International Journal of Approximate Reasoning 71 (2016) 34–49

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Multi-objective optimization method for learning thresholds in a decision-theoretic rough set model

Ruilin Pan *, Zhanchao Zhang, Yanglong Fan, Jianhua Cao, Ke Lu, Tingsheng Yang

School of Management Science and Engineering, Anhui University of Technology, Maanshan 243032, China

A R T I C L E I N F O

Article history:
Received 24 July 2015
Received in revised form 17 January 2016
Accepted 18 January 2016
Available online 21 January 2016

Keywords:
Decision-theoretic rough set model
F_measure
Multi-objective optimization
Pareto optimal set
Threshold learning

A B S T R A C T

For decision-theoretic rough sets, a key issue is determining the thresholds for the probabilistic rough set model by setting appropriate cost functions. However, it is not easy to obtain correct cost functions because of a lack of prior knowledge and few previous studies have addressed the determination of learning thresholds and cost functions from datasets. In the present study, a multi-objective optimization model is proposed for threshold learning. In our model, we integrate an objective function that minimizes the decision cost with another that decreases the size of the boundary region. The ranges of the thresholds and two types of F_measure are used as constraints. In addition, a multi-objective genetic algorithm is employed to obtain the Pareto optimal set. We used 12 UCI datasets to validate the performance of our method, where the experimental results demonstrated the trade-off between the two objectives as well as showing that the thresholds obtained by our method were more intuitive than those obtained using other methods. The classification abilities of the solutions were improved by the F_measure constraints.

1. Introduction

The Pawlak rough set proposed in the early 1980s [1] has been applied in many research fields such as data mining [2,3] and machine learning [4,5]. However, the traditional model is too strict to include objects in the approximation regions. Thus, probabilistic rough set models were introduced to loosen the extreme membership requirements of the equivalence classes in the object set [6]. As a special type of probabilistic rough set model, decision-theoretic rough set (DTRS) models [7] can be used to derive several probabilistic rough set models, e.g., the 0.5 probabilistic rough set model [8] and variable precision rough set model [9].

Previous research into DTRS models can be summarized briefly as follows. First, a series of studies addressed the extension of the DTRS model. Yao [10] studied the derivation of other probabilistic rough set models from the DTRS. Liu et al. [11] introduced three-way decision discriminant analysis into the DTRS model, while Yao and Zhou [12] introduced naive Bayesian classification into DTRSs. Based on the current status of the DTRS model, Yang and Yao [13], Zhou [14], and Li and Zhou [15] proposed multi-agent, multi-class, and multi-view DTRS models, respectively, as different DTRS model extensions.

* Corresponding author.
E-mail addresses: alltimefight@gmail.com (R. Pan), 1576252368@qq.com (Z. Zhang), 646149326@qq.com (Y. Fan), jhcao_74@163.com (J. Cao), looker@ahut.edu.cn (K. Lu), 845250374@qq.com (T. Yang).

http://dx.doi.org/10.1016/j.ijar.2016.01.002
0888-613X/

Using a Bayesian decision procedure and graded rough set, Li and Xu [16] proposed a new framework for DTRS called the double-quantitative-DTRS. By considering the new expression of evaluation information using hesitant fuzzy sets (HFSs), Liang and Liu [17] introduced HFSs into DTRSs and explored their decision mechanisms. Other studies have concentrated on the methodology of DTRS theory, such as attribute reduction and rule acquisition. In particular, Yao and Zhao [18], Min et al. [19], and Jia et al. [20] studied attribute reduction with respect to DTRS theory from different viewpoints. In [18], different classification properties, such as coverage, cost, confidence, decision-monotonicity, and generality, were regarded as important factors for attribute reduction. In [19], the new problem of minimal test cost reduction was proposed, where three metrics for evaluating the performance of the reduction algorithm were defined from a statistical viewpoint. Furthermore, in [20], a new definition was proposed for attribute reduction for DTRS models by formulating an optimization problem that aims to minimize the cost of decisions. In addition, Li et al. [21] defined a new attribute reduction method based on further investigations of its monotonicity property. Grzymala-Busse et al. [22] analyzed positive and boundary regions as well as comparing possible rules using the Modified Learning from Examples Module Version 2 algorithm. Finally, several studies have considered the application of DTRS models. Thus, Zhou et al. [23] proposed a three-way decision method for filtering spam based on a Bayesian decision procedure. Li et al. [24] proposed an instance-centric hierarchical classification framework using the three-way decision method. Yu et al. [25] used the DTRS model to formulate an efficient automatic clustering method. Liu et al. [26] applied the three-way decision method to a policy-making procedure to reduce the decision risk.

