Economic Modelling 37 (2014) 175–183

Contents lists available at ScienceDirect

Economic Modelling

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

An ensemble-based model for two-class imbalanced financial problem

Jui-Jung Liao [a], Ching-Hui Shih [b,*], Tai-Feng Chen [c], Ming-Fu Hsu [c]

[a] Department of Management Science and Engineering, Business School, Nankai University, 94, Weijin Rd., Tianjin 300071, PR China
[b] Department of Accounting and Information Systems, Asia University, 500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan, ROC
[c] Department of International Business Studies, National Chi Nan University, 1, University Rd., Puli, Nantou 545, Taiwan, ROC

ARTICLE INFO

Article history:
Accepted 7 November 2013

Keywords:
Artificial neural network
Ensemble learning
Imbalance class
Knowledge extraction
Decision making

ABSTRACT

This study proposes an ensemble-based model (EBM) for the two-class imbalanced classification problem by joining together the support vector machine (SVM), multiple feature selection combination, back-propagation neural network (BPNN) ensemble, and rough set theory (RST). To improve the significance of the rare and specific region belonging to the minority class in the decision region, we take the SVM as a pre-processor to balance the training dataset and use multiple feature selection combination grounded on ensemble learning in order to determine the most representative features from the re-sized dataset. The representative features are then fed into the BPNN ensemble to construct an effective financial pre-warning mechanism. Lacking comprehensibility and readability is one of the fatal weaknesses of an ensemble classifier and it impedes its real-life application. Thus, the study executes RST to extract knowledge from the BPNN ensemble for decision makers to make suitable judgments. Decision makers can take the decision rules as a roadmap to modify a firm's capital structure so as to survive in an extremely turbulent financial market. Empirical results reveal that the introduced EBM's prediction accuracy is very promising in financial risk mining, relative to other detection approaches in this study.

1. Introduction

In the past three decades, financial crisis prediction (FCP) has been an extremely essential topic in both academics and practical domains of corporate finance (Chen et al., 2010; Chevallier, 2012; Louzis and Vouldis, 2012; Naifar, 2012; Olson et al., 2012). The reason is that a financial crisis can incur extreme social costs affecting stock market participants, workers, suppliers, clients, central bank governors, etc. Research on corporate FCP can generate guidelines for a firm's risk precaution, because it yields pre-warning signals before the outbreak of a financial crisis. Hence, it eliminates the chances of corporate being trapped in economic obstacles and prevents in advance the deterioration of the corporate capital structure under a turbulent financial market.

The class imbalance problem has been recognized in many real world applications (Garcia et al., 2012; Japkowicz and Stephen, 2002; Sun et al., 2009; Yang and Wu, 2006) and is an evolving topic in machine learning research. It appears when the number of examples of one class is much lower than the examples of the other class (Chawla et al., 2002). Since the greater proportion of learning approaches are designed to maximize the global measure of accuracy, which is independent of class distribution, it causes a bias towards the majority class in the training of classifiers and results in a lower sensitivity for identifying the minority class examples (Garcia et al., 2012). In an imbalance dataset, the minority class is normally the more essential class. As financial crisis corporations are outnumbered by non-financial crisis corporations, the classification problem can be transformed into a class imbalance task.

The methods for solving the class imbalance task can generally be divided into two categories: sampling approaches and algorithmic approaches. The sampling approach is factually a re-sizing procedure to balance the given imbalance dataset. The algorithmic approach can be viewed as a procedure to adjust the inherent learning algorithm so that it can tackle the class imbalance problem validly. The learning algorithms include recognition-based approaches (Japkowicz and Stephen, 2002) and cost-sensitive approaches (Ting, 2002).

As of yet, no research study in the literature has taken the intelligent technique as a pre-processing approach to re-size data and to further overcome the effect of class imbalance (Farquad and Bose, 2012). The support vector machine (SVM) is one of the best intelligent techniques and has demonstrated its superior performance in pattern recognition and classification problem. By modifying the penalty parameter of the SVM to a higher value, the separating boundary moves more towards the majority class examples (non-financial crisis corporations), in turn misclassifying the majority class examples as minority class examples (financial crisis corporations). In other words, more misclassification for prevalent class examples (majority class) means a greater amount of examples for the rare class (minority class). Thus, the class imbalance problem can be solved by an intelligent pre-processing method.

Feature selection is an inevitable process in data mining and pattern recognition. The basic concept of feature selection is to determine a subset of available attributes, by discarding attributes with little or no detective information, as well as redundant attributes that are highly

* Corresponding author. Tel.: +886 7 224 1588.
E-mail addresses: rollanliao@gmail.com (J.-J. Liao), chshih168@gmail.com (C.-H. Shih), zy4@mail.stust.edu.tw (T.-F. Chen), hsumf0222@gmail.com (M.-F. Hsu).

