Information Fusion 36 (2017) 80–89

Contents lists available at ScienceDirect

Information Fusion

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

One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies

Ligang Zhou[a,*], Qingyang Wang[b], Hamido Fujita[c]

[a] School of Business, Macau University of Science and Technology, Taipa, Macau
[b] Asia-Pacific Academy of Economics and Management, University of Macau, Macau
[c] Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Japan

ARTICLE INFO

Article history:
Received 31 December 2015
Revised 10 November 2016
Accepted 10 November 2016
Available online 11 November 2016

Keywords:
Multi-class classification
One versus one
Listing status
Prediction
Decision directed acyclic graph
Optimizing

ABSTRACT

Most existing research has demonstrated the success of different decomposition and ensemble strategies for solving multi-class classification problems. This study proposes a new ensemble strategy for One-vs-One (OVO) scheme that uses optimizing decision directed acyclic graph (ODDAG) whose structure is determined by maximizing the fitness on the training set instead of by predefined rules. It makes an attempt to reduce the effect of non-competent classifiers in OVO scheme like decision directed acyclic graph (DDAG) but in another way. We test the proposed method on some public data sets and compare it to some other widely used methods to select the proper candidates and related settings for a problem with practical concern from financial industry in China, i.e. the prediction of listing status of companies. The experimental result shows that our model can outperform the benchmarked methods on this real problem. In addition, the ODDAG combined with decision tree is a white box model whose internal rules can be viewed and checked by decision makers.

1. Introduction

Due to the wide existence of multi-class classification problems in different areas, many different methods have been developed to solve such problems. The strategies behind these methods can be roughly categorized into two categories [1]. One category of extensible algorithm, is to extend some binary classification techniques by special formulations to make them applicable for multi-class classification problems, such as discriminant analysis [2–4], decision trees [5,6], k-nearest neighbors [7,8], Naive Bayes [9,10], neural networks [11–13], and support vector machines [14–17]. Another category of decomposition and ensemble methods (DEM), is to decompose a multi-class classification problem into a set of binary classification problems that can be solved by binary classifiers (BCs), and then classify a new observation by applying an aggregative strategy on the binary classifiers' predictions.

A wide variety of empirical studies have reported the decomposition and ensemble methods can increase the performance on multi-class classification problems. Garcia et al. [18] demonstrated the proposed DEM increases the performance of the noise filters studied. Sesmero [19] formalized and evaluated an ensemble of classifiers by using a specific attribute subset to train the base learners and demonstrated their model is as accurate as some well-known classification methods in most cases. Galar et al. [20–22] conducted a comprehensive investigation into different ensemble strategies to combine the outputs of the binary classifiers generated from different decomposition strategies in solving multi-class classification problems. Their empirical study showed that the performance of DEM is dependent on the selection of decomposition strategy, ensemble strategy, the binary classifier, and the characteristics of the problem.

Most existing research shows that the design or selection of decomposition and ensemble strategies play an important role in the performance of DEMs. With regard to decomposition strategies, OVO [23,24], One-vs-All (OVA) [20,25], and error-correcting output coding (ECOC) [26] are the most widely used. Lorena [27] provided a comprehensive review on different decomposition strategies for multi-class problems. Galar et al. [20] demonstrated that OVO and OVA strategies are simple but powerful.

However, one inherent flaw of OVO decomposition strategy is the problem of non-competent classifiers [28]. A binary classifier can only distinguish the observations from the two classes used

* Corresponding author.
E-mail addresses: mrlgzhou@gmail.com (L. Zhou), qywang@umac.mo (Q. Wang), issam@iwate-pu.ac.jp (H. Fujita).

http://dx.doi.org/10.1016/j.inffus.2016.11.009
1566-2535/

in training set, and therefore the binary classifier has no capability to discriminate the observations from other classes. For a new observation whose class is unknown, some binary classifiers from OVO decomposition are competent and some are non-competent. However, which binary classifiers are competent and which binary classifiers are non-competent for the new observation are unknown as well. If all outputs from both the competent and non-competent classifiers are taken into account equivalently in the ensemble stage, the non-competent classifiers may mislead the correct classification of the new observation.

Many efforts have been made in developing ensemble strategy to reduce the effect of non-competent classifiers. Weighted voting (WV), an straightforward and widely used strategy, is to classify a new observation into the class with the maximum total confidence which is the sum of confidence from all binary classifiers on the class. DDAG [29] has been proved to be another effective strategy [14]. The basic idea of DDAG is to start at the top node of the hierarchy and successively discard certain classes by nodes that have been accessed until reaching the bottom of the structure, where the final leaf node returns the class of the estimated observation. Each node in the structure corresponds to a binary classifier. Dynamic classifier selection (DCS) [28], an effective and emerging strategy, can reduce the number of non-competent classifiers in the ensemble process. The main idea of DCS is to consider a reduced subset of binary classifiers by analyzing the classes of the observation's neighbours.

