ARTICLE IN PRESS
JID: KNOSYS [m5G;December 29, 2016;10:57]
Knowledge-Based Systems 000 (2016) 1–13
Contents lists available at ScienceDirect
Knowledge-Based Systems
journal homepage: www.elsevier.com/locate/knosys

Different classes' ratio fuzzy rough set based robust feature selection

Yuwen Li[a], Shunxiang Wu[a], Yaojin Lin[b,*], Jinghua Liu[a]

[a] Department of Automation, Xiamen University, Xiamen, 361005, PR China
[b] School of Computer Science, Minnan Normal University, Zhangzhou, 363000, PR China

ARTICLE INFO

Article history:
Received 1 June 2016
Revised 9 November 2016
Accepted 28 December 2016
Available online xxx

Keywords:
Feature selection
Fuzzy rough set
Robust
Noise

ABSTRACT

In order to solve the problem that the classical fuzzy rough set (FRS) model used for feature selection is sensitive to noisy information, we propose an effective robust fuzzy rough set model, called different classes' ratio fuzzy rough set (DC_ratio FRS) model. The proposed model can reduce the influence of noisy samples on the computation of the lower and upper approximations, and recognize the noisy samples directly. Moreover, the DC_ratio FRS model is robust against noise because it ignores a noisy sample which can be identified by computing the different classes' ratio of this sample. Different classes' ratio denotes the proportion of samples belonging to different classes in the neighbors of a given sample. Then, the properties of the DC_ratio FRS model are also discussed, and sample pair selection (SPS) based on the DC_ratio FRS model is used to feature selection. Finally, extensive experiments are given to illustrate the robustness and effectiveness of the proposed model.

1. Introduction

In image annotation [19,20,40,61], medical diagnosis [28,48], and bioinformatics [42,58] domains, data usually involves a great number of features. Generally, some features are redundant and/or irrelevant for a specific learning task. In which, the redundant features may bring high computational complexity and the curse of dimensionality. In addition, the irrelevant features might confuse learning algorithms and deteriorate learning performance. Therefore, it is meaningful to select relevant and indispensable features for various learning.

In classification learning, feature selection is a key step of data preprocessing, usually also called attribute reduction. Feature selection is used to find an optimal feature subset to perform classification learning under the premise of keeping classification accuracy. There are two steps in feature selection: the first step is to construct feature evaluation function, and the other step is to employ some searching strategies to obtain optimal features. In which, feature evaluation function is used to measure the quality of the candidate features, such as dependency [49], neighborhood dependency [21] and fuzzy dependency [14,15] in the rough set theory; mutual information [23,35,38] in information theory; and sample margin [56] in statistical learning theory, respectively. In addition, searching strategy includes sequential forward selection [29], sequential backward elimination [39], and floating search [46].

Rough set was proposed by Pawlak originally [45], has been successfully applied in feature selection [10,33,34,36,37,47] and rule learning [9,11,12,31,32]. However, the classical rough set model could not work effectively on the real-valued data sets. Therefore, many extension models of rough set, such as neighborhood rough set (NRS) [21] and fuzzy rough set (FRS) [14,41,52–54,59], have been introduced to handle with this problem. Especially, fuzzy rough set can effectively deal with both fuzziness and vagueness of the data sets with continuous features [15,25,27,50,57]. However, the classical FRS restricted by the nearest neighbor sample is sensitive to noise, due to the computation of the lower and upper approximations based on the classical FRS is associated with the nearest neighbor sample for a given target sample.

To improve the robustness of the classical FRS, some robust fuzzy rough set models have been proposed. Existing robust FRS models can be roughly classified into two groups: in the first group, samples locate around classification boundary that are considered as noisy samples, such as β-precision fuzzy rough set (β-PFRS) [17], soft fuzzy rough set (SFRS) [22], soft minimum enclosing ball (SMEB)-based fuzzy rough sets (SMEB-FRS) [3], data-distribution-aware fuzzy rough set (PFRS) [2], and k-trimmed fuzzy rough set (k-trimmed FRS) [24]. The important characteristic of these models in the first group is how to select the nearest neighbor sample. The other group uses robust approximation operators, such as k-means fuzzy rough set (k-means FRS) [24], k-median fuzzy rough set (k-median FRS) [24], vaguely quantified rough set

* Corresponding author.
E-mail addresses: liyuwen2012@163.com (Y. Li), wsx1009@163.com (S. Wu), yjlin@mnnu.edu.cn, zzlinyaojin@163.com (Y. Lin), zzliujinghua@163.com (J. Liu).

http://dx.doi.org/10.1016/j.knosys.2016.12.024
0950-7051/

Please cite this article as: Y. Li et al., Different classes' ratio fuzzy rough set based robust feature selection, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.12.024

(VQRS) [13], fuzzy variable precision rough set (FVPRS) [60], and the ordered weighted average fuzzy rough set (OWA-FRS) [51].

