Knowledge-Based Systems 000 (2018) 1–11

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Attribute reduction for multi-label learning with fuzzy rough set

Yaojin Lin[a,b,*], Yuwen Li[c], Chenxi Wang[a], Jinkun Chen[d]

[a] School of Computer Science, Minnan Normal University, Zhangzhou 363000, PR China
[b] Key Laboratory of Data Science and Intelligence Application, Fujian Province Unversity, PR China
[c] Department of Automation, Xiamen University, Xiamen, 361000 PR China
[d] School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, PR China

ARTICLE INFO

Article history:
Received 1 September 2017
Revised 25 February 2018
Accepted 2 April 2018
Available online xxx

Keywords:
Attribute reduction
Fuzzy rough set
Multi-label learning
Local sampling

ABSTRACT

In multi-label learning, each sample is related to multiple labels simultaneously, and attribute space of samples is with high-dimensionality. Therefore, the key issue for attribute reduction in multi-label data is to measure the quality of each attribute with respect to a set of labels. Stimulated by fuzzy rough set theory, which allows different fuzzy relations to measure the similarity between samples under different labels. In this paper, we propose a novel fuzzy rough set model for attribute reduction in multi-label learning. Different from single-label attribute reduction, a bottleneck of fuzzy rough set for multi-label attribute reduction is to find the true different classes' samples for the target sample, which deeply affects the robustness of fuzzy upper and lower approximations. We first define the score vector of each sample to evaluate the probability of being different class's sample with respect to the target sample. Then, local sampling is leveraged to construct a robust distance between samples. It can implement the robustness against noisy information when calculating the fuzzy lower and upper approximations under the whole label space. Moreover, multi-label fuzzy rough set model is proposed, and some related properties are discussed. Finally, the significance measure of a candidate attribute is defined, and a greedy forward attribute selection algorithm is designed. Extensive experiments are carried out to verify the effectiveness of the proposed algorithm by comparing it with some state-of-the-art approaches on eight publicly available data sets.

1. Introduction

Traditional supervised learning handles tasks that every instance has only one label, namely single-label learning. In plentiful practical applications, however, every instance is possibly related to multiple labels simultaneously, such as text classification [2], automatic scene categorization [4], music emotions annotation [46], and gene function prediction [59]. For example, a news may belong to economy, sport, and entertainment [2]; an image may be related to a set of semantic classes, such as countries, trees, and sunrises [4]; and a music associates with relaxing and happy [46]. As we know, thousands or even tens of thousands of attributes are stored in multi-label data [52], and some attributes may be irrelevant and/or redundant, which usually degrade the performance of multi-label classification learning algorithms. Therefore, multi-label learning is also affected by the curse of dimensionality [32,54,61,63].

Attribute reduction, also called feature selection, is an important pre-processing step for multi-label classification learning which can mitigate the curse of dimensionality [15,30,61,63,69]. Therefore, some attribute reduction algorithms [25] are designed to improve the classification performance via selecting an optimal attribute subset from original attribute space. Moreover, attribute reduction can keep most useful information of data set, maintain the physical meaning of attribute, and give better readability and interpretability of classification model. Generally, attribute reduction techniques can be classified into supervised, semi-supervised, and unsupervised methods. In the supervised technique [44], the relevance of an attribute can be evaluated by its correlation with the label, such as Fisher score [8], ReliefF [37], and mutual information [26,29]. Unsupervised attribute reduction generates the pseudo labels before executing feature selection [45,51,67,68]. Semi-supervised attribute reduction [5] is with the so-called small-labeled-sample problem, in which the amount of data that is unlabeled can be much larger than the amount of labeled data. Furthermore, multi-label attribute reduction also could be mainly divided into three groups: filter, wrapper, and embedded [40,42,43,50]. Filter utilizes some measure criteria to sort

* Corresponding author.
E-mail address: yjlin@mnnu.edu.cn (Y. Lin).

https://doi.org/10.1016/j.knosys.2018.04.004
0950-7051/

conditional attributes or select some informative attributes[38,58]. Wrapper uses the predictive accuracy of a predetermined learning algorithm to measure the quality of selected attributes [13]. Embedded combines classifier to design attribute reduction in a single optimization process [14]. In which, the filter method is widely used in attribute reduction, due to its simplicity and interpretability. In addition, a number of filter models are proposed based on different evaluation criteria of attributes, such as dependency [7], neighborhood dependency [17], and fuzzy dependency [9,10] in different rough set models; mutual information [26,29] in information theory; and sample margin [56] in statistical learning theory, respectively.

As we know, fuzzy systems have been successfully used in many domains [36,70], including pattern classification, optimization, and machine learning. Due to the favorable properties of fuzzy systems, the scope of intelligent technical applications based on soft-computing methods is continuously expanding. As a result, the use of fuzzy system and fuzzy set, as intelligent system components, is also growing. These methods can be roughly classified into three main categories [27,28,48,58]: (1) Some methods tackle the definition of the number of fuzzy sets via describing each attribute (granularity); (2) Some methods define the distribution of fuzzy sets which used in partitions; (3) Some methods define the number of fuzzy sets and their distribution together.

