Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 177 (2007) 4493–4514
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Learning fuzzy rules from fuzzy samples based on rough set technique ☆

Xizhao Wang [a], Eric C.C. Tsang [b], Suyun Zhao [a,b,*], Degang Chen [c], Daniel S. Yeung [b]

[a] *Department of Mathematics and Computer Science, Hebei University, Baoding, Hebei 071002, China*
[b] *Department of Computing, Hong Kong Polytechnic University, Hong Hom, Kowloon, Hong Kong*
[c] *Department of Mathematics and Physics, North China Electric Power University, Beijing, China*

Received 8 November 2004; received in revised form 6 April 2007; accepted 16 April 2007

Abstract

Although the traditional rough set theory has been a powerful mathematical tool for modeling incompleteness and vagueness, its performance in dealing with initial fuzzy data is usually poor. This paper makes an attempt to improve its performance by extending the traditional rough set approach to the fuzzy environment. The extension is twofold. One is knowledge representation and the other is knowledge reduction. First, we provide new definitions of fuzzy lower and upper approximations by considering the similarity between the two objects. Second, we extend a number of underlying concepts of knowledge reduction (such as the reduct and core) to the fuzzy environment and use these extensions to propose a heuristic algorithm to learn fuzzy rules from initial fuzzy data. Finally, we provide some numerical experiments to demonstrate the feasibility of the proposed algorithm. One of the main contributions of this paper is that the fundamental relationship between the reducts and core of rough sets is still pertinent after the proposed extension.

Keywords: Fuzzy rough sets; Knowledge discovery; Knowledge reduction; Fuzzy reduct; Fuzzy core

1. Introduction

Rough set theory proposed by Pawlak [16], as the traditional rough set approach, is a useful mathematical tool for describing and modeling incomplete and insufficient information. It has been widely applied in many fields such as Machine Learning, Data Mining, Pattern Recognition, Fault Diagnostics, etc. During the last decade, a number of generalizations of the traditional rough set approach, such as generalizations of

☆ This work is supported by the National Natural Science Foundation of China (No. 60473045) and is also supported by the Hong Kong Research Grant Council under Grant B-Q826.
* Corresponding author. Address: Department of Mathematics and Computer Science, Hebei University, Baoding, Hebei 071002, China. Tel.: +86 312 5079638; fax: +86 312 5079638.
E-mail addresses: xizhaowang@ieee.org (X. Wang), zhaosy@mail.hbu.edu.cn (S. Zhao).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2007.04.010

approximation spaces, concept approximations, combination of rough sets and fuzzy sets, etc., have been proposed. Interested readers may consult Ref. [18] for a summary of these generalizations. The focus of this paper is the generalization of combining rough sets and fuzzy sets [2–6,8,9,11,12,14,15,19,21,23–26,29–33,35–38].

In the traditional rough set approach, the values of attributes are assumed to be nominal data, i.e. symbols. In many applications, however, the attribute-values can be linguistic terms (i.e. fuzzy sets). For example, the attribute ''height'' may be given the values of ''high'', ''mid'', and ''low''. The traditional rough set approach would treat these values as symbols, thereby some important information included in these values such as the partial ordering and membership degrees is ignored, which means that the traditional rough sets approach cannot effectively deal with fuzzy initial data (e.g. linguistic terms) [27]. In recent years, many models have been proposed for generalizing rough sets to the fuzzy environment. Roughly speaking, these generalization models can be classified into two types. One type of model involves the generalization of set approximation, which proposes generalizations of lower and upper approximations to approximate the fuzzy subset of a universe [3–5,14,15,23–25,29–33,35–38]. The other type of model involves fuzzy generalization of knowledge reduction, which use fuzzy rough technique to reduce fuzzy attributes and learn fuzzy rules from fuzzy samples [1,8,9,12,21,27].

Set approximation generalization models have been the focus of recent work in generalizing rough sets into the fuzzy environment. Again, two main approaches, the constructive and axiomatic, exist. The constructive approach replaces the equivalence relation (i.e. indiscernibility relation) with fuzzy similarity relation [1,6,12,15,21,22,35], fuzzy partition [4,5,35] and covers [36–38], and uses these replacements to construct the generalized set approximation. The replacement of the equivalence relation is the primary notion in the constructive approach. Unlike the constructive approach, the axiomatic approach takes the set approximation operator as the primary notion. It focuses on studying the mathematical structure of fuzzy rough sets. Moris and Yakout first generalized the axiomatic approach to fuzzy rough sets, but their work was restricted to the fuzzy T-rough set defined by fuzzy T similarity relation and the lower and upper approximation operators proposed by them are not dual. Therefore, Mi and Wu et al. [14,29,30,35] improved their generalization and proposed fuzzy rough sets defined by fuzzy binary relation. They also presented two pairs of dual generalized lower and upper approximation operators. Yeung et al. [33] further developed a unified framework of fuzzy rough sets by combining constructive and axiomatic approaches.

Comparatively, knowledge reduction is a less-studied model relative to the set approximation model in fuzzy rough sets [1,8,9,12,21,22,27]. Again, there are two main approaches to reduce knowledge. One approach is attribute reduction while the other is rule induction. Attribute reduction with the fuzzy rough set technique has been mentioned by some researchers. Pioneering work on attribute reduction with the fuzzy rough sets technique has been proposed [12,21]. The attribute reduction algorithm performed well on practical datasets (e.g. web categorization) [12,21], yet due to its poorly designed termination criteria, it did not converge on many real datasets [1]. A further drawback is that it is inefficient on large dimensional problems because the time complexity of the algorithm increases exponentially with the number of attributes. It has been suggested that the computational efficiency could be improved if the fuzzy rough sets were placed on a compact computational domain [1]. Despite this, it should be noted that none of these works offers a clear definition of attribute reduction or structure for attribute reduction based on fuzzy rough sets. Recently, the formal concept of attribute reduction with fuzzy rough sets has been proposed and an algorithm using a discernibility matrix for computing all attribute reductions has been designed [26]. This work was restricted to the fuzzy T similarity relationship defined by 'min' norm because the reasoning process relies on the 'min' norm.

Only a limited number of rule induction methods using fuzzy rough set technique to learn rules from fuzzy samples, have been proposed [8,9,22]. They are summarized as follows. Slowinski and Vanderpooten [22] induced decision rules by using a similarity-based approximation, but they did not mention the underlying concept of knowledge reduction and only proposed an induction algorithm. Hong et al. [8,9] also proposed algorithms to produce a set of maximally general fuzzy rules from noisy quantitative training data by applying the variable precision rough set model. They proposed the approximation of fuzzy lower and upper but neglected to mention the concepts of either the core or the reduct, which are the most fundamental concepts of knowledge reduction. Wang and Hong [27] proposed first transforming the fuzzy values to crisp values and then computing the corresponding reducts and core of rules. Some information hidden in fuzzy values, such as partial ordering relation and membership degree, is lost.

In this paper, we attempt to improve the performance of a fuzzy rough set approach by generalizing the underlying concepts of knowledge reduction, such as the reduct and core, to the fuzzy environment. We propose an approach, not only to reducing fuzzy attributes, but also to inducing fuzzy rules. First, we give some new definitions of fuzzy lower and upper approximations by considering the similarity between two objects. Second, we extend a number of underlying concepts of knowledge reduction (such as the reduct and core) to the fuzzy environment and use these extensions to propose a heuristic algorithm to learn fuzzy rules from initial fuzzy data. Finally, we provide some numerical experiments to demonstrate the feasibility of the proposed algorithm.

The rest of this paper is organized as follows: Section 2 outlines the fuzzy information system and proposes the fuzzy indiscernibility relation. Section 3 describes our proposed set approximation model. Section 4 extends a number of basic concepts of knowledge reduction, such as the reduct and core, to the fuzzy environment. In Section 5, we provide the algorithms for reducing attribute and learning a set of fuzzy decision rules, we also analyze the time complexity of the algorithms. In Section 6 some numerical experiments are reported to show the feasibility of the proposed approach. The last section concludes this paper.

2. Fuzzy information system

In this section, we describe the fuzzy information system or fuzzy decision table, which describes the data format in this paper and define some notations. We also describe and exemplify the notion of the fuzzy indiscernibility relation, which is the primary notion in our work.

2.1. Fuzzy information system and notations

The fuzzy information system [28] represents the formulation of a problem with fuzzy samples (samples containing fuzzy representation). Consider a set of samples $\{x_1, x_2, \ldots, x_n\}$ which is regarded as a universe of discourse U. Let $\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, \ldots, \widetilde{A}^{(m)}$ and $\widetilde{A}^{(m+1)}$ be a set of fuzzy attributes where $\widetilde{A}^{(m+1)}$ denotes a fuzzy decision attribute. Each fuzzy attribute $\widetilde{A}^{(j)}$ consists of a set of linguistic terms $F(\widetilde{A}^{(j)}) = \{\widetilde{F}_1^{(j)}, \widetilde{F}_2^{(j)}, \ldots, \widetilde{F}_{p_j}^{(j)}\}$ $(j = 1, 2, \ldots, m+1)$. All linguistic terms are defined in the same universe of discourse U. The value of the ith sample x_i with respect to the jth attribute $\widetilde{A}^{(j)}$, denoted by u_{ij}, is a fuzzy set defined in $F(\widetilde{A}^{(j)})$. In other words, the fuzzy set u_{ij} has a form of $u_{ij} = u_{ij}^{(1)}/\widetilde{F}_1^{(j)} + u_{ij}^{(2)}/\widetilde{F}_2^{(j)} + \cdots + u_{ij}^{(p_j)}/\widetilde{F}_{p^j}^{(j)}$ where $u_{ij}^{(k)}$ $(k = 1, 2, \ldots, p_j)$ denotes the corresponding membership degree. To demonstrate the fuzzy information system, we consider data in Table 1, which describes a small training set with fuzzy samples [27]. The universe of discourse is $U = \{x_1, x_2, \ldots, x_{14}\}$. There are five fuzzy attributes: $\widetilde{A}^{(1)} = outlook$, $\widetilde{A}^{(2)} = temperature$, $\widetilde{A}^{(3)} = humidity$,

Table 1
A fuzzy information table

e_j	Class											
	Outlook			Temperature			Humidity		Windy		Class	
	Sunny	Overcast	Rain	Hot	Mild	Cool	High	Normal	False	True	Positive	Negative
$J = 1$	0.9	0.1	0.0	0.9	0.1	0.0	0.8	0.2	0.7	0.4	0.4	0.7
$J = 2$	0.9	0.1	0.1	0.8	0.2	0.1	0.9	0.2	0.1	0.8	0.3	0.7
$J = 3$	0.1	0.9	0.2	0.9	0.1	0.1	0.9	0.1	0.9	0.1	0.8	0.3
$J = 4$	0.0	0.1	0.9	0.1	0.9	0.0	0.6	0.5	0.8	0.3	0.6	0.5
$J = 5$	0.1	0.0	0.9	0.0	0.1	0.9	0.0	1	0.8	0.2	0.9	0.2
$J = 6$	0.1	0.1	0.9	0.0	0.2	0.9	0.1	0.9	0.1	0.9	0.3	0.8
$J = 7$	0.0	1.0	0.0	0.0	0.1	0.9	0.1	0.9	0.2	0.9	0.9	0.3
$J = 8$	0.9	0.1	0.0	0.3	0.9	0.1	0.9	0.1	1.0	0.0	0.2	0.9
$J = 9$	0.8	0.2	0.0	0.0	0.4	0.6	0.0	1.0	1.0	0.0	0.9	0.2
$J = 10$	0.0	0.1	0.9	0.0	1.0	0.0	0.0	1.0	0.9	0.1	0.6	0.5
$J = 11$	0.9	0.1	0.0	0.0	0.9	0.1	0.1	0.9	0.0	1.0	0.8	0.3
$J = 12$	0.0	1.0	0.0	0.1	0.9	0.0	1.0	0.0	0.0	1.0	0.7	0.4
$J = 13$	0.0	0.9	0.1	1.0	0.0	0.0	0.0	1.0	0.9	0.1	0.7	0.2
$J = 14$	0.0	0.1	0.9	0.0	0.9	0.1	0.9	0.1	0.0	1.0	0.1	0.9

$\widetilde{A}^{(4)} = windy$ and $\widetilde{A}^{(5)} = class$. Fuzzy Information System with decision attributes is also called Fuzzy Decision Table.

In the following sections, we will use the following notations. Let $S = (U, \widetilde{A})$ be a fuzzy information system where $U = \{x_1, x_2, \ldots, x_n\}$ denotes the universe of discourse and $\widetilde{A}$ denotes the set of fuzzy attributes. $\widetilde{A}^{(j)}$ $(1 \leqslant j \leqslant m+1)$ denotes the jth fuzzy attribute and $\widetilde{F}_k^{(j)}$ denotes the kth linguistic term (fuzzy set) of fuzzy attribute $\widetilde{A}^{(j)}$. Moreover, for a matrix A, we will use max(A) to denote the maximum element of A and min(A) to denote the minimum element of A.

