ARTICLE IN PRESS
JID: KNOSYS [m5G;December 30, 2016;13:32]

Knowledge-Based Systems 000 (2016) 1–11

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble

Jie Sun [a,*], Hamido Fujita [b], Peng Chen [c], Hui Li [d]

[a] *Department of Accounting, Business School, Tianjin University of Finance and Economics, NO. 25 Zhujiang Road, Tianjin, 300222, PR China*
[b] *Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Japan*
[c] *School of Economics and Management, Zhejiang Normal University, NO. 688 Yingbin Dadao Road, Jinhua, 321004, Zhejiang, PR China*
[d] *College of Tourism and Service Management, Nankai University, NO. 38 Tongyan Road, Tianjin, 300350, PR China*

ARTICLE INFO

Article history:
Received 17 August 2016
Revised 18 December 2016
Accepted 19 December 2016
Available online xxx

Keywords:
Dynamic financial distress prediction
Concept drift
Time weighting
Adaboost
Support vector machine

ABSTRACT

Dynamic financial distress prediction (DFDP) is important for improving corporate financial risk management. However, earlier studies ignore the time weight of samples when constructing ensemble FDP models. This study proposes two new DFDP approaches based on time weighting and Adaboost support vector machine (SVM) ensemble. One is the double expert voting ensemble based on Adaboost-SVM and Timeboost-SVM (DEVE-AT), which externally combines the outputs of an error-based decision expert and a time-based decision expert. The other is Adaboost SVM internally integrated with time weighting (ADASVM-TW), which uses a novel error-time-based sample weight updating function in the Adaboost iteration. These two approaches consider time weighting of samples in constructing Adaboost-based SVM ensemble, and they are more suitable for DFDP in case of financial distress concept drift. Empirical experiment is carried out with sample data of 932 Chinese listed companies' 7 financial ratios, and time moving process is simulated by dividing the sample data into 13 batches with one year as time step. Experimental results show that both DEVE-AT and ADASVM-TW have significantly better DFDP performance than single SVM, batch-based ensemble with local weighted scheme, Adaboost-SVM and Timeboost-SVM, and they are more suitable for disposing concept drift of financial distress.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Since the global financial crisis in 2008, the world's economic growth began to slow down. In the unstable economic environment, enterprises have to bear great pressure, and many of them fall into financial distress or even bankruptcy due to unscientific management. Meanwhile, the financial distress of enterprises will increase the risk of banking industry to some extent. When many enterprises are unable to repay the bank loan because of financial distress, the banks' bad debts would increase and this would furthermore bring huge hidden troubles to the whole financial system. In such a situation, it needs an efficient financial distress predict (FDP) system, which can help enterprises improve risk management and help banks make scientific credit decision.

However, most FDP research is based on static sample data, which is contradictory with the dynamic sample data flow in reality. Some studies have considered financial distress concept drift (FDCD) and focused on dynamic financial distress predict (DFDP) [1–4], but they only considered the method of time window or data batch combination, ignoring the idea that new samples and old samples should have different weights in DFDP modeling. This paper attempts to propose two new approaches for DFDP in condition of FDCD, based on time weighting combined with Adaboost support vector machine (SVM) ensemble, so as to enrich the methodology system of DFDP.

2. Literature review

2.1. Literature review on FDP

Many researches have focused on the subject of FDP. Altman [5] used MDA to identify companies into known categories and concluded that bankruptcy could be explained quite completely by a combination of five financial ratios. Ohlson [6] was the first to apply logistic regression (Logit) model to predicting financial distress. Frydman et al. [7] also tried decision tree (DT) to FDP. Afterward, neural network is the most widely used machine learning method in this field. For example, Chen and Du [8] and Wu et al. [9] respectively constructed the three-layer feed-forward back-propagation neural network model and the probabilistic

* Corresponding author.
E-mail address: sunjiehit@gmail.com (J. Sun).

http://dx.doi.org/10.1016/j.knosys.2016.12.019
0950-7051/© 2016 Elsevier B.V. All rights reserved.

Please cite this article as: J. Sun et al., Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.12.019

neural network model for FDP. Besides, Sun and Hui [10] and Sun and Li [11] respectively put forward a FDP method based on similarity weighted voting CBR and a systematic data mining method of decision tree. However, among various data mining methods, SVM is proved to be a relatively good method for FDP because it is based on the structural risk minimization principle rather than the empirical risk minimization principle [12–15].

Based on the single classifier FDP methods, ensemble methods are widely applied in the field of FDP. Some researches constructed ensemble classifiers for FDP using several classification algorithms on the same training set [16–18]. Other researches constructed ensemble classifiers for FDP using only one algorithm on different data subsets [19,20] or under different parameters [21–24]. Among them, Sun et al. [23] and Kim and Upneja [24] found that AdaBoost ensemble method showed good performance in the empirical research of FDP.