Rough set theory, originally proposed by Pawlak [35], also has been widely applied in attribute reduction and rule learning [6,7,23,27,28,58]. However, the classical rough set model could not work effectively on the hybrid attributes. Therefore, many extensive models of rough set, such as neighborhood rough set (NRS) [17], and fuzzy rough set (FRS) [20,47,53], have been introduced to handle with this problem. Especially, fuzzy rough set theory provides an important theoretical tool for attribute reduction [21,31]. In the framework of fuzzy rough set, the fuzzy upper and lower approximation operators are defined based on a fuzzy similarity relation. In which, the upper approximation of a given sample is calculated by the distance between the given sample and its nearest hit sample. Reversely, the lower approximation of a given sample can be calculated by the distance between the given sample and its nearest miss sample. Finally, as an evaluation metric, the fuzzy dependency is used to select attribute subset.

For fuzzy rough set, the core work of existing attribute reduction methods concentrates on two aspects [24,49,55]: (1) Optimize fuzzy similarity relation; (2) Construct robust distance. For the first aspect, Wang et al. [48] introduced a parameterized fuzzy relation to granulate fuzzy relation of samples and used fuzzy dependency membership as a key evaluation criterion to select optimal attributes. Zeng et al. [57] developed a novel fuzzy similarity relation through incorporating a new hybrid distance in hybrid information system. Hu et al. [16] employed various kernel functions to construct kernelized fuzzy relations between samples to extend fuzzy relations of the classical fuzzy rough sets model. Zhang et al. [64] proposed diverse fuzzy relations based on different types of attributes to measure the similarity between samples. For the second aspect, several robust fuzzy rough set models have been put forward to enhance the robustness of the classical FRS, due to the classical fuzzy rough set is related to the distance between a given sample and its nearest miss sample. Consequently, the FRS model is sensitive to noise samples, which astricts their practical applications. Existing robust FRS models can be roughly classified into two groups: in the first group, samples are considered as noisy samples which locate around classification boundary, such as β-precision fuzzy rough set (β-PFRS) [11], probabilistic variable precision fuzzy rough set (P-VP-FRS) [1], soft fuzzy rough set (SFRS) [18], k-trimmed fuzzy rough set (k-trimmed FRS) [19], and data-distribution-aware fuzzy rough set (PFRS) [3]. In the other group, robust approximation operator is used to replace the minimum and maximum operators, such as k-median fuzzy rough set (k-median FRS) [19], k-means fuzzy rough set (k-means FRS) [19], and fuzzy variable precision rough set (FVPRS) [66].

In multi-label learning, there are three challenges which affect the performance of learning algorithms: (1) The integrality of label space, (2) the data are of high dimensionality, and (3) the relationship between samples is fuzzy and uncertainty. In which, the fuzzy and uncertainty denotes that two samples belong to the same group under one label but belong to different groups under other label. Therefore, in this paper, we propose a novel fuzzy rough set model to address these challenges in one shot. Recently, some multi-label attribute reduction methods based on fuzzy rough set have been presented and discussed, but their common characteristic is that these algorithms firstly transform multi-label data set into single-label data set [58,60], and then execute attribute reduction on single-label data set. In fact, these existing fuzzy rough set approaches ignore the integrality of label space, i.e., each sample belongs to several labels simultaneously.

In this paper, the proposed fuzzy rough set model can use different fuzzy relations to estimate the similarity between samples under different labels, and evaluate attributes on multi-label data directly. In order to exactly calculate the lower and upper approximation operators of multi-label fuzzy rough set, the score vectors of samples under the whole label space is firstly used to find the true different classes' samples with respect to a given target sample. Then, based on the advantages of ensemble learning, local sampling is employed to obtain more robust distance between the target sample and its nearest miss sample, which can solve the low separability of fuzzy similarity relation on high dimensional multi-label data. Finally, a multi-label fuzzy dependency function is defined, and a forward greedy attribute reduction algorithm is proposed to select optimal multi-label attribute subset. Extensive experiments are conducted to show the effectiveness of the proposed algorithm. The major contributions of the proposed model can be summarized as follows:

- We use fuzzy rough set to address multi-label attribute reduction, due to FRS can construct different fuzzy relations to estimate the similarity between samples with respect to different labels.
- Different from existing multi-label attribute reduction methods based on fuzzy rough set, the proposed multi-label fuzzy rough set model can keep the integrality of label space, i.e., it is not need to transform multi-label data into single-label data.
- For the proposed multi-label fuzzy rough set model, the score vector of sample is defined to find the true different classes' samples with respect to the target sample, and local sampling is used to construct more robust lower and upper approximations.

The rest of this paper is structured as follows. Section 2 introduces multi-label learning and fuzzy rough set. Then, we present the multi-label attribute reduction algorithm based on fuzzy rough set and report on experimental evaluations in Sections 3 and 4, respectively. Finally, some conclusions are provided in Section 5.

2. Preliminaries

2.1. Multi-label learning

$X \subset \mathbf{R}^d$ expresses a set of n samples, where a sample $x_{i(i=1,2,\cdots,n)} \in X$ denotes a d-dimensional feature vector (x_{i1}, x_{i2}, $\cdots$, x_{id}). Let $L = \{l_1, l_2, \cdots, l_m\}$ be a set of m labels in multi-label learning. Assume a sample x_i is labeled with a subset of L. If x_i has label $l_{j(j=1,2,\cdots,m)}$, then $y^j = 1$; Otherwise $y^j = 0$. Therefore, the label value of x can be depicted as a m-dimensional vector $\mathbf{y} = [y^1, y^2, \cdots, y^m]$.

Please cite this article as: Y. Lin et al., Attribute reduction for multi-label learning with fuzzy rough set, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.04.004