Expert Systems with Applications 41 (2014) 2989–3004

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Multi-label classification by exploiting label correlations

Ying Yu [a,b,c,e,*], Witold Pedrycz [b,d], Duoqian Miao [a,c]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2G7, Canada
[c] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, PR China
[d] System Research Institute, Polish Academy of Sciences, Warsaw, Poland
[e] School of Software, Jiangxi Agricultural University, Nanchang 330013, PR China

A R T I C L E   I N F O

Keywords:
Multi-label classification
Rough sets
Uncertainty
Correlation

A B S T R A C T

Nowadays, multi-label classification methods are of increasing interest in the areas such as text categorization, image annotation and protein function classification. Due to the correlation among the labels, traditional single-label classification methods are not directly applicable to the multi-label classification problem. This paper presents two novel multi-label classification algorithms based on the variable precision neighborhood rough sets, called multi-label classification using rough sets (MLRS) and MLRS using local correlation (MLRS-LC). The proposed algorithms consider two important factors that affect the accuracy of prediction, namely the correlation among the labels and the uncertainty that exists within the mapping between the feature space and the label space. MLRS provides a global view at the label correlation while MLRS-LC deals with the label correlation at the local level. Given a new instance, MLRS determines its location and then computes the probabilities of labels according to its location. The MLRS-LC first finds out its topic and then the probabilities of new instance belonging to each class is calculated in related topic. A series of experiments reported for seven multi-label datasets show that MLRS and MLRS-LC achieve promising performance when compared with some well-known multi-label learning algorithms.

1. Introduction

Nowadays, multi-label classification problem (Tsoumakas, Katakis, & Vlahavas, 2010) has received an increased attention finding applicability in various applications. For example, in text categorization, a document may belong to multiple classes simultaneously (Jiang, Tsai, & Lee, 2012). In the video indexing domain, each audio clip can have several different labels (Snoek et al., 2006). In functional genomics, a gene may have multiple functions (Vens, Struyf, Schietgat, et al., 2008). In automatic image annotation, a scene may be associated with several concepts as well (Yu, Pedrycz, & Miao, 2013). In all the cases identified above, each instance is associated with multiple labels and the classes encountered in the problem are not mutually exclusive but may overlap. This situation is different from the traditional single-label classification (i.e, multi-class) where an instance is only associated with a single label and by definition, the classes are mutually exclusive (see Fig. 1).

In what follows, we provide a formal definition of the multi-label classification problem.

Definition 1. Let $X \subset R^d$ denote a d-dimensions input domain of instances and let $Y = \{l_1, l_2, ..., l_m\}$ be an output domain of possible labels. Given a training set $T = \{(x_i, y_i)|1 \leqslant i \leqslant n, x_i \in X, y_i \subset Y\}$, the goal of learning system is to form a multi-label classifier $f{:}X \to 2^Y$ which optimizes some specific evaluation metric. For a testing instance $x \in X$, its associated label set $y \subset Y$ is expressed with the use of f.

According to Definition 1, it is clear that a single-label classification is a particular case of the multi-label classification. When the number of labels of instances is equal to 1 ($|y_i| = 1$), the multi-label classification problem transforms into a single-label classification problem.

Due to the existence of relevance and co-occurrence among labels in multi-label classification, single-label classification methods cannot be used to directly address the multi-label classification problem (Streich & Buhmann, 2008; Tsoumakas et al., 2010). A large body of research has been carried out to explore effective and efficient multi-label classification approaches which are generally grouped into two main categories: problem transformation methods and algorithm adaptation methods (Tsoumakas et al., 2010). However, most of these methods neglect a fact that there exists some uncertainty during the process of classification. The uncertainty is caused by some reasons. First, due to the finite number of training instances, we cannot acquire an exact distribution of each

* Corresponding author at: Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China.
E-mail addresses: yuyingtongzhi@hotmail.com, yuyingjx@163.com (Y. Yu).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.10.030

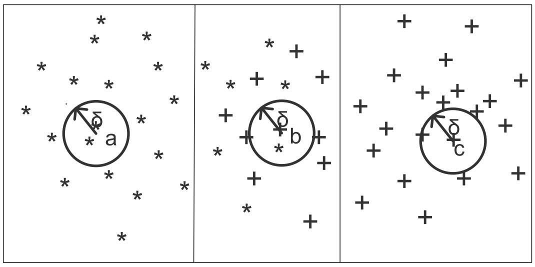

Fig. 1. Single-label example with two classes.

class. Second, because of the overlap existing among different classes, there exists some ambiguity in the feature space for a given instance. The uncertainty affects the precision of prediction. Rough sets form a conceptual vehicle to deal with ambiguous, vague, and uncertain knowledge while the neighborhood rough set model is an extension of traditional rough set model to deal with the uncertainty in the numerical data. The problems indicated above stimulate us to propose two multi-label classification algorithms based on neighborhood rough sets to cope with the uncertainty as well as the correlation among labels. The two proposed algorithms referred to as MLRS and MLRS-LC consider the global and local correlation among the labels. By introducing the concept of upper and lower approximations of neighborhood rough set model, MLRS and MLRS-LC firstly find out all the possibly related labels for a given instance and exclude all unrelated labels. Then they confirm the final labels according to the neighborhood of given instance. Experimental results concerning seven multi-label datasets show that the proposed approaches exhibit a promising performance when considering uncertainty and correlation aspects. They can not only improve the classification precision but also reduce the training time compared with other standard multi-label algorithms.

