ARTICLE IN PRESS
JID: NEUCOM [m5G;June 8, 2017;12:19]
Neurocomputing 000 (2017) 1–12
Contents lists available at ScienceDirect
Neurocomputing
journal homepage: www.elsevier.com/locate/neucom

Multi-objective optimization method for thresholds learning and neighborhood computing in a neighborhood based decision-theoretic rough set model

Ruilin Pan*, Xuemin Wang, Changsheng Yi, Zhanchao Zhang, Yanglong Fan, Wei Bao

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

ARTICLE INFO

Article history:
Received 18 June 2016
Revised 10 January 2017
Accepted 27 May 2017
Available online xxx

Communicated by Wei Chiang Hong

Keywords:
Neighborhood decision theory rough set
Neighborhood computing
Thresholds learning
Multi-objective optimization

ABSTRACT

Recently, a neighborhood based decision-theoretic rough set (NDTRS) model was proposed to deal with the general data which contained numerical values and noisy values simultaneously. However, it still suffered from the issue of granularity selection and the relationship between the thresholds and the neighborhood was also not investigated in depth. In this paper, a multi-objective optimization model for NDTRS to learn the thresholds and select the granularity (compute the neighborhood) comprehensively is proposed. In this model, three significant problems: decreasing the size of the boundary region, decreasing the overall decision cost for the three types of rules, and increasing the size of the neighborhood are taken into consideration. We use 10 UCI datasets to validate the performance of our method. With the Improved Strength Pareto Evolutionary Algorithm (SPEA2), the Pareto optimal solutions are obtained automatically. The experimental results demonstrate the trade-off among the three objectives and show that the thresholds and neighborhoods obtained by our method are more intuitive.

1. Introduction

As a useful tool for uncertain data description and knowledge representation, Pawlak rough set [1] has been used in many research fields such as data mining [2, 3], pattern recognition [4], machine learning [5–7], and big data [8]. In the traditional rough set model, an object x will be classified into category X if its equivalence class (expressed as [x]) entirely belongs to category X, so the traditional rough set is sensitive to noisy data.

The tolerance of classification error is taken into account in probabilistic rough sets (PRSs), where $p(X|[x]) = (|X \cap [x]|/|[x]|) \geq \alpha$ is used to classify the object x into the category X, while α is a threshold between 0 and 1 [9,10]. With a pair of thresholds, α and β, three disjoint regions are defined in PRSs: positive, boundary, and negative regions. As a specific type of PRS, decision-theoretic rough sets (DTRSs) provide a sound mathematical interpretation of thresholds based on the Bayesian decision procedure [11]. The thresholds in different PRS models can be deduced from appropriate cost functions given by DTRSs.

However, the traditional rough sets and PRSs cannot directly deal with continuous numerical data. When processing continuous numerical data, they always need to scatter the records, and then regard each scattered number as a nominal value of the attribute [12]. In recent years, a generalized neighborhood system defining equivalence class based on other relations, such as distance functions [13], fuzzy binary relations [14,15], dissimilarity and similarity measures [16], are used to measure and represent the continuous values.

To deal with the general data which contains numerical values and noisy values simultaneously, Li et al. [17] introduced the neighborhood relation into DTRS and proposed a neighborhood based decision-theoretic rough set (NDTRS) model under the framework of DTRS. They defined two kinds of attribute reducts: a positive region related attribute reduct and a minimum cost attribute reduct and then designed heuristic approaches to compute them. For NDTRS, Chen et al. [18] discuss reduction questions with three condition entropy measures. They proved the monotonic principles of the entropy measures and constructed the heuristic reduction algorithms in NDTRS.

Despite the success of the neighborhood system and NDTRS in attribute reduction, they still suffered from the issue of granularity selection [19–21]. In [17], Li et al. have proven several theorems about the relationship between the parameters: δ and (α, β).

* Corresponding author.
E-mail addresses: alltimefight@gmail.com, rlpan1973@gmail.com (R. Pan), 1436291095@qq.com (X. Wang), yics2005@163.com (C. Yi), 1576252368@qq.com (Z. Zhang), 646149326@qq.com (Y. Fan), asinear@163.com (W. Bao).

http://dx.doi.org/10.1016/j.neucom.2017.05.068

However, the parameters are investigated either fixing δ or fixing (α, β) in their paper.

In this regard, after reviewing some concepts related to granular computing with neighborhood rough sets (NRSs) and thresholds learning in DTRSs, we propose a multi-objective optimization model for NDTRS to learn the thresholds and select the granularity comprehensively. In this model, three significant problems: decreasing the size of the boundary region, decreasing the overall decision cost for the three types of rules [22], and increasing the size of the neighborhood are taken into consideration. We modify Hu et al.'s single objective function for granularity selection into a double objective model irrespective of the penalty in their model [19], and then add another objective of the overall decision cost regarding NDTRS into our model.

