Information Sciences 316 (2015) 54–67

Contents lists available at ScienceDirect

Information Sciences

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

Double-quantitative decision-theoretic rough set

Wentao Li [a], Weihua Xu [a,b,*]

[a] School of Mathematics and Statistics, Chongqing University of Technology, Chongqing 400054, PR China
[b] Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, PR China

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form 3 April 2015
Accepted 11 April 2015
Available online 17 April 2015

Keywords:
Bayesian decision
Decision-theoretic rough set
Double quantification
Graded rough set
Quantitative information

ABSTRACT

The probabilistic rough set (PRS) and the graded rough set (GRS) are two quantification models that measure relative and absolute quantitative information between the equivalence class and a basic concept, respectively. As a special PRS model, the decision-theoretic rough set (DTRS) mainly utilizes the conditional probability to express relative quantification. However, it ignores absolute quantitative information of the overlap between equivalence class and the basic set, and it cannot reflect the distinctive degrees of information and extremely narrow their applications in real life. In order to overcome these defects, this paper proposes a framework of double-quantitative decision-theoretic rough set (Dq-DTRS) based on Bayesian decision procedure and GRS. Two kinds of Dq-DTRS model are constructed, which essentially indicate the relative and absolute quantification. After further studies to discuss decision rules and the inner relationship between these two models, we introduce an illustrative case study about the medical diagnosis to interpret and express the theories, which is valuable for applying these theories to deal with practical issues.

1. Introduction

Rough set theory proposed by Pawlak [17], is an extension of the classical set theory and could be regarded as a mathematical and soft computing 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 effectiveness has been confirmed successful applications in many science and engineering fields, such as pattern recognition, data mining, image processing, and medical diagnosis. Rough set theory is 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. Rough set models give rise to a construct that highlights some items endowed with uncertainty [18]. The main idea of rough set is the use of a known knowledge in knowledge base to approximate the inaccurate and uncertain knowledge.

Pawlak rough set has a severe limitation. The relationship between equivalence classes and the basic set are strict that there are no fault tolerance mechanisms. Quantitative information about the degree of overlap of the equivalence classes

* Corresponding author at: School of Mathematics and Statistics, Chongqing University of Technology, Chongqing 400054, PR China. Tel.: +86 15998921583.

E-mail address: chxuwh@gmail.com (W. Xu).

http://dx.doi.org/10.1016/j.ins.2015.04.020
0020-0255/

and the basic set is not taken into consideration. Therefore, neither wider relationships nor quantitative information can be utilized. 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. The classical rough set model must be improved and expansions of the model that include quantification are of particular value. Improved models are called quantitative rough set model, and among them are the PRS model [30] and GRS model [29].

Recently, the PRS has been paid much attentions. The acceptance of PRS is merely due to the fact that they are defined by using probabilistic information and are more general and flexible. The introduction of probability enables the models to treat the universe of objects as samples from a much larger universe [19]. Probability is an important tool for describing uncertainty. The PRS model [25] has many merits, such as the measurability of the probabilistic information, the generality and flexibility of the model and its insensitivity to noise. The PRS model has been investigated extensively, and many concrete realizations of the model are available, including DTRS [24], game-theoretic rough sets [4], variable precision rough sets [39], 0.5-PRS [25], parameterized rough set [3], Bayesian rough sets [38] and fuzzy PRS [10,36]. The main differences among those models are their different, but equivalent, formulations of probabilistic approximations and interpretations of the required parameters. PRS models use conditional probability to quantify the degree of set inclusion. Notions can be interpreted in terms of probabilities or a posteriori probabilities [25]. Thresholds on the probability are used to define rough set approximations. The threshold values, known as parameters, are applied to a rough membership function or a rough inclusion to obtain probabilistic or parameterized approximations. However, when we apply the PRS models to the real-life issues, the conditional probability is calculated based on the rough membership functions. It is a kind of relative error from the form, in other words, the conditional probability reflects the relative quantitative information [12,20,33]. All of the existing PRS model using the rough membership functions are related to the relative quantitative information.

Because the problem of noisy data is substantially mitigated, the DTRS model is highly useful in data acquisition and analysis, it is an expansion of the Pawlak rough set model. The DTRS model has become increasingly popular in a variety of theoretical and practical fields, producing many thorough results. DTRS implies the ideal of three-way decisions [5,27]. Yao presented a new decision making method based on the DTRS, which is called the three-way decision theory, namely, the decision rules obtained from positive region, negative region and boundary region [15,25,27]. In fact, PRS is developed based on the Bayesian decision principle and Bayesian decision procedure, in which its parameters can be learned from a decision table. The three-way decision rules have much more superiority than both two-way decision and Pawlak's classical decision rules [26]. The DTRS can derive various rough set models through setting the thresholds. Since the DTRS was proposed by Yao in 1990 [30], it has attracted much more attentions. Professor Yao gave a decision theoretic framework for approximating concepts in 1992 [23] and later applied this model to attribute reduction [31]. Azam and Yao proposed a threshold configuration mechanism for reducing the overall uncertainty of probabilistic regions in the PRS [1]. Jia et al. proposed an optimization representation of DTRS model and raised an optimization problem by considering the minimization of the decision cost [6,7]. Liu et al. combined the logistic regression and the DTRS into a new classification approach, which can effectively reduce the misclassification rate [14]. Ma et al. explored the PRS model by considering two universes and accordingly discussed the rough entropy [15]. Yu et al. applied DTRS model for automatically determining the number of clusters with much smaller time cost [32]. Qian et al. combined the thought of multigranulation into DTRS, then proposed three kinds of multigranulation DTRS model [19]. Later on that, Li et al. developed a probabilistic rough set model by considering dominance relations other than equivalence relations [9], and then further studied multigranulation DTRS in an ordered information system [8]. Professor Zhou introduced a kind of DTRS model for an information table with more than two decision classes, which is the multi-class DTRS [37]. These studies represents a snapshot of recent achievements and developments on the DTRS theory. In particular, Greco et al. presented a generalized variable precision rough set model using the absolute and relative rough membership [3]. The inclusion degree, as a generalization of the rough membership, has been used extensively in the study of measures, reasoning, applications of uncertainty and approximate spaces.

GRS model [13,22,29] has many features in common with DTRS model and functions as a typical expansion model by including quantification. The DTRS based on Bayesian decision procedure and the GRS are two fundamental expansion models that achieve strong fault tolerance capabilities by utilizing quantitative descriptions. Since Yao and Lin explored the relationships between rough sets and modal logics, they proposed the GRS model based on graded modal logics [29]. GRS model primarily considers the absolute quantitative information regarding the basic concept and knowledge granules and is a generalization of the Pawlak rough set model. The regions of the GRS model are extensions of grade approximations. 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. Obviously, regions of the GRS model also extend the corresponding notions of the classical rough set model. They classify the universe more precisely and have their own logical meanings related to the grade quantitative index, in another way, GRS models consider absolute quantitative information between equivalence classes and the basic concept [34,35].

DTRS model and GRS model can respectively reflect relative and absolute quantitative information about the degree of overlap between equivalence classes and a basic set. The relative and absolute quantitative information are two distinctive objective sides that describe approximate space, and each has its own virtues and application environments, so that none can be neglected. Relative quantitative information and absolute quantitative information are two kinds of quantification mythology in certain applications. Usually, most researchers prefer using the relative quantitative information [7,11,15,16,24,25,38,39]. However, the absolute quantitative information is more important than or as important as the relative quantitative information in some specific fields or special cases, many corresponding examples can be found in practice. We introduce three