Neurocomputing 177 (2016) 636–642

Contents lists available at ScienceDirect

Neurocomputing

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

A nonlinear subspace multiple kernel learning for financial distress prediction of Chinese listed companies

Xiangrong Zhang [a,b,*], Longying Hu [b]

[a] School of Management, Harbin Institute of Technology, Harbin, China
[b] School of Management, Heilongjiang Institute of Technology, Harbin, China

ARTICLE INFO

Article history:
Received 5 September 2015
Received in revised form
13 November 2015
Accepted 20 November 2015
Available online 17 December 2015

Keywords:
Financial distress prediction
Multiple kernels learning
Listed company
Subspace learning

ABSTRACT

Financial distress prediction (FDP) is of great importance for managers, creditors and investors to take correct measures so as to reduce loss. Many quantitative methods have been proposed to develop empirical models for FDP recently. In this paper, a nonlinear subspace multiple kernel learning (MKL) method is proposed for the task of FDP. A key point is how basis kernels could be well explored for measuring similarity between samples while a MKL strategy is used for FDP. In the proposed MKL method, a divide-and-conquer strategy is adopted to learn the weights of the basis kernels and the optimal predictor for FDP, respectively. The optimal weights of the basis kernels in linear combination is derived through solving a nonlinear form of maximum eigenvalue problem instead of solving complicated multiple-kernel optimization. Support vector machine (SVM) is then used to generate an optimal predictor with the optimally linearly-combined kernel. In experiments, the proposed method is compared with other FDP methods on Normal and ST Chinese listed companies during the period of 2006–2013, in order to demonstrate the prediction performance. The performance of the proposed method is superior to the state-of-the-art predictor compared in the experiments.

1. Introduction

As the world's second largest economy, Chinese economic development has brought a great power to the global economic recovery, but the enterprise management mechanism remains backward state, including listed companies' financial distress study. Accurate judgment before arising the company financial distress will benefit to reduce property loss form of country and companies' themselves. The companies can take some effective measure to stop this situation spend and bring it back on track. Financial distress prediction (FDP) is using some useful methods to catch this situation, analyzing the report data from enterprises before distress arising [1–3].

In the early stages, some methods were developed for FDP, such as univariate analysis [1], multiple discriminant analysis (MDA) [2], logistic regression algorithm (Logit) [1]. With the development of some artificial intelligence methods, these methods are also used in FDP, like neural networks (NNs) [4,5], support vector machine (SVM) [5–7]. In recent years, some combinations of multiple classifiers are also present to solve the limited explanatory ability problem in single classifiers, such as Bagging method and Adaboost method [9]. Most of researches focus on the model learning for FDP but ignore the importance of financial ratios selection. Although there have already exist some state-of-the-art feature selection method which can be used for ratio selection, like Principal Component Analysis (PCA) [10], Linear Discriminate Analysis (LDA) [11], Kernel-PCA [12] and Kernel-LDA [13], those feature selection are not suitable for FDP and not have a sufficient interpretability. Recently, SVM method has present excellent nonlinear generalization ability to high dimension and small sample evaluation problem and can get upper prediction accuracy using kernel method. SVM has been recently applied for FDP task and demonstrated good performance [7,8]. The conventional SVM only use single kernel like Gaussian kernel with fixed parameters to measure similarity of samples from same class or different classes.

In recent years, the limitation of SVM with single kernel has been recognized gradually. The limitation of single kernel learning motivates researchers to develop new kind of kernel learning methods, called multiple kernel learning (MKL). The Multiple Kernel Learning (MKL) methods based on SVM framework can get a better perform, using a composite kernel effectively to increase the adaptive capacity [14–18]. Essentially, multiple basis kernels with different forms or same forms but different parameters provide more enhanced ability to measure sample similarity. Integrating the basis kernels will results in better generalization

* Corresponding author at: School of Management, Harbin Institute of Technology, 92 West Dazhi Street, Harbin 150001, China. Tel.: +86 159 451 82234.

http://dx.doi.org/10.1016/j.neucom.2015.11.078
0925-2312/

capability and better classification performance. In our previous work, a two-step multiple kernel regression (MKR) was proposed for macroeconomic data forecasting of China [19]. Those existing MKL algorithms demonstrate better performance than the conventional SVM for forecasting and FDP. However, better utilizing the potential of basis kernels for measuring sample similarity is still an open topic.

In this paper, a nonlinear subspace multiple kernel learning (MKL) method is proposed for the task of FDP. In the proposed MKL method, a divide-and-conquer strategy is adopted to separately learn the weights of the basis kernels and the optimal predictor for FDP. The optimal weights of basis kernels in linear combination is derived through solving a nonlinear form of maximum eigenvalue problem instead of solving complicated multiple-kernel optimization. Support vector machine (SVM) is then used to form an optimal predictor with the optimally linearly-combined kernel. The main contribution of this paper can be summarized as follow. In the existing MKL algorithms, two main ways to learn a linear combination of the basis kernels. One is to directly solve a complicated optimization problem which simultaneously optimizes the weights and the final classification results. The other one adopts linear subspace methods to learn the optimally linear combination of the basis kernels. Compared to the existing state-of-the-art, the main contribution of this work is to adopt more effectively nonlinear subspace methods to get a combined kernel which has excellent ability to learn samples.

