Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 161 (2010) 1871 – 1883
FUZZY sets and systems
www.elsevier.com/locate/fss

Local reduction of decision system with fuzzy rough sets

Chen Degang[a,*], Zhao Suyun[b]

[a]*Department of Mathematics and Physics, North China Electric Power University, Beijing 102206, PR China*
[b]*Department of Mathematics and Computer Science, Hebei University, Baoding, Hebei 071002, PR China*

Received 8 May 2008; received in revised form 24 August 2009; accepted 15 December 2009
Available online 23 December 2009

Abstract

One important and valuable topic in fuzzy rough sets is attribute reduction of decision system. The existing attribute reductions with fuzzy rough sets consider all decision classes together and cannot identify key conditional attributes explicitly for special decision class. In this paper we introduce the concept of local reduction with fuzzy rough sets for decision system. The local reduction can identify key conditional attribute and offer a minimal description for every single decision class. Approach of discernibility matrix is employed to investigate the structure of local reduction. At last, several experiments are performed to show that the idea of local reduction is feasible and valid.

Keywords: Decision system; Fuzzy similarity relation; Fuzzy rough sets; Local reduction

1. Introduction

Fuzzy rough sets are generalization of crisp rough sets [18] to deal with decision table with real valued conditional attributes. In existing fuzzy rough sets [5,6,18,19,23,26–28,30], a fuzzy similarity relation is employed to characterize similarity degree of two objects and develop upper and lower approximations of fuzzy sets. A primary application of fuzzy rough sets is to reduce the number of attributes in databases thereby improving the performance of applications in a number of aspects including speed, storage, and accuracy. For a decision table with real valued conditional attributes, this can be done by reducing the number of redundant conditional attributes and find a subset of the original conditional attributes that are the most informative.

Attribute reductions with fuzzy rough sets were first developed in [9] and further characterized in [1–3,8,10–14] by employing the idea to keep relative positive region of decision attribute invariant. Since relative positive region is computed as union of lower approximations of every decision class, so these attribute reductions can be viewed as global reductions. Attribute reduction with fuzzy rough sets can be viewed as a pure structural approach that only depends on dataset and need not other prior knowledge.

However, in many practical problems people always pay more attention on some special decision classes rather than other ones, and conditional attributes with closed connection to these special decision classes always draw much attention. For example, in decision-making of medical diagnosis, attributes with closed connection to the disease

* Corresponding author.
E-mail address: chengdegang@263.net (C. Degang).

0165-0114/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2009.12.010

always draw much attention than other ones, and it is clearly meaningful to identify such key attributes. Even some methods may be implicitly applied to determine which attributes influence particular decision classes [21], the existing global reductions have not explicitly considered key conditional attributes for special decision class, i.e., contribution of every conditional attribute to particular decision classes in a reduct cannot been distinguished. On the other hand, the existing research on attribute reduction with fuzzy rough sets mainly pays attention on developing algorithm to compute reducts, and less effort has been concentrated on investigation of mathematical structure of reducts. In this paper we formally introduce concept of local reduction with fuzzy rough sets for decision system. The local reduction can identify key conditional attributes and offer a minimal description for every subset of decision classes including sets of single decision class. With local reduction, attributes in the global reduction are grouped by their contribution to particular decision classes, and structure of global reduction is revealed. Furthermore, approach of discernibility matrix is employed to compute local reduction. Several experiments are performed to show that idea of local reduction is feasible and valid.

The rest of this paper is structured as follows. In Section 2 we review the existing fuzzy rough sets. In Section 3 we introduce concept of local reduction with fuzzy rough sets. In Section 4 we design algorithm to compute local reduction and perform several experiments. A conclusion is proposed following Section 4.

2. Review of fuzzy rough sets

The classical rough sets model proposed by Pawlak [22] can just deal with databases of symbolic attribute value (we omit detail review of classical rough sets and refer readers to [22]), this limits further application of rough sets. Several generalizations of classical rough sets were considered by scholars. Among these generalizations, combination of rough sets and fuzzy sets develops a powerful tool, called fuzzy rough sets, to deal with real-value attribute datasets using the similarity of samples. The existing fuzzy rough sets are mainly developed by fuzzy logic operators, we first briefly review basic contents of fuzzy logic operators found in [18,19,23,30].

A triangular norm, or shortly t-norm, is an increasing, associative and commutative mapping $T : [0,1]\times[0,1]\to[0,1]$ that satisfies the boundary condition $(\forall x\in[0,1], T(x,1)=x)$. $\vartheta_T(\alpha,\gamma)=\sup\{\theta\in[0,1] : T(\alpha,\theta)\le\gamma,\gamma\in[0,1]\}$, is called a R-implicator based on T. If T is lower semi-continuous, then ϑ_T is called residuation implication of T, or T-residuated implication. The properties of T-residuated implication ϑ_T are listed in [18,30].

A triangular conorm (shortly t-conorm) is an increasing, associative and commutative mapping $S : [0,1]\times[0,1]\to[0,1]$ that satisfies the boundary condition $(\forall x\in[0,1], S(x,0)=x)$. For a t-conorm S, in [14] σ is defined as $\sigma(a,b)=\inf\{c\in[0,1] : S(a,c)\ge b, b\in[0,1]\}$. Properties of σ are listed in [19,30].

A involutive negator N is a decreasing mapping $[0,1]\to[0,1]$ satisfying $N(0)=1$, $N(1)=0$ and $N(N(x))=x$ for all.

Now we begin to summarize existing fuzzy rough sets.

Let U be a nonempty universe. A fuzzy T-similarity relation R is a fuzzy relation R on U which is reflexive ($R(x,x)=1$), symmetric ($R(x,y)=R(y,x)$) and T-transitive ($R(x,y)\ge T(R(x,z),R(z,y))$, for every $x,y,z\in U$. The concept of fuzzy rough set was first proposed by Dubois and Prade [5,6] and then study in detail in [18,19,23,26,27,30]. These fuzzy rough sets can be summarized as the following four operators:

(1) T-upper approximation operator: $\overline{R_T}A(x)=\sup_{u\in U}T(R(x,u),A(u))$.
(2) S-lower approximation operator: $\underline{R_S}A(x)=\inf_{u\in U}S(N(R(x,u)),A(u))$.
(3) σ-upper approximation operator: $\overline{R_\sigma}A(x)=\sup_{u\in U}\sigma(N(R(x,u)),A(u))$.
(4) ϑ-lower approximation operator: $\underline{R_\vartheta}A(x)=\inf_{u\in U}\vartheta(R(x,u),A(u))$.

for every fuzzy set $A\in F(U)$, here R is an arbitrary fuzzy relation.

These fuzzy rough sets have been studied in detail in [30] from constructive, axiomatic, lattice and fuzzy topological viewpoints. Along this line, in [17] fuzzy rough sets are discussed on two different universe, while in [4,15] fuzzy rough sets are developed on different types of lattices.

