Contents lists available at ScienceDirect
International Journal of Approximate Reasoning
www.elsevier.com/locate/ijar

The uncertainty of probabilistic rough sets in multi-granulation spaces ☆

Qinghua Zhang [a,b,*], Qiang Zhang [a], Guoyin Wang [b]

[a] School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[b] Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form 5 May 2016
Accepted 1 June 2016
Available online xxxx

Keywords:
Rough sets
Probabilistic rough sets
Uncertainty measure
Multi-granulation spaces
Threshold

ABSTRACT

Pawlak's rough sets model describes an uncertain target set (concept) with two crisp boundary lines (i.e. lower and upper approximation sets) and as an effective tool has successfully been used to deal with uncertain information systems. Based on the rough sets model, a probabilistic rough sets model with a pair of thresholds was proposed to improve the fault-tolerance ability of rough sets. The uncertainty of Pawlak's rough sets model is rooted in the objects contained in the boundary region of the target concept, while the uncertainty of probabilistic rough sets model comes from three regions, because the objects in the positive or negative regions are probably uncertain, and the membership degrees of these objects are not necessary equal to 1 or 0. In this paper, a method for measuring the uncertainty of probabilistic rough sets is proposed, and the change rules of uncertainty with changing knowledge spaces are presented and analyzed. Then, for an uncertain target concept, the uncertainties of the three regions are discussed, and the related change rules of uncertainty with changing knowledge spaces are revealed and successfully proved. Finally, a comparative analysis on the uncertainty of a target concept in rough sets and probabilistic rough sets model is presented. These results are important to further enrich and improve probabilistic rough sets theory, and effectively promote the development of uncertainty artificial intelligence.

1. Introduction

The probabilistic rough sets model is an extension of Pawlak's rough sets model. It allows a tolerance of inaccuracy in the lower and upper approximation sets, and attempts to generalize the restrictive definition of the lower and upper approximations by allowing certain acceptable levels of errors [1–5]. In other words, by the probabilistic value of an object belonging to a target concept, the probabilistic rough sets model classifies the domain into three regions, namely the positive, boundary and negative regions. In this extension model, a pair of parameters α, β $(0 \le \beta < \alpha \le 1)$ is used to control the classification accuracy in order to improve the fault-tolerance ability of the rough sets model. The theories and applications are researched and many important results have been obtained [3–8].

☆ This work is supported by the National Natural Science Foundation of P.R. China (Nos. 61472056, 61272060), The Natural Science Foundation of Chongqing of P.R. China (No. CSTC2013jjB40003) and the Science and Technology Research Program of Chongqing Education Committee of P.R. China (No. KJ130534).

* Corresponding author at: Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.
E-mail address: zhangqh@cqupt.edu.cn (Q. Zhang).

http://dx.doi.org/10.1016/j.ijar.2016.06.001
0888-613X/© 2016 Elsevier Inc. All rights reserved.

In Pawlak's rough sets model [9], for a target concept, the domain will be divided into three disjoint regions by determining whether each object belongs to the target concept: if an object belongs to the target concept, this object is classified into the positive region; if an object definitely does not belong to the target concept, it is classified into negative region; if we can't certainly determine whether the object belongs to the target concept, the object is classified into the boundary region. From the viewpoint of three-way decision theory [5,10–12], the object in the positive region is the accepted state for the decision-maker, the object in the negative region is the rejected state, and the object in the boundary region is the deferred state or no commitment [13–20]. In the real world, the smaller or narrower the boundary region, the higher the classification accuracy. In contrast, if the boundary region is so big or wide that the classification can't satisfy the user's requirement, the boundary region should be subdivided according to some rules in order to expand the positive or negative regions. Based on this idea, probabilistic rough sets model was proposed to reduce the boundary region and increase the ability of fault-tolerance with a pair of parameters, at the cost of classification accuracy [21].

