ARTICLE IN PRESS
JID: KNOSYS [m5G;August 11, 2016;13:4]
Knowledge-Based Systems 000 (2016) 1–7
Contents lists available at ScienceDirect
Knowledge-Based Systems
journal homepage: www.elsevier.com/locate/knosys

Feature subset selection based on fuzzy neighborhood rough sets

Changzhong Wang[a,*], Mingwen Shao[b], Qiang He[c], Yuhua Qian[d], Yali Qi[a]

[a] Department of Mathematics, Bohai university, Jinzhou, 121000, PR China
[b] College of Computer and Communication Engineering, Chinese University of Petroleum, Qingdao, Shandong, 266580, P.R China
[c] College of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, P.R China
[d] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, P.R China

ARTICLE INFO

Article history:
Received 21 March 2016
Revised 4 August 2016
Accepted 7 August 2016
Available online xxx

Keywords:
Fuzzy neighborhood
Fuzzy decision
Feature selection
Rough set model

ABSTRACT

Rough set theory has been extensively discussed in machine learning and pattern recognition. It provides us another important theoretical tool for feature selection. In this paper, we construct a novel rough set model for feature subset selection. First, we define the fuzzy decision of a sample by using the concept of fuzzy neighborhood. A parameterized fuzzy relation is introduced to characterize fuzzy information granules for analysis of real-valued data. Then, we use the relationship between fuzzy neighborhood and fuzzy decision to construct a new rough set model: fuzzy neighborhood rough set model. Based on this model, the definitions of upper and lower approximation, boundary region and positive region are given, and the effects of parameters on these concepts are discussed. To make the new model tolerate noises in data, we introduce a variable-precision fuzzy neighborhood rough set model. This model can decrease the possibility that a sample is classified into a wrong category. Finally, we define the dependency between fuzzy decision and condition attributes and employ the dependency to evaluate the significance of a candidate feature, using which a greedy feature subset selection algorithm is designed. The proposed algorithm is compared with some classical algorithms. The experiments show that the proposed algorithm gets higher classification performance and the numbers of selected features are relatively small.

1. Introduction

Nowadays, databases expand quickly, more and more attributes are obtained in production practice. Some of attributes may be redundant or irrelevant to a classification task, they need to be removed before any further data processing can be carried out. Feature selection or attribute reduction is a technique for reducing features. Its aim is to find an optimal feature subset to predict sample categories. Feature subset selection can also facilitate data visualization and data understanding [11]. In recent years, much attention has been paid to feature selection in machine learning, data analysis and pattern recognition.

There is a key issue in feature selection process: feature evaluation. How to construct an effective evaluation function is one of the most important steps. It directly affects the performance of a classifier. A lots of feature evaluation measures, such as information entropy [8,12], dependency [3,9–11], correlation [7], and consistency [4], have been proposed for feature selection until now. In general, different evaluation measures may lead to different optimal feature subsets. However, every measure is aimed to determine the discriminating ability of a subset of features.

The classical rough set theory [17] has been proven to be an effective tool for feature selection. It employs a dependency function to evaluate the classification quality of a subset of attributes. However, this model is just applicable to nominal data. In practical problems, it is most often the case that the values of attributes may be both crisp and real-valued. The real-valued features need to be discretized before the dependency is calculated. The inherent error that exists in discretization process is of major concern. This is where the traditional rough set theory encounters a problem.

Some generalizations of the model were proposed to deal with this problem [5,6,13–16,19–25]. Neighborhood rough set and fuzzy rough set are considered two important models. Lin generalized the classical rough set with neighborhood operators and introduced a neighborhood rough set model [14]. Dubois and Prade defined fuzzy rough approximation operators by combining rough sets and fuzzy sets and proposed a fuzzy rough set model [5]. Recently, some feature selection algorithms based on the generalized models have been proposed [1–3,8–11,18,20,26].

As we know, the core idea of rough set theory is based on granulation and approximation. In a neighborhood rough set, neighborhood similarity classes are used to approximately characterize decision equivalence classes. The limitation of this model is that it

* Corresponding author.
E-mail address: changzhongwang@126.com (C. Wang).

http://dx.doi.org/10.1016/j.knosys.2016.08.009

cannot describe the fuzziness of samples in fuzzy background. In classical fuzzy rough set model, fuzzy information granules are the elemental granules. The membership degrees of a sample to different decision classes are computed by min-max operations. That is to say, the decision of a sample is based on a nearest sample. However, there may be some risks in computation of fuzzy lower approximations when a data set has noise. Data noise can destroy the accuracy of calculation of membership degrees and lead to an increase in classification error rate. To better describe sample decisions by using fuzzy information granules, a new rough set model, named fuzzy neighborhood rough set model, is introduced in this paper.

We first define the fuzzy decision of a sample and employ a parameterized fuzzy relation to characterize its fuzzy information granule. We then use the inclusion relation of them to decide whether the sample is classified into one of decision classes. Because this way of decision-making fully utilizes the classification information of multiple samples, it overcomes the disadvantage of fuzzy rough set model by using a nearest neighbor to determine the membership degree of a sample to different decision classes. The proposed model is a nature generalization of neighborhood rough sets. This is the main difference from the classical fuzzy rough set theory. As the proposed model is too strict to tolerate noise in data, a variable precision fuzzy neighborhood rough set model is introduced. This model is more effective to process the fuzzy or uncertain knowledge because it can decrease the possibility that a sample is classified into a wrong class. Finally, we define the dependency between features and decision and design a feature selection algorithm. Numerical experiments show that the proposed algorithm yields better performance.