2.2. Fuzzy indiscernibility relation

The indiscernibility relation (i.e. the equivalence relation) is a key and primitive concept of rough set theory. In this section, we define and exemplify the extension of indiscernibility relation. This extension is called fuzzy indiscernibility relation (fuzzy similarity relation).

Definition 1. $\forall \widetilde{B} \subseteq \widetilde{A}$, $\widetilde{R}(\widetilde{B})$ is the fuzzy indiscernibility relation on $\widetilde{B}$ if and only if $\widetilde{R}(\widetilde{B}) = \left\{\frac{\alpha}{(x_i, x_j)} | (x_i, x_j) \in U \times U, \alpha = \min_{\widetilde{A}^{(k)} \in \widetilde{B}} \widetilde{A}^{(k)}(x_i, x_j)\right\}$, where $\widetilde{A}^{(k)}(x_i, y_j)$ is the similarity degree of two samples x_i, x_j on $\widetilde{A}^{(k)}$.

Then α is called indiscernibility degree on $\widetilde{B}$ of x_i, x_j, denoted by $r_{ij} = \alpha = \widetilde{R}(\widetilde{B})(x_i, x_j),\ r_{ij} \in [0, 1]$.

According to Definition 1, the fuzzy indiscernibility relation $\widetilde{R}(\widetilde{B})$ is a fuzzy similarity relation. $\widetilde{R}(\widetilde{B})$ can also be represented by a fuzzy similarity matrix as follows:

$$\widetilde{R}(\widetilde{B}) = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

We set that $\widetilde{R}(\Phi) = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$ if $\widetilde{B} = \Phi$. This is because we cannot distinguish any samples of the universe of discourse (and therefore we have to regard them as indistinguishable) when the set of fuzzy attributes is empty.

From the definition of fuzzy indiscernibility relation, it is easy to see that if $\forall \widetilde{B} \subseteq \widetilde{A}$, then the element of fuzzy matrix $\widetilde{R}(\widetilde{B})$ is greater than or equal to the corresponding element of fuzzy matrix $\widetilde{R}(\widetilde{A})$, i.e.min$(\widetilde{R}(\widetilde{B}) - \widetilde{R}(\widetilde{A})) \geqslant 0$ or $\widetilde{R}(\widetilde{B}) \geqslant \widetilde{R}(\widetilde{A})$.

The following provides an example of the fuzzy indiscernibility relation:

Example 1. Let us consider Table 2, which is a part of Table 1 (with minor modification). Suppose $S' = (U', \widetilde{A})$, where $U' = \{x_1, \ldots, x_8\}$ is the set of eight samples, every sample is described by a set of fuzzy

Table 2
Subset of training set of learning from fuzzy samples

e_j	Class									
	Outlook			Temperature			Humidity		Windy	
	Sunny	Overcast	Rain	Hot	Mild	Cool	High	Normal	False	True
$J = 1$	0.9	0.1	0.0	0.9	0.1	0.0	0.8	0.2	0.7	0.4
$J = 2$	0.9	0.1	0.1	0.8	0.2	0.1	0.9	0.2	**0.8**	**0.3**
$J = 3$	0.1	0.9	0.2	0.9	0.1	0.1	0.9	0.1	0.9	0.1
$J = 4$	0.0	0.1	0.9	0.1	0.9	0.0	0.6	0.5	0.8	0.3
$J = 5$	0.1	0.0	0.9	0.0	0.1	0.9	0.0	1	0.8	0.2
$J = 6$	0.1	0.1	0.9	0.0	0.2	0.9	0.1	0.9	0.1	0.9
$J = 7$	0.0	1.0	0.0	0.0	0.1	0.9	0.1	0.9	0.2	0.9
$J = 8$	0.9	0.1	0.0	0.3	0.9	0.1	0.9	0.1	1.0	0.0

attributes $\widetilde{A} = \{\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, \ldots, \widetilde{A}^{(4)}\}$: $\widetilde{A}^{(1)} = outlook$, $\widetilde{A}^{(2)} = temperature$, $\widetilde{A}^{(3)} = humidity$, $\widetilde{A}^{(4)} = windy$, the membership degrees of every sample are given in Table 2.

Fuzzy indiscernibility relation is calculated as follows: Assume that the function $s_{i,j} = \begin{cases} 1 & i = j \\ \min_{k=1}^{m}(1 - |t_{ik} - t_{jk}|) & i \neq j \end{cases}$ is given as the similarity measure. Here $\{t_1, \ldots, t_n\}$ is the set of the samples, and every sample t_i is described by $\{t_{i1}, \ldots, t_{im}\} \in R^+$, $(i = 1, \ldots, n)$.

Then the similarity degree of the samples x_1 and x_2 on the attribute $\widetilde{A}^{(1)}$ can be calculated as follows:

$$\widetilde{A}^{(1)}(x_1, x_2) = (1 - |0.9 - 0.9|) \wedge (1 - |0.1 - 0.1|) \wedge (1 - |0 - 0.1|) = 0.9$$

Considering the samples x_1 and x_2, the indiscernibility degree can be calculated as follows:

$$r_{12} = \min_{\widetilde{A}^{(k)} \in \widetilde{A}} (\widetilde{A}^{(k)}(x_1, x_2)) = \min\{0.9, 0.9, 0.9, 0.9\} = 0.9$$

Similarly,

$$\begin{aligned} r_{13} &= \min_{\widetilde{A}^{(k)} \in \widetilde{A}} (\widetilde{A}^{(k)}(x_1, x_3)) = \min\{0.2, 0.9, 0.9, 0.7\} = 0.2 \\ r_{23} &= \min_{\widetilde{A}^{(k)} \in \widetilde{A}} (\widetilde{A}^{(k)}(x_2, x_3)) = \min\{0.2, 0.9, 0.9, 0.2\} = 0.2 \\ &\vdots \qquad\qquad\qquad \vdots \\ r_{78} &= \min_{\widetilde{A}^{(k)} \in \widetilde{A}} (\widetilde{A}^{(k)}(x_8, x_9)) = \min\{0.1, 0.2, 0.2, 0.1\} = 0.1 \end{aligned}$$

Thus, the fuzzy similarity matrix (i.e. the fuzzy indiscernibility relation) can be obtained as follows:

$$\widetilde{R}(\widetilde{A}) = \begin{bmatrix} 1 & 0.9 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 & 0.2 \\ 0.9 & 1 & 0.2 & 0.1 & 0.1 & 0.2 & 0.1 & 0.3 \\ 0.2 & 0.2 & 1 & 0.2 & 0.1 & 0.1 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.2 & 1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1 & 0.2 & 0 & 0.1 \\ 0.1 & 0.2 & 0.1 & 0.1 & 0.2 & 1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0 & 0.1 & 1 & 0.1 \\ 0.2 & 0.3 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 & 1 \end{bmatrix}$$

3. Approximation of set

In this section, we describe and exemplify the basic concepts of knowledge representation in our proposed fuzzy rough approach, specifically the lower and upper approximations, fuzzy rough set, positive region, negative region and boundary region.

After considering the similarity between two fuzzy samples, we redefine the lower and upper approximations as follows:

Definition 2. $\forall M \subset U$, the β-upper approximation of M is defined as

$$\widetilde{\widetilde{M}}_\beta = \{x \in U | \exists y \in M, \text{ such that } \widetilde{R}(\widetilde{A})(x, y) \geqslant \beta\}$$

If $\beta = 1$, the 1-upper approximation of M is $\widetilde{\widetilde{M}}_1 = \{x \in U | \exists y \in M, \text{ such that } \widetilde{R}(\widetilde{A})(x, y) = 1\}$. This is equivalent to the formula $\widetilde{\widetilde{M}}_1 = \{x \in U | [x]_{\widetilde{R}(\widetilde{A})} \cap M \neq \Phi\}$ (where $\widetilde{R}(\widetilde{A})$ is an equivalence relation and $[x]_{\widetilde{R}(\widetilde{A})}$ is one equivalence class of $\widetilde{R}(\widetilde{A})$). This shows that the 1-upper approximation of M is identical to the traditional upper approximation.

Definition 3. $\forall M \subset U$, the β-lower approximation of M is defined as

$$\underline{\widetilde{M}}_\beta = \{x \in M | \widetilde{R}(\widetilde{A})(x,y) \leqslant 1-\beta, \forall y \in U - M\}$$

Likewise, if $\beta = 1$ the 1-lower approximation of M is $\underline{\widetilde{M}}_1 = \{x \in M | \widetilde{R}(\widetilde{A})(x,y) = 0, \forall y \in U - M\}$. This is equivalent to the formula $\underline{\widetilde{M}}_1 = \{x \in M | [x]_{\widetilde{R}(\widetilde{A})} \subseteq M\}$. This shows that the 1-lower approximation of M is identical to the traditional lower approximation.

Using this pair of fuzzy lower and upper approximations $\langle \widetilde{\overline{M}}_\beta, \underline{\widetilde{M}}_\beta \rangle$, we can represent any subset M of the universe of discourse U.

Definition 4. $\forall M_1, M_2 \subset U$, if the following two formulae hold:

$$\forall x \in M_1,\ \exists y \in M_2,\ \text{such that}\ \widetilde{R}(\widetilde{A})(x,y) \geqslant \alpha \tag{1}$$

$$\forall x \in M_2,\ \exists y \in M_1,\ \text{such that}\ \widetilde{R}(\widetilde{A})(x,y) \geqslant \alpha \tag{2}$$

then we say that M_1 is α-approximately equal to M_2, denoted by $M_1 \overset{\alpha}{\cong} M_2$.

Definition 5. If $\widetilde{\overline{M}}_\beta \overset{\alpha}{\cong} \underline{\widetilde{M}}_\beta$, then we say that M is β-definable in α-approximation. Otherwise, we say that M is β-undefinable in α-approximation, and M is a *fuzzy-rough set*.

It is easy to see that if $\beta = \alpha = 1$ fuzzy-rough set proposed in this paper degenerates into the traditional rough set.

Definition 6. $\underline{\widetilde{M}}_\beta$ is called the *β-positive region* of M in $S = (U, \widetilde{A})$, $U - \widetilde{\overline{M}}_\beta$ is called the *β-negative region* of M in $S = (U, \widetilde{A})$ and $\widetilde{\overline{M}}_\beta - \underline{\widetilde{M}}_\beta$ is called the *β-boundary region* of M in $S = (U, \widetilde{A})$.The following provide an example of the definitions of the set approximation.

Example 2. Let us consider Tables 2 and 3. Here Table 3 is the discrete case of Table 2. Suppose that $M = \{x_1, x_3, x_4\}$ is the subset of U'.

(1) In Table 2, find the 0.9-lower and 0.9-upper approximations of M, the 0.9-positive region of M, the 0.9-negative region of M and 0.9-boundary region of M.According to Definition 2, the 0.9-upper approximation of M can be obtained as follows:

$$\widetilde{\overline{M}}_{0.9} = \{x \in U | \exists y \in M,\ \text{such that}\ \widetilde{R}(\widetilde{A})(x,y) \geqslant 0.9\} = \{x_1, x_2, x_3, x_4\}$$

According to Definition 3, the 0.9-lower approximation of M can be obtained as follows:

$$\underline{\widetilde{M}}_{0.9} = \{x \in M | \widetilde{R}(\widetilde{A})(x,y) \leqslant 0.1, \forall y \in U - M\} = \{x_4\}$$

Furthermore, $\underline{\widetilde{M}}_{0.9} = \{x_4\}$ is the 0.9-positive region of M, $U - \widetilde{\overline{M}}_{0.9} = \{x_5, \ldots, x_8\}$ is the 0.9-negative region of M and $\widetilde{\overline{M}}_{0.9} - \underline{\widetilde{M}}_{0.9} = \{x_1, x_2, x_3\}$ is the 0.9-boundary region of M in $S' = (U', \widetilde{A})$.

