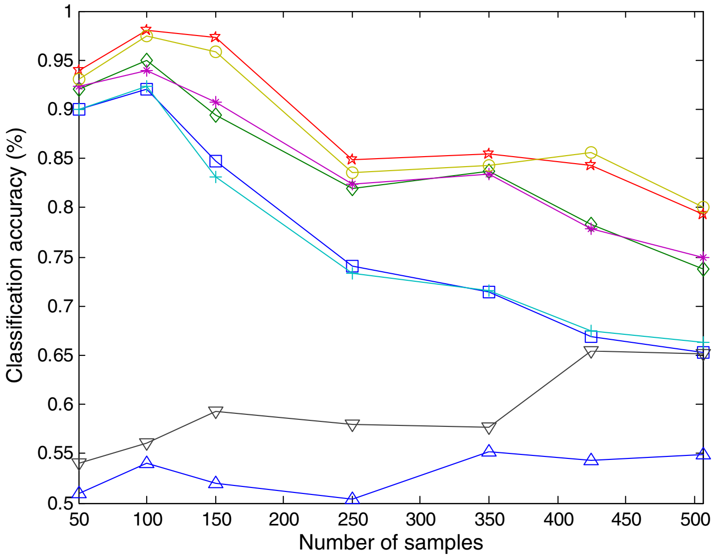

Fig. 8. Training accuracy (instruction of signs in Fig. 8: —□—: MRBPA($\eta^-=0.5,\eta^+=0.55$); —◇—: MRBPA($\eta^-=0.5,\eta^+=0.52$); —☆—: MRBPA($\eta^-=0.5,\eta^+=0.5$); —+—: MRBCA($\zeta^-=0.5,\zeta^+=0.55$); —∗—: MRBCA($\zeta^-=0.5,\zeta^+=0.52$); —o—: MRBCA($\zeta^-=0.5,\zeta^+=0.5$); —▽—: RIA($\alpha=0.85$); —△—: RIA($\alpha=0.8$)).

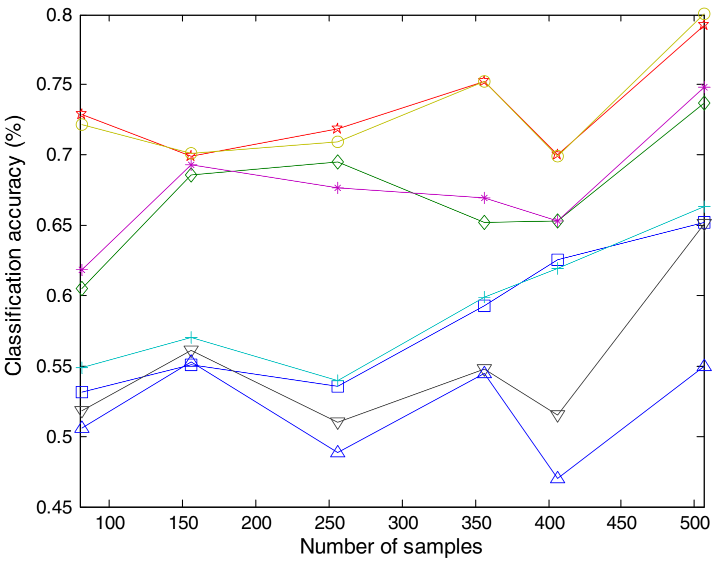

Fig. 9. Testing accuracy (instruction of signs in Fig. 9: —□—: MRBPA($\eta^-=0.5,\eta^+=0.55$); —◇—: MRBPA($\eta^-=0.5,\eta^+=0.52$); —☆—: MRBPA($\eta^-=0.5,\eta^+=0.5$); —+—: MRBCA($\zeta^-=0.5,\zeta^+=0.55$); —∗—: MRBCA($\zeta^-=0.5,\zeta^+=0.52$); —o—: MRBCA($\zeta^-=0.5,\zeta^+=0.5$); —▽—: RIA($\alpha=0.85$); —△—: RIA($\alpha=0.8$)).

Based on the new definition of dependency degree, "fuzzy-rough QUICKREDUCT algorithm" in [37] is used to reduce the attributes. According to the definition of fuzzy subsethood, the fuzzy RIA [37] is then used to generate the fuzzy rules.

