Information Sciences 290 (2015) 72–85

Contents lists available at ScienceDirect

Information Sciences

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

Automatic determination about precision parameter value based on inclusion degree with variable precision rough set model

Yusheng Cheng [a,*], Wenfa Zhan [b,1], Xindong Wu [c,d,2], Yuzhou Zhang [a,3]

[a] School of Computers and Information, Anqing Normal University, Anqing 246011, China
[b] Department of Science and Technology, Anqing Normal University, Anqing 246011, China
[c] School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China
[d] Department of Computer Science, University of Vermont, Burlington, VT 05405, USA

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form 3 August 2014
Accepted 13 August 2014
Available online 23 August 2014

Keywords:
Precision parameter value
Decision-theoretic rough set
Symmetric variable precision rough set
Decision table

ABSTRACT

The rough set theory provides a powerful approach for attributes reduction and data analysis. The variable precision rough set (VPRS) model, an extension of the original rough set approach, tolerates misclassifications of the training data to some degree, which promotes the applications of rough set theory in inconsistent information systems. However, in most existing algorithms of feature reduction based on VPRS, the precision parameter (β) is introduced as prior knowledge, which restricts their applications because it is not clear how to set the β value. By studying β-consistency in the measurement of a decision table and the threshold value of the β-consistent decision table, this paper presents an algorithm for automatic determination of the precision parameter value from a decision table based on VPRS. At the same time, the precision parameter value from our proposed method is compared with the thresholds from the decision-theoretic rough set (DTRS). The influences of the precision parameter are also discussed on attribute reduction, which shows the necessity of the estimated precision parameter from a decision table. The simulation results including VPRS and other classification methods in real data further indicate that different precision parameter values make a great difference on rules and setting a precise parameter near the threshold value of the β-consistent decision table can precisely reflect the decision distribution of the decision table.

1. Introduction

As a new mathematical tool for dealing with inexact, uncertain knowledge, the rough set theory (RST) has been successfully employed in machine learning, data mining and other fields since it was put forward by Pawlak [24]. RST can be used to model classification but the classification must be fully correct or certain, which limits the practical applicability of RST in real world applications. As a probabilistic rough set model (PRS), the decision-theoretic rough set model

* Corresponding author. Tel.: +86 18055665796.
E-mail addresses: chengyshaq@163.com (Y. Cheng), zhanwf@aqtc.edu.cn (W. Zhan), xwu@uvm.edu (X. Wu), zhangyzh@aqtc.edu.cn (Y. Zhang).
1 Tel.: +86 13955612021.
2 Tel.: +1 802 656 7839.
3 Tel.: +86 15385562628.

http://dx.doi.org/10.1016/j.ins.2014.08.034

(DTRS) [2,7,8,10–22,26,27,32–36,38] extended RST by considering probabilistic information of objects into a set of positive, negative or boundary regions, by introducing a pair of thresholds (α, β). When proper thresholds are used, we can derive several existing probabilistic rough set models, such as the variable precision rough set (VPRS) [40,41], the game-theoretic rough set (GTRS) [2,12], and the Bayesian rough set (BRS) [23] from the decision-theoretic rough set model. However, how to determine and interpret the thresholds is fundamental issues in probabilistic rough sets. At present, many methods about the thresholds have been proposed by utilizing different cost functions to minimize the overall classification cost [2,10,16]. Deng and Yao [10]proposed an approach that can obtain the required threshold parameters directly from the data by utilizing the measure of Shannon entropy, which is known as the information-theoretic method. How to find an effective threshold pair was formulated as mining the uncertainty of Shannon entropy of the three regions. The information-theoretic method only provides a symmetric threshold to search the optimal threshold in the matrix. That is, the thresholds (α, β) were converted into (α, $1 - \alpha$). In order to determine an optimal threshold in the space of all possible threshold pairs, Azam and Yao [2,12] proposed a GTRS method to obtain the required threshold parameters. The GTRS method reformulates the conditional entropy in [10] as the overall uncertainty $\Delta(\alpha, \beta)$. A repetitive game was formulated where the players cooperatively determine suitable thresholds. Jia [15,16] also proposed an optimization method by minimizing the decision cost through six cost functions. Based on the suggested minimization problem, the thresholds and cost functions can be automatically obtained from data directly. These PRS based efforts have provided methods that can learn the threshold parameters automatically by searching the minimization of uncertainty of the classification of the three regions from the data.