The paper is organized as follows. In Section 2, we review some relevant literature about neighborhood rough sets and fuzzy rough sets. In Section 3, we develop a new model: fuzzy neighborhood rough set model. In Section 4, we design a heuristic algorithm of attribute reduction. In Section 5, we verify the feasibility and stability of the proposed algorithm. Section 6 concludes the paper.

2. Literature reviews

Neighborhood is one of important concepts in classification learning and reasoning with uncertainty. A neighborhood relation can be used to generate a family of neighborhood granules characterized with numerical features [15]. In 1997, Lin pointed out that neighborhoods are more general information granules than equivalence classes and introduced neighborhood relations into rough set methodology [14]. Based on this observation, a neighborhood rough set model was constructed. Then, Wu and Zhang studied some properties of neighborhood approximation spaces [22]. Yao discussed the relationship between neighborhood operators and rough approximation operators and presented the axiomatic properties of this model [23]. In 2008, Hu employed the neighborhood rough set model to deal with feature subset selection in real-valued sample space [9]. In fact, the neighborhood model is a natural generalization of classical rough sets. The model can be used to deal with mixed numerical and categorical data within a uniform framework and overcomes the drawback of discretization of data in classical rough sets. However, it cannot describe the fuzziness of samples in fuzzy background.

Fuzzy rough sets, as proposed by Dubois and Prade [5], can also deal with numerical or continuous data sets directly. Numerical attribute values are no longer needed for discretization. In this model, a fuzzy similarity relation is defined to measure the similarity between samples. The fuzzy upper and lower approximations of a decision are then defined by using the fuzzy similarity relation. The fuzzy positive region is defined as the union of the fuzzy lower approximations of decision equivalence classes. As the fuzziness is introduced into the rough set theory, more information of continuous attribute values is easily kept. So, feature selection with fuzzy rough sets becomes another important tool in handling dataset with real-valued attributes. In recent years, a series of feature selection algorithms based on fuzzy rough sets have been proposed. Jensen introduced the dependency function in classical rough sets into fuzzy rough sets and proposed an a greedy algorithm for reducing redundant attributes [11]. Bhatt and Gopal presented the concept of compact computational domain for Jensen's algorithm to improve computational efficiency [1]. Chen used fuzzy rough sets to define fuzzy discernibility matrix by which all attribute reducts are computed [2]. For data-based attribute selection, Cornelis generalized the classical rough set model using fuzzy tolerance relations within the context of fuzzy rough set theory [3]. Hu et al. employed kernel functions to define fuzzy similarity relations and constructed a greedy algorithm for dimensionality reduction [10]. Meanwhile, the classical fuzzy rough set model was improved to analyze noisy data. Mieszkowicz Rolka introduced the model of variable precision fuzzy rough sets to deal with noisy data [19], where the fuzzy memberships of a sample to the lower and upper approximations were computed with fuzzy inclusion. Zhao et al. defined the concept of fuzzy variable precision rough sets to handle noise of misclassification and perturbation [26]. To solve the problem of data fitting in classical fuzzy rough sets, Wang proposed a fitting fuzzy rough set model to conduct feature selection [20]. However, in all kinds of fuzzy rough set models, the fuzzy upper and lower approximations of a decision is computed by using a nearest sample, there may be some risks when a data set has noise. This is the main drawback of fuzzy rough set models.

3. Fuzzy neighborhood rough set model

Let $\langle U, A, D\rangle$ be a decision table, where $U = \{x_1, x_2, \ldots, x_n\}$ is called a sample space, A is a set of attributes or features characterizing samples and D is a decision attribute. Assume that the samples are partitioned into r mutually exclusive decision classes by D, that is, $U/D = \{D_1, D_2, \ldots, D_r\}$. In this section, the fuzzy decision of a sample is defined and parameterized fuzzy information granules associated with samples are introduced. The task is to approximate the fuzzy decision classes with parameterized fuzzy information granules.

Let $B \subseteq A$ be a subset of attributes on U, and then B can induce a fuzzy binary relation R_B on U. R_B is called a fuzzy similarity relation if it satisfies

(1) Reflectivity: $R_B(x,x) = 1$, $\forall x \in U$; (2) Symmetry: $R_B(x,y) = R_B(y,x)$, $\forall x, y \in U$.

Let $a \in B$ and R_a be a fuzzy similarity relation induced by a, we denote $R_B = \bigcap_{a\in B} R_a$. For any $x \in U$, the fuzzy neighborhood of x is defined as $[x]_B(y) = R_B(x,y)$, $y \in U$.

Definition 1. Given a decision table $\langle U, A, D\rangle$, $U/D = \{D_1, D_2, \ldots D_r\}$. R_A is the fuzzy similarity relation on U induced by A, $\forall x \in U$, the fuzzy decision of x is defined as follows.

$$\tilde{D}_i(x) = \frac{\left|[x]_A \cap D_i\right|}{\left|[x]_A\right|}, \quad i = 1, 2, \ldots, r,$$

where $\tilde{D}_i$ is a fuzzy set and $\tilde{D}_i(x)$ indicates the membership degree of x to D_i. We call $\{\tilde{D}_1, \tilde{D}_2, \ldots \tilde{D}_r\}$ the fuzzy decisions of samples induced by D.

Please cite this article as: C. Wang et al., Feature subset selection based on fuzzy neighborhood rough sets, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.08.009