6.2.2. Experimental results

We randomly divide the samples into 10 subsets. One is used as the training set to find the rule set, and the remainder is used as the test sets to determine the classification accuracy. After 10 rounds, we compute the average value and variation. Rule extraction is performed using MRBPA, MRBCA, and RIA. Two indices are used to evaluate the three algorithms: total running time and classification accuracy. The threshold values of η and ζ are from [0.5, 0.5] to [0.8, 0.8] for MRBPA and MRBCA, respectively, and α in the interval [0.7, 0.9] for RIA. Figs. 1–7 show the running time using these algorithms, and indicate that the running time increases with the increase in samples. The running time of MRBPA is far less than that of RIA, while the running time of MRBCA is slightly less than that of RIA. The running time of MRBPA is significantly reduced, especially for parkinsons, housing, and concrete data sets; the results are consistent with the key idea of MRBPA; i.e., computational complexity is effectively reduced as the universe dwindles gradually. Table 5 presents the comparison of the classification accuracies with MRBPA, MRBCA, and RIA. Overall, the classification accuracies of MRBPA are slightly better than those of MRBCA. Both MRBPA and MRBCA outperform RIA. The classification accuracies of MRBPA and MRBCA are higher than those of RIA for parkinsons, concrete, image, and page data sets. Particularly for parkinsons, the classification accuracy of MRBPA is 97.44 ± 1.61%, which is a satisfactory result.

Corresponding to different thresholds and different-sized data sets, we compared the classification accuracy further by using three algorithms. The data set "housing" is reused. The selected data set is first divided into two parts: the training set composed of randomly chosen samples and the test set composed of the remainder. Thresholds η and ζ are considered parameters to control the granularity of fuzzy rules. We take the values of η and ζ from [0.5, 0.5] to [1.0, 1.0] with step 0.01. The classification accuracies vary with the thresholds. Generally, [0.5, 0.5]–[0.5, 0.7] is a candidate range for η and ζ, where both training and testing accuracies obtain good classification performance. Part training accuracies and testing accuracies corresponding to different sizes of data sets and different thresholds are enumerated in Table 6. The classification accuracies of MRBPA and MRBCA are almost equivalent. Both MRBPA and MRBCA outperform RIA. For the data set including 150 training samples and 356 test samples, the average classification accuracies for MRBPA ($\eta^- = 0.5, \eta^+ = 0.5$) are 97.333% (training accuracy) and 75.281% (testing accuracy), and 95.903% (training accuracy) and 75.225% (testing accuracy). For RIA, the accuracy is 59.333% using the training data and 54.775% for the test data ($\alpha = 0.85$). More intuitive comparisons can be found in Figs. 8 and 9.

7. Conclusions

This paper presents two fuzzy rule extraction methods for interval-valued fuzzy information systems. The main features of the methods cover four aspects. (1) Rule extraction is based on a granulation order, thus the adverse effects of attribute reduction are excluded as much as possible. (2) They can be applied to three types of interval-valued fuzzy information systems (i.e., crisp condition and interval-valued fuzzy decision, interval-valued fuzzy condition and crisp decision, and interval-valued fuzzy condition and decision). (3) For MRBPA, computational consumption can be reduced effectively as the domain gradually narrows. (4) When one interval is nested in the other, rules can still be generated. The comparative experiments show that the methods in this paper achieve better classification performances than the method based on attribute reduction.

Acknowledgments

The authors would like to thank Professor Witold Pedrycz, the Editor-in-Chief, the anonymous reviewers, Dr. Yan Zhao, Dr. Feifei Xu, and Feng Luo for their valuable comments and suggestions. This work was supported by the National Natural Science Foundation of China (Nos. 60475019, 60775036) and the Doctoral Program of Higher Education (No. 20060247039).