The existing fuzzy rough sets are mainly applied to reduce number of conditional attributes in a decision system. Attribute reduction with fuzzy rough sets was first proposed in [9], they defined dependency function to measure goodness of attributes by fuzzy rough sets proposed in [5] and designed an algorithm to compute a reduct. This algorithm was tested with some practical datasets such as web categorization and was claimed to perform well.

The existing researches on attribute reduction with fuzzy rough sets [1,8,10–14] mainly pay attention on improving method in [9]. In [1] they improved computational efficiency of algorithm in [9] on a compact computational domain. In [8] algorithm to compute reducts is developed with information entropy of fuzzy rough sets. In [10–14] algorithm in [9] was improved and a fuzzy discernibility matrix is also developed in [14]. In [3] attribute reduction with general fuzzy rough sets were defined for decision system with real valued conditional attributes and crisp decision attribute, but algorithm to compute reducts and comparisons with other methods were not proposed.

All these mentioned attribute reductions are developed along the idea to keep positive region of decision attribute invariant, and positive region is computed as union of lower approximations of every decision class, so these attribute reductions can be viewed as global reductions, and they have not implicitly considered key conditional attributes for special decision class, i.e., contribution of every conditional attribute to particular decision classes in a reduct has not been distinguished explicitly.

3. Local reduction with fuzzy rough sets

In this section we first define local reduction with fuzzy rough sets, we then investigate structure of local reduction and develop approach of discernibility matrix to compute local reduction, and finally we employ an example to illustrate the idea of local reduction.

Suppose U is a finite universe of discourse, R is a fuzzy T-similarity relation on U. The following theorem describes granular structure of fuzzy rough sets, which plays key role to investigate structure of local reduction.

Theorem 3.1 (*Chen et al. [3]*). $\underline{R_\vartheta} A = \cup\{\overline{R_T} x_\lambda : \overline{R_T} x_\lambda \subseteq A\}$, $\overline{R_T} A = \cup\{\overline{R_T} x_{A(x)} : x \in U\}$;
$\underline{R_S} A = \cup\{\overline{R_\sigma} x_\lambda : \overline{R_\sigma} x_\lambda \subseteq A\}$, $\overline{R_\sigma} A = \cup\{\overline{R_\sigma} x_{A(x)} : x \in U\}$, *here x_λ is a fuzzy set, called a fuzzy point* [16], *defined as*

$$x_\lambda(z) = \begin{cases} \lambda, & z = x \\ 0, & z \neq x \end{cases} \qquad for\ \lambda \in (0, 1].$$

According to Theorem 3.1, $\underline{R_\vartheta}$ and $\overline{R_T}$ can be grouped together since they take fuzzy sets $\overline{R_T} x_\lambda$ as basic granules, while $\underline{R_S}$ and $\overline{R_\sigma}$ can be grouped together since they take fuzzy sets $\overline{R_\sigma} x_\lambda$ as basic granules. In this paper we first consider local reduction with the pair $\underline{R_\vartheta}$ and $\overline{R_T}$, conclusion for the pair $\underline{R_S}$ and $\overline{R_\sigma}$ can be similarly obtained.

Following we consider local reduction in a decision system with real valued conditional attributes and symbolic decision attribute, conditional attributes with real values will be called fuzzy attributes. For every real valued attribute, a fuzzy T-similarity relation can be employed to measure the similar degree between every pair of objects [7]. If we substitute every real valued attribute by its corresponding fuzzy T-similarity relation and substitute the decision attribute by its corresponding equivalence relation, we can get a T-fuzzy decision system consisting of three parts, a finite universe of discourse, a family of conditional fuzzy attributes and a symbolic decision attribute. Thus every dataset with real value conditional attributes and symbolic decision attribute can be expressed as a T-fuzzy decision system so that it is convenient to deal with by techniques of fuzzy rough sets.

Suppose $\mathbf{R}$ is a finite set of fuzzy T-similarity relations called conditional attribute set, D is a decision attribute with symbolic values, $U/D = \{D_1, D_2, \ldots, D_k\}$ is the collection of decision classes with respect to D, then $(U, \mathbf{R} \cup D)$ is a T-fuzzy decision system. Denote $Sim(\mathbf{R}) = \cap\{R : R \in \mathbf{R}\}$ such that

$$Sim(\mathbf{R})(x, y) = \min\{R(x, y) : R \in \mathbf{R}\}, \tag{3.1}$$

then $Sim(\mathbf{R})$ is also a fuzzy T-similarity relation. Suppose $\mathbf{K} \subseteq U/D$, i.e., $\mathbf{K}$ is the collection of part decision classes. The positive region of $\mathbf{K}$ relative to $Sim(\mathbf{R})$ is defined as $Pos_{Sim(\mathbf{R})}\mathbf{K} = \cup_{K \in \mathbf{K}} \underline{Sim(\mathbf{R})_\vartheta}(K)$. Here $\underline{Sim(\mathbf{R})_\vartheta}(K)$ is computed by

$$\underline{Sim(\mathbf{R})_\vartheta}(K)(x) = \inf_{u \in U} \vartheta(Sim(\mathbf{R})(x, u), K(u)). \tag{3.2}$$

We will say that R is dispensable relative to $\mathbf{K}$ in $\mathbf{R}$ if $Pos_{Sim(\mathbf{R})}\mathbf{K} = Pos_{Sim(\mathbf{R}-\{R\})}\mathbf{D}$, otherwise we will say R is indispensable relative to $\mathbf{K}$ in $\mathbf{R}$. The family $\mathbf{R}$ is independent relative to $\mathbf{K}$ if each $R \in \mathbf{R}$ is indispensable relative to $\mathbf{K}$ in $\mathbf{R}$; otherwise $\mathbf{R}$ is dependent relative to $\mathbf{K}$. $\mathbf{P} \subset \mathbf{R}$ is a local reduction of $\mathbf{R}$ relative to $\mathbf{K}$ if $\mathbf{P}$ is independent relative to $\mathbf{K}$ and $Pos_{Sim(\mathbf{R})}\mathbf{K} = Pos_{Sim(\mathbf{P})}\mathbf{D}$. The collection of all the indispensable elements relative to $\mathbf{K}$ in $\mathbf{R}$ is called the local core of $\mathbf{R}$ relative to $\mathbf{K}$, denoted as $Core_{\mathbf{K}}(\mathbf{R})$. Similar to the corresponding result in classical rough sets we have $Core_{\mathbf{K}}(\mathbf{R}) = \cap Red_{\mathbf{K}}(\mathbf{R})$, $Red_{\mathbf{K}}(\mathbf{R})$ is the collection of all relative reduction of $\mathbf{R}$.

If $\mathbf{K} = U/D$, then the local reduction is just the global one in [1,3,9]. Comparing with global reductions [1,3,9], the local reduction improve the global reduction by replacing decision D by $\mathbf{K}$. This improvement seems easily achieved, but it has theoretical and practical advantages. From theoretical viewpoint, attributes in global reduction and core can be grouped by their contribution to particular decision classes, thus structure of global reduction and core can be revealed. From practical viewpoint, key attributes for interesting decision classes can be captured, and less attributes can be employed to perform learning tasks.