The ensemble strategies of WV, DDAG and DCS are predetermined and heuristic. They need not check their performance on training set or validation set and therefore have no feedback mechanism for adjusting the decision structure. Takahashi and Abe [30] demonstrated that the generalization ability of the decision acyclic graph support vector machines (DAGSVM) depends on the decision acyclic graph structure. They proposed to optimize the structure with the estimate of the generalization error defined as the ratio of the number of support vectors to the number of observations in training data set for the pairwise classes so that the class pairs with higher generalization abilities are put in the upper nodes of the tree. The advantage of the DAGSVM is that the optimized tree can be obtained by one-pass scan and therefore it is time-saving. However, it is difficult to extend the optimization of generalization error due to it dependency on the number of support vectors in support vector machines. Therefore, this would lead to hinder the extension of such optimization strategy to other DDAG with different binary classifiers. This study is to develop an ensemble strategy for OVO by optimizing the DDAG structure and apply it for predicting listing status of companies which is an import problem in Chinese stock markets.

In general, listed companies in China exhibit four different listing statuses: (1) normal status without any risk warning, (2) abnormal status with delisting risk warning, (3) abnormal status with other risk warning, and (4) delisted status. These four statuses are denoted as "A" "D", "B" and "X", respectively. A company can switch from one listing status to another with the exception that delisted status cannot switch back to any other statuses.

Since different listing statuses indicates different levels of overall risk, correct prediction of the listing status of a listed company is helpful for the company's investors, creditors and other stakeholders to assess the company's risk. Although predicting the listing status of Chinese listed companies is a topic with practical significance, existing studies tend to consider it as a financial distress prediction problem which simplifies the listing status prediction problem as a binary classification problem [31–33].

Zhou et al. [34] introduced OVO and OVA based methods to predict the listing status of Chinese listed company. Although their experimental results demonstrate the efficiency of the proposed methods, but they did not explore the improvement on the aggregative strategies. This paper is to present a new learning architecture based on DDAG for predicting the listing status of Chinese listed companies and compare it to DDAG, DCS and other ensemble methods.

The rest of this paper is organized as follows. Section 2 presents the related decomposition and ensemble strategies. The new learning architecture based on DDAG is given in Section 3. The empirical studies of the proposed learning architecture on some public data sets and the big-scale data set of listing status prediction are presented in Section 4. Section 5 draws conclusions and discusses future research directions.

2. Related decomposition and ensemble strategies

Multi-class classification models aim at assigning a class label for each input observation. Given a training data set $\{(\boldsymbol{X}_1, y_1), \dots, (\boldsymbol{X}_N, y_N)\}$, where $\boldsymbol{X}_i \in \mathbb{R}^m$ denotes the ith observation feature vector, and $y_i \in \{1, \dots, K\}$ is the class label of the ith observation. A multi-class classification model is a map function F: $\boldsymbol{X} \to \{1, \dots, K\}$ inferred from the labeled training data set through a training process.

2.1. Decomposition strategies

2.1.1. One-vs-one decomposition strategy

OVO [20] approach is to divide the multi-class problem with K classes into $C_K^2 = K \times (K-1)/2$ binary classification problems. One binary classifier is constructed for each binary classification problem for discriminating each pair of classes. Let the binary classifier that discriminates the pairwise classes of i and j be denoted by B_{ij}, $i < j$, the output of binary classifier B_{ij}, denoted by p_{ij}, is the posterior probability defined as follows.

$$p_{ij} = f_{ij}(y = i|\boldsymbol{X}), \quad i < j, \quad p_{ij} \in [0, 1], \tag{1}$$

where f_{ij}: $\boldsymbol{X} \to \{i, j\}$ is the map function of binary classifier B_{ij}.

The p_{ij} can be taken as the confidence of binary classifier B_{ij} classifying an observation with feature vector $\boldsymbol{X}$ as class i. Since classifier B_{ij} is only used to discriminate two classes, if the classifier classes an observation into class i with probability p_{ij}, it classes the observation into another class j with probability $1 - p_{ij}$. The outputs of all $K \times (K-1)/2$ binary classifiers can be represented by following score matrix P:

$$P = \begin{pmatrix} - & p_{12} & \dots & p_{1K} \\ 1-p_{12} & - & \dots & p_{2K} \\ \vdots & \vdots & \vdots & p_{iK} \\ 1-p_{1K} & 1-p_{2K} & \dots & - \end{pmatrix} \tag{2}$$

2.1.2. One-vs-all decomposition strategy

OVA approach is to divide the multi-class problem with K classes into K binary classification problems. Each binary classifier is constructed to discriminate one class from all other classes. The binary classifier B_i is trained by all training samples in which the observations of non class i are relabeled as 0.

The output of binary classifier B_i, denoted by p_i, $p_i \in [0, 1]$, is defined similar to Eq. (1). The value of p_i measures the confidence of the classifier B_i classifying an observation as class i. The outputs of all K binary classifiers can be represented by the following vector P:

$$P = (p_1, p_2, \dots, p_K) \tag{3}$$

2.2. Ensemble strategies

2.2.1. Maximum confidence (MC) strategy

The class selected by the MC for OVA is the class voted by a classifier with the maximum confidence and is defined as follows: