Applied Soft Computing 19 (2014) 57–67

Contents lists available at ScienceDirect

Applied Soft Computing

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

Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine

Hui Li [a,b,*,1], Chang-Jiang Li [a], Xian-Jun Wu [c], Jie Sun [a]

[a] School of Economics and Management, Zhejiang Normal University, P.O. Box 62, 688 YingBinDaDao, Jinhua, Zhejiang 321004, PR China
[b] College of Engineering, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210, USA
[c] School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan, Hubei 430070, PR China

ARTICLE INFO

Article history:
Received 14 March 2010
Received in revised form 26 March 2011
Accepted 18 January 2014
Available online 5 February 2014

Keywords:
Financial distress identification (FDI)
Support vector machine (SVM)
Statistics-based feature selection
Wrapper

ABSTRACT

Support vector machine (SVM) is an effective tool for financial distress identification (FDI). However, a potential issue that keeps SVM from being efficiently applied in identifying financial distress is how to select features in SVM-based FDI. Although filters are commonly employed, yet this type of approach does not consider predictive capability of SVM itself when selecting features. This research devotes to constructing a statistics-based wrapper for SVM-based FDI by using statistical indices of ranking-order information from predictive performances on various parameters. This wrapper consists of four levels, i.e., data level, model level based on SVM, feature ranking-order level, and the index level of feature selection. When data is ready, predictive accuracies of a type of SVM model, i.e., linear SVM (LSVM), polynomial SVM (PSVM), Gaussian SVM (GSVM), or sigmoid SVM (SSVM), on various pairs of parameters are firstly calculated. Then, performances of SVM models on each candidate feature are transferred to be ranking-order indices. After this step, the two statistical indices of mean and standard deviation values are calculated from ranking-order information on each feature. Finally, the feature selection indices of SVM are produced by a combination of statistical indices. Each feature with its feature selection index being smaller than half of the average index is selected to compose the optimal feature set. With a dataset collected for Chinese FDI prior to 3 years, we statistically verified the performance of this statistics-based wrapper against a non-statistics-based wrapper, two filters, and non-feature selection for SVM-based FDI. Results from unseen dataset indicate that GSVM with the statistics-based wrapper significantly outperformed the other SVM models on the other feature selection methods and two wrapper-based classical statistical models.

1. Introduction

Financial distress identification (FDI) is an effective tool of risk management. This area received lots of focuses from academic and industrial views [1–3,5,6,11,16,19,27,29,33–37,39–42,44,48]. Identification on whether or not a company will fail helps financial institutions, managers, employees, investors and government officials to control risk in their decisions. Predictive accuracy of the tool is a key index indicating whether it is helpful in real-world life. Basically, a predictive model is assumed to be more useful if it is more accurate. A whole dataset is commonly partitioned into training dataset, validating dataset, and testing dataset. The identification of new problems is simulated by using the model constructed on labelled data to predict unlabeled data. This partition is commonly repeated for lots of times in order to provide statistical analysis on significance. Under this assumption, support vector machine (SVM) is an important technique for FDI for the following two reasons: (1) SVM is constructed from mature statistical learning theory [10,46]; (2) previous evidence shows that SVM produced dominating predictive performance in FDI [12,17,18,26,30,31,38,44,47].

Feature selection is a process that chooses information-rich features and retains the meaning of original features [14,23]. Filters and wrappers are two chief methods of feature selection. Filters refer to the use of an algorithm to search through the space of possible features and then to evaluate each subset by running a filter function on the subset. The so-called filter function is not the same as the model used for prediction or classification. Thus, the feature selection approach does not consider preference of the model. Wrappers are similar to filters, but evaluate against the current model instead of a filter function.

A common drawback of previous researches of SVM-based FDI is that they used either filters or genetic algorithm to select optimal feature subsets for SVM. Wrappers are supposed to yield a feature subset that helps model produce dominating predictive performance. Greedy hill climbing, which finds the optimal feature

* Corresponding author at: School of Economics and Management, Zhejiang Normal University, P.O. Box 62, 688 YingBinDaDao, Jinhua, Zhejiang 321004, PR China. Tel.: +86 579 8229 8602.
E-mail addresses: li.1999@osu.edu, lihuihit@gmail.com (H. Li).
[1] Young Researcher of World Federation on Soft Computing.

1568-4946/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2014.01.018

subset by iteratively evaluating a candidate subset of features, is commonly used in wrappers. Genetic algorithm belongs to this type. However, the drawback of the use of genetic algorithm in wrappers is that the outputted feature subset is not the same when the approach is implemented several times.

This research attempts to construct a novel stable wrapper for SVM to identify financial distress. Two key issues in application of SVM include kernel selection and parameter optimization. This new wrapper is constructed on the base of each of the following type of SVM, including: linear SVM (LSVM), polynomial SVM (PSVM), Gaussian SVM (GSVM) and sigmoid SVM (SSVM). Lots of SVM models are produced by using various pairs of parameters after kernel function is selected. Predictions of various SVM models on each candidate feature are transferred into ranking-order information of each feature. The two statistical indices of mean and standard deviation computed from the ranking-order information are combined to calculate a feature selection index of SVM. This index is used to select optimal features.

