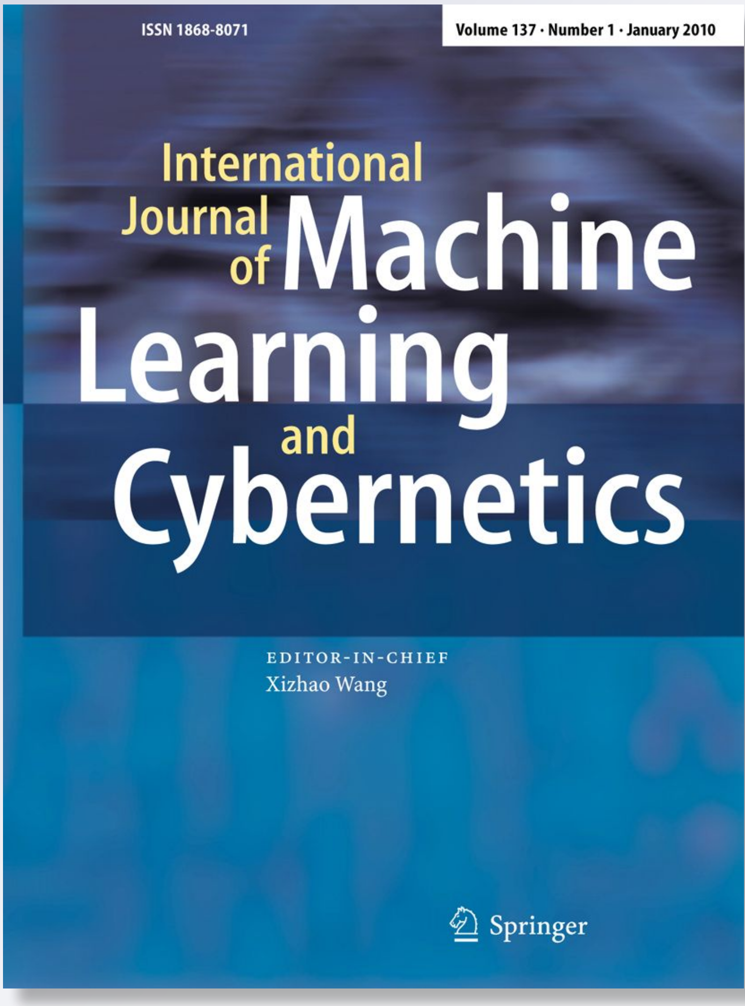
Consistency-preserving attribute reduction in fuzzy rough set framework

Yuhua Qian, Jiye Liang & Wei Wei

International Journal of Machine Learning and Cybernetics

ISSN 1868-8071

Int. J. Mach. Learn. & Cyber.
DOI 10.1007/s13042-012-0090-z



Int. J. Mach. Learn. & Cyber.
DOI 10.1007/s13042-012-0090-z

ORIGINAL ARTICLE

Consistency-preserving attribute reduction in fuzzy rough set framework

Yuhua Qian · Jiye Liang · Wei Wei

Received: 21 September 2010 / Accepted: 24 March 2012

Abstract Attribute reduction (feature selection) has become an important challenge in areas of pattern recognition, machine learning, data mining and knowledge discovery. Based on attribute reduction, one can extract fuzzy decision rules from a fuzzy decision table. As consistency is one of several criteria for evaluating the decision performance of a decision-rule set, in this paper, we devote to present a consistency-preserving attribute reduction in fuzzy rough set framework. Through constructing the membership function of an object, we first introduce a consistency measure to assess the consistencies of a fuzzy target set and a fuzzy decision table, which underlies a foundation for attribute reduction algorithm. Then, we derive two attribute significance measures based on the proposed consistency measure and design a forward greedy algorithm (ARBC) for attribute reduction from both numerical and nominal data sets. Numerical experiments show the validity of the proposed algorithm from search strategy and heuristic function in the meaning of consistency. Number of the selected features is the least for a given threshold of consistency measure.

Keywords Fuzzy rough set · Fuzzy decision tables · Consistency measure · Attribute reduction

Y. Qian (✉) · J. Liang · W. Wei
Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China
e-mail: jinchengqyh@126.com

J. Liang
e-mail: ljy@sxu.edu.cn

W. Wei
e-mail: weiwei@sxu.edu.cn

1 Introduction

Data mining and knowledge discovery from large-scale data sets is a challenging problem. In recent years, we encounter databases in which both the number of objects becomes higher and their dimensionality (number of attributes) gets larger as well. Tens, hundreds, and even thousands of attributes are stored in many real-world application databases. Attributes that are irrelevant to knowledge discovery tasks may deteriorate the performance of learning algorithms [19, 24, 25, 32, 43]. In other words, storing and processing all attributes (both relevant and irrelevant) could be computationally very expensive and impractical. To deal with this issue, as was pointed out by Hu et al. [12], some attributes can be omitted, which will not seriously impact the resulting classification (recognition) error, cf. Therefore, the omission of some attributes could not only be tolerable but even desirable relatively to the costs involved in such cases [26]. This task is often called attribute reduction or feature selection.

Rough set theory proposed by Pawlak is a popular mathematical framework for pattern recognition, image processing, feature selection, neuro computing, conflict analysis, decision support, data mining and knowledge discovery from large data sets [9, 30, 31, 32, 39, 42, 45]. Rough-set-based data analysis starts from a data table, called information systems. The information systems contain data about objects of interest, characterized by a finite set of attributes. It is often interesting to discover some dependency relationships (patterns). An information system where condition attributes and decision attributes are distinguished is called a decision table (or a decision information system). From a decision table one can induce some patterns in form of "*if*..., *then*..." decision rules. More exactly, the decision rules say that if some condition

Published online: 11 April 2012

attributes have given values, then some decision attributes have other given values.

In Pawlak's rough set model, incompletion, fuzziness and probability are not taken into consideration. Pawlak's rough set model just works in nominal data domain, for crisp equivalence relations and equivalence classes are the foundation of the model [27]. However, there are usually incompletion, fuzziness and probability of data in real-world applications. To deal with these cases, in the past 10 years, several extensions of the rough set model have been proposed in terms of various requirements, such as variable precision rough set (VPRS) model [1, 24, 54], rough set model based on tolerance relation [17, 18], Bayesian rough set model [37], fuzzy rough set model [3, 5, 40, 41, 49, 51], rough fuzzy set model [5], covering generalized rough sets [46, 52, 53], rough set model under dynamic granulation [21, 32], and multi-granulations rough set model (MGRS) [28, 29]. In a recently published paper, Hu et al. [11] established a new rough set framework called fuzzy probabilistic approximation spaces, which introduces probability into fuzzy rough set model and leads to a tool for dealing with randomness, roughness and fuzziness in real-world applications. In the fuzzy rough set model, for a fuzzy approximation space, if condition attributes and decision attributes are distinguished, then it is called a fuzzy decision table (or a fuzzy decision information system).

Attribute reduction in rough set theory offers a systematic theoretic framework for consistency-based feature selection, which does not attempt to maximize the class separability but rather attempts to retain the discernible ability of original features for the objects from the universe [15, 39] . Through using various attribute reductions, one can obtain the corresponding feature subsets and discover corresponding knowledge. Therefore, how to evaluate the decision performance of an attribute reduction approach becomes a very important issue. The solution of this problem may be helpful for determining which of knowledge discovery approaches is preferred for a practical problem in the context of fuzzy decision tables.

In this paper, from the viewpoint of decision performance evaluation, we will construct an attribute reduction approach in the framework of fuzzy rough sets, which can preserve the consistency of a given fuzzy decision table. We will also perform a series of experimental analyses for illustrating the validity of the proposed approach from search strategy and heuristic function.

In the next section, we summarize existing research on decision performance evaluation and attribute reduction in the context of rough set theory. In Sect. 3, some preliminary concepts are briefly reviewed. In Sect. 4, a consistency measure is defined in the context of a general fuzzy decision table and the relationship between the consistency measure and inclusion degree is established. In Sect. 5, based on the consistency measure, we develop a consistency-preserving attribute reduction approach in fuzzy rough set framework. In Sect. 6, we present a series of experimental studies that focus on the quantification of consistency of the selected attributes from search strategy and heuristic function on six public data sets. Finally, Sect. 7 concludes this paper by bring some remarks and discussions.

