Information Sciences 328 (2016) 222–236

Contents lists available at ScienceDirect

Information Sciences

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

Predicting the listing status of Chinese listed companies with multi-class classification models

Ligang Zhou [a,*], Kwo Ping Tam [a], Hamido Fujita [b]

[a] School of Business, Macau University of Science and Technology, Taipa, Macau
[b] Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Japan

ARTICLE INFO

Article history:
Received 3 June 2015
Revised 20 August 2015
Accepted 21 August 2015
Available online 29 August 2015

Keywords:
Prediction
One-vs-all
One-vs-one
Multi-class classification
Listing status

ABSTRACT

In China's stock markets, a listed company's different listing statuses are signals for different risk levels. It is therefore vital for investors and other stakeholders to predict the listing status of listed companies due to the difficulty of providing sufficient measurement of such risks. Existing studies tend to classify listing status into two categories for simple measurement purposes by applying binary classification models; however, such classification models cannot provide accurate risk management. Considering the existence of four different listing statuses of Chinese listed companies in practice, this study introduces three different types of multi-class classification models to predict listing status in order to achieve better performance in terms of accuracy measures. These three types of models are based on One-versus-One and One-versus-All with parallel and hierarchy strategies. The performances of the three different models with two different types of feature selection strategies are compared. Further, the effectiveness and accuracy of the models' performance are tested on a large test dataset. The achieved accuracy measures could provide better risk prediction for listed companies.

1. Introduction

The Chinese stock market, which initially opened in 1990, is an emerging market. To make individual investors aware of the risk of different listed companies, in 1998, China's Securities Regulatory Commission (CSRC) issued special listing rules regarding disclosing the risk level of listed companies. According to the listing rules, when abnormal financial or other specified conditions arise in a listed company and these abnormal conditions increase the company's risk of being terminated from the exchange or render investors unable to judge the company's prospects and consequently hurt investors' interests, the companys stock is given a special treatment (ST) label as a risk warning.

Under the listing rules, an ST-labeled company can have the ST label removed if its financial health has improved to an extent that satisfies certain specified requirements. In China, a healthy company (without the special treatment label) may become financially troubled and thus receive the delisting risk (or some other risk) warning, or may even be delisted, while a financially troubled company may regain its status as a healthy company. In general, listed companies in China exhibit four different listing statuses: (1) normal status without any risk warning, (2) abnormal status with other risk warning, (3) abnormal status with delisting risk warning, and (4) delisted status. These four statuses are denoted as "A", "B", "D" and "X", respectively. Listing status can switch from one level to another with the exception that delisted status cannot be reversed.

* Corresponding author. Tel.: +85388972903.
E-mail addresses: mrlgzhou@gmail.com (L. Zhou), kptam@must.edu.mo (K.P. Tam), issam@iwate-pu.ac.jp (H. Fujita).

http://dx.doi.org/10.1016/j.ins.2015.08.036

Correctly predicting the listing status of a listed company is very important for the company's stakeholders, including investors, creditors, suppliers and customers. Companies with different listing statuses are associated with different levels of overall risk. Predicting a change in listing status can help investors to manage their stock portfolio risk and aid creditors, suppliers and customers to accurately evaluate a company's credit risk. In addition, there is a daily plus and minus 5% cap on the trading price for special treatment companies, while a daily 10% cap is set for healthy listed companies. Consequently, stocks with different listing statuses are associated with different levels of overall risk, such as volatility, liquidity, and delisting risks, which are all major concerns for investors in making their investment decisions.

The existing literature on predicting the listing status of Chinese listed companies mainly focuses on predicting whether a healthy listed company will maintain a normal status or fall into financial distress. Ding et al. [6] introduced support vector machines to predict the financial distress of a company and defined a company in financial distress as a company that had received special treatment in the Chinese stock markets. Zhang et al. [32] developed a Z-score model to forecast whether a firm would receive special treatment. Existing studies [12,16,19,20,28,31] commonly consider listing status prediction for Chinese listed companies as a binary classification problem and develop binary classification models to predict the listing status as "normal" or "financially distressed" in a given period.

Given that there are in fact four possible listing statuses for a Chinese listed company (different listing states mean different risk levels), it is of practical significance to apply a multi-class classification approach to predict the listing status of listed companies in China.

The most popular strategy to solve multi-class classification problems is to transform the problem of multi-class classification into multiple binary classification problems [10], which carries with it two important issues. One is how to decompose a multi-class classification problem into a series of binary class classification problem; the other is how to assemble the results obtained by the multiple binary classifiers. The methods for managing the former and latter procedures are termed "decomposition" and "ensemble strategies", respectively.

There are two common decomposition strategies: One-vs-One (OVO) [18] and One-vs-All (OVA) [3]. Suppose there are K classes in a multi-class classification problem. The OVO approach is to divide the problem into $C_K^2 = K(K-1)/2$ binary classification problems, after which one binary classifier is trained to discriminate classes in each pair. The outputs from all C_K^2 binary classifiers are aggregated to predict the output class. The OVA method divides the problem into $K-1$ binary classification problems such that each binary classifier distinguishes one class from all other classes. Some existing studies show the successful application of OVA in multi-class classification problems [7,8,15]. Rifkin and Kautau [26] claimed that the OVA strategy is as accurate as any other approach when the base classifiers are well-tuned, while others [11,17] have demonstrated that the OVO strategy is also a useful alternative to multi-class classification problems with a performance superior to that of OVA. Galar et al. [10] conducted a comprehensive investigation into different ensemble methods for binary classifiers in multi-class problems with OVO and OVA strategies. Their empirical study showed that the performance of OVO and OVA is sensitive to the selection of the ensemble strategy and that the best aggregation within a problem depends on the base classifier and the characteristic of the problem.

For a listed company, there is a large amount of information that can be used to predict the company's listing status, such as the company's characteristics, financial performance and market information. Existing literature has shown that financial and marketing information is effective in predicting a company's financial status [1,5,25]. The selection of features will affect the performance of the classification models.

The number of companies with different listing statuses in each observed year varies and the listing status prediction problem is a typically highly imbalanced multi-class classification problem.

This study introduces multi-classification models to predict the listing status of Chinese listed companies by integrating the selection of features and samples into the OVA and OVO with parallel and hierarchy strategies, respectively. It also investigates the effect of different binary classifiers in the OVA and OVO strategies. The remainder of this paper is organized as follows. The frameworks based on OVO and OVA with two different ensemble strategies are explained in detail in Section 2. Section 3 presents the empirical results of the multi-class classification models. Section 4 draws conclusions and summarizes major findings.

2. One-vs-All and One-vs-One aggregative models

The OVA aggregative models (OVAAM) integrate the process of feature selection, sampling and the OVA strategies. Two different ensemble strategies, parallel and hierarchy, are employed in OVA in this study. The training and testing processes in OVAAM with parallel and hierarchical ensemble strategies are shown in Figs. 1 and 2, respectively. The training and testing processes in the One-vs-One (OVO) aggregative models (OVOAM) are shown in Fig. 3.

2.1. Feature selection method

Feature selection can speed up learning, facilitate data understanding, and improve prediction performance. A variety of feature selection methods have been proposed and examined, including filter methods, wrapper methods and embedded methods [14]. This study employs a hybrid feature selection method that combines the filter method based on a two-sample t-test with variance inflation factor (VIF) analysis. The hybrid method takes advantage of the rapidity of the filter method, while VIF analysis can assure lower level of dependence among the selected features.