This paper is organized as follows. Section 2 gives a brief review on feature selection and parameter searching methods used in previous researches of SVM-based FDI. Section 3 presents the new wrapper for SVM-based FDI. Section 4 designs an experiment to testify the efficiency and feasibility of the statistics-based wrapper. Section 5 discusses the experimental results. Section 6 makes conclusion.

2. Feature selection and parameter search in previous researches of SVM-based FDI

When SVM was firstly applied to identify financial distress, Shin et al. [38] employed a two-stage feature selection process, which is composed of *t*-test and stepwise multivariate discriminant analysis (MDA) in consecutive sequence. This type of feature selection belongs to the family of filters. The comparison on predictive performance between SVM and back-propagation neural networks (NN) indicated that SVM produced more accurate ratios than NN. Gaussian kernel was used and its parameters were searched by using the technique of grid-search. Meanwhile, Min and Lee [30] used principal components analysis for SVM-based FDI. This method is a filter-like approach. For comparative models, stepwise logit was used to select optimal features. Both of the feature selection and extraction approaches do not consider preference of SVM in feature selection. They also used grid-search technique to find optimal parameter values of SVM. The comparisons between SVM, MDA, logit, and NN also indicated that SVM outperformed all the three models. For SVM, they tried all of the four kernel functions, i.e., linear kernel, Gaussian kernel, polynomial kernel and sigmoid kernel, and found that Gaussian kernel was the optimal. Min et al. [31] combined genetic algorithm with SVM for feature selection and parameter optimization. This type of feature selection belongs to the family of wrappers. However, the so-called optimal feature subsets generated from different independent implementations of generic algorithm are different. Meanwhile, the use of genetic algorithm for feature selection has the risk of over-fitting. Gaussian kernel was used, and the results showed that SVM with genetic algorithm outperformed NN, logit and SVM with filters. Hui and Sun [18] investigated the feasibility of SVM-based FDI of Chinese listed companies. They used Gaussian kernel for SVM. Feature subset was selected by using the filter of stepwise MDA. The results indicated that SVM outperformed NN, MDA and logit. Wu et al. [47] employed nineteen features that had been used in previous researches as significant ratios in predicting business failure when constructing SVM model. SVM with Gaussian kernel was used, and genetic algorithm was used for parameter optimization. The findings indicated that the SVM was superior to logit, MDA, NN and pure SVM. The method for feature selection belongs to experience-based approach, which is very similar to a filter with the only difference that the assessing function is judgement of human beings. Hua et al. [17] applied SVM to identify financial distress by integrating SVM with logit. They used a two-stage approach for feature selection and grid-search for parameter search. The feature selection approach integrates *t* test and univariate discriminant approach in sequence. Ding et al. [12] used a hybrid filter approach of *t* test and stepwise logit in applying SVM for FDI. Kernel parameters of SVM were searched by grid-search. The findings showed that SVM with Gaussian kernel outperformed NN, MDA and logit. Tsai [44] employed all available features to build predictive models of SVM for FDI. This result means that all available features are regarded significant. Thus, this approach is a filter-like method that use experience as the assessing function. For parameter search, $C = 1$, {1,2,3,4,5} for p of polynomial kernel, and {10,15,20,25,30} for gamma of Gaussian kernel were used. The conclusion indicated that SVM was at least not worse than NN for FDI. Feature selection and parameter optimization methods used in previous researches of SVM-based FDI are summarized in Table 1.

From Table 1 we can find that previous researches of SVM-based FDI did not employ stable wrappers. Eighty-eight percent of previous researches used filter or filter-like approaches. One research used genetic algorithm as a wrapper approach to select feature subset for SVM. However, genetic algorithm is not stable. Meanwhile, it yields to over-fit data, which may limit the generation of SVM for FDI. Thus, this research is motivated to construct a stable wrapper for SVM-based FDI. The constructed wrapper is supposed to be stable in outputting optimal feature subset and not to be bothered by the problem of over-fitting.

3. The statistics-based wrapper for SVM

3.1. Kernels and parameters of SVM when constructing the approach

Wrapper for SVM evaluates against predictive performance of SVM itself. Kernel functions and parameters of SVM must be set up before constructing a wrapper. There are four commonly used kernel functions for SVM, i.e., linear kernel (u′*v), polynomial kernel ((gamma*u′*v)p̂), Gaussian kernel (exp(-gamma*|u-v|2̂)) and sigmoid kernel (tanh(gamma*u′*v)) [7,8]. Each one of the four commonly kernel functions can be employed in the wrapper. The reason

Table 1
Feature selection and parameter searching methods used in previous SVM-based FDI.

	Literature	Feature selection or extraction method	Wrapper or filter	Parameter search
1	Shin et al. [38]	*t*-test and stepwise MDA	Filter	Grid-search
2	Min and Lee [30]	Principal components analysis	Filter-like	Grid-search
3	Min et al. [31]	Genetic algorithm	Wrapper but unstable	Genetic algorithm
4	Hui and Sun [18]	Stepwise MDA	Filter	Grid-search
5	Wu et al. [47]	Experience-based selection	Filter-like	Genetic algorithm
6	Hua et al. [17]	*t* test and stepwise MDA	Filter	Grid-search
7	Ding et al. [12]	*t* test and stepwise logit	Filter	Grid-search
8	Tsai [43]	Experience-based selection	Filter-like	Experience-based selection