2 Relative works

Generally speaking, a set of decision rules can be generated from a decision table by adopting any kind of rule extraction methods [7, 8, 10, 14]. In recent years, how to evaluate the decision performance of a decision rule has become a very important issue in rough set theory [14]. Based on information entropy, Düntsch and Gediga [4] suggested some uncertainty measures of a decision rule and proposed three criteria for model selection. Greco, Pawlak and Slowinski applied some well-known confirmation measures within the rough set approach to discover relationships in data in terms of decision rules. For a decision rule set consisting of every decision rule induced from a decision table, three parameters are traditionally associated: the strength, the certainty factor and the coverage factor of the rule [9]. In many practical decision problems, we always adopt several rule-extraction methods for the same decision table. In this case, it is very important to check whether or not each of the rule-extraction approaches adopted is suitable for the given decision table. In other words, it is desirable to evaluate the decision performance of the decision-rule set extracted by each of the rule-extraction approaches. This strategy can help a decision maker to determine which of rule-extraction methods is preferred for a given decision table. However, all of the above measures for this purpose are only defined for a single decision rule and are not suitable for evaluating the decision performance of a decision-rule set. There are two more kinds of measures [27], which are approximation accuracy for decision classification and consistency degree for a decision table. Although these two measures, in some sense, could be regarded as measures for evaluating the decision performance of all decision rules generated from a complete decision table, they have some limitations. For instance, the certainty and consistency of a rule set could not be well characterized by the approximation accuracy and consistency degree when their values reaches zero. To overcome the shortcomings of the existing measures, [30, 31, 33] systemically investigate how to calculate the consistency of each of three kinds of decision tables, which are complete decision tables, incomplete decision tables and the decision tables in the context of maximal consistent blocks. To date, however, how to assess the decision performance of a decision-rule set extracted from a fuzzy decision table has not been reported. Like the measures (α, β and γ), the certainty, consistency and support of a decision-rule set extracted from a fuzzy decision table should be also studied to assess their decision performance. In fact, the approximation accuracy and approximation quality can be extended for evaluating the decision performance of a fuzzy decision table. Nevertheless, these two extensions have the same limitations, which still cannot give elaborate depictions of decision performance of a decision-rule set extracted from a fuzzy decision table. To overcome this drawback, this paper will present a new consistency measure for evaluating the decision performance of a fuzzy decision table.

As we know, the basic idea of rough set theory is to unravel an optimal set of decision rules from a decision table via an objective knowledge induction process which determines the necessary and sufficient attributes constituting the rules for decision making [44]. Attribute reduction is thus an outstanding contribution made by rough set research to data analysis. For further developments, as follows, we briefly review some attribute reduction approaches from decision tables. Many types of attribute reduction have been proposed in the analysis of information systems and decision tables, each of them aimed at some basic requirements. The concept of the β-reduct proposed by Ziarko provides a suite of reduction methods in the variable precision rough set model [54]. An attribute reduction method was proposed for knowledge reduction in random information systems [42]. Five kinds of attribute reducts and their relationships in inconsistent systems were investigated by Kryszkiewicz [18], Li et al. [20] and Mi et al. [24], respectively. By eliminating some rigorous conditions required by the distribution reduct, a maximum distribution reduct was introduced by Mi et al. [24]. In order to obtain all attribute reducts of a given data set, Skowron proposed a discernibility matrix method [34], in which any two objects determine one feature subset that can distinguish them. According to the discernibility matrix viewpoint, Qian et al. [29] provided a technique of feature selection for multi-granulation rough set model. The above feature selection methods are usually time consuming and intolerable to process large-scale data. To support efficient feature selection, many heuristic attribute reduction methods have been developed in rough set theory, cf. [12, 13, 22, 23, 38]. Each of these methods preserves a particular property of a given information system. For convenience, we review only four representative heuristic attribute reduction methods. Hu and Cercone [13] proposed a heuristic feature selection method, called positive-region reduction, which keeps the positive region of target decision unchanged. Shannon's information entropy also can be used to search reducts in the classical rough set model. Liang et al. [22, 23] defined new information entropy to measure the uncertainty of an information system and applied the entropy to reduce redundant features.

In fuzzy rough set framework, heuristic attribute reduction have been also examined by several researchers. These attribute reduction approaches are mainly based on two strategies: fuzzy positive region and fuzzy information entropy. Shen and Jensen [15] generalized the dependency function defined in classical rough set model into the fuzzy case and presented a series of attribute reduction algorithms. Bhatt and Gopal [2] gave the concept of fuzzy rough sets on compact computational domain, which is then utilized to improve computational efficiency. As Shannon's information entropy was introduced to search reducts in classical rough set model, Hu et al. [11, 12] extended the entropy to measure the information quantity in fuzz sets, called fuzzy information entropy, and applied the proposed measure to attribute reduction of hybrid data. The attribute reduction proposed by Shen and Jensen is based on the concept of fuzzy positive region, which can keep the fuzzy positive region of a given fuzzy decision table in a corresponding attribute reduct. From the viewpoint of consistency, this approach only keeps the consistency of fuzzy positive region, not but that of the entire fuzzy decision table. As to fuzzy information entropy, it only reflects that the information quality of a fuzzy decision table is changeless in the attribute reduction process and can not ensure the invariability of consistency of the fuzzy decision table.

Since the concept of consistency can characterize the decision performance of a fuzzy decision table, the issue of attribute reduction that preserves the consistency is of particular importance for data analysis using fuzzy rough set technique. Moreover, based on the proposed consistency measure, we also need to construct a heuristic attribute reduction algorithm (ARBC) such that the generating feature subset keeps the consistency of a given fuzzy decision table. To verify the validity of ARBC algorithm, we will present a series of experimental studies that focus on the quantification of consistency of the selected attributes from search strategy and heuristic function on six public data sets. To facilitate our discussion, we first discuss relevant notions in fuzzy rough set framework in Sect. 3.

3 Preliminaries

In this section, we briefly introduce basic concepts and denotations related to fuzzy rough set model, which include fuzzy equivalence relations, fuzzy approximation spaces, fuzzy decision tables and fuzzy rough approximation.

Given a nonempty finite set $U, \tilde{R}$ is a fuzzy binary relation over U, denoted by a matrix

$$M(\tilde{R}) = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}, \tag{1}$$

where $r_{ij} \in [0, 1]$ is the relation value between x_i and x_j.

We say $\tilde{R}$ is a fuzzy equivalence relation if $\forall x, y, z \in U, \tilde{R}$ satisfies

1. Reflexivity: $\tilde{R}(x, x) = 1$;
2. Symmetry: $\tilde{R}(x, y) = \tilde{R}(y, x)$;
3. Transitivity: $\tilde{R}(x, z) \geq min_y\{\tilde{R}(x, y), \tilde{R}(y, z)\}$.

Some operations of relation matrices are defined as

1. $\tilde{R}_1 = \tilde{R}_2 \Leftrightarrow \tilde{R}_1(x, y) = \tilde{R}_2(x, y)$;
2. $\tilde{R} = \tilde{R}_1 \cup \tilde{R}_2 \Leftrightarrow \tilde{R} = max\{\tilde{R}_1(x, y), \tilde{R}_2(x, y)\}$;
3. $\tilde{R} = \tilde{R}_1 \cap \tilde{R}_2 \Leftrightarrow \tilde{R} = min\{\tilde{R}_1(x, y), \tilde{R}_2(x, y)\}$;
4. $\tilde{R}_1 \subseteq \tilde{R}_2 \Leftrightarrow \tilde{R}_1(x, y) \leq \tilde{R}_2(x, y)$.

A fuzzy equivalence relation generates a fuzzy partition of the universe and a series of fuzzy equivalence classes, which are also called fuzzy knowledge granules [6, 11, 35].

The fuzzy partition of the universe generated by a fuzzy equivalence relation $\tilde{R}$ is defined as

$$\frac{U}{\tilde{R}} = \{[x_i]_{\tilde{R}}\}_{i=1}^{n}, \tag{2}$$

where $[x_i]_{\tilde{R}} = \{(r_{i1}/x_1) + (r_{i2}/x_2) + \cdots + (r_{in}/x_n)\}$. $[x_i]_{\tilde{R}}$ is the fuzzy equivalence class containing x_i and r_{ij} is the degree of x_i equivalent to x_j. Here, "+" means the union of elements.

The cardinality of the fuzzy equivalence class $[x_i]_{\tilde{R}}$ can be calculated with

$$|[x_i]_{\tilde{R}}| = \sum_{j=1}^{n} r_{ij}, \tag{3}$$

which appears to be a natural generalization of the cardinality of a crisp set.

In this case, $[x_i]_{\tilde{R}}$ is a fuzzy set and the family of $[x_i]_{\tilde{R}}$ forms a fuzzy concept system of the universe. This system will be used to approximate the object subset of the universe.

Definition 1 A two-tuple $\langle U, \tilde{R} \rangle$ is a fuzzy approximation space or a fuzzy information system, where U is a nonempty and finite set of objects, called the universe, and $\tilde{R}$ is a family of fuzzy equivalence relations defined on U.

Let $\tilde{X}$ be a fuzzy set. Then, it can be represented as

$$\tilde{X} = \left\{ \frac{\mu_{\tilde{X}}(x_1)}{x_1} + \frac{\mu_{\tilde{X}}(x_2)}{x_2} + \cdots + \frac{\mu_{\tilde{X}}(x_n)}{x_n} \right\}, \tag{4}$$

where $\mu_{\tilde{X}}(x_j)$ denotes the membership degree of the object x_j in X.

For convenience, we give another denotation of a fuzzy information system. From now, denoted by $S = (U, \tilde{A})$ be a fuzzy information system and $\tilde{A}$ a fuzzy attribute set in U. It can generates a fuzzy equivalence relation $\tilde{R}_A$ on U. The fuzzy relation matrix $M(\tilde{R}_A)$ is denoted by

$$M(\tilde{R}_A) = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}, \tag{5}$$

where $r_{ij} \in [0, 1]$ is the relation value between x_i and x_j.