If $\mathbf{K} = \{K\}$, for every $x \in K$, $\underline{Sim(\mathbf{R})_\vartheta}(K)(x)$ is the certain degree of $x \in K$ according to $\mathbf{R}$, thus local reduction relative to $\mathbf{K} = \{K\}$ is just the minimal subset of $\mathbf{R}$ to keep this certain degree invariant for every $x \in K$, while every attribute in the local core is a key one to keep certain degree of special $x \in K$.

Theorem 3.2. *If* $\mathbf{K} = \{K_1, K_2\}$, *then* $Core_{\mathbf{K}}(\mathbf{R}) = Core_{\{K_1\}}(\mathbf{R}) \cup Core_{\{K_2\}}(\mathbf{R})$.

Proof. Clearly we have $Pos_{Sim(\mathbf{R})}\{K_1\} \cap Pos_{Sim(\mathbf{R})}\{K_2\} = \phi$ since $K_1 \cap K_2 = \phi$, this implies $Pos_{Sim(\mathbf{R})}\mathbf{K} = Pos_{Sim(\mathbf{R})}\{K_1\} \cup Pos_{Sim(\mathbf{R})}\{K_2\}$. Thus we have $a \in Core_{\mathbf{K}}(\mathbf{R})$ if and only if there exists $x \in K_1$ such that $\underline{Sim(\mathbf{R} - \{a\})_\vartheta}(K_1)(x) < \underline{Sim(\mathbf{R})_\vartheta}(K_1)(x)$ or $x \in K_2$ such that $\underline{Sim(\mathbf{R} - \{a\})_\vartheta}(K_2)(x) < \underline{Sim(\mathbf{R})_\vartheta}(K_2)(x)$, and if and only if $a \in Core_{\{K_1\}}(\mathbf{R})$ or $a \in Core_{\{K_2\}}(\mathbf{R})$, and if and only if $a \in Core_{\{K_1\}}(\mathbf{R}) \cup Core_{\{K_2\}}(\mathbf{R})$. □

Theorem 3.2 also holds when there are more than two elements in $\mathbf{K}$. According to Theorem 3.2, every attribute in local core is indispensable for certain decision class. So if we pay more attention to special decision class, then local reduction relative to this special decision class can offer minimal description, and local core offers key conditional attributes for this special decision class.

Following we study under what conditions that $\mathbf{P} \subset \mathbf{R}$ could be a local reduction of $\mathbf{R}$ relative to $\mathbf{K}$. The structure of $\underline{R_\vartheta}(K)$ is clear by Theorem 3.1, and we will apply this fact in the following discussion. For $y \notin K$, clearly $\underline{R_\vartheta}(K)(y) = 0$ holds. For $y \in K$ the following theorem develops a sufficient and necessary condition for $\overline{R_T} y_\lambda \subseteq \underline{R_\vartheta}(K)$.

Theorem 3.3. *Suppose* $y \in K$, $\overline{R_T} y_\lambda \subseteq \underline{R_\vartheta}(K)$ *if and only if* $\overline{R_T} y_\lambda(z) = 0$ *for* $z \notin K$.

Proof. If $\overline{R_T} y_\lambda \subseteq \underline{R_\vartheta}(K)$, then $\overline{R_T} y_\lambda(z) = 0$ for $z \notin K$ is clear. Conversely, suppose $\overline{R_T} y_\lambda(z) = 0$ for $z \notin K$. We have $\overline{R_T} y_\lambda \subseteq K$ by $K(u) = 1$ for every $u \in K$, this implies $\overline{R_T} y_\lambda \subseteq \underline{R_\vartheta}(K)$ hold. □

Theorem 3.3 is key point to develop sufficient and necessary condition to characterize local reduction of $\mathbf{R}$ relative to $\mathbf{K}$.

Theorem 3.4. *Suppose* $\mathbf{P} \subset \mathbf{R}$, $Pos_{Sim(\mathbf{R})}\mathbf{K} = Pos_{Sim(\mathbf{P})}\mathbf{K}$ *if and only if* $\overline{Sim(\mathbf{P})_T} x_{\lambda(x)} \subseteq K$ *for every* $x \in K \in \mathbf{K}$, *here* $\lambda(x) = \underline{Sim(\mathbf{R})_\vartheta}(K)(x)$.

Proof. Since every two different decision classes in $\mathbf{K}$ have empty overlap, we know to keep $Pos_{Sim(\mathbf{R})}\mathbf{K} = Pos_{Sim(\mathbf{P})}\mathbf{K}$ is equivalent to keep $\underline{Sim(\mathbf{R})_\vartheta}(K) = \underline{Sim(\mathbf{P})_\vartheta}(K)$ for every $K \in \mathbf{K}$, and the latter statement is equivalent to $\overline{Sim(\mathbf{P})_T} x_{\lambda(x)} \subseteq K$ for every $x \in K \in \mathbf{K}$. □

We have the following two theorems to characterize local reduction by Theorems 3.3 and 3.4.

Theorem 3.5. *Suppose* $\mathbf{P} \subset \mathbf{R}$, *then* $\mathbf{P}$ *contains a local reduction of* $\mathbf{R}$ *relative to* $\mathbf{K}$ *if and only if* $\overline{Sim(\mathbf{P})_T} x_{\lambda(x)}(z) = 0$ *for* $x \in K \in \mathbf{K}$ *and* $z \notin K$, *here* $\lambda(x) = \underline{Sim(\mathbf{R})_\vartheta}(K)(x)$.

Theorem 3.6. *Suppose* $\mathbf{P} \subset \mathbf{R}$, *then* $\mathbf{P}$ *contains a local reduction of* $\mathbf{R}$ *relative to* $\mathbf{K}$ *if and only if there exists* $P \in \mathbf{P}$ *such that* $T(P(x, z), \lambda(x)) = 0$ *for* $x \in K \in \mathbf{K}$ *and* $z \notin K$.

Proof. For $x \in K \in \mathbf{K}$ and $z \notin K$, $\overline{Sim(\mathbf{P})_T} x_{\lambda(x)}(z) = \sup_{y \in U} T(Sim(\mathbf{P})(z, y), x_{\lambda(x)}(y)) = \min\{T(P(x, z), \lambda(x)) : P \in \mathbf{P}\}$, thus we finish the proof. $\square$

Clearly $\mathbf{P}$ is a local reduction of $\mathbf{R}$ relative to $\mathbf{K}$ if and only if $\mathbf{P}$ is the minimal subset of $\mathbf{R}$ satisfying conditions in Theorems 3.5 and 3.6. With above discussion, we can design an algorithm to compute all local reduction of $\mathbf{R}$ relative to $\mathbf{K}$. Suppose $U = \{x_1, x_2, \ldots, x_n\}$, $\mathbf{R} = \{R_1, R_2, \ldots, R_m\}$. By $M_{\mathbf{K}}(U, \mathbf{R})$ we denote a $n \times n$ matrix (c_{ij}), called the discernibility matrix of $(U, \mathbf{R} \cup \mathbf{K})$, such that

$$(1)\ c_{ij} = \{R \in \mathbf{R} : T(R(x_i, x_j), \lambda(x_i)) = 0\} \quad \text{if } x_j \notin [x_i]_D \in \mathbf{K}; \tag{3.3}$$

(2) $c_{ij} = \phi$, otherwise.