References

[1] R.B. Bhatt, M. Gopal, On the compact computational domain of fuzzy-rough sets, Pattern Recognition Letters 26 (2005) 1632–1640.
[2] R.B. Bhatt, M. Gopal, On fuzzy-rough sets approach to feature selection, Pattern Recognition Letters 26 (2005) 965–975.
[3] C. Cornelis, G.H. Martin, R. Jensen, D. Slezak, Feature selection with fuzzy decision reducts, in: G. Wang et al. (Eds.), Proc. RSKT'08, Lecture Notes in Artificial Intelligence, vol. 5009, Springer, Berlin, 2008, pp. 284–291.
[4] T.Q. Deng, Y.M. Chen, W.L. Xu, Q.H. Dai, A novel approach to fuzzy rough sets based on a fuzzy covering, Information Sciences 177 (2007) 2308–2326.
[5] D. Dubois, H. Prade, Twofold fuzzy sets and rough sets-some issues in knowledge representation, Fuzzy Sets and Systems 23 (1987) 3–18.
[6] D. Dubois, H. Prade, Fuzzy rough sets and fuzzy rough sets, International Journal of General Systems 17 (2–3) (1990) 191–209.
[7] Z.T. Gong, B.Z. Sun, D.G. Chen, Rough set theory for the interval-valued fuzzy information systems, Information Sciences 178 (2008) 1968–1985.
[8] B. Gorzalczany, Interval-valued fuzzy controller based on verbal modal of object, Fuzzy Sets and Systems 28 (1988) 45–53.
[9] S. Greco, M. Inuiguchi, R. Slowinski, Fuzzy rough sets and multiple-premise gradual decision rules, International Journal of Approximate Reasoning 41 (2006) 179–211.
[10] J.W. Grzymala-Busse, M. Hu, A comparison of several approaches to missing attribute values in data mining, in: W. Ziarko, Y. Yao (Eds.), RSCTC'00, Lecture Notes in Artificial Intelligence, vol. 2005, Springer, Berlin, 2001, pp. 378–385.
[11] J.W. Grzymala-Busse, W.J. Grzymala-Busse, An experimental comparison of three rough set approaches to missing attribute values, J.F. Peters et al. (Eds.), Transactions on Rough Sets VI, LNCS, vol. 4374, 2007, pp. 31–50.