A fuzzy decision table is a fuzzy information system $S = (U, \tilde{C} \cup \tilde{d})$, where $\tilde{C}$ is called a fuzzy condition attribute set and $\tilde{d}$ is called a fuzzy decision attribute. In practical decision-making issues, in general, the decision attribute $\tilde{d}$ can induce an equivalence partition, i.e., a crisp classification. In this paper, we only focus on this kind of fuzzy decision tables.

Definition 2 [5] Let $\langle U, \tilde{R} \rangle$ be a fuzzy approximation space and $\tilde{X}$ a fuzzy subset of U. The lower approximation and upper approximation are denoted by $\underline{\tilde{R}}\tilde{X}$ and $\overline{\tilde{R}}\tilde{X}$, respectively. Then, the membership degree of x to $\tilde{X}$ are defined as

$$\begin{cases} \mu_{\underline{\tilde{R}}\tilde{X}}(x) = \wedge\{\mu_{\tilde{X}}(y) \vee (1 - \tilde{R}(x, y)) : y \in U\}, \ x \in U \\ \mu_{\overline{\tilde{R}}\tilde{X}}(x) = \vee\{(\mu_{\tilde{X}}(y) \wedge \tilde{R}(x, y)) : y \in U\}, \ x \in U \end{cases}. \tag{6}$$

where $\wedge$ and $\vee$ mean *min* and *max* operators, respectively, and $\mu_{\tilde{X}}(y)$ means the membership of y to $\tilde{X}$. The order pair $\langle \underline{\tilde{R}}\tilde{X}, \overline{\tilde{R}}\tilde{X} \rangle$ is called a fuzzy rough set.

If $\underline{\tilde{R}}\tilde{X} = \overline{\tilde{R}}\tilde{X}$, then we say the fuzzy set $\tilde{X}$ is a definable set on the fuzzy approximation space. In fact, if $\tilde{X}$ is a definable set, then $\mu_{\underline{\tilde{R}}\tilde{X}}(x) = \mu_{\overline{\tilde{R}}\tilde{X}}(x), \forall x \in U$.

It is easy to see that the fuzzy approximation space can degrade to the corresponding Pawlak's approximation space when the equivalence relation and the object subset to be approximated are both crisp.

4 Consistency measures in fuzzy decision tables

As we know, fuzziness exists in many real-world applications. However, Pawlak's approximation spaces only work on the domain where crisp equivalence relations are defined. To overcome this shortcoming, Dübois et al. [5] presented the definitions of fuzzy approximation spaces, which integrates fuzziness and roughness together. In this section, we discuss how to measure the consistencies of a fuzzy target concept and a fuzzy decision table, and establish the relationships between the consistency measure and the inclusion degree in fuzzy decision tables.

4.1 A fuzzy partial relation

For our further investigation, we define a fuzzy partial relation $\preceq$ on the fuzzy attribute set as follows:

$\tilde{P} \preceq \tilde{Q} \Leftrightarrow \tilde{R}_P \subseteq \tilde{R}_Q,$

that is $\tilde{R}_P(x,y) \leq \tilde{R}_Q(x,y), \forall x,y \in U$.

For a fuzzy decision table $S=(U,\tilde{C}\cup\tilde{d})$, if $\tilde{C} \preceq \tilde{d}$, we say that S is consistent; if $\tilde{d} \preceq \tilde{C}$, we say that S is conversely consistent; otherwise, it is called a mixed fuzzy decision table.

From the definition of partial relation $\preceq$, one can obtain the following theorem.

Theorem 1 *Denoted by* $\boldsymbol{P}(\tilde{A})$ *be the power set of fuzzy attribute set* $\tilde{A}$ *in a fuzzy information system* $S=(U,\tilde{A})$*. Then,* $(\boldsymbol{P}(\tilde{A}), \preceq)$ *is a poset.*

Proof Let $\tilde{P},\tilde{Q},\tilde{T} \in \boldsymbol{P}(\tilde{A})$, $U/\tilde{P}=\{[x]_{\tilde{P}}, x\in U\}$, $U/\tilde{Q}=\{[x]_{\tilde{Q}}, x\in U\}$ and $U/\tilde{T}=\{[x]_{\tilde{T}}, x\in U\}$.

1. For any $x\in U$, it is obvious that $\tilde{R}_P(x,y) \leq \tilde{R}_P(x,y), \forall x,y\in U$. That is, $\tilde{R}_P \subseteq \tilde{R}_P$. Hence, $\tilde{P} \preceq \tilde{P}$.
2. Suppose that $\tilde{P} \preceq \tilde{Q}$ and $\tilde{Q} \preceq \tilde{P}$. When $\tilde{P} \preceq \tilde{Q}$, it follows from the definition of $\preceq$ that for any $x,y\in U$, one has that $\tilde{R}_P(x,y) \leq \tilde{R}_Q(x,y)$. When $\tilde{Q} \preceq \tilde{P}$, one also can obtain that $\tilde{R}_Q(x,y) \leq \tilde{R}_P(x,y), \forall x,y\in U$. Therefore, for any $x,y\in U$, we have that $\tilde{R}_P(x,y)=\tilde{R}_Q(x,y)$. Hence, $\tilde{P}=\tilde{Q}$.
3. Suppose that $\tilde{P} \preceq \tilde{Q}$ and $\tilde{Q} \preceq \tilde{T}$. When $\tilde{P} \preceq \tilde{Q}$, it follows from the definition of $\preceq$ that for any $x,y\in U$, one has that $\tilde{R}_P(x,y) \leq \tilde{R}_Q(x,y)$. When $\tilde{Q} \preceq \tilde{T}$, one can obtain that $\tilde{R}_Q(x,y) \leq \tilde{R}_T(x,y), \forall x,y\in U$. Therefore, for any $x,y\in U$, we have that $\tilde{R}_P(x,y) \leq \tilde{R}_Q(x,y) \leq \tilde{R}_T(x,y)$. Hence, $\tilde{P} \preceq \tilde{T}$.

Thus, $(\mathbf{P}(\tilde{A}), \preceq)$ is a poset. This completes the proof.

4.2 Consistency measure of a fuzzy decision table

As follows, we discuss the consistency of a fuzzy target concept $\tilde{X}$ in a given fuzzy decision table.

In the rough set literature, rough membership function introduced in can be used to measure degrees of inclusion of decision classes into subsets of the universe [27]. Let $S=(U,C\cup D)$ be a complete decision table, X an equivalence class and $U/D=\{[x]_D : x\in U\}$. For any object $x\in U$, the membership function of x in X is denoted by

$$\delta_X(x)=\frac{|X\cap[x]_D|}{|X|}, \tag{7}$$

where $\delta_X(x)$ $(0\leq\delta_X(x)\leq 1)$ represents a fuzzy concept.

In fact, if $\delta_X(x)=1$, then X can be said to be consistent with respect to $[x]_D$. In other words, if X is a consistent set with respect to $[x]_D$, then one has $X\subseteq[x]_D$. It can generate a fuzzy set $F_X^D=\{(x,\delta_X(x)) \mid x\in U\}$ on the universe U. Like the above discussion, one can define the rough membership function of an object x in a fuzzy target concept $\tilde{X}$ in a fuzzy decision table.

Let $S=(U,\tilde{C}\cup\tilde{d})$ be a fuzzy decision table, $\tilde{X}=\{\frac{\mu_{\tilde{X}}(x_1)}{x_1}+\frac{\mu_{\tilde{X}}(x_2)}{x_2}+\cdots+\frac{\mu_{\tilde{X}}(x_n)}{x_n}\}$ a fuzzy target concept and $U/\tilde{d}=\{[x]_{\tilde{d}} : x\in U\}$. For any object $x\in U$, the rough membership function of x in the fuzzy target concept $\tilde{X}$ is defined as

$$\delta_{\tilde{X}}(x)=\frac{|\tilde{X}\cap[x]_{\tilde{d}}|}{|\tilde{X}|}=\frac{\sum_{i=1}^{n}(\mu_{\tilde{X}}(x_i)\wedge\mu_{[x]_{\tilde{d}}}(x_i))}{\sum_{i=1}^{n}\mu_{\tilde{X}}(x_i)}, \tag{8}$$

where $\mu_{[x]_{\tilde{d}}}(x_i)$ is the membership function of x_i within the fuzzy equivalence class $[x]_{\tilde{d}}$ and $\delta_{\tilde{X}}(x)$ $(0\leq\delta_{\tilde{X}}(x)\leq 1)$ represents a fuzzy concept.

As we know, a fuzzy equivalence class also can be regarded as a fuzzy target concept. In the following, we investigate the consistency of a fuzzy equivalence class $[x_i]_{\tilde{C}}$ $(i\in\{1,2,\ldots,|U|\})$ included in the condition part $U/\tilde{C}$ in a fuzzy decision table.

