Knowledge-Based Systems 91 (2016) 102–113

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A fuzzy multigranulation decision-theoretic approach to multi-source fuzzy information systems

Guoping Lin [a,b,*], Jiye Liang [b], Yuhua Qian [b], Jinjin Li [a]

[a] School of Mathematics and Statistics, Minnan Normal University, Zhangzhou, Fujian 363000, China
[b] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China

ARTICLE INFO

Article history:
Received 1 September 2014
Revised 8 September 2015
Accepted 16 September 2015
Available online 30 October 2015

Keywords:
Decision-theoretic rough set
Multigranulation
Gaussian kernel
Hadamart product
Fuzzy set
Granular computing

ABSTRACT

Decision-theoretic rough set theory (*DTRS*) is becoming one of the important research directions for studying set approximations using Bayesian decision procedure and probability theory in rough set community. In this paper, a novel model, fuzzy multigranulation decision-theoretic rough set model (*FM-DTRS*), is proposed in terms of inclusion measure of fuzzy rough sets in the viewpoint of fuzzy multigranulation. Gaussian kernel is used to compute the similarity between objects, which induces a fuzzy equivalence relation, and then we make use of T_p-norm operator with the property of Hadamart product to aggregate the multiple induced fuzzy equivalence relations. We employ the aggregated relation to fuzzily partition the universe and then obtain multiple fuzzy granulations from the multi-source information system. Moreover, some of its properties are addressed. A comparative study between the proposed fuzzy multigranulation decision-theoretic rough set model and Qian's multigranulation decision-theoretic rough set model is made. An example is employed to illustrate the effectiveness of the proposed method which may provide an effective approach for multi-source data analysis in real applications.

1. Introduction

Rough set theory, originated by Pawlak [27,28], is a mathematical tool to deal with uncertainty in a wide variety of applications [2–4,9,18,19,29,30,42,43,50]. In the past 10 years, several extensions of Pawlak rough set model have been proposed in terms of various requirements, such as the decision-theoretic rough set model [43], the variable precision rough set (*VPRS*) model [56], the rough set model based on tolerance relation [12,36], the Bayesian rough set model [39], the fuzzy rough set model [6] and the probabilistic rough sets [44]. The probabilistic rough sets, as an important research direction in rough set community, have been paid close attentions [10,11,13,14,44–47,49]. Specially, Yao [44] presented a new rule induction method based on the decision-theoretic rough set allowing for error tolerance through setting the thresholds: α and β, which is constructed by positive region, boundary region and negative region, respectively. Since then, the decision-theoretic rough sets have attracted more and more concerns. Azam and Yao [1] proposed a threshold configuration mechanism for reducing the overall uncertainty of probabilistic regions in the probabilistic rough sets. Jia et al. [10] developed an optimization representation of decision-theoretic rough set model, and gave a heuristic approach and a particle swarm optimization approach for implementing an attribute reduction with a minimum cost. Liu et al. [13,15] combined the logistic regression with the decision-theoretic rough set to form a new classification approach and investigated the three-way decision procedure with incomplete information combining the incomplete information table and loss function table together. Yu et al. [49] applied decision-theoretic rough set model to automatically determining the number of clusters with much smaller time cost. Li et al. [20] developed a sequential strategy in a decision process, which based on a formal description of granular computing.

In the view of granular computing (proposed by Zadeh [51]), in the existing rough set models, a concept described by a set is always characterized via the so-called upper and lower approximations under a single granulation, i.e., the concept is depicted by known knowledge induced from a single binary relation on the universe. Conveniently, this kind of rough set models is called single granulation rough sets. Based on a user's different requirements, Qian et al. [32] developed the multigranulation rough set which provides a new perspective for decision making analysis based on the rough set theory. Since the multigranulation rough set was proposed, the theoretical framework has been largely enriched, and many extended

* Corresponding author at: School of Mathematics and Statistics, Minnan Normal University, Zhangzhou, Fujian 363000, china. Tel.: +86 5962528109.
E-mail addresses: guoplin@163.com (G. Lin), ljy@sxu.edu.cn (J. Liang), jinchengqyh@126.com (Y. Qian), jinjinli@mnnu.edu.cn (J. Li).

http://dx.doi.org/10.1016/j.knosys.2015.09.022
0950-7051/

multigranulation rough set models and relative applications have also been proposed and studied [17,21–24,32,33,38,40,41]. For example, Qian et al. [33] have first proposed a new multigranulation rough set model through combining MGRS and the decision-theoretic rough sets together, called a multigranulation decision-theoretic rough set model. Sang et al. [31] proposed a new decision-theoretic rough set model based on the local rough set and the dynamic granulation principle, called a decision-theoretic rough set under dynamic granulation (DG-DTRS) which satisfies the monotonicity of the positive region of a target concept (or decision).