[12] Q.H. Hu, D.R. Yu, Entropies of fuzzy indiscernibility relation and its operations, International Journal of Uncertainty Fuzziness Knowledge Based System 12 (5) (2004) 575–589.
[13] Q.H. Hu, D.R. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognition Letters 27 (2006) 414–423.
[14] H. Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recognition 40 (2007) 3509–3521.
[15] R. Jensen, Q. Shen, Fuzzy-rough attribute reduction with application to web categorization, Fuzzy Sets and Systems 141 (2004) 469–485.
[16] R. Jensen, Q. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Transactions on Fuzzy Systems 15 (1) (2007) 73–89.
[17] J.S. Jiang, C.X. Wu, D.G. Chen, The product structure of fuzzy rough sets on a group and the rough T-fuzzy group, Information Sciences 175 (2005) 97–107.
[18] T. Kohonen, Se!f-Organization and Associative Memory, Springer, Berlin, 1988.
[19] I. Kononenko. I. Bratko. E. Roskar. Experiments in automatic learning of medical diagnostic rules, Technical Report, Jozef Stefan Institute, Ljubljana, Yugoslavia, 1984.
[20] J.Y. Liang, Y.H. Qian, C.Y. Chu, D.Y. Li, J.H. Wang, Rough set approximation based on dynamic granulation, in: D. Slezak et al. (Eds.), RSFDGrC'05, Lecture Notes in Artificial Intelligence, vol. 3641, Springer, Berlin, 2005, pp. 701–708.
[21] H.W. Liu, Generated construction of interval valued fuzzy sets, Journal of Shandong University of Technology 30 (2000) 547–551.
[22] D.Q. Miao, D.G. Li, S.D. Fan, Fuzzy rough set and its improvement, in: Proceedings of the 1st International Conference on Granular Computing, Beijing'05, China, IEEE Press, 2005, pp. 1247–1251.
[23] M. Moshkov, A. Skowron, Z. Suraj, Maximal consistent extensions of information systems relative to their theories, Information Sciences 178 (2008) 2600–2620.
[24] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341–356.
[25] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.
[26] Y.H. Qian, J.Y. Liang, C.Y. Dang, Consistency measure, inclusion degree and fuzzy measure in decision tables, Fuzzy Sets and Systems 159 (2008) 2353–2377.
[27] Y.H. Qian, J.Y. Liang, C.Y. Dang, H.Y. Zhang, J.M. Ma, On the evaluation of the decision performance of an incomplete decision table, Data & Knowledge Engineering 65 (3) (2008) 373–400.
[28] Y.H. Qian, J.Y. Liang, D.Y. Li, H.Y. Zhang, C.Y. Dang, Measures for evaluating the decision performance of a decision table in rough set theory, Information Sciences 178 (1) (2008) 181–202.
[29] Y.H. Qian, J.Y. Liang, C.Y. Dang, Interval ordered information systems, Computers & Mathematics with Applications 56 (2008) 1994–2009.
[30] Y.H. Qian, J.Y. Liang, C.Y. Dang, Converse approximation and rule extraction from decision tables in rough set theory, Computers and Mathematics with Applications 55 (2008) 1754–1765.
[31] H.N. Qin, Completing the default in non-completed data tables using valued tolerance relation, Journal of Wuyi University 19 (2005) 42–44.
[32] A.M. Radzikowska, E.E. Kerre, A comparative study of fuzzy rough sets, Fuzzy Sets and Systems 126 (2002) 137–155.
[33] B.B. Rajen, M. Gopal, On fuzzy-rough sets approach to feature selection, Pattern Recognition Letters 26 (2005) 965–975.
[34] J. Richard, S. Qiang, Fuzzy-rough sets for descriptive dimensionality reduction, Fuzzy Systems, in: Proceedings of the Fuzz-IEEE'02, 2002, pp. 29–34.
[35] K.P. Sankar, Soft data mining, computational theory of perceptions, and rough-fuzzy approach, Information Sciences 163 (2004) 5–12.
[36] Q. Shen, A. Chouchoulas, A rough-fuzzy approach for generating classification rules, Pattern Recognition 35 (2002) 2425–2438.
[37] Q. Shen, R. Jensen, Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, Pattern Recognition 37 (2004) 1351–1363.
[38] B.Z. Sun, Z.T. Gong, D.G. Chen, Fuzzy rough set theory for the interval-valued fuzzy information systems, Information Sciences 178 (2008) 2794–2815.
[39] G.C.Y. Tsang, D.G. Chen, E.C.C. Tsang, J.W.T. Lee, D.S. Yeung, On attribute reduction with fuzzy rough sets, in: Proceedings of the SMC'05, 2005, pp. 2775–2780.
[40] E.C.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, J.W.T. Lee, Attribute reduction using fuzzy rough sets, IEEE Transaction on Fuzzy Systems 16 (5) (2008) 1130–1141.
[41] B. Turksen, Interval-valued fuzzy sets based on normal forms, Fuzzy Sets and Systems 20 (1986) 191–210.
[42] G.Y. Wang, Rough Set Theory and Data Mining, Xi'an Jiaotong University Press, Xi'an, 2001.
[43] X.Z. Wang, Y. Ha, D.G. Chen, On the reduction of fuzzy rough sets, in: Proceedings of the ICMLC'05, IEEE Press, 2005, pp. 3174–3178.
[44] X.Z. Wang, E.C.C. Tsang, S.Y. Zhao, D.G. Chen, D.S. Yeung, Learning fuzzy rules from fuzzy samples based on rough set technique, Information Sciences 177 (2007) 4493–4514.
[45] X.Z. Wang, J.H. Zhai, S.X. Lu, Induction of multiple fuzzy decision trees based on rough set technique, Information Sciences 178 (2008) 3188–3202.
[46] Y.F. Wang, Mining stock price using fuzzy rough set system, Expert Systems with Applications 24 (2003) 13–23.
[47] D.S. Yeung, D.G. Chen, E.C.C. Tsang, J.W.T. Lee, X.Z. Wang, On the generalization of fuzzy rough sets, IEEE Transactions on Fuzzy Systems 13 (2005) 343–361.
[48] Y.F. Yuan, M.J. Shaw, Introduction of fuzzy decision tree, Fuzzy Sets and Systems 69 (1995) 125–139.
[49] S.Y. Zhao, E.C.C. Tsang, On fuzzy approximation operators in attribute reduction with fuzzy rough sets, Information Sciences 178 (2008) 3163–3176.
[50] W.X. Zhang, W.Z. Wu, J.Y. Liang, D.Y. Li, The theory and methodology of rough Sets (in Chinese), The Science Press, Beijing, 2001.
[51] J.Y. Zhao, Z.L. Zhang, Fuzzy-rough data reduction based on information entropy, in: Proceedings of the ICMLC'07, 2007, pp.3708–3712.
[52] S.Y. Zhao, E.C.C. Tsang, X.Z. Wang, D.G. Chen, D.S. Yeung, Fuzzy matrix computation for fuzzy information system to reduce attributes, in: Proceedings of the ICMLC'06, IEEE Press, 2006, pp. 2300–2304.
[53] L. Zhou, W.Z. Wu, On generalized intuitionistic fuzzy rough approximation operators, Information Sciences 178 (2008) 2448–2465.
[54] W. Zhu, W. Zhang, Y. Fu, An incomplete data analysis approach using rough set theory, Proceedings of the International Conference on Intelligent Mechatronics and Automation (2004) 332–338.
[55] <http://www.ics.uci.edu/~mlearn/MLRepository.html>.