Let $S=(U,\tilde{C}\cup\tilde{d})$ be a fuzzy decision table, $[x_i]_{\tilde{C}}\in U/\tilde{C}$ a fuzzy equivalence class and $U/\tilde{d}=\{[x]_{\tilde{d}} : x\in U\}$. For any object $x\in U$, the rough membership function of x in the fuzzy equivalence class $[x_i]_{\tilde{C}}$ is defined as

$$\delta_{[x_i]_{\tilde{C}}}(x)=\begin{cases}\frac{\sum_{j=1}^{n}(r_{ij}\wedge\mu_{[x]_{\tilde{d}}}(x_j))}{\sum_{j=1}^{n}r_{ij}}, & \text{if } x=x_i;\\ 0, & \text{if } x\neq x_i.\end{cases} \tag{9}$$

where r_{ij} is the relation value between x_i and x_j in the fuzzy equivalence class $[x_i]_{\tilde{C}}\in U/\tilde{C}$ and $r_{ij}\wedge\mu_{[x]_{\tilde{d}}}(x_j)$ is equivalent to $min\{r_{ij},\mu_{[x]_{\tilde{d}}}(x_j)\}$. It is clear that $\delta_{[x_i]_{\tilde{C}}}(x)$ $(0\leq\delta_{[x_i]_{\tilde{C}}}(x)\leq 1)$ denotes a fuzzy concept.

If $\delta_{[x_i]_{\tilde{C}}}(x)=1$, then the fuzzy equivalence class $[x_i]_{\tilde{C}}$ can be said to be consistent with respect to the decision attribute $\tilde{d}$. In other words, if $[x_i]_{\tilde{C}}$ is a consistent set with respect to $\tilde{d}$, then $[x_i]_{\tilde{C}}\subseteq[x_i]_{\tilde{d}}$. It can generate a fuzzy set $F_{[x_i]_{\tilde{C}}}^{\tilde{d}}=\{(x,\delta_{[x_i]_{\tilde{C}}}(x)) \mid x\in U\}$ on the universe U. Through using the concept of a consistent set, in the following, we give a definition of consistency measure of a fuzzy equivalence class in fuzzy decision tables.

Definition 3 Let $S=(U,\tilde{C}\cup\tilde{d})$ be a fuzzy decision table, $[x_i]_{\tilde{C}}\in U/\tilde{C}$ a fuzzy equivalence class and $U/\tilde{d}=\{[x]_{\tilde{d}}:\ x\in U\}$. A consistency measure of $[x_i]_{\tilde{C}}$ with respect to $\tilde{d}$ is defined as

$$C(F^{\tilde{d}}_{[x_i]_{\tilde{C}}})=\sum_{x\in U}\delta_{[x_i]_{\tilde{C}}}(x), \tag{10}$$

where $0\le C(F^{\tilde{d}}_{[x_i]_{\tilde{C}}})\le 1$.

It is easy to see that the following theorem holds.

Theorem 2 *The consistency measure of a consistent fuzzy equivalence class in a fuzzy decision table is one.*

In the following, based on the above discussion, we research the consistency between two fuzzy attribute subsets in a fuzzy decision table.

Definition 4 Let $S=(U,\tilde{C}\cup\tilde{d})$ be a fuzzy decision table, $U/\tilde{C}=\{[x_i]_{\tilde{C}}\mid x_i\in U,i\le|U|\}$ and $U/\tilde{d}=\{[x]_{\tilde{d}}:\ x\in U\}$. A consistency measure of $\tilde{C}$ with respect to $\tilde{d}$ is defined as

$$C(\tilde{C},\tilde{d})=\frac{1}{|U|}\sum_{i=1}^{|U|}\sum_{x\in U}\delta_{[x_i]_{\tilde{C}}}(x), \tag{11}$$

where $0\le C(\tilde{C},\tilde{d})\le 1$ and $\delta_{[x_i]_{\tilde{C}}}(x)$ is the rough membership function of $x\in U$ in the fuzzy equivalence class $[x_i]_{\tilde{C}}$.

This consistency measure can be seen as an extension of the consistency from the literature [33], which is used to evaluate the consistency degree of a decision-rule set. From the definition, it is seen that this kind of consistencies are different with traditional consistencies in rough set theory. This definition is illustrated by the following example.

Example 1 Consider a fuzzy decision table $S=(U,\tilde{C}\cup\tilde{d})$, the relation matrix induced by the condition attribute set $\tilde{C}$ is $M(\tilde{R}_C)=\begin{pmatrix}1&0.9&0\\0.9&1&0\\0&0&1\end{pmatrix}$, and that induced by the decision attribute set $\tilde{d}$ is $M(\tilde{R}_d)=\begin{pmatrix}1&0&0\\0&1&1\\0&1&1\end{pmatrix}$.

From the formula (10), it can be calculated that $C(F^{\tilde{d}}_{[x_1]_{\tilde{C}}})=\frac{1+0+0}{1+0.9+0}=\frac{10}{19}$, $C(F^{\tilde{d}}_{[x_2]_{\tilde{C}}})=\frac{0+1+0}{0.9+1+0}=\frac{10}{19}$ and $C(F^{\tilde{d}}_{[x_3]_{\tilde{C}}})=\frac{0+0+1}{0+0+1}=1$.

According to the formula (11), we have that

$$C(\tilde{C},\tilde{d})=\frac{1}{|U|}\sum_{i=1}^{|U|}\sum_{x\in U}\delta_{[x_i]_{\tilde{C}}}(x)=\frac{1}{3}\left(\frac{10}{19}+\frac{10}{19}+1\right)=0.6842.$$

Therefore, the consistency measure of $\tilde{C}$ with respect to $\tilde{d}$ in this decision table S is 0.6842.

Theorem 3 *Let $S=(U,C\cup d)$ be an incomplete decision table, $U/SIM(C)=\{S_C(x_1),S_C(x_2),\ldots,S_C(x_{|U|})\}$ and $U/SIM(d)=\{S_d(x):x\in U\}$. Then, the consistency measure of C with respect to d degenerates into the following formula* [31]

$$C(C,d)=\frac{1}{|U|}\sum_{i=1}^{|U|}\sum_{x\in U}\delta_{S_C(x_i)}(x). \tag{12}$$

This implies that the proposed consistency measure in the context of fuzzy decision tables also can measure the consistency of an incomplete/complete decision table, which is the natural generalization of those measures in classical decision tables.

From Definition 3, one can obtain the following Theorem 4 and Corollary 1.

Theorem 4 *The consistency measure of a consistent fuzzy decision table is one.*

Proof Let $S=(U,\tilde{C}\cup\tilde{d})$ be a fuzzy decision table, $U/\tilde{C}=\{[x_i]_{\tilde{C}}\mid x_i\in U,i\le|U|\}$ and $U/\tilde{d}=\{[x]_{\tilde{d}}:x\in U\}$. If S is consistent, then, for any $x\in U$, one has $[x_i]_{\tilde{C}}\subseteq[x_i]_{\tilde{d}}$, i.e., $r_{ij}\le\mu_{[x]_{\tilde{d}}}(x_j),\forall j\le|U|$. Hence, when $x=x_i$, we have

$$\delta_{[x_i]_{\tilde{C}}}(x)=\frac{\sum_{j=1}^{n}(r_{ij}\wedge\mu_{[x]_{\tilde{d}}}(x_j))}{\sum_{j=1}^{n}r_{ij}}=\frac{\sum_{j=1}^{n}r_{ij}}{\sum_{j=1}^{n}r_{ij}}=1;$$

otherwise, $\delta_{[x_i]_{\tilde{C}}}(x)=0$. Therefore,

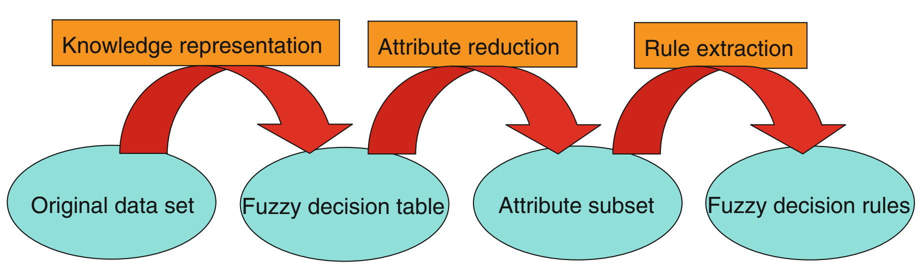


Fig. 1 Knowledge discovery process from a data set using fuzzy rough set approach

$$C(\tilde{C},\tilde{d}) = \frac{1}{|U|}\sum_{i=1}^{|U|}\sum_{x\in U}\delta_{[x_i]_{\tilde{C}}}(x) = \frac{1}{|U|}\sum_{i=1}^{|U|}(1\times 1+(|U|-1)\times 0) = 1.$$

Thus, the consistency measure of a consistent fuzzy decision table is one. This completes the proof.

Theorem 5 *If $C(\tilde{d},\tilde{C})=1$, then the fuzzy decision table S is conversely consistent.*

Proof It can be proved from the definition of converse consistency in fuzzy decision tables and Definition 3.