The paper is organized as follows. In Section 2, basic notation and evaluation metrics used in multi-label classification are briefly introduced. Section 3 provides some background material on multi-label classification and neighborhood rough sets respectively. Sections 4 and 5 respectively introduce the proposed approaches MLRS and MLRS-LC. Section 6 contains experimental results obtained by applying the proposed algorithms and other multi-label learning algorithms to multi-labeled datasets. Finally, Section 7 concludes the study and identifies some future research directions.

2. Preliminaries

In this section, we present the formal notation to be used throughout the paper.

We assume that $X \subset R^d$ denotes an input domain of instances and any instance is represented as a d-dimensional vector $x = [x^1, x^2, \ldots, x^d]$, $(x \in X)$. Let $Y = \{l_1, l_2, \ldots, l_m\}$ be a finite domain of possible labels. Each instance is associated with a subset of Y and this subset is described as an m-dimensional vector $y = [y^1, y^2, \ldots, y^m]$ where $y^j = 1$ only if instance x has label l_j and 0 otherwise.

Let $T = \{(x_i, y_i) | 1 \leqslant i \leqslant n, x_i \in X, y_i \subset Y\}$ be a training set composed of n labeled instances and $D = \{(x_i, y_i) | 1 \leqslant i \leqslant q, x_i \in X, y_i \subset Y\}$be a testing set composed of q labeled instances. The subscript in this description is used to avoid the confusion with the label dimension. Therefore, y_i^j corresponds to the binary relevance of the jth label belonging to the ith instance.

The performance evaluation of the multi-label classification system is different from that of the singe-label classification system. In multi-label classification, the evaluation is much more complicated. In experimental evaluation, we consider some measures proposed in literature Schapire & Singer (2000) and Godbole & Sarawagi (2004).

(1) *Hamming loss* (Schapire & Singer, 2000): this measure evaluates how many times an instance-label pair is misclassified considering the predicted set of labels y' and the ground-truth set of labels y.

$$hloss = 1 - \frac{1}{mq} \sum_{i=1}^{q} \sum_{j=1}^{m} 1_{y_i'^j = y_i^j}$$

(2) *Average precision* (Schapire & Singer, 2000): this measure evaluates the average fraction of labels ranked above a particular label $\lambda \in y_i$ which actually are in y_i.

$$avgprec = \frac{1}{q} \sum_{i=1}^{q} \frac{1}{|y_i|} \sum_{\lambda \in y_i} \frac{|\{\lambda' \in y_i : r_i(\lambda') \leqslant r_i(\lambda)\}|}{r_i(\lambda)}$$

where $r_i(l)$ denotes the rank of label $l \in Y$ predicted by the algorithm for a given instance x_i.

(3) *Accuracy* (Godbole & Sarawagi, 2004): the measure gives an average degree of similarity between the predicted and the ground truth label sets of all testing examples.

$$accuracy = \frac{1}{q} \sum_{i=1}^{q} \frac{|y_i \cap y_i'|}{|y_i \cup y_i'|}$$

(4) *F1-measure* (Godbole & Sarawagi, 2004): for completeness of the analysis, we include the *F1-measure.* This is the harmonic mean between precision and recall, common to information retrieval. It can be calculated from true positives, true negatives, false positives and false negatives based on the predictions and the corresponding actual values.

$$F1 = \frac{1}{q} \sum_{i=1}^{q} \frac{|y_i \cap y_i'|}{|y_i| + |y_i'|}$$

Smaller values of *Hamming loss* correspond to higher classification quality, while larger values of *average precision*, *accuracy* and *F-measure* relate to higher classification quality.

3. Related work

Before embarking on an introduction of MLRS and MLRS-LC presented in this paper, let us review some existing works on multi-label learning and neighborhood rough sets.

3.1. Multi-label classification

As mentioned in Section 1, multi-label classification algorithms can be categorized into two different groups: (i) problem transformation methods and (ii) algorithm adaption methods. The first group includes methods that are algorithm independent. They transform the multi-label problem into one or more single-label problems. The representative problem transformation methods include binary relevance method (BR) Boutell, Luo, Shen, et al. (2004), binary pair wise classification approach (PW) Hüllermeier, Fürnkranz, Cheng, et al. (2008) and label combination or label power-set method (LC) Tsoumakas & Vlahavas (2007). The second group includes methods that extend specific learning algorithms in order to handle multi-label data directly. Well-known approaches include Adaboost (Schapire & Singer, 2000), BP-MLL (Zhang & Zhou, 2006), lazy methods (Denœux, Younes, & Abdallah, 2010; Spyromitros, Tsoumakas, & Vlahavas, 2008; Zhang & Zhou, 2007) and others.

BR (Boutell et al., 2004) is a popular problem transformation method that learns m binary classifiers for each different label in Y. Then each binary model is trained to predict the relevance of one of labels. Although BR is mentioned throughout the literature,