However, the decision and most of knowledge in the real life applications are often fuzzy and one often encounters a kind of special information system in which data come from different sources, such an information system is called a multi-source information system. Therefore it is necessary to introduce the fuzzy rough methodology into the classical *DTRS* for wider applications. The researchers [4,6,16,52–55] dealt with the real-value data sets by applying a fuzzy rough technique to solving problem. For example, Chen et al. [4] and Zhao et al. [53] used fuzzy rough sets to propose novel methods for attribute reduction and rule induction. Liang et al. [16] has proposed the triangular fuzzy decision-theoretic rough set by considering the losses being expressed by triangular fuzzy numbers. However, they still cannot be used to analyze data in the context of fuzzy multigranulation, which limits its further applications in many problems under the framework of the fuzzy environment. This motivates us to develop a new approximate strategy based on decision-theoretic rough sets to analyze data from the multi-source fuzzy information system.

Kernel methods have been proven to be an important methodology which is widely discussed in pattern recognition and machine learning domains. It maps data into a higher dimensional feature space in order to simplify classification tasks and make them linear [37,42]. In the rough set field, Hu et al. [7,8] found a high level of similarity between kernel methods and rough sets and made use of kernel to extract fuzzy relations for rough sets based data analysis. As an example, in this paper, Gaussian kernel is used to generate a fuzzy binary relation which satisfies reflexive, symmetric and transitive. However, the existing study is based on data coming from only a single source and little attention was paid to deal with data which come from different sources. To address this issue, in this study, we will introduce Gaussian kernel to extract a fuzzy equivalence relation between objects from a multi-source information system. It can be proven that the similarity matrix induced by Gaussian kernel is both reflexive and positive semi-definitive. Then we employ the T_P-norm operator: $T_P(a,b) = a \cdot b$ to aggregate multiple induced fuzzy relations to get an aggregation matrix which is called Hadamard product matrix [35]. It can induce a new fuzzy T_{cos}-equivalence relation because it is still both reflexive and positive semi-definitive, and then it will be used to partition a universe into a family fuzzy information granules forming a fuzzy granular structure for one of the sources from the multi-source information system. By the same way, one can obtain multiple fuzzy granular structures which constitute the fundamentals of the novel model in the paper.

From the above, based on different fuzzy granular structures generating from a multi-source fuzzy information system, the aim of this paper is to present a new approach to approximate the decision class with a certain level of tolerance for errors through inclusion measure between two fuzzy granules. This approach called fuzzy multigranulation decision-theoretic rough sets (*FM-DTRS*) combines the multigranulation decision-theoretic idea with fuzzy set theory. Some of its properties are addressed. A comparative study between the proposed and Qian's multigranulation decision-theoretic rough set model is made.

The rest of this paper is organized as follows. Some basic concepts of classical rough sets, variable precision fuzzy rough sets, and multigranulation rough sets are briefly reviewed in Section 2. In Section 3, we first investigate two fuzzy multigranulation decision-theoretic rough set forms that include the optimistic fuzzy multigranulation decision-theoretic rough sets, and the pessimistic fuzzy multigranulation decision-theoretic rough sets. Then, we analyze the loss function and the entire decision risk in the context of fuzzy multigranulation. When the thresholds have a special constraint, the multigranulation decision-theoretic rough sets will produce one of various variables of multigranulation rough sets. In Section 4, an example is used to illustrate our method. Finally, Section 5 concludes this paper by bringing some remarks and discussions.