Consequently, the consistency of a fuzzy decision table can be measured through using some fuzzy concepts and it also can be induced to a fuzzy measure.

In order to establish the relationship between the consistency measure and inclusion degree [47, 49, 50], we introduce three existing types of definitions of inclusion degrees as follows.

Definition 5 Let $(X,\leq)$ be a poset. A corresponding number $D(y/x)$ $(\forall x,y\in X)$ is called the inclusion degree—if the following conditions hold

1. $0\leq D(y/x)\leq 1, \quad (x,y\in X)$;
2. $x\leq y\Rightarrow D(y/x)=1, \quad (x,y\in X)$;
3. $z\leq x\leq y\Rightarrow D(z/y)\leq D(z/x), \quad (x,y,z\in X)$.

If we modify condition (3) as

(3′) $x\leq y\Rightarrow \forall z\in X, D(z/y)\leq D(z/x), \quad (x,y\in X)$, D is called a strong inclusion degree denoted as type S_1 [34]. If D is an inclusion degree and further satisfies the condition

4. $x\leq y\Rightarrow \forall z\in X, D(x/z)\leq D(y/z), \quad (x,y\in X)$, then D is a strong inclusion degree denoted as type S_2 [34].

Generally, type S_1 and type S_2 are special cases of inclusion degrees; type S_1 is not all type S_2 and type S_2 is not all type S_1.

The following theorem establishes the relationship between the type 2 inclusion degree and the consistency measure proposed by this paper.

Theorem 6 *Let $S=(U,\tilde{A})$ be a fuzzy information system and $\tilde{P},\tilde{Q}\in\mathbf{P}(\tilde{A})$, where $\mathbf{P}(\tilde{A})$ is the power set of $\tilde{A}$. Then,*

$$D(\tilde{Q}/\tilde{P}) = \frac{1}{|U|}\sum_{i=1}^{|U|}\sum_{x\in U}\delta_{[x_i]_{\tilde{P}}}(x) \tag{13}$$

is a type 2 inclusion degree on the poset $(\mathbf{P}(\tilde{A}),\preceq)$.

Proof From the definition of inclusion degree, we have that:

1. Let $\tilde{P},\tilde{Q}\in\mathbf{P}(\tilde{A})$. From $0\leq\frac{\sum_{j=1}^{n}(r_{ij}\wedge\mu_{[x]_{\tilde{Q}}}(x_j))}{\sum_{j=1}^{n}r_{ij}}\leq 1$, it follows that $\delta_{[x_i]_{\tilde{P}}}(x)=0$ $(x\neq x_i)$ or $\delta_{[x_i]_{\tilde{P}}}(x)=\frac{\sum_{j=1}^{n}(r_{ij}\wedge\mu_{[x]_{\tilde{Q}}}(x_j))}{\sum_{j=1}^{n}r_{ij}}$ $(x=x_i)$. Hence, $0\leq\sum_{x\in U}\delta_{[x_i]_{\tilde{P}}}(x)\leq 1$. Thus, $0\leq D(\tilde{Q}/\tilde{P})\leq 1$.
2. When $P\preceq Q$, one can obtain that $[x_i]_{\tilde{P}}\subseteq[x_i]_{\tilde{Q}}$, $\forall i\in\{1,2,\cdots,|U|\}$. So, for any $i\leq|U|$, if $x=x_i$, then $\delta_{[x_i]_{\tilde{P}}}(x)=1$; otherwise $\delta_{[x_i]_{\tilde{P}}}(x)=0$. Hence, $\sum_{x\in U}\delta_{[x_i]_{\tilde{P}}}(x)=1+(|U|-1)\cdot 0=1$. Therefore, one has that $D(\tilde{Q}/\tilde{P})=\frac{1}{|U|}\sum_{i=1}^{|U|}1=1$.
3. Let $\tilde{P},\tilde{Q},\tilde{R}\in\mathbf{P}(\tilde{A})$ with $\tilde{P}\preceq\tilde{Q}$. Hence, it follows that $[x_i]_{\tilde{P}}\subseteq[x_i]_{\tilde{Q}}$, i.e., $\mu_{[x_i]_{\tilde{P}}}(x_j)\leq\mu_{[x_i]_{\tilde{Q}}}(x_j)$, $\forall i,j\in\{1,2,\cdots,|U|\}$. Thus,

$$\begin{aligned} D(\tilde{P}/\tilde{R}) &= \frac{1}{|U|}\sum_{i=1}^{|U|}\sum_{x\in U}\delta_{[x_i]_{\tilde{R}}}(x) \\ &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{\sum_{j=1}^{n}(\mu_{[x_i]_{\tilde{R}}}(x_j)\wedge\mu_{[x_i]_{\tilde{P}}}(x_j))}{\sum_{j=1}^{n}\mu_{[x_i]_{\tilde{R}}}(x_j)} \\ &\leq \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{\sum_{j=1}^{n}(\mu_{[x_i]_{\tilde{R}}}(x_j)\wedge\mu_{[x_i]_{\tilde{Q}}}(x_j))}{\sum_{j=1}^{n}\mu_{[x_i]_{\tilde{R}}}(x_j)} \\ &= D(\tilde{Q}/\tilde{R}). \end{aligned}$$

Therefore, $D(\tilde{Q}/\tilde{P})$ is a type S_2 inclusion degree on the poset $(\mathbf{P}(\tilde{A}),\preceq)$. This completes the proof.

5 Consistency-preserving attribute reduction

Rule extraction from real-world data sets is one kind of important knowledge discovery task. If we adopt fuzzy rough set model for this purpose, it first needs to transform original data sets into fuzzy information decision tables. Then, through using attribute reduction and rule extraction, one can extract a set of fuzzy decision rules. The overall procedure for a knowledge discovery method from a given data set is displayed in Fig. 1.

Given an attribute set, in rough set theory, the task of attribute reduction can be seen as a search for an "optimal" attribute subset through the competing candidate subsets. The definition of what an optimal subset is may vary depending on the problem to be solved. Although an exhaustive method may be used for this purpose in theory, this is quite impractical for most data sets. Usually attribute reduction algorithms involve heuristic or random search strategies in an attempt to avoid this prohibitive complexity.

As we know, it is very important to keep the consistency of a given data set for efficient knowledge discovery. Recent researches showed that a well-devised feature selection algorithm would significantly improve the efficiency of knowledge discovery because attribute reduction lessens the impact of the "curse of dimensionality" and speeds up the training and test process [15]. The key problem of this task is how to get a subset of attributes that keeps the consistency of an original data set.

Let $S = (U, \tilde{C} \cup \tilde{d})$ be a fuzzy decision table, where $\tilde{C}$ and $\tilde{d}$ be the condition attribute set and decision attribute. $\tilde{B} \subseteq \tilde{C}, \forall \tilde{a} \in \tilde{B}$, we define a coefficient

$$Sig_{in}(\tilde{a}, \tilde{B}, \tilde{d}) = C(\tilde{B}, \tilde{d}) - C(\tilde{B} - \tilde{a}, \tilde{d}) \quad (14)$$

as the significance of attribute $\tilde{a}$ in $\tilde{B}$ relative to decision $\tilde{d}$. $Sig_{in}(\tilde{a}, \tilde{B}, \tilde{d})$ reflects the changes of the consistency if attribute $\tilde{a}$ is eliminated from $\tilde{B}$. Accordingly, we also can define

$$Sig_{out}(\tilde{a}, \tilde{B}, \tilde{d}) = C(\tilde{B} \cup \tilde{a}, \tilde{d}) - C(\tilde{B}, \tilde{d}), \quad (15)$$

$\forall \tilde{a} \in \tilde{C} - \tilde{B}$. $Sig_{out}(\tilde{a}, \tilde{B}, \tilde{d})$ measures the increment of the consistency if attribute $\tilde{a}$ is introduced in $\tilde{B}$. This measure can be used in a forward feature selection algorithm, while $Sig_{in}(\tilde{a}, \tilde{B}, \tilde{d})$ is applicable to determine the significance of every attribute in the context of consistency.

Starting with the attribute with maximal consistency, we take the attribute with the maximal significance into the attribute subset in each loop until the entire consistency of this feature subset satisfies the target requirement (or is over a given threshold), and then we can get a feature subset. Formally, a forward greedy feature selection algorithm can be written as follows.

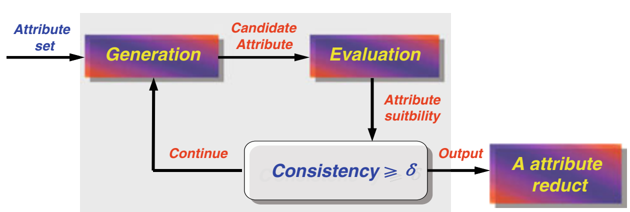


Fig. 2 The attribute reduction process in ARBC algorithm

Through using ARBC algorithm, the time complexity to obtain a feature subset from a fuzzy decision table is polynomial. At step 1, the time complexity for computing the significance of each of attributes $O(|\tilde{C}||U|^2)$. For steps 3–7, since $|\tilde{C}| - 1$ is the maximum value for the circle times, the time complexity is

$$O((|\tilde{C}| - 1)|U|^2 + (|\tilde{C}| - 2)|U|^2 + \cdots + |U|^2) = O(|\tilde{C}|^2|U|^2)$$

Therefore, the time complexity of ARBC is $O(|\tilde{C}|^2|U|^2)$.

