International Journal of Approximate Reasoning 59 (2015) 41–67

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Monotonic uncertainty measures for attribute reduction in probabilistic rough set model

Guoyin Wang [a], Xi'ao Ma [a,b,*], Hong Yu [a]

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

[b] School of Information Science and Technology, Southwest Jiaotong University, Chengdu, 610031, PR China

ARTICLE INFO

Article history:
Received 1 June 2014
Received in revised form 5 January 2015
Accepted 19 January 2015
Available online 21 January 2015

Keywords:
Uncertainty measure
Approximation accuracy
Attribute reduction
Probabilistic rough set model
Pawlak rough set model

ABSTRACT

Attribute reduction is one of the most fundamental and important topics in rough set theory. Uncertainty measures play an important role in attribute reduction. In the classical rough set model, uncertainty measures have the monotonicity with respect to the granularity of partition. However, the monotonicity of uncertainty measures does not hold when uncertainty measures in classical rough set model are directly extended into probabilistic rough set model, which makes it not so reasonable to use them to evaluate the uncertainty in probabilistic rough set model. Moreover, the monotonicity is very important for constructing attribute reduction algorithms because the monotonicity of uncertainty measures can simplify the algorithm design. This paper focuses on constructing monotonic uncertainty measures in probabilistic rough set model. Firstly, we analyze the non-monotonicity problem of uncertainty measures in probabilistic rough set model. Secondly, we propose three basic uncertainty measures and three expected granularity-based uncertainty measures, the monotonicity of these measures is proved to be held and the relationship between these measures and corresponding uncertainty measures in classical rough set model is also obtained. Finally, a new attribute reduct is defined based on the proposed monotonic uncertainty measure, and the corresponding heuristic reduction algorithms are developed. The results of experimental analysis are included to validate the effectiveness of the proposed uncertainty measures and new reduct definition.

1. Introduction

Rough set theory introduced by Pawlak [29] is a valid mathematical theory that deals well with imprecise, vague and uncertain information, and it has been applied in many research fields successfully, such as machine learning, data mining, knowledge discovery, intelligent data analyzing [12,13,23,34]. Attribute reduction is one of the most fundamental and important researches in rough set theory [11,26,28,37]. Attribute reduction can improve the classification performance of learning algorithm, simplify data description and avoid overfitting.

In rough set theory, uncertainty measures play a vital role in attribute reduction. Uncertainty measures can supply new points of view for analyzing data and help us to disclose the substantive characteristics of data sets [6]. Therefore, uncer-

* Corresponding author at: School of Information Science and Technology, Southwest Jiaotong University, Chengdu, 610031, PR China. Tel.: +86 13608348037.

E-mail addresses: wanggy@ieee.org (G. Wang), maxiao73559@163.com (X. Ma), hongyu.cqupt@gmail.com (H. Yu).

http://dx.doi.org/10.1016/j.ijar.2015.01.005
0888-613X/

tainty measures have been receiving increasing attention from researchers [7,5,15,20,33,40]. Generally, the methodologies dealing with uncertainty measurement problem in rough set theory can be classified into two categories: the pure rough set approach and the information theory approach [7].

For pure rough set approach, some numerical measures are used as uncertainty measures. For example, Pawlak [30] presented three numerical measures, including accuracy measure, roughness measure and approximation accuracy measure in a bid to evaluate uncertainty of a rough set and a rough classification. Liang et al. [22] revealed the limitations of accuracy measure, roughness measure and approximation accuracy measure, and modified them by multiplying by a measure of granularity of partitions. Dai and Xu [7] investigated three types of lower and upper approximations and their corresponding accuracy measures, roughness measures and approximation accuracy measures in incomplete information system.

For information theory approach, Shannon's entropy and its variants are applied to study the uncertainty measurement problem. For example, Duntsch and Gediga [9] proposed three types of model selection criteria by using information entropy for uncertainty measures of rough set prediction. Beaubouef et al. [2] addressed the measurement of uncertainty in rough sets and rough relational databases by introducing a measurement based on information theory. Wierman [39] presented a well justified measure of uncertainty for rough set theory as well as an axiomatic derivation. Liang et al. [20] defined a new information entropy, which is called the complementary entropy, based on the complement behavior of information gain for measuring both uncertainty and fuzziness in rough set theory. Qian et al. [32] introduced the combination entropy, the conditional combination entropy and the mutual information to evaluate the uncertainty in rough set theory. Cattaneo et al. [3,4] analyzed the information entropy and granulation co-entropy of partitions and coverings and discussed their monotonicity, which gives some ideas with respect to the definition of monotonic information measures.