The key feature of the DTRS is a sound mathematical interpretation of thresholds based on the Bayesian decision procedure. Using learned thresholds, three pairwise disjoint regions can be defined in the probabilistic rough set: positive, boundary, and negative regions. As a new semantic interpretation of the three regions, Yao [27,28] introduced the concept of a three-way decision comprising positive, negative, and boundary rules. However, it is not easy to obtain effective decision cost functions for the DTRS model because of a lack of prior knowledge. To overcome this problem, only a few studies have addressed the problem of learning the decision cost functions and thresholds from datasets automatically. In particular, Deng and Yao [29] and Jia et al. [30] proposed different single objective optimization models for automatically learning thresholds from datasets, where the former determined the optimal thresholds by aiming to minimize the uncertainty induced by the three regions, whereas the latter focused on minimizing the decision cost for learning optimal thresholds. However, a major challenge regarding probabilistic rough set models was ignored in these models because they did not formulate a method for decreasing the size of the boundary region by further exploration of the data [31]. In particular, in the model proposed by Jia et al. [30], they used penalties to control the size of the boundary region, but the penalties were provided by users and they could not be selected easily. Herbert and Yao [31] proposed a game-theoretic rough set (GTRS) model to decrease the size of the boundary region and to calculate the required thresholds within a game-theoretic environment. In a related study [32], the configuration of probabilistic thresholds was interpreted as a decision-making problem in a competitive game involving multiple criteria, such as accuracy, generality, confidence, and coverage. In a recent study of GTRS theory, Azam and Yao [33] constructed a mechanism for analyzing the uncertainties of rough set regions with the aim of determining effective threshold values. A competitive game was formulated between the regions to modify the thresholds in order to improve their respective uncertainty levels. By playing the game repeatedly and utilizing its results to update the thresholds, a learning mechanism was proposed to automatically tune the thresholds based on the data. The games based on accuracy and generality consider the three regions, and the uncertainty-based games consider the individual regions. However, in the GTRS model, users have to provide initial possible increases/decreases in the threshold values to set up the game. In addition, the games between decision cost and boundary regions were not investigated further.

In the present study, we propose a multi-objective optimization model for automatic threshold learning. We consider two significant problems regarding DTRS theory: decreasing the size of the boundary region and decreasing the overall decision cost for three types of rules. Using the model proposed by Jia et al. [30], we modify the formulae irrespective of the penalties in our first objective and constraint, before adding a simple but very meaningful objective, $\alpha-\beta$, which intuitively represents the goal of decreasing the size of the boundary region. In contrast to our method, Li and Zhou [15] proposed a three-way view decision model where optimistic, pessimistic, and equable decisions are made according to the cost of misclassification. The thresholds for probabilistic inclusion are calculated based on the minimal risk cost under the respective decision bias. Similarly, Min et al. [19] posited a minimal test cost reduction problem, which constitutes a new but more general problem than the classical reduction one. It is also quite different from our method.

The multi-objective problem is regarded as a game in our method to investigate the trade-off that exists between these two objectives. This game gives rise to a set of Pareto optimal solutions, among which one cannot be said to be better than the other. Excluding the Pareto optimal solutions, no other outcome makes each player (objective) at least as well off and at least one objective better off. In addition, two types of F_measure constraints are used to improve the classification ability of the selected solutions in our model. The first type of F_measure, which is called the F1_measure, is used at the end of the algorithm to preserve the solutions with better classification performance. The second, which is called the F2_measure, is applied during the iteration procedure. Using the self-correcting mechanism in the F2_measure, we modify α/β in non-feasible individuals to satisfy the F2_measure constraint.

We used 12 representative UCI datasets [34] to validate the performance of our model, where the experimental results demonstrated several advantages of our method. First, compared with the other methods mentioned above, a set of Pareto optimal thresholds is learned automatically and the penalties in Jia et al.'s model [30] can be neglected. Second, the newly added objective function to decrease the size of the boundary region is represented simply and it is easy to understand.