The shadowed set proposed by Pedrycz [22–25] is an important approximation of a fuzzy set, which reduces low-membership grades to 0 and elevating high-membership grades to 1. It reveals interesting conceptual and algorithmic relationships existing between the rough and fuzzy sets. A shadowed set divides the domain into three disjoint regions by modifying the membership degree of a fuzzy set into ternary $\{0, [0, 1], 1\}$, which can be treated as the degrees of membership of elements to a target concept. Objects with membership grade 1 constitute the core of the shadowed set, and objects with membership grade $[0, 1]$ constitute the shadow of the shadowed set. In general, the uncertainty of the object elevating the membership degree to 1 or reducing the membership degree to 0 is reduced to 0. In order to balance the uncertainty of the shadowed sets, the condition 'Reduction of membership + Elevation of membership = Shadow' is commonly satisfied [26]. In other words, the common feature of both the probabilistic rough sets model and the shadowed sets model is reducing the boundary region and increasing the amount of certain objects, in order to simplify the computational model to process the uncertain problems. The two models focus on reducing uncertainty by elevating the membership degree of some objects with a high membership degree to 1, and reducing the membership degree of the objects with a low membership degree to 0.

Research on the uncertainty of rough sets is important in order to acquire approximate rules from information systems. Many typical uncertain measurements, such as roughness, approximation accuracy, rough entropy, fuzzy entropy, fuzziness and so on, were proposed gradually [27–34]. Of course, measuring the uncertainty of a probabilistic rough sets model has attracted many researchers' attention, because this analysis can show the quality of acquiring rules from uncertain information systems. Based on game-theoretic rough sets, Azam and Yao [35] analyzed the uncertainties of probabilistic rough sets with three aspects: the positive, negative and boundary regions. In [35], Azam and Yao used the game-theoretic rough sets model to construct a mechanism for analyzing the uncertainties of rough sets, in the aim of finding effective threshold values. They found that a competitive game was formulated between the regions that modify the thresholds in order to improve their respective uncertainty levels, so a learning mechanism was proposed that automatically tunes the thresholds based on the data itself. Wang and Ma [36] constructed the monotonic uncertainty measures method for attribute reduction in probabilistic rough sets in order to acquire approximate rules with incremental knowledge acquisition methods. Ma and Sun [37] presented a limitation of the uncertainty measure for the traditional method and proposed the Shannon entropy of covering (based on the universe) through analyzing many properties of the uncertainty measure, such as the roughness and accuracy for probabilistic rough sets over two universes. This research effectively improved the probabilistic rough sets model and extended the application range of rough sets [38–40].

The analysis on the uncertainties of three regions is helpful in improving classification quality. Compared with Pawlak's rough sets model, the probabilistic rough sets model classifies a domain into three regions and has a good fault-tolerance ability. With the increase of attribute information, not only the objects in the boundary region, but also the objects in the positive or negative regions may be reclassified. In many application fields, such as disease diagnosis, screening suspect, risk decision and so on, the uncertainty analysis of three regions is important to further judge uncertain objects. Particularly when we use a three-way decision model to deal with risk decision problems, the uncertainty analysis of three regions is very useful in making a low risk decision.

A boundary region is a key factor for generating the uncertainty of probabilistic rough sets, but it is not the only factor, because the uncertainty of probabilistic rough sets probably comes from the positive or negative regions. Many scholars divided the uncertainty of probabilistic rough sets into three parts: uncertainty of the positive region, uncertainty of the boundary region and uncertainty of the negative region [35]. This is different to the traditional rough sets model because the uncertainty of Pawlak's rough sets only comes from the boundary region. With the increase of new attributes or information, the granules (namely, equivalence classes) in Pawlak's knowledge space will be subdivided into finer granules. In the view of multi-granulation rough sets (MGRS) [41], with the increase of condition attributes in an information system the equivalence classes will become small gradually. A variety of multi-granulation extended rough sets models were proposed and developed. Liu [42] proposed four types of multi-granulation covering rough sets (MGCRS) models under a covering approximation space. Xu [43–46] presented two types of multi-granulation tolerance rough sets (MGTRS) models and multi-granulation fuzzy rough sets (MGFRS) models based on tolerance relations, and he established two new types of multi-granulation rough sets models based on ordered information system. Based on Bayesian decision-theoretic rough sets, Qian [47] developed a multi-granulation decision-theoretic rough sets model, and Li [48] analyzed the multi-granulation decision-theoretic rough sets model in ordered information systems to extend the theory of decision-theoretic rough sets. Yang [49,50] defined the optimistic and pessimistic fuzzy rough sets respectively in a multi-granulation space,

Please cite this article in press as: Q. Zhang et al., The uncertainty of probabilistic rough sets in multi-granulation spaces, Int. J. Approx. Reason. (2016), http://dx.doi.org/10.1016/j.ijar.2016.06.001