In this paper, the multi-objective problem is regarded as a game to investigate the trade-off existing among these three objectives. This game gives rise to a set of Pareto optimal solutions, among which one cannot be said to be better than the others. Excluding the Pareto optimal solutions, no other outcome makes each player (objective) at least as well off and at least one objective better off. We use 10 representative UCI datasets [23] to validate the performance of our model. By solving the optimization problem, a set of Pareto optimal solutions including proper granularities (neighborhoods) and thresholds can be learned automatically. What's more, the feasible results are offered for users to select with the consideration of neighborhood and thresholds comprehensively.

The remainder of this paper is organized as follows. A brief introduction about the related work is given in Section 2. Section 3 introduces the basic concepts of DTRS, NRS, and NDTRS. In Section 4, a multi-objective optimization model for thresholds learning and neighborhood computing regarding the NDTRS model is proposed, and the model is solved by the Improved Strength Pareto Evolutionary Algorithm (SPEA2). Section 5 presents the experimental results and provides some remarks. In Section 6, we give our conclusions.

2. Related work

In this section, we will briefly introduce the related work on DTRS, neighborhood systems, and multi-objective optimization algorithms.

In recent years, DTRS has attracted much attention [24] and two main issues regarding the DTRS model itself: attribute reduction and thresholds learning were investigated in depth. On the attribute reduction of the model, it was thoroughly discussed by many researchers [25–28]. While on the thresholds learning of the model, only a few studies have addressed the problem. Deng and Yao [29] proposed a single objective optimization model to determine the optimal thresholds by aiming to minimize the uncertainty induced by the three regions. Different from their work, Jia et al. [30] proposed a single objective optimization model focused on minimizing the decision cost for learning optimal thresholds automatically. Based on Jia et al.'s work, Pan et al. [22] proposed a multi-objective optimization method to learn thresholds in DTRS automatically. Li and Zhou [31] proposed a three-way view decision model where optimistic, pessimistic, and equable decisions were made according to the cost of misclassification. They calculated the thresholds based on the minimal risk cost under the respective decision bias. Herbert and Yao [32] 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 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.

In the studies of neighborhood systems, using distance functions is a very common and useful method [34–37]. Hu et al. [38] adopted three kinds of distance functions and proposed a NRS model, which was easy to understand and implement. On the neighborhood computing of the model, Lin et al. [39] developed a neighborhood based multi-granulation rough set in the framework of multi-granulation rough sets. Zhu et al. [21] explored ensemble learning techniques for adaptively evaluating and combined the models derived from multiple granularities. Chen et al. [40] set up a connection between neighborhood-covering rough sets and evidence theory to establish a basic framework of numerical characterizations of attribute reduction. Lin et al. [41] presented a new feature selection method that selected distinguishing features by fusing neighborhood multi-granulation. Chen et al. [42] introduced the neighborhood entropy to firstly evaluate the uncertainty of a neighborhood information system.

Regarding with the development of learning algorithms and multi-objective optimization and clustering, a lot of studies are proposed. Pettersson et al. [10] used a genetic algorithm based multi-objective optimization technique to complete the training process of a feed forward neural network. Giri et al. [43] developed a new Bi-objective Genetic Programming (BioGP) technique which creatively attempted to minimize training error through a single objective procedure. They also compared the meta-models constructed for simulated moving bed (SMB) reactors with those obtained from an evolutionary neural network (EvoNN) developed earlier and with a polynomial regression model [44]. Bevilacqua et al. [45] incorporate in the original BioGP an adaptive mechanism that automatically tunes the mutation rate. To deal with the multi objective optimization, Jiao et al. proposed a modified objective function method with feasible-guiding strategy on the basis of NSGA-II to handle co-evolutionary multi-objective optimization (CMOPs) [46] and a direction vectors based co-evolutionary multi-objective optimization algorithm regarding with the decomposition idea [47]. Shang et al. proposed a novel immune clonal algorithm [48] and a novel feature selection algorithm for clustering, named self-representation based dual-graph regularized feature selection clustering (DFSC) [49]. They simultaneously maximized the between-class scatter matrix and minimized the within-class scatter matrix to enhance the discriminating power [50].

3. Preliminaries

In this section, we recall the basic notions related to DTRS [51], NRS [38], and NDTRS [17].

3.1. Decision-theoretic rough set model

In this subsection, the basic notions of DTRS model and three-way decision theory are presented.

Suppose a knowledge representation scheme defined by a decision table S is expressed as the tuple [51]:

$$S = (U, A_t = C \cup \{D\}, \{V_a | a \in A_t\}, \{I_a | a \in A_t\}), \quad (1)$$

where U, A_t, and C are finite nonempty sets of objects, attributes, and condition attributes describing the objects, respectively, and D is a decision attribute that indicates the categories of the objects. Furthermore, V_a is a non-empty set of values for attribute $a \in A_t$, and I_a: $U \rightarrow V_a$ is an information function that maps an object in U to exactly one value in V_a.

Given a subset of attributes $B \subseteq A_t$, an indiscernibility relation $ind(B)$ is defined as follows [51]:

$$x \; ind(B) \; y \leftrightarrow \forall a \in B \; [I_a(x) = I_a(y)], \quad (2)$$

where two objects x and y are indiscernible with respect to B if and only if they have exactly the same value for each attribute in