$M_{\mathbf{K}}(U, \mathbf{R})$ may not be symmetric and clearly $c_{ii} = \phi$. $R \in c_{ij}$ implies $\overline{R_T}((x_i)_{\lambda(x_i)})(x_j) = 0$, thus c_{ij} is the collection of conditional attributes to ensure $\overline{R_T}((x_i)_{\lambda(x_i)})(x_j) = 0$ for $x_j \notin [x_i]_D \in \mathbf{K}$.

Furthermore, if $\mathbf{K} = \{K\}$, then for $x_j \in K$, we have every element in the *j*th line of $M_{\mathbf{K}}(U, \mathbf{R}\}$ is ϕ. Thus if $\mathbf{K} = \{K_1, K_2\}$, and $M_{\mathbf{K}}(U, \mathbf{R}\} = (c_{ij})$, $M_{\{K_1\}}(U, \mathbf{R}\} = (c'_{ij})$ and $M_{\{K_2\}}(U, \mathbf{R}\} = (c''_{ij})$, then we have $c_{ij} = c'_{ij} \cup c''_{ij}$ and $c'_{ij} \cap c''_{ij} = \phi$. This statement can be examined by Example 3.11. Generally, the union of two local reductions related to $\{K_1\}$ and $\{K_1\}$, respectively, may not be, but must include a local reduction related to $\mathbf{K} = \{K_1, K_2\}$. These results also holds when there are more than two elements in $\mathbf{K}$, and they will be examined by Example 3.11.

We have the following theorem for the relative core to compute indispensable attributes for particular decision classes.

Theorem 3.7. $Core_{\mathbf{K}}(\mathbf{R}) = \{R : c_{ij} = \{R\}\}$ *for some* $1 \le i, j \le n$.

Proof. $R \in Core_{\mathbf{K}}(\mathbf{R}) \Leftrightarrow Pos_{Sim(\mathbf{R})}\mathbf{K} \ne Pos_{Sim(\mathbf{R}-\{R\})}\mathbf{K} \Leftrightarrow$ there exists $x_i \in [x_i]_D \in \mathbf{K}$ such that $\underline{Sim(\mathbf{R})_\vartheta}([x_i]) \ne \underline{Sim(\mathbf{R} - \{R\})_\vartheta}([x_i]_D) \Leftrightarrow$ there exists $x_j \in U$ such that $T(R(x_i, x_j), \lambda(x_i)) = 0$, and $T(R'(x_i, x_j), \lambda(x_i)) > 0$ holds for any other $R' \in \mathbf{R} \Leftrightarrow c_{ij} = \{R\}$.

The statement $c_{ij} = \{R\}$ implies that R is the unique attribute to maintain $T(R(x_i, x_j), \lambda(x_i)) = 0$ for $x_j \notin [x_i]_D \in \mathbf{K}$. $\square$

Theorem 3.8. *Suppose* $\mathbf{P} \subset \mathbf{R}$, *then* $\mathbf{P}$ *contains a local reduction of* $\mathbf{R}$ *relative to* $\mathbf{K}$ *if and only if* $\mathbf{P} \cap c_{ij} \ne \phi$ *for every* $c_{ij} \ne \phi$.

The proof is straightforward by Theorem 3.6 and definition of c_{ij}.

Theorem 3.9. *Suppose* $\mathbf{P} \subset \mathbf{R}$, *then* $\mathbf{P}$ *contains a local reduction of* $\mathbf{R}$ *relative to* $\mathbf{K}$ *if and only if* $\mathbf{P}$ *is the minimal set satisfying* $\mathbf{P} \cap c_{ij} \ne \phi$ *for every* $c_{ij} \ne \phi$.

A discernibility function $f_{\mathbf{K}}(U, \mathbf{R})$ for $(U, \mathbf{R} \cup \mathbf{K})$ is a Boolean function of m Boolean variables $\overline{R_1}, \overline{R_2}, \ldots, \overline{R_m}$ corresponding to the fuzzy attributes $R_1, R_2, \ldots, R_m$, respectively, and defined as follows: $f_{\mathbf{K}}(U, \mathbf{R})(\overline{R_1}, \overline{R_2}, \ldots, \overline{R_m}) = \wedge\{\vee(c_{ij}) : \underline{c_{ij}} \ne \phi\}$, where $\vee(c_{ij})$ is the disjunction of all variables $\overline{R}$ such that $R \in c_{ij}$. In the sequel we shall write R_i instead of $\overline{R_i}$ when no confusion can arise.

Let $g_{\mathbf{K}}(U, \mathbf{R})$ be the reduced disjunctive form of $f_{\mathbf{K}}(U, \mathbf{R})$ obtained from $f_{\mathbf{K}}(U, \mathbf{R})$ by applying the multiplication and absorption laws as many times as possible. Then there exist l and $\mathbf{R}_k \subseteq \mathbf{R}$ for $k = 1, 2, \ldots, l$ such that $g_{\mathbf{K}}(U, \mathbf{R}) = (\wedge\mathbf{R}_1) \vee \cdots \vee (\wedge\mathbf{R}_l)$ where every element in $\mathbf{R}_k$ only appears one time. We have the following theorem.

Theorem 3.10. $Red_{\mathbf{K}}(\mathbf{R}) = \{\mathbf{R}_1, \ldots, \mathbf{R}_l\}$.

Proof. For every $t = 1, 2, \ldots, l$ and $c_{ij} \ne \phi$, we have $\wedge\mathbf{R}_t \le \vee c_{ij}$ by $\vee_{k=1}^{l}(\wedge\mathbf{R}_t) = \wedge\{\vee c_{ij} : c_{ij} \ne \phi\}$, so $\mathbf{R}_t \cap c_{ij} \ne \phi$ for every $c_{ij} \ne \phi$. Let $R \in \mathbf{R}_t$ and $\mathbf{R}'_t = \mathbf{R}_t - \{R\}$, then $g_{\mathbf{K}}(U, \mathbf{R}) < \vee_{r=1}^{t-1}(\wedge R_r) \vee (\wedge R'_t) \vee (\vee_{r=t+1}^{l}(\wedge R_r))$. If for every $c_{ij} \ne \phi$, we have $\mathbf{R}'_t \cap c_{ij} \ne \phi$, then $\wedge\mathbf{R}'_t \le \vee c_{ij}$ for every $c_{ij} \ne \phi$. This implies $g_{\mathbf{K}}(U, \mathbf{R}) \ge \vee_{r=1}^{t-1}(\wedge R_r) \vee (\wedge R'_t) \vee (\vee_{r=t+1}^{l}(\wedge R_r))$

Table 1
Samples of credit card evaluation problem.