Table 3
The crisp data transformed from the data in Table 2

e_j	Class									
	Outlook			Temperature			Humidity		Windy	
	Sunny	Overcast	Rain	Hot	Mild	Cool	High	Normal	False	True
$J = 1$	1	0	0	1	0	0	1	0	1	0
$J = 2$	1	0	0	1	0	0	1	0	**1**	**0**
$J = 3$	0	1	0	1	0	0	1	0	1	0
$J = 4$	0	0	1	0	1	0	1	0	1	0
$J = 5$	0	0	1	0	0	1	0	1	1	0
$J = 6$	0	0	1	0	0	1	0	1	0	1
$J = 7$	0	1	0	0	0	1	0	1	0	1
$J = 8$	1	0	0	0	1	0	1	0	1	0

(2) In Table 3, find the 1-lower and 1-upper approximations of M, the traditional lower and upper approximations.In Table 3, the fuzzy similarity matrix (i.e. fuzzy indiscernibility relation) degenerates into a Boolean matrix as follows:

$$R(\widetilde{A}) = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

According to Definitions 2 and 3, the 1-upper and 1-lower approximations of M can be obtained from Table 3. They are respectively $\overline{\widetilde{M}}_1 = \{x_1, x_2, x_3, x_4\}$ and $\underline{\widetilde{M}}_1 = \{x_3, x_4\}$.The traditional upper and lower approximations of M are respectively $\{x_1, x_2, x_3, x_4\}$ and $\{x_3, x_4\}$.This example shows that the 1-lower and 1-upper approximations of M are identical to the traditional lower and upper approximations, respectively.

4. Fuzzy reduct and fuzzy core

On the basis of the fuzzy indiscernibility relation, we extend the basic concepts of knowledge reduction, such as the reduct and core, to the fuzzy environment.

4.1. Fuzzy reduct and fuzzy core

First, we extend the concept of indispensability, based on which the concepts of the fuzzy reduct and core can be proposed.

Definition 7. $\forall \widetilde{A}^{(j)} \in \widetilde{B}$, let $\beta = \max(\widetilde{R}(\widetilde{B} - \widetilde{A}^{(j)}) - \widetilde{R}(\widetilde{B}))$, then we say that $\widetilde{A}^{(j)}$ is β-indispensable in $\widetilde{B}$, or we say that the indispensability degree of $\widetilde{A}^{(j)}$ in $\widetilde{B}$ is β.

Definition 8. For $\widetilde{B}' \subseteq \widetilde{B}$ and the given threshold $\beta, \beta \in [0,1]$, if the following two formulae hold:

$$\max(\widetilde{R}(\widetilde{B}') - \widetilde{R}(\widetilde{B})) \leqslant 1 - \beta \tag{3}$$

$$\forall \widetilde{A}^{(j)} \in \widetilde{B}', \max(\widetilde{R}(\widetilde{B}' - \widetilde{A}^{(j)}) - \widetilde{R}(\widetilde{B}')) > 1 - \beta \tag{4}$$

Then $\widetilde{B}'$ is called the β-reduct of $\widetilde{B}$, denoted by $Reduct^{\beta}(\widetilde{B})$.

Note that β-reduct of $\widetilde{B}$ is not unique.

Definition 9. Given the threshold β, $\beta \in [0,1]$, the set $\widetilde{P}$ consists of the fuzzy attributes whose indispensability degrees in $\widetilde{B}$ are greater than $1 - \beta$. Then $\widetilde{P}$ is called the β-core of $\widetilde{B}$, denoted by $Core^{\beta}(\widetilde{B})$.

In the following, we discuss the threshold β in definitions 8 and 9. If $\beta = 1$, definitions 8 and 9 degenerate into the following:

Definition 8A. For $\widetilde{B}' \subseteq \widetilde{B}$, if the following two formulae hold:

$$\max(\widetilde{R}(\widetilde{B}') - \widetilde{R}(\widetilde{B})) \leqslant 0 \tag{5}$$

$$\forall \widetilde{A}^{(j)} \in \widetilde{B}', \max(\widetilde{R}(\widetilde{B}' - \widetilde{A}^{(j)}) - \widetilde{R}(\widetilde{B})) > 0 \tag{6}$$

Then $\widetilde{B}'$ is called the 1-reduct of $\widetilde{B}$, denoted by $Reduct^{1}(\widetilde{B})$.

Definition 9A. The set $\widetilde{P}$ is composed of fuzzy attributes whose indispensability degrees in $\widetilde{B}$ are greater than 0. Then $\widetilde{P}$ is called the 1-core of $\widetilde{B}$, denoted by $Core^1(\widetilde{B})$.

In Definition 8A, the formula (5) is equivalent to the formula $ind(\widetilde{B}') = ind(\widetilde{B})$, and the formula (6) corresponds to the fact that the set $\widetilde{B}'$ is independent. All these show that if $\beta = 1$ Definition 8 degenerates into the traditional definition of the reduct.

Likewise, if $\beta = 1$, Definition 9 degenerates into the traditional definition of the core.

Furthermore, we set that $\beta > \min(\widetilde{R}(\widetilde{B}))$,

If $\beta \leqslant \min(\widetilde{R}(\widetilde{B}))$, then we have

$$\max(\widetilde{R}(\widetilde{B}') - \widetilde{R}(\widetilde{B})) \leqslant 1 - \min(\widetilde{R}(\widetilde{B})) \tag{7}$$

$$\forall \widetilde{A}^{(j)} \in \widetilde{B}, \quad \max(\widetilde{R}(\widetilde{B}' - \widetilde{A}^{(j)}) - \widetilde{R}(\widetilde{B})) \leqslant 1 - \min(\widetilde{R}(\widetilde{B})) \tag{8}$$

$$1 - \min(\widetilde{R}(\widetilde{B})) \leqslant 1 - \beta \tag{9}$$

From the formulae (7)–(9), it is easy to see that formula (3) always holds, whereas formula (4) is always wrong. In this case, no subset of the fuzzy attributes set $\widetilde{B}$ satisfying Definition 8 exists. In other words, for the set of fuzzy attributes $\widetilde{B}$, no two samples can be distinguished if the threshold $\beta \leqslant \min(\widetilde{R}(\widetilde{B}))$.

The fuzzy reduct and fuzzy core satisfy the following theorem:

Theorem 1. *For an arbitrary threshold* $\beta > \min(\widetilde{R}(B)), \beta \in [0, 1]$, *the formula* $Core^\beta(\widetilde{B}) = \cap Reduct^\beta(\widetilde{B})$ *always holds.*

Proof. First we prove $Core^\beta(\widetilde{B}) \supseteq (\cap Reduct^\beta(\widetilde{B}))$. We will complete the proof by contradiction.

Suppose that $\exists \widetilde{A}^{(k)} \in \cap Reduct^\beta(\widetilde{B})$, and $\widetilde{A}^{(k)} \notin Core^\beta(\widetilde{B})$. According to Definition 9, we obtain $\max(\widetilde{R}(\widetilde{B} - \widetilde{A}^{(k)}) - \widetilde{R}(\widetilde{B})) \leqslant 1 - \beta$. Then the subset of the fuzzy attributes set $\widetilde{B}' \subseteq \widetilde{B} - \widetilde{A}^{(k)}$ must exist, where $\widetilde{B}'$ satisfies $\max(\widetilde{R}(\widetilde{B}') - \widetilde{R}(\widetilde{B})) \leqslant 1 - \beta$, and $\forall \widetilde{B}^{(j)} \in \widetilde{B}'$, the formula $\max(\widetilde{R}(\widetilde{B}' - \widetilde{B}^{(j)}) - \widetilde{R}(\widetilde{B}')) > 1 - \beta$ always holds. This shows that $\widetilde{B}'$ is the β-reduct of $\widetilde{B}$ and $\widetilde{A}^{(k)} \notin \widetilde{B}'$. This result contradicts our supposition. We thus find that if $\widetilde{A}^{(k)} \in \cap Reduct^\beta(\widetilde{B})$, then $\widetilde{A}^{(k)} \in Core^\beta(\widetilde{B})$, i.e. $Core^\beta(\widetilde{B}) \supseteq (\cap Reduct^\beta(\widetilde{B}))$.

Next, we prove $Core^\beta(\widetilde{B}) \subseteq (\cap Reduct^\beta(\widetilde{B}))$. That is to say we need to prove that if $\widetilde{A}^{(k)} \notin \cap Reduct^\beta(\widetilde{B})$, then $\widetilde{A}^{(k)} \notin Core^\beta(\widetilde{B})$.

Suppose that $\widetilde{B}'$ is β-reduct of $\widetilde{B}$. If $\widetilde{B}' = \widetilde{B}$, then $\widetilde{B}$ has only one β-reduct. According to the definitions of fuzzy reduct and fuzzy core, we know that $\widetilde{B}$ is also the β-core of $\widetilde{B}$, i.e. $Core^\beta(\widetilde{B}) = \cap Reduct^\beta(\widetilde{B})$. If $\widetilde{B}' \subset \widetilde{B}$ and $\widetilde{A}^{(a)} \in \widetilde{B} - \widetilde{B}'$, then we have $\widetilde{A}^{(a)} \notin \cap Reduct^\beta(\widetilde{B})$. According to the definition of fuzzy indiscernibility relation, we know that if $\widetilde{B}' \subseteq \widetilde{B} - \{\widetilde{A}^{(a)}\}$ and $\widetilde{B}' \subseteq \widetilde{B}$, then we have

$$\min(\widetilde{R}(\widetilde{B}') - \widetilde{R}(\widetilde{B} - \{\widetilde{A}^{(a)}\})) \geqslant 0 \tag{10}$$

$$\max(\widetilde{R}(\widetilde{B}') - \widetilde{R}(\widetilde{B})) \leqslant 1 - \beta \tag{11}$$

From the formulae (10) and (11), we know that the formula $\max(\widetilde{R}(\widetilde{B} - \{\widetilde{A}^{(a)}\}) - \widetilde{R}(\widetilde{B})) \leqslant 1 - \beta$ holds, i.e. $\widetilde{A}^{(a)} \notin Core^\beta(\widetilde{B})$. This shows that $Core^\beta(\widetilde{B}) \subseteq (\cap Reduct^\beta(\widetilde{B}))$. □

The significance of attributes is also an important concept in knowledge reduction of rough set theory. In the following, the concept of the significance of the degree of fuzzy attributes is given:

Definition 10. For $\widetilde{B}' \subset \widetilde{B}$, let $\gamma = \frac{\sum_{i<j} r_{ij}}{n(n-1)/2}$, where $r_{ij} \in \widetilde{R}(\widetilde{B}')$, n is the size of the universe, then we say that the significance degree of $\widetilde{B}'$ in $\widetilde{B}$ is γ.

Example 3. Consider Table 2 again, suppose that $M = \{x_1, x_3, x_4\}$ is the subset of U'.

(1) Calculate fuzzy indiscernibility degree of $\widetilde{A}^{(1)}$ in $\widetilde{A}$.Fuzzy similarity matrix $\widetilde{R}(\widetilde{A} - \widetilde{A}^{(1)})$ can be obtained as follows:

$$\widetilde{R}(\widetilde{A}-\widetilde{A}^{(1)}) = \begin{bmatrix} 1 & 0.9 & 0.7 & 0.2 & 0.1 & 0.1 & 0.1 & 0.2 \\ 0.9 & 1 & 0.8 & 0.3 & 0.1 & 0.2 & 0.2 & 0.3 \\ 0.7 & 0.8 & 1 & 0.2 & 0.1 & 0.1 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.2 & 1 & 0.1 & 0.1 & 0.1 & 0.6 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1 & 0.3 & 0.3 & 0.1 \\ 0.1 & 0.2 & 0.1 & 0.1 & 0.3 & 1 & 0.9 & 0.1 \\ 0.1 & 0.2 & 0.1 & 0.1 & 0.3 & 0.9 & 1 & 0.1 \\ 0.2 & 0.3 & 0.2 & 0.6 & 0.1 & 0.1 & 0.1 & 1 \end{bmatrix}$$

According to Definition 7, the indiscernibility degree of $\widetilde{A}^{(1)}$, $\widetilde{A}^{(2)}$, $\widetilde{A}^{(3)}$ and $\widetilde{A}^{(4)}$ in $\widetilde{A}$ is as follows:

$$\beta = \max(\widetilde{R}(\widetilde{A}-\widetilde{A}^{(1)}) - \widetilde{R}(\widetilde{A})) = \max \begin{bmatrix} 0 & 0 & 0.5 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0.2 & 0 & 0 & 0.1 & 0 \\ 0.5 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 \\ 0 & 0.1 & 0 & 0 & 0.3 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 \end{bmatrix} = 0.6$$

In the same way, the indiscernibility degree of $\widetilde{A}^{(2)}$, $\widetilde{A}^{(3)}$ and $\widetilde{A}^{(4)}$ can be calculated. They are 0.4, 0.1 and 0, respectively.

