International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Distance-based double-quantitative rough fuzzy sets with logic operations

Wentao Li [a,b], Witold Pedrycz [b], Xiaoping Xue [a,*], Weihua Xu [c], Bingjiao Fan [d]

[a] Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, PR China
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, T6R 2V4, Canada
[c] School of Mathematics and Statistics, Southwest University, Chongqing, 400715, PR China
[d] Faculty of Information Technology, Macau University of Science and Technology, Taipa, Macau

ARTICLE INFO

Article history:
Received 18 March 2018
Received in revised form 12 July 2018
Accepted 18 July 2018
Available online xxxx

Keywords:
Distance matrix
Double quantification
Fuzzy similarity relation
Logic operation
Rough fuzzy set

ABSTRACT

Based on various requirements, many generalized rough set models have been developed to alleviate the limitations of generic Pawlak rough set theory and tackle different categories of information systems. One of the limitations is that rough set models based on equivalence relation are only applicable to discrete data information systems, and not suitable for dealing with real-valued continuous data without any prior processing. Another limitation is that "classical" rough sets do not consider the quantitative information about the degree of overlap between equivalence classes and the basic set, so they cannot cope well with the quantification problems. In this paper, we propose a framework of distance-based double-quantitative rough fuzzy set (Db-Dq-RFS) with logic operation by forming a distance-based fuzzy similarity relation in an information system with continuous data to simultaneously solve the two limitations. It is presented how to construct the distance-based fuzzy similarity relation in a normalized information system, and how to use this fuzzy similarity relation to generate distance-based single-quantitative rough fuzzy set (Db-Sq-RFS) models and the Db-Dq-RFS models with logic operation. The proposed Db-Dq-RFS models can overcome certain limitations of the classical rough set model. Following further studies to discuss the decision rules with parameters variation in the four kinds of Db-Dq-RFS models, we present an illustrative example to interpret the proposed developments and to verify the effect of parameters variation on decision rules. To illustrate the effectiveness of the parameters variation on decision rules, experimental evaluation is performed using five datasets coming from the University of California–Irvine (UCI) repository.

1. Introduction

Rough set theory [39] is an extension of set theory and a such could be regarded as a mathematical tool to handle imprecision, vagueness and uncertainty in data analysis. This relatively new soft computing methodology has received great attention in recent years, and its usefulness has been confirmed through successful applications in many areas science and engineering, such as pattern recognition, data mining, image processing, and medical diagnosis. Rough set theory is

* Corresponding author.
E-mail addresses: drliwentao@gmail.com (W. Li), wpedrycz@ualberta.ca (W. Pedrycz), xiaopingxue@263.net (X. Xue), chxuwh@gmail.com (W. Xu), bjfanmust@gmail.com (B. Fan).

https://doi.org/10.1016/j.ijar.2018.07.007
0888-613X/

built on the basis of the classification mechanism, it is classified as the equivalence relation in a specific universe, and the equivalence relation constitutes a partition of the universe. A concept, or more precisely the extension of a concept, is represented by a subset of a universe of objects and is approximated by a pair of definable concepts of a logic language. However, classical rough set models exhibit limitations in real-life applications. These limitations are mainly embodied and concluded as follows. Firstly, Pawlak rough sets do not deal well with the quantification problems. The relationship between equivalence classes and the basic set is so strict that there are no fault tolerance mechanisms available, and the quantitative information about the degree of overlap of the equivalence classes and the basic set is not taken into consideration. In fact, there are some degrees of inclusion relations between sets, and the extent of overlap of sets is important information to consider in applications. Secondly, the requirement imposed on an equivalence relation in Pawlak rough set model is a stringent condition that has limited the application domains of the theory. When we use the rough set to deal with numerical data, the rough set theory based on equivalence relations is mainly applicable to the information systems with discrete data. In case of continuous numerical data, a discretization process has to be completed, and this preprocessing will commonly result in loss of information, reducing the classification accuracy.