	C_1	C_2	C_3	C_4	C_5	C_6
x_1	0.8	0.1	0.1	0.5	0.2	0.3
x_2	0.3	0.5	0.2	0.8	0.1	0.1
x_3	0.2	0.2	0.6	0.7	0.3	0.2
x_4	0.6	0.3	0.1	0.2	0.5	0.3
x_5	0.3	0.4	0.3	0.3	0.6	0.1
x_6	0.2	0.3	0.5	0.3	0.5	0.2
x_7	0.3	0.3	0.4	0.2	0.6	0.2
x_8	0.3	0.4	0.3	0.1	0.4	0.5
x_9	0.3	0.2	0.5	0.4	0.4	0.2

which is a contradiction. Hence there exists $c_{i_0 j_0} \neq \phi$ such that $\mathbf{R}'_t \cap c_{i_0 j_0} = \phi$ which implies $\mathbf{R}_t$ is a relative local reduction of $\mathbf{R}$.

For every $\mathbf{X} \in Red_{\mathbf{K}}(\mathbf{R})$, we have $\mathbf{X} \cap c_{ij} \neq \phi$ for every $c_{ij} \neq \phi$, this implies $\wedge\mathbf{X} \leq g_{\mathbf{K}}(U, \mathbf{R})$. Suppose for every $t = 1, 2, \ldots, l$ we have $\mathbf{R}_t - \mathbf{X} \neq \phi$, then for every $t = 1, 2, \ldots, l$ one can find $R_t \in \mathbf{R}_t - \mathbf{X}$. By $g(U, \mathbf{R}) = (\wedge\mathbf{R}_1) \vee \cdots \vee (\wedge\mathbf{R}_l)$ we have $g(U, \mathbf{R}) = (\vee_{k=1}^{l} R_t) \wedge \Phi$, here Φ is a conjunction of disjunctions of elements belong to different $\mathbf{R}_k$ which implies $\wedge\mathbf{X} \leq \vee_{t=1}^{l} R_t$. So there is R_{k_0} such that $\wedge\mathbf{X} \leq R_{k_0}$, this implies $R_{k_0} \in \mathbf{X}$ which is a contradiction. Thus there exists k' such that $\mathbf{R}_{k'} - \mathbf{X} = \phi$, which implies $\mathbf{R}_{k'} \subseteq \mathbf{X}$. Since both $\mathbf{R}_{k'}$ and $\mathbf{X}$ are relative reductions, we have $\mathbf{R}_{k'} = \mathbf{X}$. Hence we have $Red_{\mathbf{K}}(\mathbf{R}) = \{\mathbf{R}_1, \ldots, \mathbf{R}_l\}$. $\square$

Now we set up a framework of local reduction with the pair approximation operators $\underline{R_\vartheta}$ and $\overline{R_T}$, definition of local reduction and results for the pair $\underline{R_S}$ and $\overline{R_\sigma}$ can be obtained similarly as Theorems 3.3–3.10 by taking fuzzy sets $\overline{R_\sigma} x_\lambda$ as basic granules.

Following we employ an example to illustrate our idea in this section.

Example 3.11. Let us consider an evaluation problem of credit card applicants. Suppose $U = \{x_1, \ldots, x_9\}$ is a set of nine applicants, every applicant is described by six fuzzy attributes: C_1 = best education, C_2 = better education, C_3 = good education, C_4 = high salary, C_5 = middle salary and C_6 = low salary. The membership degrees of every applicant are given in Table 1.

Every fuzzy attribute C_k can define a T_L-fuzzy similarity relation R_k as $R_k(x_i, x_j) = 1 - |C_k(x_i) - C_k(x_j)|$, here $T_L(x, y) = \max\{0, x + y - 1\}$, $\vartheta_L(x, y) = \min\{1, 1 - x + y\}$ $Sim(\mathbf{R})$ is computed as follows:

$$(Sim(\mathbf{R})(x_i, x_j)) = \begin{pmatrix} 1 & 0.5 & 0.4 & 0.7 & 0.5 & 0.4 & 0.5 & 0.5 & 0.5 \\ & 1 & 0.6 & 0.4 & 0.5 & 0.5 & 0.4 & 0.3 & 0.6 \\ & & 1 & 0.5 & 0.6 & 0.6 & 0.5 & 0.4 & 0.7 \\ & & & 1 & 0.7 & 0.6 & 0.7 & 0.7 & 0.6 \\ & & & & 1 & 0.8 & 0.9 & 0.6 & 0.8 \\ & & & & & 1 & 0.9 & 0.7 & 0.9 \\ & & & & & & 1 & 0.7 & 0.8 \\ & & & & & & & 1 & 0.7 \\ & & & & & & & & 1 \end{pmatrix}.$$

Suppose a decision partition is $A = \{x_1, x_2, x_4, x_7\}$, $B = \{x_3, x_5, x_6, x_8, x_9\}$, then

$$\underline{Sim(\mathbf{R})_{\vartheta_L}}(A)(x) = \begin{cases} 0.5, & x = x_1, \\ 0.4, & x = x_2, \\ 0.3, & x = x_4, \\ 0.1, & x = x_7, \\ 0, & \text{otherwise}, \end{cases} \quad \underline{Sim(\mathbf{R})_{\vartheta_L}}(B)(x) = \begin{cases} 0.4, & x = x_3, \\ 0.1, & x = x_5, \\ 0.1, & x = x_6, \\ 0.3, & x = x_8, \\ 0.2, & x = x_9, \\ 0, & \text{otherwise} \end{cases}$$

and we have discernibility matrixes $M_{\{A\}}(U, \mathbf{R})$, $M_{\{B\}}(U, \mathbf{R})$ and $M_{\{A,B\}}(U, \mathbf{R})$ as follows:

$$M_{\{A\}}(U,\mathbf{R}) = (c_{ij})$$
$$= \begin{pmatrix} \phi & \phi & \{1\} & \phi & \{1\} & \{1\} & \phi & \{1\} & \{1\} \\ \phi & \phi & \{3\} & \phi & \{4,5\} & \{4,5\} & \phi & \{4,6\} & \{4\} \\ \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \\ \phi & \phi & \{1,3,4\} & \phi & \{1\} & \{1,3\} & \phi & \{1\} & \{1,3\} \\ \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \\ \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \\ \phi & \phi & \{1,2,3,4,5\} & \phi & \{2,3,4,6\} & \{1,3,4,5\} & \phi & \{2,3,4,5,6\} & \{2,3,4,5\} \\ \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \\ \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \end{pmatrix},$$