(2) Find all 0.9-reducts and the 0.9-core of $\widetilde{A}$ in $S' = (U', \widetilde{A})$.
According to Definition 9, the 0.9-core of $\widetilde{A}$ is the subset of condition attributes: $\{\widetilde{A}^{(1)}, \widetilde{A}^{(2)}\}$.
Let $\widetilde{A}' = \{\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, \widetilde{A}^{(3)}\}$, then

$$\max(\widetilde{R}(\widetilde{A}') - \widetilde{R}(\widetilde{A}^{(4)})) = 0 < 0.1 \qquad \max(\widetilde{R}(\widetilde{A}'-\widetilde{A}^{(1)}) - \widetilde{R}(\widetilde{A})) = 0.9 > 0.1$$
$$\max(\widetilde{R}(\widetilde{A}'-\widetilde{A}^{(2)}) - \widetilde{R}(\widetilde{A})) = 0.7 > 0.1 \qquad \max(\widetilde{R}(\widetilde{A}'-\widetilde{A}^{(3)}) - \widetilde{R}(\widetilde{A})) = 0.6 > 0.1$$

According to Definition 8, the subset $\widetilde{A}' = \{\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, \widetilde{A}^{(3)}\}$ is 0.9-reduct of $\widetilde{A}$.

In the same way, all 0.9-reducts of $\widetilde{A}$ can be found. They are $\widetilde{A}' = \{\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, \widetilde{A}^{(3)}\}$ and $\widetilde{A}'' = \{\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, \widetilde{A}^{(4)}\}$. Their intersections are $\widetilde{A}' \cap \widetilde{A}'' = \{\widetilde{A}^{(1)}, \widetilde{A}^{(2)}\}$. This result demonstrates that the fundamental relation between reducts and core in rough set theory is maintained after the proposed extension.

4.2. Fuzzy relative reduct and fuzzy relative core

The concepts of fuzzy reduct and fuzzy core are limited to many applications such as fuzzy decision table. We, therefore, generalize them to fuzzy relative reduct and fuzzy relative core, which is also the generalization of the traditional relative reduct and relative core. In this section, we describe and exemplify basic concepts of fuzzy relative attribute reduction that is the inconsistence degree between two samples, the relative indispensability degree, fuzzy relative reduct and fuzzy relative core.

First, the inconsistence degree is given as follows:

Definition 11. For $x_i, x_j \in U$, let

$$t = \widetilde{R}(\widetilde{B})(x_i, x_j) - \widetilde{R}(\widetilde{A}^{(m+1)})(x_i, x_j), \quad r_{ij} = \begin{cases} t & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then r_{ij} is called the inconsistence degree of $x_i x_j$ on $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$.

Let $T_{\widetilde{A}^{(m+1)}}(\widetilde{B}) = (r_{ij})_{n\times n}$, then $T_{\widetilde{A}^{(m+1)}}(\widetilde{B})$ is called the inconsistence matrix of fuzzy condition attributes set $\widetilde{B}$ with respect to fuzzy decision attribute $\widetilde{A}^{(m+1)}$ on U.

In the following, we extend the concepts of relative indispensability, fuzzy relative reduct and core based on the fuzzy inconsistence degree.

Definition 12. $\forall \widetilde{A}^{(j)} \in \widetilde{B}$, let $\beta = \max(T_{\widetilde{A}^{(m+1)}}(\widetilde{B} - \{\widetilde{A}^{(j)}\}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B}))$, then we say that $\widetilde{A}^{(j)}$ is β-indispensable in $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$, or we say that the relative indispensability degree of $\widetilde{A}^{(j)}$ in $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$ is β.

Definition 13. For $\widetilde{B}' \subseteq \widetilde{B}$ and the given threshold β, $\beta \in [0,1]$, if the following two formulae hold:

$$\max(T_{\widetilde{A}^{(m+1)}}(\widetilde{B}') - T_{\widetilde{A}^{(m+1)}}(\widetilde{B})) \leqslant 1 - \beta \tag{12}$$

$$\forall \widetilde{A}^{(j)} \in \widetilde{B}', \quad \max(T_{\widetilde{a}^{(m+1)}}(\widetilde{B}' - \{\widetilde{A}^{(j)}\}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B})) > 1 - \beta \tag{13}$$

Then $\widetilde{B}'$ is called β-reduct of $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$, denoted by $Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$.

Note that the β-reduct of $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$ is not unique.

Definition 14. Given the threshold β, $\beta \in [0,1]$, the set $\widetilde{P}$ consists of fuzzy attributes whose relative indispensability degrees with respect to $\widetilde{A}^{(m+1)}$ are greater than $1-\beta$. Then $\widetilde{P}$ is called the β-core of $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$, denoted by $Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$.

Definition 15 gives the concept of the significance degree of fuzzy attributes with respect to fuzzy decision attribute.

Definition 15. For $\widetilde{B}' \subset \widetilde{B}$, let $\gamma = \frac{\sum_{i<j} t_{ij}}{n(n-1)/2}$, where $t_{ij} \in \widetilde{T}_{\widetilde{A}^{(m+1)}}(\widetilde{B}')$ and n is the size of the universe, we say that the significance degree of $\widetilde{B}'$ in $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$ is γ.

In the following, we analyze the threshold β in Definitions 13 and 14.

If $\beta = 1$, Definition 13 degenerates into the traditional definition of the relative reduct and Definition 14 degenerates into the traditional definition of relative core.

Likewise, we set that $\beta > \min(T_{\widetilde{A}^{(m+1)}}(\widetilde{B}))$.

Fuzzy relative reduct and fuzzy relative core satisfy the following theorem:

Theorem 2. *For an arbitrary threshold* $\beta > \min(T_{\widetilde{A}^{(m+1)}}(\widetilde{B}))$, $\beta \in [0,1]$, *the formula* $Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B}) = \cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$ *always holds.*

Proof. First we prove that $Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B}) \supseteq \cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$. We complete the proof by contradiction.

Suppose that $\exists \widetilde{A}^{(j)} \in \cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$, and $\widetilde{A}^{(j)} \notin Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$. According to Definition 13, it is easy to see that the formula $\max(T_{\widetilde{A}^{(m+1)}}(\widetilde{B} - \{\widetilde{A}^{(j)}\}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B})) \leqslant 1 - \beta$ holds. Then, the subset of fuzzy attributes set $\widetilde{Q} \subseteq \widetilde{B} - \{\widetilde{A}^{(j)}\}$ must exist, where $\widetilde{Q}$ satisfies $\max(T_{\widetilde{A}^{(m+1)}}(\widetilde{Q}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B})) \leqslant 1 - \beta$. And $\forall \widetilde{A}^{(k)} \in \widetilde{Q}$, the formula $\max(T_{\widetilde{A}^{(m+1)}}(\widetilde{Q} - \{\widetilde{A}^{(k)}\}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B})) > 1 - \beta$ always holds. This shows that $\widetilde{Q}$ is the β-reduct of $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$ and $\widetilde{A}^{j} \notin \widetilde{Q}$. This contradicts our supposition. Thus, we have shown that if $\widetilde{A}^{(j)} \in \cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$, then $\widetilde{A}^{(j)} \in Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$, i.e. $Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B}) \supseteq \cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$.

Next, we prove $Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B}) \subseteq \cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$. That is to say, we need to prove that if $\widetilde{A}^{(j)} \notin \cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$, then $\widetilde{A}^{(j)} \notin Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$.

Suppose that $\widetilde{Q}$ is β-reduct of $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$. Let $\widetilde{A}^{(j)} \in \widetilde{B} - \widetilde{Q}$, then we have $\widetilde{A}^{(j)} \notin \cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$. According to the definition of inconsistence matrix, we get

$$\max(T_{\widetilde{A}^{(m+1)}}(\widetilde{Q}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B})) \leqslant 1 - \beta \tag{14}$$

$$\min(T_{\widetilde{A}^{(m+1)}}(\widetilde{Q}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B} - \{\widetilde{A}^{(j)}\})) \geqslant 0 \tag{15}$$

From the formulae (14) and (15), the formula $\max(T_{\widetilde{A}^{(m+1)}}(\widetilde{B}-\{\widetilde{A}^{(j)}\})-T_{\widetilde{A}^{(m+1)}}(\widetilde{B}))\leqslant 1-\beta$ holds, i.e. $\widetilde{A}^{(j)}\notin Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$. This shows that $Core^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})\subseteq\cap Reduct^{\beta}_{\widetilde{A}^{(m+1)}}(\widetilde{B})$, which completes the proof. □

Example 4. Consider Table 4, which is the subset of training set of learning from fuzzy samples in Table 1. Suppose $U'=\{x_1,\ldots,x_8\}$ is the set of eight samples, every sample is described by a set of fuzzy attributes $\{\widetilde{B},\widetilde{A}^{(5)}\}$ where $\widetilde{B}=\{\widetilde{A}^{(1)},\widetilde{A}^{(2)},\ldots,\widetilde{A}^{(4)}\}$ is the set of fuzzy condition attributes, and $\widetilde{A}^{(5)}=class$ is fuzzy decision attribute. The membership degrees of every sample are given in Table 4.

(1) Calculate the inconsistent degrees of the sample pairs x_1 and x_2, x_2 and x_3 on $\widetilde{B}$ with respect to $\widetilde{A}^{(5)}$.

We can calculate that

$$t_{12}=\widetilde{R}(\widetilde{B})(x_1,x_2)-\widetilde{R}(\widetilde{A}^{(5)})(x_1,x_2)=0.9-0.6=0.3$$
$$t_{23}=\widetilde{R}(\widetilde{B})(x_2,x_3)-\widetilde{R}(\widetilde{A}^{(5)})(x_2,x_3)=0.2-0.9=-0.7$$

According to Definition 11, the inconsistence degree of x_1, x_2 on $\widetilde{B}$ with respect to $\widetilde{A}^{(5)}$ is $r_{12}=0.3$. The inconsistence degree of x_2, x_3 on $\widetilde{B}$ with respect to $\widetilde{A}^{(5)}$ is $r_{23}=0$.

(2) Calculate the relative indiscernibility degree of $\widetilde{A}^{(1)}$, $\widetilde{A}^{(2)}$, $\widetilde{A}^{(3)}$ and $\widetilde{A}^{(4)}$ in $\widetilde{B}$ with respect to $\widetilde{A}^{(5)}$.

The inconsistence matrix of fuzzy condition attributes set $\widetilde{B}$ with respect to fuzzy decision attribute $\widetilde{A}^{(5)}$ is

$$T_{\widetilde{A}^{(5)}}(\widetilde{B})=(r_{ij})_{8\times 8}\begin{bmatrix}0&0.3&0&0&0&0&0&0\\0.3&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\end{bmatrix}$$

According to Definition 12, the relative indispensability degree of $\widetilde{A}^{(1)}$ in $\widetilde{B}$ with respect to $\widetilde{A}^{(5)}$ is

$$\beta=\max(T_{\widetilde{A}^{(5)}}(\widetilde{B}-\{\widetilde{A}^{(1)}\})-T_{\widetilde{A}^{(5)}}(\widetilde{B}))=\max\begin{bmatrix}0&0&0.1&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0.1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0.5&0\\0&0&0&0&0&0.5&0&0\\0&0&0&0&0&0&0&0\end{bmatrix}=0.5$$

Table 4
Another subset of training set of learning from fuzzy samples

e_j	Class											
	Outlook			Temperature			Humidity		Windy		Class	
	Sunny	Overcast	Rain	Hot	Mild	Cool	High	Normal	False	True1	Positive	Negative
$J=1$	0.9	0.1	0.0	0.9	0.1	0.0	0.8	0.2	0.7	0.4	0.4	0.7
$J=2$	0.9	0.1	0.1	0.8	0.2	0.1	0.9	0.2	**0.8**	**0.3**	**0.7**	**0.3**
$J=3$	0.1	0.9	0.2	0.9	0.1	0.1	0.9	0.1	0.9	0.1	0.8	0.3
$J=4$	0.0	0.1	0.9	0.1	0.9	0.0	0.6	0.5	0.8	0.3	0.6	0.5
$J=5$	0.1	0.0	0.9	0.0	0.1	0.9	0.0	1	0.8	0.2	0.9	0.2
$J=6$	0.1	0.1	0.9	0.0	0.2	0.9	0.1	0.9	0.1	0.9	0.3	0.8
$J=7$	0.0	1.0	0.0	0.0	0.1	0.9	0.1	0.9	0.2	0.9	0.9	0.3
$J=8$	0.9	0.1	0.0	0.3	0.9	0.1	0.9	0.1	1.0	0.0	0.2	0.9

The relative indispensability degrees of $\widetilde{A}^{(2)}$, $\widetilde{A}^{(3)}$ and $\widetilde{A}^{(4)}$ in $\widetilde{B}$ with respect to $\widetilde{A}^{(5)}$ can be similarly calculated. They are 0.3, 0 and 0.5, respectively.

(3) Find one relative 0.9-reduct and relative 0.9-core of $\widetilde{B}$ with respect to $\widetilde{A}^{(5)}$.