It is important to note that the afore-mentioned researches on the uncertainty measurement have been conducted on the classical rough set model. Regrettably, the classical rough set model assumes the classification must be fully correct or certain. Hence, it cannot effectively deal with data sets which have noisy data and latent useful knowledge. To alleviate this problem, some excellent probabilistic rough set models [25,27,36,47], such as 0.5 probabilistic rough set model [31], variable precision rough set model [46], Bayesian rough set model [35], game rough set model [14] and decision-theoretic rough set model [41], are proposed by introducing probabilistic threshold values.

So far, however, few studies have been done on uncertainty measurement in probabilistic rough set model [38]. In this paper, we focus on the uncertainty measurement problem in probabilistic rough set model, and intend to construct effective uncertainty measures by pure rough set approach. As is known to all, in the classical rough set model, the lower and upper approximations are defined by requiring the set inclusion must be fully correct or certain, namely, the definitions of the lower and upper approximations do not allow any tolerance of errors. Therefore, the accuracy measure, the roughness measure and the approximation accuracy measure, which are used as uncertainty measures, based on the lower and upper approximations are monotonous with respect to the granularity of partitions. In other words, the finer the partition is, the more the available knowledge is, and correspondingly the smaller the uncertainty is. The coarser the partition is, the less the available knowledge is, and correspondingly the greater the uncertainty is. This process is much similar to the reasoning of human's mind. Hence, it is reasonable to use these measures as uncertainty measures in classical rough set model [7].

Unfortunately, we will find the parallel extended versions of accuracy measure, roughness measure and approximation accuracy measure are not monotonic with respect to the granularity of partitions because the lower and upper approximations are defined by introducing the probabilistic threshold values in probabilistic rough set model, which makes it not so reasonable to use them as uncertainty measures in probabilistic rough set model. Furthermore, the uncertainty measures are often used to construct the attribute reduction methods, and the monotonicity of uncertainty measures is very important for developing attribute reduction algorithms. However, the non-monotonicity of uncertainty measures will complicate the algorithm design. Thus, it is necessary to further study on monotonic uncertainty measures in probabilistic rough set model.

This paper aims to propose some monotonic uncertainty measures in the probabilistic rough set model. To begin with, three basic uncertainty measures are proposed in probabilistic rough set model. Furthermore, three modified uncertainty measures are further presented based on the expected granularity of blocks in a partition which is proposed by Yao and Zhao [42] to provide a unified framework for measures of granularity of a partition. The proposed measures are proven to have the monotonicity with respect to the granularity of partitions as well as to be natural generalizations of the corresponding uncertainty measures in the classical rough set model. The monotonicity experiments have been conducted on some real-life data sets to test and verify the validity of the proposed measures.

Furthermore, we provide attribute reduction approaches based on monotonic uncertainty measures in probabilistic rough set model. Firstly, a new attribute reduct is defined by using the proposed monotonic uncertainty measure. Secondly, the core and attribute importance are defined. Thirdly, heuristic algorithms are developed based on deletion strategy and addition–deletion strategy [44]. Finally, some experiments are conducted to show the effectiveness of the proposed definition of attribute reduct.

The rest of the paper is organized as follows. Section 2 presents some preliminary knowledge regarding uncertainty measures in classical rough set model. In Section 3, the non-monotonicity of uncertainty measures are analyzed in probabilistic rough set model. Some monotonic uncertainty measures are proposed and their properties are also investigated in Section 4. In Section 5, a new attribute reduct is defined by using proposed monotonic uncertainty measure, and the corresponding heuristic algorithms are constructed. Simulation experiments are conducted to verify the validity of the proposed uncertainty measures and to show the effectiveness of the proposed definition of attribute reduct in Section 6. Finally, conclusions are discussed in Section 7.