$$M_{\{B\}}(U,\mathbf{R}) = (c_{ij})$$
$$= \begin{pmatrix} \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \\ \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \\ \{1,3\} & \{3\} & \phi & \{1,3,4\} & \phi & \phi & \{4\} & \phi & \phi \\ \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \\ \{1,2,3,4,5,6\} & \{2,3,4,5\} & \phi & \{1,2,3,4,5,6\} & \phi & \phi & \{2,3,4,6\} & \phi & \phi \\ \{1,2,3,4,5,6\} & \{1,2,3,4,5,6\} & \phi & \{1,3,4,6\} & \phi & \phi & \{1,3,4,5\} & \phi & \phi \\ \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi & \phi \\ \{1,2,4\} & \{4,5,6\} & \phi & \{1\} & \phi & \phi & \{6\} & \phi & \phi \\ \{1,3,5\} & \{2,3,4,5\} & \phi & \{1,3,4\} & \phi & \phi & \{4,5\} & \phi & \phi \end{pmatrix},$$

$$M_{\{A,B\}}(U,\mathbf{R}) = (c_{ij})$$
$$= \begin{pmatrix} \phi & \phi & \{1\} & \phi & \{1\} & \{1\} & \phi & \{1\} & \{1\} \\ \phi & \phi & \{3\} & \phi & \{4,5\} & \{4,5\} & \phi & \{4,6\} & \{4\} \\ \{1,3\} & \{3\} & \phi & \{1,3,4\} & \phi & \phi & \{4\} & \phi & \phi \\ \phi & \phi & \{1,3,4\} & \phi & \{1\} & \{1,3\} & \phi & \{1\} & \{1,3\} \\ \{1,2,3,4,5,6\} & \{2,3,4,5\} & \phi & \{1,2,3,4,5,6\} & \phi & \phi & \{2,3,4,6\} & \phi & \phi \\ \{1,2,3,4,5,6\} & \{1,2,3,4,5,6\} & \phi & \{1,3,4,6\} & \phi & \phi & \{1,3,4,5\} & \phi & \phi \\ \phi & \phi & \{1,2,3,4,5\} & \phi & \{2,3,4,6\} & \{1,3,4,5\} & \phi & \{2,3,4,5,6\} & \{2,3,4,5\} \\ \{1,2,4\} & \{4,5,6\} & \phi & \{1\} & \phi & \phi & \{6\} & \phi & \phi \\ \{1,3,5\} & \{2,3,4,5\} & \phi & \{1,3,4\} & \phi & \phi & \{4,5\} & \phi & \phi \end{pmatrix},$$

where $i \in c_{ij}$ means $R_i \in c_{ij}$, $i = 1, \ldots, 6$.

We can get that $\{R_1, R_3, R_4\}$ is the only local reduction of $\mathbf{R}$ relative to $\{A\}$, and $\{R_1, R_3, R_4, R_6\}$ is the only local reduction of $\mathbf{R}$ relative to $\{B\}$ and $\{A, B\}$. Clearly every c_{ij} in $M_{\{A,B\}}(U, \mathbf{R})$ is the union of corresponding ones in $M_{\{A\}}(U, \mathbf{R})$ and $M_{\{B\}}(U, \mathbf{R})$.

4. Algorithm to find local reductions and numerical experiments

In this section we propose one detail algorithm to find local reduction and then we present some numerical experiments to illustrate the propose method in this paper is feasible and necessary in real applications.

4.1. Algorithm to find local reductions

In this subsection we design one algorithm (Heuristic) to find one local reduction by approach of discernibility matrix proposed in Section 3.

Table 2
The information of some datasets from UCI.

Datasets	Objects	Data type (condition attributes)	Attributes	Fuzzified attributes	Decision classes
Ionosphere	351	Real number	34	99	2
Tae	151	Real number	6	15	3
Yeast	1484	Real number	9	24	9[a]
Iris	150	Real number	5	12	3
New_thyroid	215	Real number	6	15	3
WDBC	569	Real number	31	91	2
Wine	178	Real number	14	39	3

[a] We omit one decision class from this dataset since only five objects belong to the deleted decision class.

Algorithm 4.1. To find one reductions for certain decision class K in $(U, \mathbf{R}, D)$.

Input: $reduct \leftarrow \{\}$
Step 1: Compute the similarity relation of the set of all condition attributes: $SIM(\mathbf{R})$ by formulae (3.1);
Step 2: Compute $\underline{SIM(\mathbf{R})_{\vartheta}}(K)(x)$ for every $x \in U$ by formula (3.2);
Step 3: Compute c_{ij} by formulae (3.3);
Step 4: Compute $Core_K(\mathbf{R}) = \{a : c_{ij} = \{a\}\}$; Delete those c_{ij} with nonempty overlap with $Core_K(\mathbf{R})$;
Step 5: Let $Reduct = Core_K(\mathbf{R})$;
Step 6: Add the element *a* whose frequency of occurrence is maximum in all c_{ij} and into *Reduct*; and delete those c_{ij} with nonempty overlap with *Reduct*;
Step 7: If there still exist some $c_{ij} \neq \phi$, go to Step 6; Otherwise, goto Step 8;
Step 8: If *Reduct* is not independent, delete the redundant elements in *Reduct*;
Step 9: Output *Reduct*.

This algorithm is similar to the standard Johnson Reducer in crisp rough sets [20]. However, it is interesting to design other algorithms to compute reducts and will be our future work since this section mainly focuses on examination of feasibility of our idea of local reduction.

The computational complexity of this algorithm is $O(|U|^2 \times |A|)$. Here $|U|$ is the size of the universe, $|A|$ is the number of attributes. By using this computational complexity analysis, we find that to find one local reduction is similar to find one global reduction from the viewpoint of computational complexity. Here global reduction represents the reduction find based on all decision classes.

4.2. Numerical experiments

We select several datasets from UCI Machine Learning Repository [25], whose information is summarized in Table 2, to illustrate the proposed method in this paper. A simple algorithm [29] is used to generate triangular membership function for every conditional attribute so that we get the fuzzified attribute.

The experiments are setup as follows.

Experimental setup:

Dataset: Six datasets, e.g. 'Iris', 'Wine' are selected.

Classifier: Nearest neighbor algorithm in [24] is used to be the classifier.

Dataset Split: In process of classification, dataset after attribute reduction is split into two parts. The randomly chosen 50% of objects are used as the training set. The remainder is used as the testing set. Classification result is the average of 20 times training and testing.

Indices: They are (1) number of selected attributes in the reduct, (2) classification accuracy of reduct.

Triangular norm: Since "MIN" operator is often used in attribute reduction methods [1,8,9], we select it in the following experiments.

Accuracy: The accuracy in this paper means the rate of accurate classification on specific decision class. It is calculated by *A/B*. Here *A* is the number of accurate classification on certain decision class of certain dataset, *B* is the number of objects belonging to this specific decision class of this specific dataset.

Table 3
The comparison on attribute selection between local reduction and global reduction.