This algorithm is a forward search algorithm. Figure 2 shows the process of attribute reduction in ARBC algorithm, which is helpful for us more clearly understanding the mechanism of the algorithm.

Input: $S = (U, \tilde{C} \cup \tilde{d})$ and a threshold of consistency measure δ
Output: An attribute subset $\tilde{B}$ of $\tilde{C}$.
1. Compute the significance of each attribute with $Sig_{in}(\tilde{a}, \tilde{C}, \tilde{d}), \forall \tilde{a} \in \tilde{C}$
2. $\tilde{B} \leftarrow \tilde{a}_0$, $Sig_{in}(\tilde{a}_0, \tilde{C}, \tilde{d}) = \max_i \{Sig_{in}(\tilde{a}_i, \tilde{C}, \tilde{d})\}$
3. Do
4. $\quad \forall \tilde{a} \in \tilde{C} - \tilde{B}$, compute $Sig_{out}(\tilde{a}, \tilde{B}, \tilde{d})$
5. $\quad\quad$ If $Sig_{out}(\tilde{a}_j, \tilde{B}, \tilde{d}) = \max_i \{Sig_{out}(\tilde{a}_i, \tilde{B}, \tilde{d})\}$
6. $\quad\quad\quad \tilde{B} = \tilde{B} \cup \{\tilde{a}_j\}$
7. Until $C(\tilde{B}, \tilde{d}) \geq \delta$
8. Return $\tilde{B}$

Algorithm 1 Forward attribute reduction based onthe consistency measure (ARBC)

Table 2 Variation of consistencies with four forward search methods on six data sets

	Search methods				
	Data sets	Heuristic	MAX–MIN	Random	MIN–MAX
1	Pima	0.6784	0.6761	0.6629	0.6108
2	Glass	0.5251	0.5148	0.4675	0.4163
3	Yeast	0.3501	0.3484	0.3058	0.2799
4	Ecoli	0.7138	0.6912	0.6221	0.5067
5	Zoo	0.9586	0.9240	0.8298	0.7079
6	Cancer	0.9844	0.9798	0.9375	0.9346
	Average	0.7017	0.6891	0.6376	0.5670

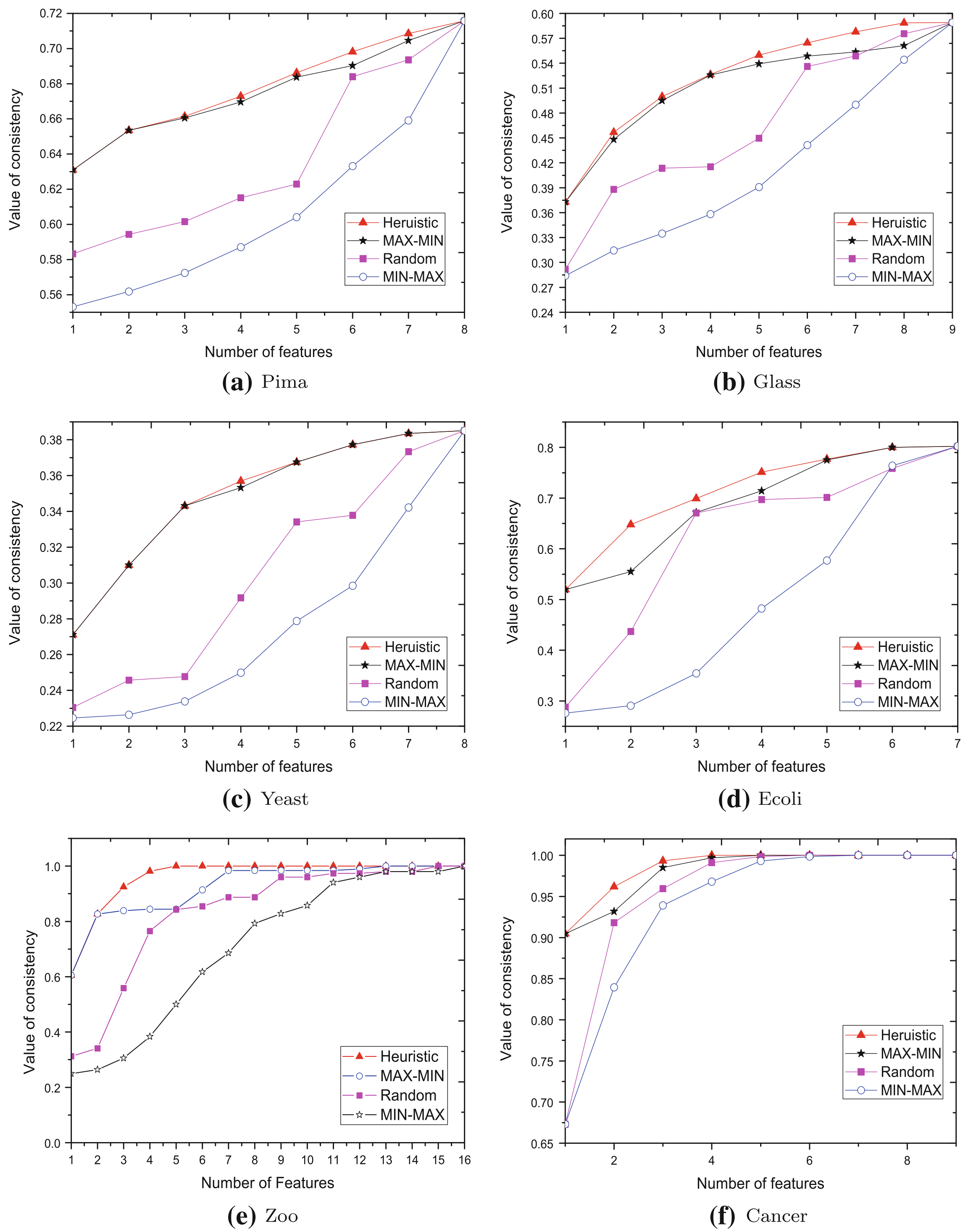
Table 1 Data description

	Data sets	Samples	Features	Data type	Classes
1	Pima-indians-diabetes (Pima)	768	8	Numeric	2
2	Glass (Glass)	214	9	Numeric	7
3	Yeast (Yeast)	1,484	8	Numeric	10
4	E.coli (Ecoli)	336	7	Numeric	8
5	Zoo (Zoo)	101	16	Nominal	7
6	Breast-cancer-wisconsin (Cancer)	683	9	Nominal	2

6 Experimental analysis

Six data sets from the University of California at Irvine (UCI) Machine Learning Repository are used in the empirical study. The information about these six data sets is shown in Table 1. The objective of these experiments is to show the power of the proposed method to select numerical or nominal attributes.

For numeric data, firstly, we normalize the numerical attribute x into the interval $[0, 1]$ with [12]



(a) Pima **(b)** Glass **(c)** Yeast **(d)** Ecoli **(e)** Zoo **(f)** Cancer

Fig. 3 Variation of consistencies with four forward search methods on six data sets

$$a' = \frac{a - a_{min}}{a_{max} - a_{min}}.$$

The value of the fuzzy similarity degree r_{ij} between objects x_i and x_j with respect to numerical attribute a is computed as

$$r_{ij} = \begin{cases} 1 - 4 \times |x_i - x_j|, & |x_i - x_j| \le 0.25, \\ 0, & otherwise. \end{cases}$$

As $r_{ij} = r_{ji}$ and $r_{ii} = 1, 0 \le r_{ij} \le 1$, the matrix $M = (r_{ij})_{n \times n}$ is a fuzzy similarity relation. We can get a fuzzy equivalence relation from M with max–min transitivity operation. In practice the operation cannot be conducted and we directly search a feature subset with a similarity relation [12].

For nominal data, given a nominal decision table $S = (U, C \cup d)$, we have that

$$r_{ij} = \begin{cases} 1, f(x_i) = f(x_j), & \forall a \in C \\ 0, & otherwise. \end{cases}$$

The matrix $M = (r_{ij})_{n \times n}$ is clearly an equivalence relation.

As to consistency-based feature selection, in this section, we perform experimental analysis for numeric data sets (Pima, Glass, Yeast and Ecoli) and nominal data sets (Zoo and Cancer), respectively. In order to emphasize the advantage of ARBC algorithm, three kind of search methods (MAX–MIN, Random and MIN–MAX) are employed to comparison analysis. MAX–MIN starts with an empty set, and adds one attribute with maximum significance $Sig_{in}(\tilde{a}, \tilde{C}, \tilde{d})$ into a pool each time until the consistency does not increase, MIN–MAX presents a sequence of attributes with the increasing order of significance, and Random gives a rank of attributes randomly.