According to Definition 14, the relative 0.9-core of $\widetilde{A}$ is $\{\widetilde{A}^{(1)},\widetilde{A}^{(2)},\widetilde{A}^{(4)}\}$. According to Definition 13, $\{\widetilde{A}^{(1)}, \widetilde{A}^{(2)},\widetilde{A}^{(4)}\}$ is the relative 0.9-reduct of $\widetilde{A}$.

4.3. Fuzzy reduct and fuzzy core of rule

The fuzzy decision table requires the reduction, not just of fuzzy attributes, but also of the superfluous fuzzy attribute-values. Reducing fuzzy attribute-values is equivalent to reducing every initial fuzzy rule. This is because a fuzzy sample corresponds to an initial fuzzy rule in the fuzzy decision table. In this section, we introduce some concepts to reduce the initial fuzzy rules.

First, we define the concept of fuzzy indispensability in the initial fuzzy rule.

Definition 16. For the initial fuzzy rule x_i and the fuzzy set (linguistic term) $\widetilde{F}_k^{(j)}$ $(1 \leqslant k \leqslant p_j)$ describing fuzzy attribute $\widetilde{A}^{(j)}$, let

$$\beta = \max(T_{\widetilde{A}^{(m+1)}}(\widetilde{B} - \{\widetilde{A}^{(j)} - \{\widetilde{F}_k^{(j)}\}\}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B}))(x_i, \cdot)$$

Then we say that fuzzy attribute-value $\widetilde{F}_k^{(j)}(x_i)$ is β-indispensable in the initial fuzzy rule x_i, or we can say that the indispensability degree of fuzzy attribute-value $\widetilde{F}_k^{(j)}(x_i)$ in the initial fuzzy rule x_i is β.

In the following, we assume that $\widetilde{B}'$ is composed of several modified fuzzy condition attributes. Here the modified fuzzy attribute is obtained by removing one or several fuzzy sets describing it.

$\widetilde{A}'^{(m+1)}$ is obtained by removing one or several fuzzy sets from fuzzy decision attribute $\widetilde{A}^{(m+1)}$. We replace $\widetilde{A}^{(m+1)}$ by $\widetilde{A}'^{(m+1)}$ in the following definitions if the following two formulae hold:

$$\max\left(T_{\widetilde{A}'^{(m+1)}}(\widetilde{B}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B})\right)(x_i, \cdot) \leqslant 1 - \beta \tag{16}$$

For the kth fuzzy set (linguistic term) $\widetilde{F}_k^{(m+1)}$ of fuzzy attribute $\widetilde{A}'^{(m+1)}$,

$$\max(T_{\{\widetilde{A}'^{(m+1)} - \{\widetilde{F}_k^{(m+1)}\}\}}(\widetilde{B}) - T_{\widetilde{A}^{(m+1)}}(\widetilde{B}))(x_i, \cdot) > 1 - \beta \tag{17}$$

Definition 17. For the initial fuzzy rule x_i and the given threshold β, $\beta \in [0, 1]$, if the following two formulae hold:

$$\max(T_{\widetilde{A}'^{(m+1)}}(\widetilde{B}') - T_{\widetilde{A}'^{(m+1)}}(\widetilde{B}))(x_i, \cdot) \leqslant 1 - \beta \tag{18}$$

for the kth fuzzy set (linguistic term) $\widetilde{F}_k^{(j)}$ $(1 \leqslant k \leqslant p_j)$ of fuzzy attribute $\widetilde{A}^{(j)}$,

$$\max(T_{\widetilde{A}'^{(m+1)}}(\widetilde{B}' - \{\widetilde{A}^{(j)} - \{\widetilde{F}_k^{(j)}\}\}) - T_{\widetilde{A}'^{(m+1)}}(\widetilde{B}))(x_i, \cdot) > 1 - \beta \tag{19}$$

then $\widetilde{B}'(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$ is called β-reduct rule of the initial fuzzy rule x_i, denoted by $Reduct^{\beta}_{\widetilde{A}'^{(m+1)}}(\widetilde{B})(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$.

Note that β-reduct rule of the initial fuzzy rule x_i is not unique.

Definition 18. For the initial fuzzy rule x_i and the given threshold β, $\beta \in [0, 1]$, the set $\widetilde{P}$ consists of fuzzy attribute-values whose indispensability degrees in initial fuzzy rule x_i are greater than $1 - \beta$. Then $\widetilde{P} \to \widetilde{A}'^{(m+1)}(x_i)$ is called the β-core rule of the initial fuzzy rule x_i, denoted by $Core^{\beta}_{\widetilde{A}'^{(m+1)}}(\widetilde{B})(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$.

Likewise, we set that $\beta > \max_{i=1}^{n}(\min(T_{\widetilde{A}'^{(m+1)}}(\widetilde{B}))(x_i, \cdot))$. If $\beta \leqslant \min(T_{\widetilde{A}'^{(m+1)}}(\widetilde{B}))(x_i, \cdot)$, $i = 1, \ldots, n$, then there exists no β-reduct rule of the initial fuzzy rule x_i satisfying Definition 17.

In the following, we define the concept of significance degree of fuzzy attribute-values.

Definition 19. For the initial fuzzy rule x_i and $\widetilde{B}' \subset \widetilde{B}$, let $\gamma = \frac{\sum_{j=1}^{n} t_{ij}}{n(n-1)/2}$, where $t_{ij} \in \widetilde{T}_{\widetilde{A}^{(m+1)}}(\widetilde{B}')(x_i,\cdot)$ and n is the size of the universe, we say that the significance degree of $\widetilde{B}'$ in $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$ of the initial fuzzy rule x_i is γ.

Theorem 3. *For an arbitrary threshold* $\beta > \max_{i=1}^{n}(\min(T_{\widetilde{A}'^{(m+1)}}(\widetilde{B}))(x_i,\cdot))$, $\beta \in [0,1]$, *the formula* $Reduct^{\beta}_{\widetilde{A}'^{(m+1)}}(\widetilde{B})(x_i) = Core^{\beta}_{\widetilde{A}'^{(m+1)}}(\widetilde{B})(x_i)$, $1 \leqslant i \leqslant n$ *always holds.*

The proof is similar to Theorem 2. We omit the details of the proof.

To reduce initial fuzzy rules, we need to propose the following new definitions such as covering and rough covering of fuzzy rules, etc.

Definition 20. For $i, j = 1, 2, \ldots, n$ and $j \neq i$, if the similarity degree between the conditional attributes of the initial rules x_i and x_j is greater than or equal to λ, then we say that the rules x_i and x_j λ-rough cover each other.

Definition 21. For $i, j = 1, 2, \ldots, n$ and $j \neq i$, if the similarity degree between the conditional attributes of the initial rules x_i and x_j is greater than or equal to λ, and the similarity degree between the decision attributes of the initial rule x_i and x_j is greater than or equal to λ too, then we say that the rules x_i and x_j λ-cover each other.

Definition 22. For the kth β-reduct rule of an initial rule x_i, find all initial rules which are λ-rough covered by this reduct rule and add the number of these initial rules to RN_{ik}, we say that the sum RN_{ik} is the λ-rough covering degree of the kth β-reduct rule of the initial rule x_i.

Definition 23. For the kth β-reduct rule of an initial rule x_i, find out all initial rules which are λ-covered by this reduct rule and count these initial rules and assign the number to N_{ik}, we say that the number N_{ik} is the λ-covering degree of the kth β-reduct rule of the initial rule x_i.

Definition 24. For any fuzzy rule, count the linguistic terms (fuzzy sets), which are included in the antecedents of this fuzzy rule, and assign the number to δ, then δ is called the rank of this rule.

Definition 25. For any fuzzy rule x_i, ϕ is the λ-rough covering degree of x_i, and φ is the λ-covering degree of this fuzzy rule x_i, then we say that φ/ϕ is the true degree of the fuzzy rule x_i.

In this section, our approach to the reduction of initial fuzzy rules has been to ignore the small perturbation of fuzzy data as represented by the value $1-\beta$. The smaller the threshold $1-\beta$, the smaller and simpler the decision rule set. But a smaller threshold is not necessarily better. Rather, the choice of a threshold is dependent on the tolerance of the small perturbations of fuzzy data.

Example 5. Consider Table 4 again.

(1) Calculate the indiscernibility degree of fuzzy attribute-value $\widetilde{F}_1^{(4)}(x_2)$ of the condition attribute $\widetilde{A}^{(4)}$ in the initial fuzzy rule x_2.

Consider the fuzzy attribute-value $\widetilde{F}_1^{(4)}(x_2)$ (i.e. Windy (false, 0.8)) which is the attribute-value of the fuzzy set "false" of the condition attribute $\widetilde{A}^{(4)}$ (i.e. Windy) in the initial fuzzy rule x_2 in Table 4. The indiscernibility degree of this value is calculated as follows:

$$\beta = \max(T_{\widetilde{A}^{(5)}}(\widetilde{B} - \{\widetilde{A}^{(4)} - \{\widetilde{F}_1^{(4)}\}\}) - T_{\widetilde{A}^{(5)}}(\widetilde{B}))(x_2,\cdot) = \max((0.3\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) - (0.3\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0)) = 0$$

(2) Calculate the 0.9-reduct rule of the initial fuzzy rule x_1 in Table 4.

Suppose $\widetilde{B}'$ is composed of several modified fuzzy attributes which are $\widetilde{A}'^{(1)} = \text{outlook(overcast)}$ and $\widetilde{A}'^{(3)} = \text{humidity(high)}$. $\widetilde{A}'^{(5)}$ (i.e. class(Negative)) is obtained by removing one fuzzy set from the fuzzy decision attribute $\widetilde{A}^{(5)}$. Replace $\widetilde{A}^{(5)}$ by $\widetilde{A}'^{(5)}$ in the initial fuzzy rules x_1.

Consider the rule $\widetilde{B}'(x_1) \to \widetilde{A}'^{(5)}(x_1)$ (i.e. outlook(overcast, 0.1) ∧ humidity(high, 0.8) → class(Negative, 0.7)), we have

$$\max(T_{\widetilde{A}'^{(5)}}(\widetilde{B}') - T_{\widetilde{A}'^{(5)}}(\widetilde{B}))(x_1, \cdot) = 0.1 \leqslant 0.1$$

For the fuzzy set $\widetilde{F}_2^{(1)}$ (i.e. outlook(overcast)) from $\widetilde{B}'$,

$$\max(T_{\widetilde{A}'^{(5)}}(\widetilde{B}' - \{\widetilde{A}^{(1)} - \{\widetilde{F}_2^{(1)}\}\}) - T_{\widetilde{A}'^{(5)}}(\widetilde{B}))(x_1, \cdot) = 0.3 > 0.1$$

For the fuzzy set $\widetilde{F}_1^{(3)}$ (i.e. humidity(high)) from $\widetilde{B}'$,

$$\max(T_{\widetilde{A}'^{(5)}}(\widetilde{B}' - \{\widetilde{A}'^{(3)} - \{\widetilde{F}_1^{(3)}\}\}) - T_{\widetilde{A}'^{(5)}}(\widetilde{B}))(x_1, \cdot) = 0.4 > 0.1$$

According to Definition 17, the rule outlook (overcast, 0.1) ∧ humidity(high, 0.8) → class(Negative, 0.7) is 0.9-reduct rule of the initial fuzzy rule x_1.

5. The procedure and algorithms of learning from fuzzy decision table and an illustrative example

In this section, we first give the procedure of learning from the fuzzy decision table and then provide five algorithms to learning fuzzy rules. We also give an illustrative example to demonstrate the learning result.

5.1. The procedure of learning from fuzzy decision table

The procedure of learning fuzzy rules from the fuzzy decision table is provided in Fig. 1.

(1) *Preprocessing*: First, identify the condition and decision attributes. Next, transform the fuzzy data into decision table format. Finally, use a fuzzy similarity matrix or fuzzy inconsistent matrix to represent fuzzy information in the fuzzy decision table.

(2) *Fuzzy attribute reduction*: Remove the superfluous attributes from a fuzzy decision table by ignoring the small perturbations of the fuzzy inconsistent matrix. That is to say, compute the attribute reducts of initial fuzzy dataset.

(3) *Reducing fuzzy attribute-values*: Remove the superfluous and irrelevant attribute-values from the dataset by ignoring the small perturbation of fuzzy inconsistent matrix and compute all reduct rules of each initial fuzzy rule.

(4) *Fuzzy rules induction*: From all reduct rules of every initial rule, induce the rules which are most representative so as to form a decision rule set.

(5) *Classification*: The rule set can be used in classification. In the following $\widetilde{B} = \{\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, \ldots, \widetilde{A}^{(m)}\}$ denotes the set of fuzzy condition attributes, $\widetilde{A}^{(m+1)}$ denotes fuzzy decision attribute. Re_{ik} denotes the kth β-reduct rule of the initial rule x_i. K_i denotes the number of β-reduct rules of the initial rule x_i.