	Decision class 1	Decision class 2	Decision class 3
Iris			
Local reduction	[8 10 4 11]	[8 7 4 10 11 1]	[5 8 11 10 1]
Global reduction	[8 10 4 11 7 5 1]	[8 7 4 10 11 1 5]	[5 8 11 10 1 4 7]
New_thyroid			
Local reduction	[8 11 14 4 1 5 2 13]	[2 5 7 4]	[2 5 13 10 4]
Global reduction	[8 11 14 4 1 5 2 13 7 10]	[2 5 7 4 14 1 8 11 13 10]	[2 5 13 10 4 14 11 8 1 7]
Wine			
Local reduction	[37 34 1 14 31 29 2 5]	[1 29 5 34 20 31 14 2 37 13]	[20 32 17 35 29 1 14]
Global reduction	[37 34 1 14 31 29 2 5 20 13]	[1 29 5 34 20 31 14 2 37 13]	[20 32 17 35 29 1 14 37 5 2 10]

Experimental results on these datasets are summarized in Tables 3–6. Here global reduction represents the reductions obtained based on all decision classes.

We first take dataset New_thyroid to illustrate the advantages of local reduction. There are 15 fuzzified attributes with this dataset. Numbers attributes in local reductions relative to Decision class 1, 2 and 3 are 8, 4 and 5, respectively, these three local reductions generate the same global reduction with 10 attributes. Thus attributes in the global reduction are grouped according to their contribution to particular decision class. Certainly these groups may have overlaps. On the other hand, classification accuracies with local reductions are similar to the ones with global reduction and are clearly better than the ones with remainder attributes, this demonstrates the advantages of local reduction.

Table 3 presents that detailed comparison on attribute reduction between local reduction and global reduction. In Table 3 each global reduction is obtained in this way: first to find the local reduction for certain decision class, then check whether this local reduction is a global reduction: if yes, stop; if no, add some attributes until it become a global reduction. For example, one local reduction of decision class 1 on dataset 'Iris' is [8 10 4 11]. Since this local reduction is not a global reduction, we add some attributes [7 5 1] and thus [8 10 4 11 7 5 1] become one global reduction on dataset 'Iris'.

Table 3 shows that local reductions are often different from global reductions. In most cases, local reduction is subset of certain global reduction. That is to say, number of attributes in local reduction is often smaller than the one of global reduction. Table 3 also shows that local reductions of different decision classes are different. All these show that it is necessary to study local reductions for different decision classes.

Next, we compare local reduction with the remainder attributes. The comparison results are summarized in Table 4. 'Remainder' in Table 4 represents remained condition attributes after removing the attributes in local reduction.

Table 4 shows that average accuracy of local reduction (i.e. 68.49039) is significantly higher than average accuracy of remainder attributes (i.e. 51.84212). What is more, number of attributes of local reduction (9.36) is far less than number of remainder attributes (24.4). All these show that local reduction is the vital part to keep information contained in fuzzy decision table invariant.

We also need to note that local reduction does not always have higher accuracy than remainder part. As part of Table 4, Table 5 shows that some times accuracy of remainder part is comparable with the one of local reduction. This is because if the core is empty, then there may have another local reduction in remainder part. For example, remainder part of data 'WDBC' may contain another local reduction in the remaining 76 attributes.

Finally, we compare local reduction with global reduction and comparison results are summarized in Table 6. In Table 6 the 'if equal ?' represents whether local reduction are equal to global reduction. One should notices that accuracies may be different even the local reduction are equal to global reduction, this is because the training samples are selected randomly, so it is possible different training samples are employed even for the same attribute set.

Table 6 shows that for most datasets local reduction does not equal to global reduction. For those datasets that local reduction does not equal to global reduction, local reduction has similar classification accuracy with global reduction, whereas number of selected attributes in local reduction is obviously less than the one of global reduction. For example, average of classification accuracy of local reduction (68.49039) is close to average of classification accuracy of global

Table 4
The comparison between the local reduction and remainder.

Dataset	Accuracy		Number of attributes	
	Local reduct	Remainder	Local reduct	Remainder
Ionosphere				
Class 1	97.176	70	17	85
Class 2	55.86	20	17	85
WDBC				
Class 1	91.566	93.212	14	76
Class 2	96.893	97.704	14	76
Tae				
Class 1	57.837	40	8	7
Class 2	28.639	25	7	8
Class 3	56.185	25	8	7
Yeast				
Class 1	63.423	45	14	10
Class 2	47.568	45	13	11
Class 3	50.988	20	12	12
Class 4	71.321	5	11	13
Class 5	22.008	0	9	15
Class 6	47.314	0	9	15
Class 7	0.9127	0	7	17
Class 8	43.472	0	7	17
Class 9	35.952	0	8	16
Iris				
Class 1	100	98.542	4	8
Class 2	95.626	85.427	6	6
Class 3	88.63	93.046	5	7
New_thyroid				
Class 1	98.3	98.25	8	7
Class 2	85.674	81.804	4	11
Class 3	86.749	76.698	5	10
Wine				
Class 1	100	94.815	8	31
Class 2	91.501	82.765	10	29
Class 3	98.665	98.79	7	32
Average	***68.49039***	51.84212	***9.36***	24.4

Table 5
Some special cases between the local reduction and remainder.

Dataset	Accuracy		Number of attributes	
	Local reduct	Remainder	Local reduct	Remainder
Ionosphere				
Class 1	97.176	***70***	17	85
Class 2	55.86	***20***	17	85
WDBC				
Class 1	91.566	***93.212***	14	76
Class 2	96.893	***97.704***	14	76

reduction (69.56336), whereas average of number of selected attributes of local reduction (9.36) is obviously less than average of number of attributes of global reduction (12.2). It is easy to get that local reduction is comparable to global reduction on the specific decision class with less number of attributes.

Table 6
Comparison between local reduction and global reduction.

Dataset	Accuracy		Number of attributes		If equal?
	Local reduct	Global reduct	Local reduct	Global reduct	
Ionosphere					
Class 1	97.176	96.885	17	17	Yes
Class 2	55.86	57.474	17	17	Yes
WDBC					
Class 1	91.566	90.359	14	14	Yes
Class 2	96.893	96.768	14	14	Yes
Tae					
Class 1	57.837	60.669	8	8	Yes
Class 2	28.639	35.403	7	8	No
Class 3	56.185	53.335	8	8	Yes
Yeast					
Class 1	63.423	60.288	14	15	No
Class 2	47.568	46.242	13	15	No
Class 3	50.988	52.826	12	15	No
Class 4	71.321	71.88	11	15	No
Class 5	22.008	27.904	9	15	No
Class 6	47.314	64.699	9	15	No
Class 7	0.9127	0	7	15	No
Class 8	43.472	32.547	7	15	No
Class 9	35.952	48.29	8	15	No
Iris					
Class 1	100	100	4	7	No
Class 2	95.626	96.251	6	8	No
Class 3	88.63	90.535	5	7	No
New_thyroid					
Class 1	98.3	98.487	8	10	No
Class 2	85.674	80.952	4	10	No
Class 3	86.749	88.658	5	10	No
Wine					
Class 1	100	100	8	10	No
Class 2	91.501	92.339	10	10	Yes
Class 3	98.665	96.293	7	11	No
Average	68.49039	69.56336	***9.36***	12.2	

To demonstrate this conclusion, we select SVM instead of Nearest neighbor algorithm to repeat the above experiments. Apply SVM-KM MATLAB Toolbox (http://asi.insa-rouen.fr/enseignants/∼arakotom/toolbox/index.html) to five selected databases. The kernel is selected as Gaussian kernel in SVM. We select 1 as the parameter of Gaussian Kernel, which is the default value (the use of the SVM-KM MATLAB Toolbox). Table 7 shows the comparisons of accuracies between local and global reduct with SVM, and these comparisons can also support our conclusion.