Table 2 shows the comparisons of consistencies with four search strategies on six data sets. In this table, the value from each lattice is the mean value of sum of consistencies on various features, and the last row displays the mean value of consistencies induced by each search strategy on six data sets. More detailed change trendline of each search strategy on each of these six data sets is displayed in Fig. 3. Note that the consistency measure of a numeric decision table may not achieve the maximum value one (see sub-figures a, b, c and d in Fig. 3), while a consistent nominal decision table can achieve the maximum value one (see sub-figures e and f in Fig. 3).

From Table 2 and Fig. 3, we can find that the forward search strategy based on the consistency measure is consistently better than the other three as to the six data sets (numerical data sets and nominal data sets). What is more, the MAX–MIN (descend) strategy is far better than random and MIN–MAX (ascending) strategy, which shows the search strategy greatly influences the result of feature selection. Without considering the dependency degree between two attributes subsets, selecting the attribute with bigger consistency as a selected attribute will be a good solution. Whereas, forward search heuristic algorithm ARBC is consistently better than MAX–MIN strategy in the experiments.

Interestingly, in the above four search strategies, it can be seen that the consistencies of attribute sets monotonously increase with the number of selected features becoming bigger. This monotonicity provides us an approach to selecting a feature subset according to user's requirements in real applications. As we know, the consistency measure of a numerical decision table is always less than the maximum value one. In this situation, a decision maker can give a threshold δ as the target of feature selection. For example, considering feature selection from Pima, let $\delta = 0.6696$. For this feature selection issue, ARBC gets four features, MAX–MIN obtains five features, Random needs six features, and MIN–MAX requires all features for satisfying the threshold. From the other five data sets, we can scan the advantage of ARBC algorithm. In particular, for a consistent decision table, we can set the threshold equals one. From the consistent nominal data set Zoo, we can find that ARBC only selects five features, MAX–MIN needs thirteen features, Random gets fifteen features, and MIN–MAX obtains sixteen features. ARBC algorithm demonstrates the great advantage. From the data set Cancer, it also can been verified.

Table 3 shows the comparisons of consistencies with three heuristic functions (consistency, fuzzy information entropy and positive region) on six data sets. Fuzzy information entropy and positive region are two often heuristic functions for feature selection from high-dimensional data [2, 11, 15]. In this experiment, forward search strategy is employed to investigate the performance of each of consistency measure, fuzzy information entropy and positive region feature selection. In this table, the value

Table 3 Variation of consistencies with three heuristic functions on six data sets

	Heuristic functions			
	Data sets	Consistency	Fuzzy information entropy	Positive region
1	Pima	0.6784	0.6624	0.6579
2	Glass	0.5251	0.4989	0.4845
3	Yeast	0.3501	0.3278	0.3010
4	Ecoli	0.7138	0.6418	0.5609
5	Zoo	0.9586	0.9446	0.7496
6	Cancer	0.7017	0.6744	0.9474
	Average	0.7017	0.6891	0.6169

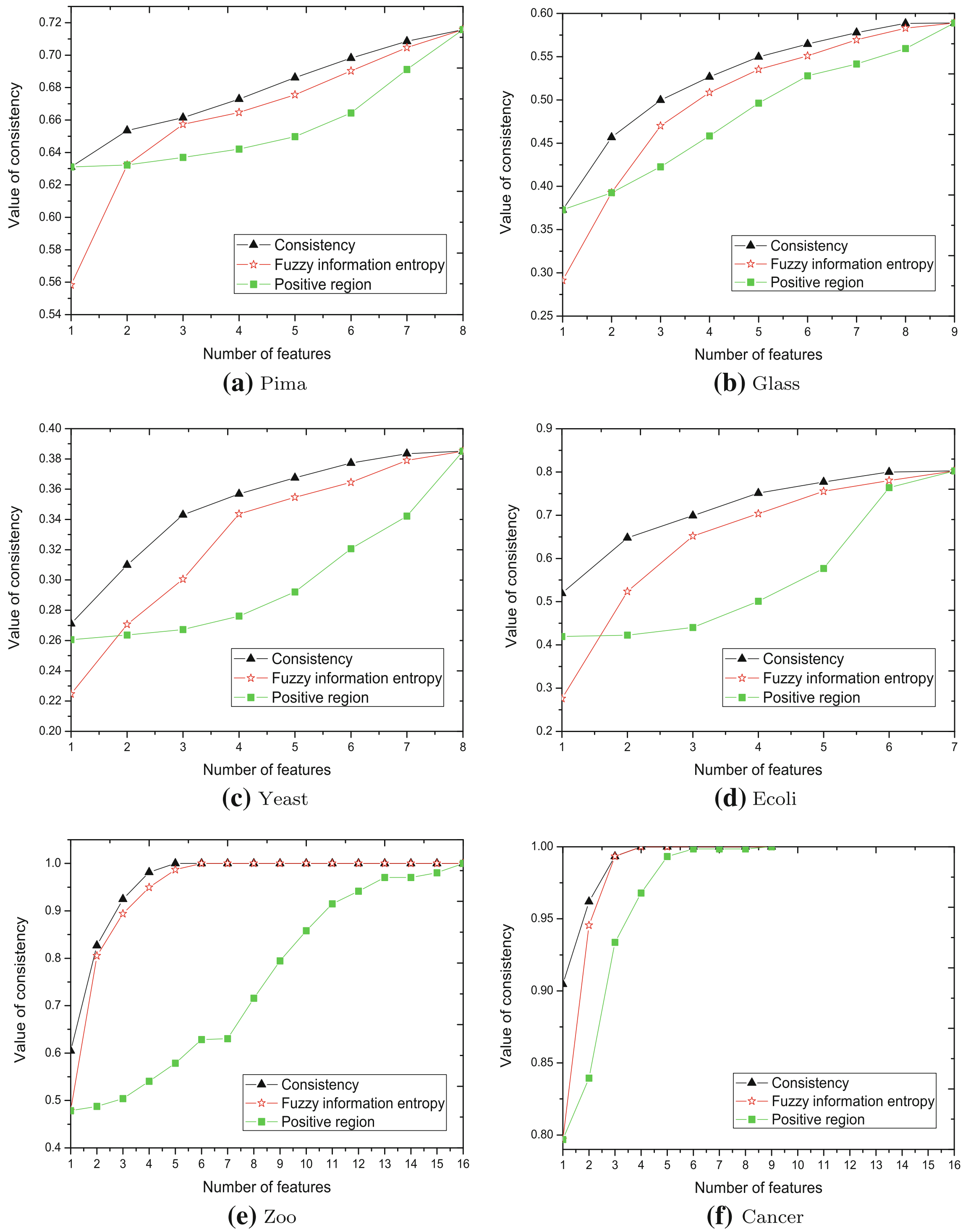


Fig. 4 Variation of consistencies with three heuristic functions on six data sets

from each lattice is the mean value of sum of consistencies on various features, and the last row presents the mean value of consistencies induced by each heuristic function on six data sets. More detailed change trendline of each heuristic function on each of these six data sets is displayed in Fig. 4.

From Table 3 and Fig. 4, we can find that the ARBC algorithm based on the consistency measure is consistently better than that based on the fuzzy information entropy and that based on positive region as to the six data sets in the meaning of consistency. Furthermore, the feature selection based on the fuzzy information entropy is better than that based on the positive region, which illustrates the heuristic function also greatly influences the result of feature selection. From the viewpoint of consistency-preserving, the proposed algorithm based on the consistency measure outperforms the other heuristic functions.

It is deserved that for these three heuristic functions, the consistencies of attribute sets monotonously increase with the number of selected features becoming bigger. Considering feature selection from Pima, let $\delta = 0.6696$, we know that ARBC gets four features, the fuzzy information entropy obtains five features, and the positive region requires seven features for satisfying the threshold. From the other five data sets, we also can draw the advantage of feature selection based on the consistency measure. In particular, for a consistent decision table, let the threshold equals one. From the consistent nominal data set Zoo, we can find that ARBC only selects five features, the fuzzy information entropy needs thirteen features, and the positive region obtains sixteen features. ARBC algorithm demonstrates the great advantage in the meaning of consistency. From the data set Cancer, we can draw the same conclusion.

7 Conclusions

Fuzzy rough set framework is a rational generalization of Pawlak's rough set theory, which integrates two types of uncertainties (roughness and fuzziness) into one rough set framework. Through fuzzy knowledge representation, both numerical data and nominal data can be uniformly represented as a fuzzy decision table. Using fuzzy rough set method, one can acquire fuzzy decision rules and make a fuzzy rough decision from a fuzzy decision table. Attribute reduction (feature selection) plays an important role in this purpose.

Considering consistency is one of several criteria for evaluating the decision performance of a decision-rule set, in this paper, we have introduced a consistency measure to assess the consistencies of a fuzzy target set and a fuzzy decision table, which underlies a foundation for attribute reduction algorithm. Using the proposed consistency measure, we have defined two attribute significance measures which are used to select candidate attributes in attribute reduction process. Based on these strategies,we have presented a consistency-preserving attribute reduction in fuzzy rough set framework. This approach does not require discretizing the numerical data in fuzzy rough set framework, called ARBC, and can select an attribute subset from both numerical and nominal data.