In recent years, DFDP became an important branch of FDP research. It focuses on how to update the FDP model dynamically when the new sample data batches gradually emerge and FDCD happens over time [25]. For FDP modeling based on a lateral data set composed of many distress and non-distress sample companies, Sun and Li [1] applied the instance selection mechanism based on time windowing, and Sun et al. [2] proposed the adaptive and dynamic ensemble of SVM based on data batch combination. For FDP modeling based on the longitudinal financial data stream of a certain company, the process of financial condition evaluation and the process of FDP can be dynamically integrated. For example, Sun et al. [3] attempted sequential floating forward selection with principal component neural network optimized by genetic algorithm, and Sun et al. [4] put forward an approach based on the entropy-based weighting, SVM and vertical sliding time window.

2.2. Literature review on concept drift

Schlimmer and Granger [26] defined concept drift as changes in the target concept, and these changes are induced by changes in the hidden context. Widmer and Kubat [27] further categorized concept drift into the real and the virtual, and considered the concept drift defined by Schlimmer and Granger [26] as the real concept drift. Furthermore, they considered the virtual concept drift as the change in the underlying data distribution when the target concept remains the same. In terms of the degree of virtual concept drift, it is further divided into sudden concept drift and gradual concept drift [28].

When concept drift happens, the model trained on the old data set may become ineffective for the new concept environment. To solve this problem, three types of methodology are proposed. The first is sample selection with time windowing. For example, Hulten [29] introduced a method of Concept-adapting Very-Fast Decision Tree learner, to build a time window to select samples and solve concept drift in data flowing. Sample selection with time windowing is the most simple approach to treat concept drift. However, it completely abandons the old samples outside the time window, and this leads to the limitation that some old samples that are informative are excluded. The second is sample weighting. For instance, Koychev [30] introduced a gradual forgetting algorithm that considered sample age for adaptation to concept drift. Klinkenberg [31] brought us two time weighting methods based on exponential function, namely global scheme and local scheme. Sample weighting can avoid the limitation of sample selection to some extent. But for such a special situation as concept repeat, certain sampling weighting scheme may discord with the importance of samples and its effectiveness shows instability in such a condition. The third is classifier ensemble. Kyosuke et al. [32] proposed an adaptive classifier ensemble method by adding online learners into the ensemble system. Wu et al. [33] proposed an online bagging strategy to construct classifier ensemble model that can update itself by considering inflow of new samples. Song et al. [34] constructed an ensemble model named as Dynamic Clustering Forest to classify textual data stream with concept drift. These classifier ensemble methods deal with concept drift by adding new base classifiers or updating their weights with emergence of new samples. But they do not consider the different roles of older samples and newer samples in the process of base classifier training. Since classifier ensemble has the advantages such as stability and better classification performance than single classifier, it is necessary to integrates sample weighting with classifier ensemble for disposing concept drift.

3. Background theory

3.1. Time weighting

Time weighting has been used widely to solve concept drift. The basic assumption it relies on is that the new samples are always more important than the older ones. With time flowing, the importance of an old sample gradually decreases. Klinkenberg [31] introduced a time weighting method, with the basic opinion that the importance of data sample batches is monotonically decreasing with time flowing. Its function is shown in formula (1):

$$w_t = \exp(-\lambda t) \tag{1}$$

In this formula, t is the time variable that denotes the time point t time steps ago, and λ is the time weighting parameter. The larger λ is, the sooner an example becomes irrelevant. For $\lambda \to \infty$, the model is learned only on the newest examples, and for $\lambda = 0$, all examples share an equal weight.

3.2. Support vector machine

SVM is a statistical learning method [35]. For a classification sample data set expressed by $D = \{x_i, y_i\}_{i=1}^{N}$ where$y \in \{1, -1\}$, SVM learning is to construct the optimal separating hyper-plane with the biggest margin width between the two classes. It is an optimization problem expressed as formula (2).

$$\begin{aligned} &\min \quad \tfrac{1}{2} w^{\mathrm{T}} w \quad + C \sum_{i=1}^{N} \xi_i \\ &\text{s.t.} \\ &\begin{cases} y_i[w^{\mathrm{T}}\Phi(x_i) + b] \geq 1 - \xi_i & (i = 1, 2, \ldots, N\ \ ) \\ \xi_i \geq 0 \quad (i = 1, 2, \ldots, N) \end{cases} \end{aligned} \tag{2}$$

In the above formula, ξ_i are the slack variables, which are used to allow certain degree of training misclassification, $C \in R^{+}$ is a tuning parameter that weights the importance of classification error with the margin width, and w^{T} represents the weight vector. The kernel function $\Phi(x_i)$ is used to map the original feature space into a high-dimensional space, and there are mainly four types of kernel functions, i.e. linear kernel function, polynomial kernel function, radial basis kernel function (RBF), and sigmoid kernel function [36].

3.3. Batch-based ensemble with local weighted scheme (BE-LWS)

To solve concept drift problem, Klinkenberg proposed a local weighted scheme (LWS) [31], whose algorithm is illustrated in Table 1, in which *Batch T* is the most recent training data batch and *Batch* 1 is the oldest available training data batch.

In the above algorithm, it uses the most recent data batch to build a classifier, so as to test the suitability of the rest data batches for current modeling. This is based on the assumption that the distribution of the recent data batch is supposed to be the most similar to the future testing data set. Therefore, after

Download English Version:

https://daneshyari.com/en/article/4946160

Download Persian Version:

https://daneshyari.com/article/4946160

Daneshyari.com