The rest of the paper is divided into five sections. Section 2 briefly describes the kernel learning and MKL. Section 3 represents the proposed MKL method and its' flowchart for task of FDP. Section 4 gives a detailed description of the test data, *i.e.* the Chinese listed companies' ratios data and analysis of the experimental result. The last section provides conclusion.

2. The proposed MKL method

In the proposed MKL method, the basis kernels are firstly generated by means of Gaussian kernels with different bandwidth parameters. The optimal weights of the basis kernels in linear combination form are learned via subspace learning manner. In other words, searching the optimal weights is converted into a subspace learning problem. In order to solve the subspace learning problem, an eigenvalue decomposition method is performed on basis kernels in Reproduced Kernel Hilbert Space (RKHS). The eigenvector which responds to the maximum eigenvalue is just the optimal weights of the basis kernels.

2.1. Conventional MKL

Given a set of training data set $\{\mathbf{x}_i, y_i\}_{i=1}^N$, for binary classification, $y \in \{+1, -1\}$. As we know, the dual optimization of the conventional SVM can be written as the following form

$$\max\left\{L(\alpha_i,\alpha_j)=\sum_{i=1}^N\alpha_i-\frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N\alpha_i\alpha_jy_iy_jK(\mathbf{x}_i,\mathbf{x}_j)\right\}$$
$$\text{s.t.}\begin{cases}\sum_{i=1}^N\alpha_iy_i=0\\ \alpha_i,\alpha_j\in[0,C],\ \forall i,j=1,2,\cdots,N\end{cases}\tag{1}$$

where α_i and α_j are Lagrange multipliers, and if α_i is nonzero, the corresponding $\mathbf{x}_i$ is called support vector which determines the decision hyperplane.

In Eq. (1), K is the kernel matrix which can be denoted by $(K)_{i,j}=k(\mathbf{x}_i,\mathbf{x}_j)$. In aspect of forms of kernel mapping, Gaussian radial basis function (RBF) kernel which is widely used in various tasks of signal processing and pattern recognitions, is considered in this paper. The Gaussian kernel function is given below

$$k(\mathbf{x},\mathbf{z})=\exp\left(-\frac{\|\mathbf{x}-\mathbf{z}\|^2}{2\sigma^2}\right)\tag{2}$$

Here, the Gaussian kernel is used as the kernel function. The Gaussian kernel is the most representative kernel with some merits such as translation invariability. In the proposed MKL method, the Gaussian kernels with different bandwidths are fixed as the predefined basis kernels. As far as MKL is concerned, K can be substitute by convex linear combination of predefined basis kernels with different kernel forms or different kernel parameters. The linear combination of the basis kernels can be expressed as follows

$$K=\sum_{m=1}^M d_mK_m\tag{3}$$

where $\{K_m\}_{m=1}^M$ is a set of the positive semidefinite (PSD) basis kernel matrices with bounded trace and $\left\{d_m|d_m\ \geq 0,\ \sum_{m=1}^M d_m=1\right\}_{m=1}^M$ are the corresponding weights.

By integrating Eq. (3) into Eq. (1), the dual problem of MKL under the optimization routine of SVM can be represented as follows

$$\max\left\{L(\alpha_i,\alpha_j)=\sum_{i=1}^N\alpha_i-\frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N\alpha_i\alpha_jy_iy_j\sum_{m=1}^M d_mK_m(\mathbf{x}_i,\mathbf{x}_j)\right\}$$
$$\text{s.t.}\left\{\sum_{i=1}^N\alpha_iy_i=0\right.$$
$$\alpha_i,\alpha_j\in[0,C],\ \forall i,j=1,2,\cdots,N$$
$$d_m\geq 0,\text{and}\sum_{m=1}^M d_m=1.\tag{4}$$

By means of the optimization routine of the conventional SVM, MKL refers to achieve the simultaneous optimization of the kernel combination and learning performance.

2.2. Subspace learning in MKL

Given the Gram matrices of M basis kernels $\mathcal{K}_0=\{K_m, m=1,2,\cdots,M, K_m\in\mathbb{R}^{N\times N}\}$. We can reformulate Eq. (3) as follows

$$K=\mathbf{d}^T\mathbf{K}\tag{5}$$

where $\mathbf{d}\geq 0$, $\sum_{i=1}^M d_i=1$ and $\mathbf{K}=[K_1K_2\cdots K_M]^T$.

It is easy to find that the Eq. (5) is a typical form of subspace projection. Now we build a loss function as follows

$$\mathcal{L}(U,Z)=\|\mathbf{K}-UZ\|_F^2\tag{6}$$

where U is the projection matrix whose columns, Z is the projected matrix in the linear subspace spanned by U, and $\|\cdot\|_F$ denotes Frobenius norm of matrix.

According to projection theorem, minimizing the loss function $\mathcal{L}(U,Z)$ is only determined by $Z=U^T\mathbf{K}$. Furthermore, minimizing the problem for $\mathcal{L}(U,Z)$ can be converted to the following dual problem under a rank-one constraint

$$\underset{U}{\text{argmax}}\left\|U^T\Sigma U\right\|_F=\underset{U}{\text{argmax}}\left\|U^T\mathbf{K}\right\|_F$$