Different from the existing robust fuzzy rough set models, in this paper, the proposed model considers the influence of the nearest sample with respect to a given target sample, and the information of the neighbors around the nearest sample, simultaneously. Now, an example is used to illustrate the effectiveness of the DC_ratio FRS model. Assume a decision table with two classes, where x belongs to the class 1 (D_1) and y_1, y_2, y_3 belong to the class 2 (D_2), respectively. For x, the lower approximation of x with respect to D_1 not only considers the dissimilarity between x and y_1, y_2, y_3, in which y_2 is the nearest different classes sample of x, but also considers the influence of the neighbors of y_2. Based on the neighbors of y_2, we can compute the different classes' ratio of y_2, i.e., y_2 is recognized to be noise which depends on the different classes' ratio of y_2. In which, different classes' ratio of y_2 is computed by the ratio of samples belonging to D_1 vs. all neighbors of y_2. If the different classes' ratio of y_2 is higher than a given threshold, y_2 is recognized to be a noise. Reversely, if the different classes' ratio of y_2 is lower than a given threshold, y_2 is recognized to be a normal sample. Based on the above analysis, we propose the DC_ratio FRS model, which ignores some samples that are identified as noises via the computation of different classes' ratio. Moreover, the proposed model not only can reduce the influence of noise samples, but also can identify the noise samples directly. Finally, extensive experiments are conducted to show the robustness and effectiveness of the proposed model. The main contributions of the proposed model can be summarized as follows:

- The proposed model not only can find out the nearest sample with respect to a target sample, but also can measure the influence of the neighbors of the nearest sample.
- A simple and intuitive metric is proposed to recognize the noisy samples.
- Experimental results show that our proposed method outperforms some other models in classification performance and statistical analysis, respectively.

This paper is organized as follows. In Section 2, related work about robust FRS models is given. Section 3 discusses the classical FRS model, and the feature selection method based on fuzzy dependency, discernibility matrix and sample pair selection. Section 4 introduces the DC_ratio FRS model, and analyzes the related properties. Section 5 designs the feature selection algorithm with the DC_ratio FRS model based on sample pair selection. Section 6 discusses the experimental results. Finally, the conclusions are given in Section 7.

2. Related work about robust FRS models

The existing robust fuzzy rough set models include β-PFRS [17], SFRS [22], SMEB-FRS [3], PFRS [2], k-trimmed FRS [24], k-means FRS [24], k-median FRS [24], VQRS [13], FVPRS [60], and OWA-FRS [51]. The target of robust FRS models mentioned above is to reduce the impact of the noisy samples on the lower approximation. These robust FRS models can be divided into two groups. In the first group, they enlarge the lower approximation by ignoring some nearest samples, such as β-PFRS [17], SFRS [22], SMEB-FRS [3], PFRS [2], and k-trimmed FRS [24]. The other group replaces the statistics of minimum with robust approximation operators to reduce the impact of the nearest noisy samples, such as k-means FRS [24], k-median FRS [24], VQRS [13], FVPRS [60], and OWA-FRS [51]. The detailed robustness analysis of the above models is as follows.

β-PFRS [17] with the concept of β-precision aggregation achieves robustness to noise by overlooking the some nearest neighbors of a sample x from different classes in computing the lower approximation. In which, β is the direct parameter used for determining the number of ignored samples. The smaller the value of β, the more samples from different classes are neglected. Thus, it can be calculated the number of samples neglected in computing lower approximation of β-PFRS by setting the value of β. Hu et al. [22] proposed SFRS as a robust fuzzy rough set model by the idea of soft-margin SVM. SFRS improves the computation of the lower and upper approximations, where the membership is not calculated with the nearest sample from different classes, but the $k-$th sample, where k is determined by tradeoff between the number of misclassified samples and the augmentation of membership. By this way, SFRS model is robust to the noisy samples. With the SMEB-FRS [3], the lower and upper approximations of a sample are computed with the nearest sample in the soft minimum enclosing ball of samples from other classes or the same class, respectively. The samples outside the soft minimum enclosing ball are considered to be noise and ignored. This reduces the uncertainty brought by noisy information. So the SMEB-FRS model improves the robustness of the classical FRS. An et al. [2] proposed PFRS that considers distribution information and incorporates it in computing lower and upper fuzzy approximations. When computing the lower approximation membership of a sample x with respect to a class L, a trade-off was established between the similarity of x and y and the probability density values y, where y belongs to different classes with x. Similarly, when computing the upper approximation membership of a sample x with respect to a class L, a trade-off was established between the similarity of x and y and the probability density values y, where y belongs to the same class with x. In this way, the boundary sample points are determined as noise and ignored when computing lower and upper approximations.

K-trimmed FRS, k-means FRS, and k-median FRS models were proposed in [24]. Three robust statistics are introduced to substitute the operators of minimum and maximum in the classical FRS model to achieve robustness. The three models do not compute the lower and upper approximations with respect to the nearest samples as they might be outliers. These models use k-trimmed or the mean or the median of k nearest samples to compute the membership of fuzzy approximations. This way, the variation of approximations caused by outliers is expected to be reduced. Therefore, these models may be robust.

VQRS [13] was used as a robust method of feature selection by introducing vague quantifiers like “some” or “most” into the definition of lower and upper approximation. In the VQRS model, a sample x belongs to the lower approximation of a class L to the extent that most elements related to x are in L. Similarly, x belongs to the upper approximation to the extent that some elements related to x are in L. The use of vague quantifiers “most” and “some”, as opposed to the traditionally used crisp quantifiers “all” and “at least one” makes the model more robust. FVPRS [60] was constructed by combining the fuzzy rough sets and variable precision rough sets. A threshold is introduced into the definition of the fuzzy lower and upper approximation proposed by fuzzy rough set. By this threshold, some memberships of the fuzzy set are ignored which make the FVPRS non-sensitive to the noisy data. As a result the generalization ability of the FVPRS model is robust. OWA-FRS [51] was presented which replace the crisp min and max operators by softer OWA operators. It was noted that the traditional fuzzy rough approximations are highly susceptible to noise, as they use the crisp min and max operators, such that the outlier can drastically influence the approximation values, and therefore OWA-FRS is robust against noisy data.

Through the above analysis, different from the existing noise-tolerant FRS models, the proposed model considers the influence of the nearest sample with respect to the target sample, and other information of the neighbors around the nearest sample. Specially, although the DC_ratio FRS model is similar with SFRS, the pro-