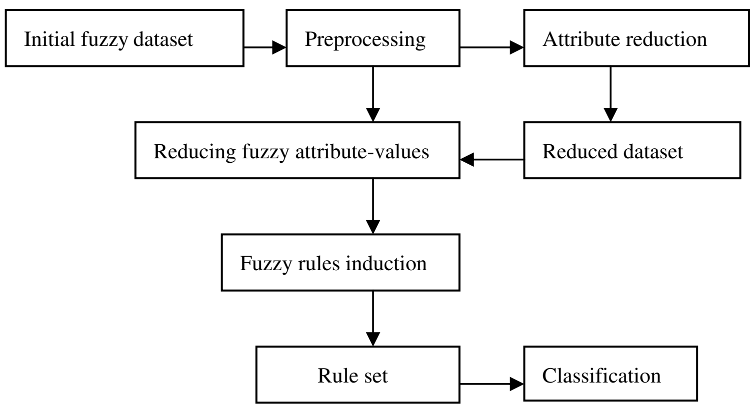

Fig. 1. The program of learning fuzzy rules from fuzzy sample

Algorithm 1 (*Finding all fuzzy relative reducts*).

Input: $S=(U,\widetilde{A})$ is the initial fuzzy dataset, $\widetilde{B}=\{\widetilde{A}^{(1)},\widetilde{A}^{(2)},\ldots,\widetilde{A}^{(m)}\}$ and $\widetilde{A}^{(m+1)}$ are the set of all fuzzy condition and fuzzy decision attributes, respectively. H is the set of fuzzy attributes, which are β-indispensable in $\widetilde{B}$. l is a positive integer and β is the given threshold.
Output: R is the set of all β-reducts in $\widetilde{B}$ with respect to $\widetilde{A}^{(m+1)}$.
Step 1: Initialize $l=1$, $R=\Phi$, $H=\Phi$, $\widetilde{B}=\{\widetilde{A}^{(1)},\widetilde{A}^{(2)},\ldots,\widetilde{A}^{(m)}\}=\{\widetilde{F}_1^{(j)},\widetilde{F}_2^{(j)},\ldots,\widetilde{F}_{p_j}^{(j)}|j=1,2,\ldots,m\}$ and $\widetilde{A}^{(m+1)}$.
Step 2: According to Definition 12, compute the indispensability degree with respect to $\widetilde{A}^{(m+1)}$ of each linguistic term $\widetilde{F}_k^{(j)}$ $(j=1,2,\ldots,m,\ k=1,2,\ldots,p_j)$. Add the fuzzy linguistic terms, which are β-indispensable with respect to $\widetilde{A}^{(m+1)}$, to H. Let $Q=\widetilde{B}-H$.
Step 3: According to Definition 13, if H is the β-reduct with respect to $\widetilde{A}^{(m+1)}$, then add it to the set R and **Stop**; Otherwise, go to **Step 4**.
Step 4: Compute the power set P of Q.
Step 5: If such a set $N_l\in P$, which has l elements, exists, go to **Step 7**; Otherwise, go to **Step 6.**
Step 6: Let $l=l+1$, go to **Step 5**.
Step 7: Select a set $M_l\in P$ which has l elements. Let $H'=H\cup M_l$.
Step 8: According to Definition 13, if H' is the β-reduct with respect to $\widetilde{A}^{(m+1)}$, then add it to the set R and let $P=P-\{M\in P|M_l\subset M\}$, go to **Step 9**; Otherwise, let $P=P-\{M_l\}$, go to **Step 9**.
Step 9: If $P\neq\Phi$, go to **Step 5**; Otherwise, **Stop**.

According to algorithm 1, the time complexity of finding all fuzzy relative reducts is $O(2^{|LA|}\times|U|^2)$, where $|LA|$ is the number of the linguistic terms of all condition attributes, $|U|$ is the cardinality of the universe of discourse U. Because the time complexity of the algorithm increases exponentially with the total number of linguistic terms, it is impossible to find all reducts in large dimensional datasets and it is unnecessary to find all reducts in real applications. For this reason, we describe a more effective heuristic algorithm for finding the close-to-minimal attribute reducts.

Algorithm 2 ((*Heuristic*): *Finding the close-to-minimal fuzzy reduct*).

Input: $S=(U,\widetilde{A})$ is the initial fuzzy dataset, $\widetilde{B}=\{\widetilde{A}^{(1)},\widetilde{A}^{(2)},\ldots,\widetilde{A}^{(m)}\}$ and $\widetilde{A}^{(m+1)}$ are the set of all fuzzy condition and fuzzy decision attributes, respectively. H is the set of fuzzy attributes which are β-indispensable with respect to $\widetilde{A}^{(m+1)}$ in $\widetilde{B}$. β is the given threshold.
Output: R is the close-to-minimal β-reduct.
Step 1: Initialize $R=\Phi$, $H=\Phi$, $\widetilde{B}=\{\widetilde{A}^{(1)},\widetilde{A}^{(2)},\ldots,\widetilde{A}^{(m)}\}=\{\widetilde{F}_1^{(j)},\widetilde{F}_2^{(j)},\ldots,\widetilde{F}_{p_j}^{(j)}|j=1,2,\ldots,m\}$ and $\widetilde{A}^{(m+1)}$.
Step 2: According to Definition 12, compute the indispensability degree with respect to $\widetilde{A}^{(m+1)}$ of each linguistic term $\widetilde{F}_k^{(j)}$ $(j=1,2,\ldots,m,\ k=1,2,\ldots,p_j)$. Add those linguistic terms which are β-indispensable with respect to $\widetilde{A}^{(m+1)}$, to H. Let $Q=\widetilde{B}-H$.
Step 3: According to Definition 13, if H is the β-reduct with respect to $\widetilde{A}^{(m+1)}$, add it to the set R and **Stop**; Otherwise, go to **Step 4**.
Step 4: Add an arbitrary linguistic term $M_t\in Q$ $(t=1,2,\ldots,|Q|$. Here $|Q|$ is the cardinality of $Q)$ into H, denoted as $H_t'=H\cup M_t$.
Step 5: According to Definition 15, compute the significance degree of fuzzy attributes subset H_t' $(t=1,2,\ldots,|Q|)$.
Step 6: If the significance degree of $H_k'=H\cup M_k$ is the maximum one, let $H=H\cup M_k$.
Step 7: According to Definition 13, if the subset H is the β-reduct with respect to $\widetilde{A}^{(m+1)}$, then let $R=H$ and **Stop**; Otherwise, let $Q=Q-\{M_k\}$, go to **Step 4**.

Using the above heuristic Algorithm 2, it is possible to obtain the close-to-minimal β-reduct. For the worst case scenario, the time complexity of this heuristic algorithm is $O\left(\frac{(|LA|^2+|LA|)}{2}\times|U|^2\right)$.

The algorithms for finding all fuzzy reduct rules from each initial fuzzy rule are given as follows:

Algorithm 3 (Finding all fuzzy reduct rules of the initial fuzzy rule).

Input: x_i is the initial fuzzy rule of $S = (U, \widetilde{A})$, $F(x_i)$ and $\widetilde{A}^{(m+1)}(x_i)$ are the set of all fuzzy condition attribute-values and the set of all fuzzy decision attribute-values of the rule x_i, respectively. $H(x_i)$ is the set of fuzzy attribute-values which are β-indispensable in the rule x_i. l is a positive integer and β is the given threshold.

Output: R_i is the set of the β-reduct rules of the initial rule x_i.

Step 1: Initialize $l=1$, $R_i=\Phi$, $H(x_i)=\Phi$, $F(x_i) = \{\widetilde{F}_1^{(j)}(x_i), \widetilde{F}_2^{(j)}(x_i), \ldots, \widetilde{F}_{p_j}^{(j)}(x_i) | j = 1,2,\ldots,m\}$ and $\widetilde{A}^{(m+1)}(x_i) = \{\widetilde{F}_1^{(m+1)}(x_i), \widetilde{F}_2^{(m+1)}(x_i), \ldots, \widetilde{F}_{p_{m+1}}^{(m+1)}(x_i)\}$.

Step 2: According to the formulae (14) and (15), replace $\widetilde{A}^{(m+1)}(x_i)$ by $\widetilde{A}'^{(m+1)}(x_i)$ in the initial fuzzy rules x_i.

Step 3: For the rule $F(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$, compute the indispensability degree of fuzzy attribute-values $\widetilde{F}_k^{(j)}(x_i)$ ($j = 1,2,\ldots,m$, $k = 1,2,\ldots,p_j$). Add the fuzzy attribute-values, which are β-indispensable, to $H(x_i)$. Let $Q(x_i) = F(x_i) - H(x_i)$.

Step 4: According to Definition 17, if the rule $H(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$ is the β-reduct rule of the initial rule x_i, add it to the set R_i and **Stop**; Otherwise, go to **Step 5**.

Step 5: Compute the power set P of $Q(x_i)$.

Step 6: If there exists a set $N_l \in P$ which has l elements, go to **Step 8;** Otherwise, go to **Step 7**.

Step 7: Let $l = l + 1$, go to **Step 6**.

Step 8: Select a set $M_l \in P$ which has l elements. Let $H'(x_i) = H(x_i) \cup M_l$.

Step 9: According to Definition 17, if the rule $H'(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$ is the β-reduct rule of the initial rule x_i, add it to the set R_i and let $P = P - \{M \in P | M_l \subset M\}$, go to **Step 10**; Otherwise, let $P = P - \{M_l\}$, go to **Step 10**.

Step 10: If $P \neq \Phi$, go to **Step 6**; Otherwise, **Stop**.

Algorithm 4 ((Heuristic): Finding the close-to-minimal fuzzy reduct rule of the initial fuzzy rule).

Input: x_i is the initial fuzzy rule of $S = (U, \widetilde{A})$, $F(x_i)$ and $\widetilde{A}^{(m+1)}(x_i)$ are the set of all fuzzy condition attribute-values and the set of all fuzzy decision attribute-values of the rule x_i, respectively. $H(x_i)$ is the set of fuzzy attribute-values which are β-indispensable in the rule x_i. β is the given threshold.

Output: R_i is the close-to-minimal β-reduct rule of the initial fuzzy rule x_i.

Step 1: Initialize $l=1$, $R_i=\Phi$, $H(x_i)=\Phi$, $F(x_i) = \{\widetilde{F}_1^{(j)}(x_i), \widetilde{F}_2^{(j)}(x_i), \ldots, \widetilde{F}_{p_j}^{(j)}(x_i) | j = 1,2,\ldots,m\}$ and $\widetilde{A}^{(m+1)}(x_i) = \{\widetilde{F}_1^{(m+1)}(x_i), \widetilde{F}_2^{(m+1)}(x_i), \ldots, \widetilde{F}_{p_{m+1}}^{(m+1)}(x_i)\}$.

Step 2: According to the formulae (14) and (15), we replace $\widetilde{A}^{(m+1)}(x_i)$ by $\widetilde{A}'^{(m+1)}(x_i)$ in the initial fuzzy rules x_i.

Step 3: For the rule $F(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$, compute the indispensability degree of fuzzy attribute-values $\widetilde{F}_k^{(j)}(x_i)$ ($j = 1,2,\ldots,m$, $k = 1,2,\ldots,p_j$). Add those fuzzy attribute-values, which are β-indispensable, to $H(x_i)$. Let $Q(x_i) = F(x_i) - H(x_i)$.

Step 4: According to Definition 17, if the rule $H(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$ is the β-reduct rule of the initial rule x_i, then add it to the set R_i and **Stop**; Otherwise, go to **Step 5**.

Step 5: Add an arbitrary attribute-value $M_t(x_i) \in Q\ (x_i)$ $(t = 1,2,\ldots,|Q(x_i)|)$ into $H(x_i)$, denoted as $H'(x_i) = H(x_i) \cup M_t(x_i)$. Here $|Q(x_i)|$ is the number of the attributes values in $Q(x_i)$.

Step 6: According to Definition 19, compute the significance degree of fuzzy attribute-values subset $H'_t(x_i)$.

Step 7: If the significance degree of $H'_k(x_i) = H(x_i) \cup M_k(x_i)$ is the maximum, let $H(x_i) = H(x_i) \cup M_k(x_i)$.

Step 8: According to Definition 17, if the rule $H(x_i) \to \widetilde{A}'^{(m+1)}(x_i)$ is the β-reduct rule of the initial rule x_i, add it to the set R_i and **Stop**; Otherwise, let $Q(x_i) = Q(x_i) - \{M_k(x_i)\}$, go to **Step 5**.