To overcome the first limitation, improving the Pawlak rough set model by incorporating quantitative information is a promising direction and model expansions that include such quantification are of particular relevance [26,34,61,67,72,73]. The improved models are called quantitative rough set models; they include probabilistic rough set (PRS) model [59,60] and graded rough set (GRS) model [62]. As pointed out in [72,73], PRS and GRS are the two different and typical single-quantitative rough set models. As to the second limitation, one of the main directions of research is to develop generalized rough sets by using the non-equivalence relation to take the place of the equivalence relation [4,29]. Many researchers have presented the notion of approximation operators by using tolerance relation [21,57], neighborhood relation [17,51], similarity relation [3,44], and others [18,19,25–28,31,52] to solve the second limitation. Pawlak rough set model can be extended to a fuzzy domain by replacing the equivalence relation with fuzzy equivalence relation [18,37]. Fuzzy equivalence relation satisfies reflexivity, symmetry and transitivity. As a more general and extensive relation than fuzzy equivalence relation, the similarity relation only satisfies the properties of reflexivity and symmetry.

PRS model and its generalizations can be formulated based on the notion of rough membership functions and rough inclusion. Threshold values, serving as parameters, are applied to a rough membership function or a rough inclusion to obtain probabilistic or parameterized approximations. Three probabilistic rough set models have been proposed and studied intensively, which are decision-theoretic rough set (DTRS) model [9,22,23,33,43,46,63], variable precision rough set model [76], and Bayesian rough set model [75]. The main differences among these models are their different, but equivalent, formulations of probabilistic approximations and interpretations of the required parameters. Since Yao and Lin explored the relationships between rough sets and modal logics, they proposed the GRS model based on graded modal logics [62]. GRS model primarily considers the absolute quantitative information regarding the basic concept and knowledge granules, and it is also a generalization of the Pawlak rough set model. The regions of the GRS model also extend the corresponding notions used in the classical rough set models. Because the inclusion relation of the grade approximations does not hold any longer, positive and negative regions, upper and lower boundary regions are naturally proposed. They classify the universe more precisely and exhibit their own logical meanings related to the grade quantitative index. GRS model considers absolute quantitative information between equivalence classes and the basic concept [32,72,73]. PRS model and GRS model can reflect relative quantitative information and absolute quantitative information about the degree of overlap between equivalence classes and a basic concept, respectively. The relative and absolute quantitative information are two distinct objective sides that describe approximate space, and each has its own virtues and pertinent application environments, so that none can be neglected. Relative quantitative information and absolute quantitative information are two kinds of quantification mythologies encountered in certain applications. From the examples reported in [26,72], both quantification indexes exhibit a close, supplementary, and dialectical relationship, and each one actually has its own representation virtues and application environments. It should be noted that the existing models regarding to double quantification studied in [11,26,54,64,68–73] are all based on equivalence relations. That is to say, these studies can only overcome the first limitation, while cannot address the second limitation.

As it has been pointed out before, the second limitation can be eliminated by replacing the equivalence relation with similarity relation. From the results presented in [5,15,17,53], the distance provides a comprehensible perspective for characterizing the difference between two objects in a metric space. In other words, distances between objects can describe similarities between them, which means that the distance matrices between objects can be used to induce similarity relations. Many different kinds of distance functions have been proposed to work for numerical attribute values in the research field of statistics [6], pattern cognition [8], and cognitive computing [38,50,55]. Recently, Cook et al. presented a general framework for distance-based consensus in ordinal ranking models [7]. Gesu and Starovoitov investigated the distance function for the application of image comparison [13]. Angiulli et al. studied a distance-based detection and prediction of outliers [1]. Khalifeh et al. covered some new results on distance-based graph invariants [20]. Luxburg and Bousquet used the Lischitz functions to make the distance-based classifications in information system [36]. Song et al. investigated the self-similarity of objects in the complex networks [45]. Leu et al. proposed a distance-based fuzzy time series model for exchange rates forecasting [24]. Yu et al. developed a distance-based group decision-making methodology for multi-person multi-criteria emergency decision support [65]. Luukka proposed the similarity classifier based on modified probabilistic equivalence relations [35]. Liang et al. introduced a set distance to understand measures from rough set theory from the viewpoint of distance [30]. Inspired by the above distance-based studies, a novel fuzzy similarity relation based on the

Please cite this article in press as: W. Li et al., Distance-based double-quantitative rough fuzzy sets with logic operations, Int. J. Approx. Reason. (2018), https://doi.org/10.1016/j.ijar.2018.07.007