2. Preliminaries

In this section, we introduce some basic notions and redescribe some related rough set models by inclusion degree, which are Pawlak rough sets, variable precision fuzzy rough sets, and multigranulation decision-theoretic rough sets, repestively [7,27,34,52,56]. Throughout this paper, let U be a finite non-empty set called the universe of discourse. The class of all fuzzy sets in U will be denoted as $F(U)$. For a set A, $|A|$ denotes the cardinality of the set A.

Definition 1. Assumed $\widetilde{R}$ is a fuzzy equivalence relation induced by a numerical attribute or fuzzy attribute. For any $x, y, z \in U$, it satisfies:

(1) reflexivity: $\widetilde{R}(x,y) = 1$;
(2) Symmetry: $\widetilde{R}(x,y) = \widetilde{R}(y,x)$; and
(3) Transitivity: $\widetilde{R}(x,z) \geq min(\widetilde{R}(x,y), \widetilde{R}(y,z))$.

The relation can be written as a matrix as

$$M(\widetilde{R}) = (\widetilde{r}_{ij})_{n\times n} = \begin{bmatrix} \widetilde{r}_{11} & \widetilde{r}_{12} & \cdots & \widetilde{r}_{1n} \\ \widetilde{r}_{21} & \widetilde{r}_{22} & \cdots & \widetilde{r}_{2n} \\ \vdots & & \ddots & \\ \widetilde{r}_{n1} & \widetilde{r}_{n2} & \cdots & \widetilde{r}_{nn} \end{bmatrix}$$

where $\widetilde{r}_{ij}$ is the similarity degree between x_i and x_j.

If condition (3) is replaced by $T(\widetilde{R}(x,y), \widetilde{R}(y,z)) \leq \widetilde{R}(x,z)$ called (T-transitivity), then $\widetilde{R}$ is said to be a fuzzy T-equivalence relation Kerre and Ovchinnikov where T is some triangular norm .

Definition 2. The fuzzy equivalence class $S_{\widetilde{R}}(x_i)$ of x_i induced by a fuzzy equivalence relation $\widetilde{R}$ is defined as

$$S_{\widetilde{R}}(x_i) = \frac{\widetilde{r}_{i1}}{x_1} + \frac{\widetilde{r}_{i2}}{x_2} + \cdots + \frac{\widetilde{r}_{in}}{x_n},$$

where '$+$' means the union operation. Obviously, $S_{\widetilde{R}}(x_i)$ is a fuzzy information granule containing x_i. $\widetilde{r}_{ij}$ is the degree of x_i equivalent to x_j. Obviously, a crisp equivalence class $[x]_R$ is a special fuzzy granule with $\widetilde{r}(x,y) = 1, y \in [x]_R$.

As we know, a fuzzy equivalence relation generates a family of fuzzy information granules from the universe, which composes a fuzzy equivalence granular structure, written by $K(\widetilde{R}) = \{S_{\widetilde{R}}(x_1), S_{\widetilde{R}}(x_2), \ldots, S_{\widetilde{R}}(x_n)\}$. Particularly, if $\widetilde{r}_{ii} = 1$ and $\widetilde{r}_{ij} = 0, j \neq i, i, j < n$, then $S_{\widetilde{R}}(x_i) = 1, i \leq n$, and $\widetilde{R}$ is called a fuzzy identity relation, and we write it as $\widetilde{R} = \varphi$; if $\widetilde{r}_{ij} = 1, i, j < n$, then $|S_{\widetilde{R}}(x_i)| = |U|, i \leq n$ and $\widetilde{R}$ is called a fuzzy universal relation, which is written as $\widetilde{R} = \delta$.

Definition 3. [7] Let A and B be two fuzzy granules in the universe U, the inclusion measure $I(A, B)$) is defined as

$$I(A,B) = \frac{|A \wedge B|}{|A|},$$

where '$\wedge$' means the operation 'min' and $|A| = \sum_{x\in U} \mu_A(x)$.

We denote $A \subset_{\varepsilon} B$, meaning $I(A,B) \geq \varepsilon$.

Pawlak rough set is based on the two fundamentals concepts: an equivalence relation R and a family of equivalence classes which is a partition of a finite non-empty universe U. If $U = \{x_1, x_2, \ldots, x_n\}$ is characterized with a collection of attribute, each attribute generates an indiscernible relation R on U. Then $< U, R >$ is called an