The final step in learning from fuzzy samples based on rough set technique is to learn rules from all fuzzy reduct rules. A heuristic algorithm of inducing fuzzy rules from the fuzzy decision table is given as follows:

Algorithm 5 ((*Heuristic*): *Inducing fuzzy rules*).

Input: P is the set of initial fuzzy rules, R is the set of fuzzy reduct rules of every initial rule, and β, λ are the given thresholds.

Output: Q is the close-to-minimal fuzzy rule set.

Step 1: Initialize $P = U$, $R = \Phi$ and $Q = \Phi$.

Step 2: For each initial fuzzy rule x_i ($1 \leqslant i \leqslant n$) in P, compute all β-reduct rules and add them to R.

Step 3: For each β-reduct rule Re_{ik} ($1 \leqslant i \leqslant n$, $1 \leqslant k \leqslant K_i$) in R, find all initial fuzzy rules in P, which are λ-covered by the rule Re_{ik}, and compute the λ-covering degree N_{ik} of the rule Re_{ik}.

Step 4: For each β-reduct rule $Re_{ik}(1 \leqslant i \leqslant n, 1 \leqslant k \leqslant K_i)$ in R, find all initial rules in P, which are λ-rough covered by the rule Re_{ik}, and compute the λ-rough covering degree RN_{ik} of the rule Re_{ik}.

Step 5: Compute $CR_{ik} = \frac{N_{ik} \times N_{ik}}{RN_{ik}}$ as the criterion to select the most informative β-reduct rule.

Step 6: Select the most informative β-reduct rules. Among them, choose the β-rule $Re^{\bullet}$ whose rank is the smallest. Remove this rule $Re^{\bullet}$ from R and add it to Q.

Step 7: From P, remove those initial fuzzy rules which are β-rough covered by the rule $Re^{\bullet}$.

Step 8: If $P \neq \Phi$, go to **Step 3**; Otherwise, **Stop**.

In **Step 2**, it is also feasible to find the close-to-minimal fuzzy reduct rule instead of finding all β-reduct rules.

For the algorithms mentioned above, If $\beta = \lambda = 1$ the algorithms degenerate into the traditional algorithms proposed in [16,17]. That is to say, when the threshold β and λ are set to 1, the algorithms proposed in this paper can be use to learn from nominal data.

5.2. An illustrative example

Using the heuristic algorithms mentioned above, when $\beta = \lambda = 0.7$ we learn the close-to-minimal rule set from Table 1 as follows:

Rule 1-1: humidity(high, 0) $\rightarrow$ class(positive, 0.8) true degree: 0.8571
Rule 1-2: outlook(overcast, 0.1) $\wedge$ humidity(normal, 0.1) $\rightarrow$ class(negative, 0.9) true degree: 1
Rule 1-3: outlook(overcast, 1) $\rightarrow$ class(positive, 0.8) true degree: 1
Rule 1-4: outlook(overcast, 1) $\rightarrow$ class(positive, 0.6) true degree: 1

5.3. Analysis about the time complexity of the method proposed in this paper

In this paper, we present a method of learning fuzzy rules from fuzzy samples based on rough set technique. The fuzzy information is represented in the form of a fuzzy matrix. The families of fuzzy indiscernibility relations proposed in this paper and the families of fuzzy matrices are isomorphic. That is to say, the method proposed in this paper is fuzzy matrix computation for fuzzy information systems. So, all those families of time complexities in this paper (such as the time complexity of finding fuzzy indispensability relation, the time complexity of finding fuzzy core) are equal to the corresponding time complexities in the matrix computation of the traditional rough set theory, respectively [7]. In this paper, the time complexities of the method are analyzed as follows: First, the time complexity of computing fuzzy indiscernibility relation is equal to the time complexity of finding the corresponding fuzzy matrix $O(|A| \times |U|^2)$. Here $|U|$ is the cardinality of the universe of discourse U and $|A|$ is the cardinality of the set of attributes. Secondly, the time complexity of computing fuzzy indiscernibility degree of certain attribute is $O(|A| \times |U|^2)$. The time complexity of finding fuzzy core attributes is $O(|A|^2 \times |U|^2)$. Thirdly, according to Algorithm 1, the time complexity of finding all fuzzy attribute reducts is $O(2^{|LA|} \times |U|^2)$. Here, $|LA|$ is the total number of the linguistic terms of all condition attributes. According to Algorithm 3, the time complexity of finding all fuzzy reduct rules of each initial rule is $O(2^{|LA|} \times |U|)$. It is unnecessary to find all fuzzy attribute reducts or all fuzzy reduct rules of each initial rule. Then, the heuristic Algorithms 2 and 4 are given. For the worst case, the time complexity of finding the close-to-minimal attribute reduct is $O\left(\frac{(|LA|^2+|LA|)}{2} \times |U|^2\right)$. The time complexity of finding the close-to-minimal reduct rule of each initial rule is $O\left(\frac{(|LA|^2+|LA|)}{2} \times |U|\right)$. Finally, according to the heuristic Algorithm 5, the time complexity of extracting the close-to-minimal rule set from all fuzzy reduct rules is $O(|LA| \times |U| \times nr)$, where nr is the number of the reduct rules of every initial rule.

6. Experimental comparison of proposed with other methods

In this section, we describe some experiments. Some compare the proposed method with the attribute reduct method using fuzzy rough sets (seen in [1]). Others compare the proposed method with the method of learning fuzzy rules from fuzzy samples (seen in [8,27]).

6.1. Fuzzification of the dataset

A simple algorithm [34] is used to generate triangular membership functions on a numerical dataset. The triangular membership functions are defined as follows:

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

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

$$T_i(x) = \begin{cases} 0 & x \geqslant m_{i+1} \\ (m_{i+1} - x)/(m_{i+1} - m_i) & m_i \leqslant x < m_{i+1} \\ (x - m_{i-1})/(m_i - m_{i-1}) & m_{i-1} < x < m_i \\ 0 & x \leqslant m_{i-1} \end{cases} \quad i = 2, 3, \ldots, k-1$$

In this case, the centers $m_i, i = 1, \ldots, k$ can be calculated by using Kohonen's feature-maps algorithm [13].

For the datasets selected in this paper, only inputs need to be fuzzified, outputs are nominal classification sets.

6.2. Comparison with a method of attribute reduction

As mentioned previously, this paper proposes one attribute reduction method based on the fuzzy rough set technique which is different with Ref. [1]. The method proposed in [1] builds on the notion of fuzzy lower approximation to enable reduction of attributes, while the method proposed in this paper considers the change in degree of fuzzy inconsistence to introduce the concepts of fuzzy attribute reduction. In the following, we compare the difference between these two methods experimentally.

6.2.1. Experimental setup

We use the datasets, "Wine", "Haberman" and "New_thyroid", from the UCI Machine Learning repository [10] to compare our proposed method with the method proposed in [1]. The information from these datasets is summarized in Table 8. The classification performance of the selected attributes has been measured by applying ID3 Algorithm [20,34].

Four indices have been used for comparison of these methods: (1) Number of the selected attributes; (2) Total running time to select attributes; (3) Training accuracy of classification; (4) Testing accuracy of classification. The comparison results are summarized in Tables 5–7.

6.2.2. Experimental analysis

Considering the effect of attribute reduction on classification performance in Tables 5–7, the method proposed in this paper exhibits better classification accuracy. The highest test classification accuracy (0.9101 and 0.9113 in Tables 5 and 7, respectively) is obtained by the method proposed in this paper. However, the method proposed in [1] finds the attribute reduction quicker than our proposed method. The time complexity associated with identification of the attribute reduction is $O\left(\frac{(|LA|^2+|LA|)}{2} \times |U|^2\right)$ in our proposed method, while the time complexity of the method proposed in [1] is $O\left(\frac{(|LA|^2+|LA|)\times|U|}{2}\right)$.

Table 5
Comparison of attribute reduction on dataset "Wine"

Fuzzified dataset "Wine"	Selected attributes	Number of the selected attributes	Total running time to select attributes (CPU seconds)	Training accuracy of classification	Testing accuracy of classification
All initial attributes	$1,2,3,\ldots,39$	39	–[a]	0.9866	0.8640
Reduct by this paper	$1,3,4,13,19,22,28,34,37$	9	5.7813	0.9912	**0.9101**
Reduct by [1]	$4,8,10,13,14,16,17,25,29,32,38$	11	1.8420	0.9619	0.8370

[a] We need no running-time because we have selected all condition attributes.

Table 6
Comparison of attribute reduction on dataset "Haberman"

Fuzzified dataset "Haberman"	Selected attributes	Number of the selected attributes	Total running time to select attributes (CPU seconds)	Training accuracy of classification	Testing accuracy of classification
All initial attributes	$1,2,3,\ldots,9$	9	–[a]	0.7896	0.7189
Reduct by this paper	$1,3,4,7$	4	1.6406	0.7915	**0.7268**
Reduct by [16]	$1,4,5,6$	4	0.7010	0.7424	**0.7234**

[a] We need no running-time because we have selected all the condition attributes.

Table 7
Comparison of attribute reduction on dataset "New_thyroid"

Fuzzified dataset "New_thyroid"	Selected attributes	Number of the selected attributes	Total running time to select attributes (CPU seconds)	Training accuracy of classification	Testing accuracy of classification
All initial attributes	$1,2,3,\ldots,15$	15	–[a]	0.9821	0.8955
Reduct by this paper	$1,3,4,5,6,7,10,13,15$	9	2.8125	0.9804	**0.9113**
Reduct by [1]	$1,3,4,7,10,11,12,14,15$	9	1.0910	0.9544	0.8890

[a] We need no running-time because we have selected all the condition attributes.

The method proposed in [1] and the method in this paper share several common attributes. This shows that some informative attributes can been found by each of the both methods.

From the above results, it is concluded that the method proposed in this paper is feasible and effective to find the attribute reduction in fuzzy decision systems.

6.2.3. Comparison with methods of learning fuzzy rules from fuzzy samples

The method proposed in this paper can find the attribute reduction and the close-to-minimal rule set. In this section, we compare our method with the methods described in [8,27], both of which learn fuzzy rules from fuzzy samples using fuzzy rough set technique.

6.2.4. Experimental setup

We use the datasets, "New_thyroid", "Haberman" and "Diabetes", from the UCI Machine Learning Repository [10] summarized in Table 8 to compare the methods of rules induction using fuzzy rough sets.

Table 8
Information of some datasets from UCI

Dataset	Number of objects	Number of original attributes column	Number of attributes column after fuzzification	Decision classes
New_thyroid	215	6	16	3
Diabetes	768	9	25	2
Haberman	306	4	10	2
Wine	178	14	40	3

Table 9
Results of comparisons of the learning fuzzy rules from fuzzy samples

Dataset		Method proposed in this paper	Method proposed by Hong [8]	Method proposed by Wang [27]
New_thyroid	Number of selected rules	4	7	11
	Total running time (CPU seconds)	2.1090	**0.28125**	5.3594
	Classification accuracy	**0.9256**	0.8000	0.9070
Haberman	Number of selected rules	5	4	17
	Total running time (CPU seconds)	2.6410	**0.8410**	2.1875
	Classification accuracy	**0.7682**	0.7418	**0.8693**

Table 10
Comparison of three methods on dataset "Diabetes"*

Diabetes_part (composed of 100 objects randomly chosen from Diabetes)	Method proposed in this paper	Method proposed by Hong [8]	Method proposed by Wang [27]
Number of selected rules	9	6	27
Total running time (CPU seconds)	**15.2652**	1229.2	248.7656
Classification accuracy	**0.7682**	0.6953	0.7370

* Only half of objects in diabetes are selected to use in this comparison.

These datasets have a different number of decision classes and all condition attributes in these datasets shown in Table 8 are numerical.

Three indices have been used for comparison: (1) The size of the rule set which is learned from fuzzy samples; (2) Total running time needed to find the close-to-minimal rule set; and (3) Accuracy of classification using the selected rule set. Results of comparison are summarized in Tables 9 and 10.

6.2.5. Comparison and experimental analysis

In the following, we analyze the comparison results in three areas: (1) the capability to find both reduct and rule set (2) the quality of rule set, i.e. classification performance and size of rule set, and (3) the time complexity.

For the capability to find both reduct and rules, it should be pointed out that the methods in [8,27] cannot find the attribute reduction which is not mentioned in [8]. Using the method in [8], only a set of maximally general fuzzy rules for an approximate coverage of training samples can be found. The method in [27] cannot find the attribute reduction either; only the concepts of rule reduction are proposed and discussed in [27].