As one of the most important branches in the rough set theory,VPRS was proposed by Ziarko in 1993. The majority inclusion relation and the weak dependence were the core of VPRS [40,41]. So our proposed method based on VPRS with the inclusion degree can also obtain the optimal threshold automatically by only searching the maximization of the inclusion degree of the positive or negative region [37].

VPRS deals with partial classification by introducing a probability value i.e. the precision parameter β. This parameter represents a bound on the conditional probability of a proportion of objects in a condition class that are classified into the same decision class. An et al. [1] used the symbol β to denote the proportion of correct classifications, the appropriate value range of which is (0.5 1.0]. However, Ziarko [1,40] considered β as a classification error, defined in the domain of [0.0 0.5). In this paper, we use the original meaning of precision parameter β [1,40], which is equivalent to the β parameter of the threshold pair (α, β) in DTRS. Therefore, we also provide a symmetric threshold ($1 - \beta$, β) method from the viewpoint of DTRS [2,7,8,10–22,26,27,32–36,38].

On the other hand, the VPRS model is parametric and it is difficult to set the precision parameter value (β) in some applications. Therefore, relevant β values are often given subjectively and they need to be adjusted in order to improve the quality of classification. At present, in most VPRS-related studies, β is introduced as prior knowledge, which runs against the prominent advantage of RST—'Let data speak by itself', i.e. avoid any other information outside the underlying information system. If β could be generated from the data to be processed during the reduction, it will surely play an important role in promoting the development and application of the VPRS model. Su et al. [30]provided a method to calculate the precision parameter value based on the least upper bound of the data misclassification error. Ziarko [40] proposed that the precision parameter value be specified by the decision maker. Beynon [3] proposed two methods of selecting a β-reduct without such a known β value. Slezak [28,29] put forward a Bayesian rough set model, in which the parameter was defined by a prior probability. Beynon [4,5] proposed an allowable β value range to be an interval, where the quality of classification may be known prior to determining the β value range. Katzberg et al. [17] allowed asymmetric bounds l and u to be used, with which the restrictions $l < 0.5$ and $u = 1 - l$ must hold. Zhou et al. [39] proposed a method to determine the precision parameter value based on the least upper bound of the data misclassification error. Beynon [6] utilized the ($l - u$)-graphs to choose the values of l and u based on the associated levels of the quality of classification and the degree of dependency. Despite that all these methods aim at searching for an optimal precision parameter objectively, they still focus on a special β value and attribute reduction anomalies will occur inevitably under the classic reduct definition [39]. Three-region decision about DTRS [2,7,8,10–22,26,27,32–36,38] is currently a hot topic. Yao [10,34–36], Jia [15,16] and others [2,12,19,20] have provided several methods for finding the parameters. In our previous study [9], we also discussed the influences of the precision parameter on attribute reduction.

As a flexible tool for analyzing the optimal thresholds, the information-theoretic method requires tremendous computations and is difficult to implement. Inspired by the discernibility matrices proposed by Skowron [25], the β-consistency about an information system is proposed in this paper, which develops the β-discernibility about a set X. The threshold about a β-consistent decision table, can be obtained by computing the threshold about the set from different decision classes. In order to facilitate the realization about our method, we further propose a combined probability matrix of a decision information system. The time complexity of our method is significantly smaller than DTRS methods.

Moreover, the paper is further concerned with the relations between β and the β-consistent decision table and influences of β on attribute reduction such as β lower distribution reduction, β upper distribution reduction and β lower approximate reduction.

The rest of the paper is organized as follows. Section 2 gives some basic notions related to VPRS and the relative discernibility of a set. Section 3 introduces the measurement approach of the β-consistent decision table and how to compute the value of β from a decision table. Our proposed method is compared with the thresholds from DTRS in Section 4. In Section 7, we prove that for some special threshold values, β lower distribution reduction is equivalent to β upper distribution reduction, whereas β lower approximate reduction can obtain classification decision rules with certainty under a given