We have employed 6 UCI data sets for evaluating the proposed method, in which four data sets are numerical and two data sets are nominal. series of experiments have been also conducted from search strategy and heuristic function in the meaning of consistency. he results show that in terms of search strategy, the forward greedy search strategy is consistently better than each of MAX–MIN, Random and MIN–MAX, and in terms of heuristic function, the proposed ARBC algorithm is much better than each of attribute reduction based on fuzzy information entropy and that based on fuzzy positive region. We can draw such a conclusion that number of the selected features using ARBC algorithm is the least for a given threshold of consistency measure.

Acknowledgments This work was supported by the national natural science foundation of China (No. 60903110, 60773133, 70971080), the national high technology research and development program of China (No. 2007AA01Z165), the natural science foundation of Shanxi province, China (No. 2008011038, 2009021017-1).

References

1. Beynon M (2001) Reducts within the variable precision rough sets model: a further investigation. Eur J Oper Res 134(3): 592–605
2. Bhatt RB, Gopal M (2005) On the compact computational domain of fuzzy rough sets. Pattern Recognit Lett 26:965–975
3. Cock MD, Cornelis C, Kerre EE (2007) Fuzzy rough sets: the forgotten step. IEEE Trans Fuzzy Syst 15(1):121–130
4. Düntsch I, Gediga G (1998) Uncertainty measures of rough set prediction. Artif Intell 106:109–137
5. Dübois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. Int J Gen Syst 17:191–209
6. Fan SQ, Zhang WX, Xu W (2006) Fuzzy inference based on fuzzy concept lattice. Fuzzy Sets Syst 157:3177–3187
7. Feng S, Xu L (1999) Decision support for fuzzy comprehensive evaluation of urban development. Fuzzy Sets Syst 105(1):1–12
8. Feng QR, Miao DQ, Cheng Y (2009) Hierarchical decision rules mining. Expert Syst Appl 37(3):2081–2091
9. Greco S, Pawlak Z, Slowinski R (2004) Can Bayesian confirmation measures be useful for rough set decision rules? Eng Appl Artif Intell 17:345–361
10. Hong TP, Wang TT, Wang SL (2007) Mining fuzzy β-certain and β-possible rules from quantitative data based on the variable precision rough-set model. Expert Syst Appl 32:223–232
11. Hu QH, Yu DR, Xie ZX, Liu JF (2006) Fuzzy probabilistic approximation spaces and their information measures. IEEE Trans Fuzzy Syst 14(2):191–201
12. Hu QH, Xie ZX, Yu DR (2007) Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. Pattern Recognit 40:3509–3521
13. Hu XH, Cercone N (1995) Learning in relational databases: a rough set approach. Int J Comput Intell 11(2):323–338
14. Huysmans J, Baesens B, Vanthienen J (2007) A new approach for measuring rule set consistency. Data Knowl Eng 63(1):167–182
15. Jensen R, Shen Q (2007) Fuzzy-rough sets assisted attribute selection. IEEE Trans Fuzzy Syst 17(1):73–89
16. Jensen R, Shen Q (2004) Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. IEEE Trans Knowl Data Eng 16(12):1457–1471
17. Kryszkiewicz M (1998) Rough set approach to incomplete information systems. Inf Sci 112:39–49

18. Kryszkiewicz M (1999) Rule in incomplete information systems. Inf Sci 113:271–292
19. Kwak N, Choi CH (2002) Input feature selection for classification problems. IEEE Trans Neural Netw 13:143–159
20. Li DY, Zhang B, Leung Y (2004) On knowledge reduction in inconsistent decision information systems. Int J Uncertain Fuzziness Knowl Based Syst 12(5):651–672
21. Liang JY, Qian YH, Chu CY, Li DY, Wang JH (2005) Rough set approximation based on dynamic granulation. Lect Notes Comput Sci 3641:701–708
22. Liang JY, Dang CY, Chin KS, Yam Richard CM (2002) A new method for measuring uncertainty and fuzziness in rough set theory. Int J Gen Syst 31(4):331–342
23. Liang JY, Shi ZZ, Li DY, Wierman MJ (2006) The information entropy, rough entropy and knowledge granulation in incomplete information system. Int J Gen Syst 35(6):641–654
24. Mi JS, Wu WZ, Zhang WX (2004) Approaches to knowledge reduction based on variable precision rough set model. Inf Sci 159:255–272
25. Mi JS, Wu WZ, Zhang WX (2003) Comparative studies of knowledge reductions in inconsistent systems. Fuzzy Syst Math 17(3):54–60
26. Pavlenko T (2003) On feature selection, curse-of-dimensionality and error probability in discriminant analysis. J Stat Plan Inference 115:565–584
27. Pawlak Z (ed) (1991) Rough Sets: theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht
28. Qian YH, Liang JY (2006) Rough set model based on multi-granulations. In: Proceedings of 5th IEEE Conference on Cognitive Informatics. China I, pp 297–304
29. Qian YH, Liang JY, Dang CY (2010) Incomplete multi-granulation rough set. IEEE Trans Syst Man Cybern Part A 40(2):420–431
30. Qian YH, Dang CY, Liang JY, Zhang HY, Ma JM (2008) On the evaluation of the decision performance of an incomplete decision table. Data Knowl Eng 65:373–400
31. Qian YH, Liang JY, Li DY, Zhang HY, Dang CY (2008) Measures for evaluating the decision performance of a decision table in rough set theory. Inf Sci 178:181–202
32. Qian YH, Liang JY, Dang CY (2008) Converse approximation and rule extracting from decision tables in rough set theory. Comput Math Appl 55:1754–1765
33. Qian YH, Liang JY, Dang CY (2008) Consistency measure, inclusion degree and fuzzy measure in decision tables. Fuzzy Sets Syst 159:2353–2377
34. Qiu GF, Li HZ, Xu LD, Zhang WX (2003) A knowledge processing method for intelligent systems based on inclusion degree. Expert Syst 20(4):187–195
35. Shao MW, Liu M, Zhang WX (2007) Set approximations in fuzzy formal concept analysis. Fuzzy Sets Syst 158:2627–2640
36. Skowron A (1995) extracting laws from decision tables: a rough set approach. Comput Intell 11:371–388
37. Slezak D, Ziarko W (2005) The investigation of the Bayesian rough set model. Int J Approx Reason 40:81–91
38. Slezak D (2002) Approximate entropy reducts. Found Inf 53(3–4):365–390
39. Swiniarski RW, Skowron A (2003) Rough set methods in feature selection and recognition. Pattern Recognit Lett 24:833–849
40. Wang XZ, Zhai JH, Lu SX (2008) Induction of multiple fuzzy decision trees based on rough set technique. Inf Sci 178(16):3188–3202
41. Wang XZ, Tsand E, Zhao SY, Chen DG, Yeung D (2007) Learning fuzzy rules from fuzzy examples based on rough set techniques. Inf Sci 177(20):4493–4514
42. Wu WZ, Zhang M, Li HZ, Mi JS (2005) Knowledge reduction in random information systems via dempster-shafer theory of evedence. Inf Sci 174:143–164
43. Wu WZ, Leung Y, Mi JS (2009) On generalized fuzzy belief functions in infinite spaces. IEEE Trans Fuzzy Syst 17(2):385–397
44. Wu WZ, Leung Y, Mi JS (2009) Granular computing and knowledge reduction in formal contexts. IEEE Trans Knowl Data Eng 21(10):1461–1474
45. Xu BW, Zhou YM, Lu HM (2005) An improved accuracy measure for rough sets. J Comput Syst Sci 71:163–173
46. Xu WH, Zhang WX (2007) Measuring roughness of generalized rough sets induced by a covering. Fuzzy Sets Syst 158:2443–2455
47. Xu ZB, Liang JY, Dang CY, Chin KS (2002) Inclusion degree: a perspective on measures for rough set data analysis. Inf Sci 141:227–236
48. Yu DR, Hu QH, Wu CX (2007) Uncertainty measures on fuzzy relations and their applications. Appl Soft Comput 7:1135–1143
49. Zhang WX, Leung Y (eds) (1996) The uncertainty reasoning principles. Xian Jiaotong University Press, China
50. Zhang WX, Leung Y (1996) Theory of inclusion degrees and its applications to uncertainty inferences. Soft Comput IEEE Intell Syst Inf Process, New York, pp 496–501
51. Zheng P, Germano R, Ariën J, Qin KY, Xu Y (2006) Interpreting and extracting fuzzy decision rules from fuzzy information systems and their inference. Inf Sci 176:1869–1897
52. Zhu W, Wang FY (2003) Reduction and axiomization of covering generalized rough sets. Inf Sci 152:217–230
53. Zhu W, Wang SP (2011) Matroidal approaches to generalized rough sets based on relations. J Mach Learn Cybern. doi: 10.1007/s13042-011-0027-y
54. Ziarko W (1993) Variable precision rough set model. J Comput Syst Sci 46:39–59