For the quality of the rule set, we can find the comparison results from Table 9. Table 9 shows that the classification accuracy of the method in [27] is relatively high at the cost of a large rule set. The reason may be hidden in the process of transforming fuzzy data into crisp data [27]. Membership degrees of attribute values (fuzzy sets) are not exploited in the process of learning from fuzzy samples in [27], and so, much information of membership degrees may be lost. The method in [8] is also quick to find the close-to-minimal rule set. The set of rules is smaller than the method in [27]. Because the uncovered samples are focused in each iterative induction process, the number of rules required to cover instances are thus reduced [8]. The classification accuracy of the method in [8] is slightly lower than our proposed method. Compared with the methods in [8,27], high classification accuracy and the compact rule set are the advantages of the method proposed in this paper. The reason may be that each membership degree is sufficiently applied in the process of inducing the rule set.

Tables 9 and 10 show the comparison results of the running time. The method in [8] is the quickest one to find the rule set. In [8], only the uncovered samples are focused in each iterative induction process, the execution time and number of the rules required to cover instances are thus reduced. Comparatively, the total running time of our proposed method is higher. The method proposed in this paper builds on the notion of fuzzy similarity relation to find the close-to-minimal rule set. Therefore, the time complexity of our proposed method increases with the square of the size of the universe (i.e.$|U|^2$, here $|U|$ is the cardinality of the universe U). We may conclude that the method proposed in this paper is suitable for a case where the size of the universe is small.

Table 10 shows that the running time of methods in [8,27] is far slower than of the method proposed in this paper in datasets that have a large number of fuzzified attributes (there are 25 fuzzified attributes in dataset 'diabetes'). The time complexity of computing fuzzy equivalence relation in [8] increases exponentially with the number of original attributes. According to the algorithm used to identify the minimal reduct rule of each initial fuzzy rule in [27], the time complexity of finding a minimal reduct rule in [27] also increases exponentially with the number of original attributes. It implies that the methods in [8,27] are also suitable for a case where the attribute number is not very large.

7. Conclusions

This paper makes a number of contributions. First, we show that the underlying relationship between the reduct and core of the traditional rough set approach is still pertinent, even after the proposed extension. Second, to represent the initial information, we propose a fuzzy similarity matrix as a replacement for the partition of the universe of discourse. Third, by introducing a threshold to the concepts of the reducts and core, we produce a learning result, which is less sensitive to small perturbations of attribute-values. Finally, the fuzzy model proposed in this paper can also be used to learn from nominal data where the thresholds β and λ are set to 1.

Future work should focus on improving the fuzzy model proposed in this paper to process an information system whose attribute-values are continuous data. We consider realizing the reduction process by not implementing the discretization or fuzzification of continuous data.

Appendix

In [8], the concepts of fuzzy equivalence relation are defined as follows:

When the same linguistic term R_{ik} of an attribute A_i exists in two fuzzy objects $obj^{(i)}$ and $obj^{(r)}$ with membership values f^i_{jk} and f^r_{jk} larger than zero, $obj^{(i)}$ and $obj^{(r)}$ are said to have a fuzzy indiscernibility relation (or fuzzy equivalence relation) on attribute A_i with membership value $\min(f^i_{jk}, f^r_{jk})$ If the same linguistic terms of an attribute subset B exist in both $obj^{(i)}$ and $obj^{(r)}$ with membership values larger than zero, $obj^{(i)}$ and $obj^{(r)}$ are said to have a fuzzy indiscernibility relation (or a fuzzy equivalence relation) on attribute subset B with a membership value equal to the minimum of all the membership values. These fuzzy equivalence relations thus partition the fuzzy object set U into several fuzzy subsets that may overlap, and the result is denoted by U/B.

Suppose that each attribute has three linguistic terms.

Let U be the universe of discourse, $\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, \ldots, \widetilde{A}^{(m)}$ and $\widetilde{A}^{(m+1)}$ be a set of fuzzy attributes. Suppose each fuzzy attribute $\widetilde{A}^{(j)}$ consists of three linguistic terms: $F(\widetilde{A}^{(j)}) = \{\widetilde{F}^{(j)}_L, \widetilde{F}^{(j)}_N, \widetilde{F}^{(j)}_H\}$ $(j = 1, 2, \ldots, m+1)$. According to the definition of fuzzy equivalence relation in [8,9], the fuzzy equivalence relation with single attribute $\widetilde{A}^{(j)}$ partitions the universe of discourse Uinto three fuzzy subsets:

$$U/\widetilde{A}^{(j)} = \{(obj(\widetilde{F}^{(j)}_L), \mu^{(j)}_L), (obj(\widetilde{F}^{(j)}_N), \mu^{(j)}_N), (obj(\widetilde{F}^{(j)}_H), \mu^{(j)}_H)\} \quad (j = 1, 2, \ldots, m+1)$$

Here $obj(\widetilde{F}^{(j)}_L)$ is the set of objects which have the same term $\widetilde{F}^{(j)}_L$ in the attribute $\widetilde{A}^{(j)}$. $\mu^{(j)}_L$ is the corresponding membership value. The fuzzy equivalence relation with two attributes $B = \{\widetilde{A}^{(j)}, \widetilde{A}^{(k)}\}$ partitions the universe of discourse U into 3×3 fuzzy subsets: $U/B = \{(obj(\widetilde{F}^{(j)}_L \widetilde{F}^{(k)}_L), \mu^{jk}_{LL}), (obj(\widetilde{F}^{(j)}_N \widetilde{F}^{(k)}_L), \mu^{jk}_{NL}), (obj(\widetilde{F}^{(j)}_H \widetilde{F}^{(k)}_L), \mu^{jk}_{HL}), (obj(\widetilde{F}^{(j)}_L \widetilde{F}^{(k)}_N), \mu^{jk}_{LN}), (obj(\widetilde{F}^{(j)}_N \widetilde{F}^{(k)}_N), \mu^{jk}_{NN}), (obj(\widetilde{F}^{(j)}_H \widetilde{F}^{(k)}_N), \mu^{jk}_{HN}), (obj(\widetilde{F}^{(j)}_L \widetilde{F}^{(k)}_H), \mu^{jk}_{LH}), (obj(\widetilde{F}^{(j)}_N \widetilde{F}^{(k)}_H), \mu^{jk}_{NH}), (obj(\widetilde{F}^{(j)}_H \widetilde{F}^{(k)}_H), \mu^{jk}_{HH})\}$ $(j, k = 1, 2, \ldots, m+1$ and $j \neq k)$. Here $obj(\widetilde{F}^{(j)}_L \widetilde{F}^{(k)}_L)$ is the set of objects which have the same terms $\widetilde{F}^{(j)}_L$ and $\widetilde{F}^{(k)}_L$ in the attributes set $B = \{\widetilde{A}^{(j)}, \widetilde{A}^{(k)}\}$. $\mu^{(jk)}_{LL}$ is the corresponding membership value. Similarly, for the worst case scenario, fuzzy equivalence relation with n attributes partitions the universe of discourse set into 3^n fuzzy subsets. This means that the time complexity of computing fuzzy equivalence relation increases exponentially with the number of original attributes.

References

[1] R.B. Bhatt, M. Gopal, On fuzzy rough sets approach to feature selection, Pattern Recognition Letters 26 (2005) 1632–1640.
[2] K. Chakrabarty, R. Biswas, S. Nanda, Fuzziness in rough sets, Fuzzy Sets and Systems 110 (2000) 247–251.
[3] D.G. Chen, X.Z. Wang, D.S. Yeung, E.C.C. Tsang, Rough approximations on a complete completely distributive lattice with applications to generalized rough sets, Information Sciences 176 (2006) 1829–1848.
[4] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–208.
[5] D. Dubois, H. Prade, Putting rough sets and fuzzy sets together, in: R. Slowinski (Ed.), Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, 1992.
[6] S. Greco, B. Matarazzo, R. Slowinski, Rough set processing of vague information using fuzzy similarity relations, in: C.S. Calude, G. Paun (Eds.), Finite Versus Infinite-Contributions to an Eternal Dilemma, Springer-Verlag, London, 2002, pp. 149–173.
[7] J.W. Guan, D.A. Bell, Z. Guan, Matrix computation for information system, Information Sciences 131 (2001) 129–156.
[8] T.P. Hong, Learning approximate fuzzy rules from training examples, in: The Tenth IEEE International Conference on Fuzzy Systems, 2001.
[9] T.P. Hong, T.T. Wang, S.L. Wang, B.C. Chien, Learning a coverage set of maximally general fuzzy rules by rough sets, Expert Systems with Applications 19 (2000) 97–103.
[10] http://www.ics.uci.edu/~mlearn/MLRepository.html.
[11] M. Inuiguchi, S. Greco, R. Slowinski, T. Tanino, Possibility and necessity measure specification using modifiers for decision making under fuzziness, Fuzzy Sets and Systems 137 (2003) 151–175.
[12] R. Jensen, Q. Shen, Fuzzy-rough attributes reduction with application to web categorization, Fuzzy Sets and Systems 141 (2004) 469–485.
[13] T. Kohonen, Fuzzy entropy and conditioning, Information Sciences 30 (1986) 165–174.
[14] J.S. Mi, W.X. Zhang, An axiomatic characterization of a fuzzy generalization of rough sets, Information Sciences 160 (2004) 235–249.
[15] N.N. Morsi, M.M. Yakout, Axiomatics for fuzzy rough sets, Fuzzy Sets and Systems 100 (1998) 327–342.
[16] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341–356.
[17] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data [M], Kluwer Academic Publishers, Dordrecht, 1991.
[18] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (2007) 28–40.
[19] M. Quafafou, a-RST: A generalization of rough set theory, Information Sciences 124 (2000) 301–316.
[20] J.R. Quinlan, Induction of decision trees, Machine Learning 1 (1986) 81–106.
[21] Q. Shen, R. Jensen, Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, Pattern Recognition 37 (2004) 1351–1363.
[22] R. Slowinski, D. Vanderpooten, Similarity relation as a basis for rough approximations, in: P.P. Wang (Ed.), Advances in Machine Intelligence and Soft-Computing, Department of Electrical Engineering, Duke University, Durham, NC, USA, 1997, pp. 17–33.
[23] H. Thiele, Fuzzy rough sets versus rough fuzzy sets—an interpretation and a comparative study using concepts of modal logics, University of Dortmund, Tech. Report No. CI-30=98, 1998, also in Proc. EUFIT'1997, pp. 159–167.
[24] H. Thiele, On axiomatic characterization of fuzzy approximation operators I, the fuzzy rough set based case, in: RSCTC 2000 Conference Proceedings, Banff Park Lodge, Bariff, Canada, 19 October 2000, pp. 239–247.
[25] H. Thiele, On axiomatic characterization of fuzzy approximation operators II, the rough fuzzy set based case, in: Proceedings of the 31st IEEE International Symposium on Multiple-Valued Logic, 2001, pp. 330–335.
[26] E.C.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, J.W.T. Lee, Attributes reduction using fuzzy rough sets, IEEE Transaction on Fuzzy System, in press.
[27] X.Z. Wang, J.R. Hong, Learning optimization in simplifying fuzzy rules, Fuzzy Sets and Systems 106 (1999) 349–356.
[28] X.Z. Wang, D.S. Yeung, E.C.C. Tsang, A comparative study on heuristic algorithms for generating fuzzy decision trees, IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics 31 (2001) 215–236.
[29] W.Z. Wu, J.S. Mi, W.X. Zhang, Generalized fuzzy rough sets, Information Sciences 151 (2003) 263–282.
[30] W.Z. Wu, W.X. Zhang, Constructive and axiomatic approaches of fuzzy approximation operators, Information Sciences 159 (2004) 233–254.
[31] Y.Y. Yao, Generalized rough set model, in: L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery. 1. Methodology and Applications, Physica-Verlag, Heidelberg, 1998, pp. 286–318.
[32] Y.Y. Yao, Combination of rough and fuzzy sets based an a-level sets, in: T.Y. Lin, N. Cercone (Eds.), Rough Sets and Data Mining: Analysis for Imprecise Data, Kluwer Academic Publishers, Boston, 1997, pp. 301–321.
[33] 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.
[34] Y.F. Yuan, M.J. Shaw, Introduction of fuzzy decision tree, Fuzzy Sets and Systems 69 (1995) 125–139.
[35] W.X. Zhang, W.Z. Wu, J.Y. Liang, D.Y. Li, The Theory and Methodology of Rough Sets, The Science Press, Beijing, 2001 (in Chinese).
[36] W. Zhu, Topological approaches to covering rough sets, Information Sciences 177 (2007) 1499–1508.
[37] W. Zhu, F.Y. Wang, Some results on the covering generalized rough sets, International Journal of Pattern Recognition and Artificial Intelligence 5 (2002) 6–13.
[38] W. Zhu, F.Y. Wang, Reduction and axiomization of covering generalized rough sets, Information Sciences 152 (2003) 217–230.