5. Conclusion

In many practical problems, conditional attributes with closed connection to special decision class always draw more attention. The existing attribute reduction with fuzzy rough sets lack formalism to consider key conditional attributes for special decision class, and this can be improved by the proposed local reduction with fuzzy rough sets in this

Table 7
Comparison between local reduction and global reduction with SVM.

Dataset	Accuracy		Number of attributes		If equal?
	Local reduct	Global reduct	Local reduct	Global reduct	
Ionosphere					
Class 1	97.0259	97.78126	17	17	Yes
Class 2	56.34618	53.7013	17	17	Yes
WDBC					
Class 1	94.6331	93.43666	14	14	Yes
Class 2	96.80534	96.624	14	14	Yes
Tae					
Class 1	56.68175	57.81034	8	8	Yes
Class 2	44.7536	42.13828	7	8	No
Class 3	53.00799	51.3458	8	8	Yes
Iris					
Class 1	100	100	4	7	No
Class 2	91.57113	95.30042	6	8	No
Class 3	90.92006	89.601	5	7	No
New_thyroid					
Class 1	97.93284	98.07887	8	10	No
Class 2	84.60529	88.30415	4	10	No
Class 3	90.04466	92.62506	5	10	No

paper. Experimental results imply our idea of local reduction is feasible and valid and can be applied to practical problems.

Acknowledgements

The authors are thankful to the anonymous referees for their constructive suggestions which help to improve the clarity and the completeness of the paper. This paper is supported by a grant of NSFC(70871036) and a grant of North China Electric Power University.

References

[1] R.B. Bhatt, M. Gopal, On fuzzy rough sets approach to feature selection, Pattern Recognition Letters 26 (7) (2005) 965–975.
[2] D.G. Chen, E.C.C. Tsang, S.Y. Zhao, Attributes reduction with T_L fuzzy rough sets, in: 2007 IEEE Internat. Conf. on Systems, Man, and Cybernetics, Vol. 1, 2007, pp. 486–491.
[3] D.G. Chen, X.Z.Wang, S.Y. Zhao, Attributes Reduction based on Fuzzy Rough Sets, in: RSEISP 2007, Lecture Notes in Artificial Intelligence, Vol. 4585, 2007, pp. 381–390.
[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 (11) (2007) 2308–2326.
[5] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–209.
[6] 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, Dordrecht, 1992.
[7] J.M. Fernandez Salido, S. Murakami, Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations, Fuzzy Sets and Systems 139 (2003) 635–660.
[8] Q.H. Hu, D.R. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognition Letters 27 (5) (2006) 414–423.
[9] R. Jensen, Q. Shen, Fuzzy-rough attributes reduction with application to web categorization, Fuzzy Sets and Systems 141 (2004) 469–485.
[10] R. Jensen, Q. Shen, Fuzzy-rough data reduction with ant colony optimization, Fuzzy Sets and Systems 149 (1) (2005) 5–20.
[11] Q. Shen, R. Jensen, Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, Pattern Recognition 37 (7) (2004) 1351–1363.
[12] R. Jensen, Q. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Transactions on Fuzzy Systems 15 (1) (2007) 73–89.
[13] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: rough and fuzzy rough based approaches, IEEE Transactions on Knowledge Data Engineering 16 (12) (2004) 1457–1471.

[14] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, IEEE Transactions on Fuzzy Systems 17 (4) (2009) 824–838.
[15] G.L. Liu, Generalized rough sets over fuzzy lattices, Information Sciences 178 (6) (2008) 1651–1662.
[16] Y.M. Liu, M.K. Luo, Fuzzy Topology, World Scientific, Singapore, Hong Kong, 1997.
[17] T.J. Li, W.X. Zhang, Rough fuzzy approximations on two universes of discourse, Information Sciences 178 (3) (2008) 892–906.
[18] N.N. Morsi, M.M. Yakout, Axiomatics for fuzzy rough sets, Fuzzy Sets and Systems 100 (1998) 327–342.
[19] J.S. Mi, W.X. Zhang, An axiomatic characterization of a fuzzy generalization of rough sets, Information Sciences 160 (1–4) (2004) 235–249.
[20] A. Ohrn, Discernibility and rough sets in medicine: tools and applications, Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway, 1999, p. 239.
[21] N. Mac Parthalain, R. Jensen, Q. Shen, Finding fuzzy-rough reducts with fuzzy entropy, in: Proc. 17th Internat. Conf. on Fuzzy Systems (FUZZ-IEEE'08), 2008, pp. 1282–1288.
[22] Z. Pawlak, Rough Sets Internat, Journal of Computer and Information Science 11 (5) (1982) 341–356.
[23] A.M. Radzikowska, E.E. Kerre, A comparative study of fuzzy rough sets, Fuzzy Sets and Systems 126 (2002) 137–155.
[24] D.G. Stork, Y.-T. Elad, Computer Manual in MATLAB to Accompany Pattern Classification, second ed., Wiley, Hoboken, NJ, 2004.
[25] UCI, Machine Learning Repository [online]. Available: ⟨http://www.ics.uci.edu/∼mlearn/MLRepository.html⟩, 2005.
[26] W.Z. Wu, J.S. Mi, W.X. Zhang, Generalized fuzzy rough sets, Information Sciences 151 (2003) 263–282.
[27] W.Z. Wu, W.X. Zhang, Constructive and axiomatic approaches of fuzzy approximation operators, Information Sciences 159 (3–4) (2004) 233–254.
[28] X.P. Yang, Minimization of axiom sets on fuzzy approximation operators, Information Sciences 177 (18) (2007) 3840–3854.
[29] Y.F. Yuan, M.J. Shaw, Introduction of fuzzy decision tree, Fuzzy Sets and Systems 69 (1995) 125–139.
[30] D.S. Yueng, D.G. Chen, E.C.C. Tsang, W.T.L. John, X.Z. Wang, On the generalization of fuzzy rough sets, IEEE Transactions on Fuzzy Systems 13 (3) (2005) 343–361.