0264-9993/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.econmod.2013.11.013

correlated (Chen, 2012; Gupta and Modise, 2012; Wu and Hsu, 2012). The data undergoing the feature selection process can facilitate data visualization and data understanding, decrease the computational time and cost, increase the classifier's performance, and defy the curse of dimensionality. However, prior work on feature selection is usually based on one chosen approach only. If we can execute a number of dissimilar feature selection approaches and sequentially combine the selection outcomes, then we not only are able to complement the error made by the singular approach, but also improve the classifier's performance. The fundamental idea of the multiple feature selection combination strategy is grounded on classifier ensemble (Kittler et al., 1998). The combination strategies can be divided into three different structures (original, intersection, and combination). The re-sized dataset is injected into a multiple feature selection procedure to determine the representative features. The representative features are then utilized to construct an effective warning mechanism.

Due to its importance of a risk warning mechanism, there is a growing research stream about financial risk assessment. Numerous statistical methods and optimization algorithms, such as discriminant analysis (Fisher, 1936), logistic regression (Wiginton, 1980), and linear programming (Glover, 1990) have been extensively performed on risk management and crisis assessment. Machine learning techniques, such as case-based reasoning and the decision tree, have shown very inspiring results in their empirical application to FCP. Although the aforementioned approaches can be utilized to evaluate the financial risk, the ability to distinguish non-financial crisis corporations from financial crisis corporations is still worth further improvement. The motivation of an ensemble is to make use of numerous information outputs by multiple dissimilar base classifiers instead of utilizing the result of a traditional singular classifier, so that prediction performance can be improved and the possibility of improper decision making can be eliminated (Sun and Li, 2012).

In accordance with previous domain researchers (Ruta and Gabrys, 2005), the effectiveness of the ensemble method depends on two characteristics (accuracy and diversity) of base classifiers. If the base classifiers with inferior prediction performance are embedded into the ensemble mechanism, then the combination based on them is useless due to incorrect input information. If the base classifiers without diversity are embedded into the ensemble mechanism, then the combination of multiple classifiers can only increase the mechanism's computational cost instead of ameliorating performance. Based on these two aspects – (1) the back-propagation neural network (BPNN) can be viewed as a 'universal approximator' and (2) the BPNN can provide flexible mapping between inputs and outputs – this study selects the BPNN ensemble with acceptable prediction performance (Chi, 2009; Yu et al., 2008).

Over the past decade, the BPNN ensemble has demonstrated superior generalization performance over numerous prediction techniques in a variety of research domains. However, the 'black-box' nature of BPNN has an inability to generate a readable justification or intuitive explanation for the decision it reaches. This has been shown to be one of the fatal drawbacks impeding its real-life application.

Several approaches have therefore been proposed for knowledge extraction from BPNN. This present study is based on the pedagogical structure in order to extract an informative decision rule from the BPNN ensemble. This structure treats BPNN as a black-box and extracts comprehensive rules that describe the relationship between the model's inputs and outputs (Martens et al., 2007). The basic idea is to generate artificially labeled instances where the target class of training data is replaced by the class forecasted by the BPNN ensemble. Sequentially, the artificial dataset is performed with another machine learning method with explanation ability such as tree-shaped learners, which acquire what the BPNN ensemble learned (Barakat and Bradley, 2010).

Knowledge has the nature of granularity and may be incomplete, imprecise, or even conflicting (Khoo and Zhai, 2001). Some notions can only be defined vaguely as they cannot be accurately and precisely illustrated. As a result, knowledge extraction methods are often forced to handle uncertain, imprecise or vague information. This imprecise nature of knowledge is the biggest problem to rule generation. Pawlak (1982) proposed a novel method – the rough set theory (RST) – to tackle uncertainty and vagueness. The method puts much emphasis on the discovery of patterns in an incomplete data structure and can be utilized as a basis to execute formal reasoning under uncertainty and rule discovery (Pawlak, 1984; Slowinski and Stefanowski, 1989). The knowledge generation from RST is represented in a logic statement (If–then) for decision makers to make a proper decision. The decision makers can take decision rules as a guideline to allocate limited economic resources and to modify the capital structure in order to survive in a highly competitive financial environment.

The rest of the study is organized as follows. Section 2 states the theoretical framework of intelligent preprocessing, multiple feature selection, and BPNN ensemble. Section 3 illustrates the experimental results and analysis. Section 4 provides the conclusion.

2. Methodologies

2.1. Support vector machine

Vapnik (1995) introduced support vector machine (SVM) which is a supervised learning approach based on statistical learning theory for classification and regression. Unlike other statistical learning approach (such as decision trees) which commonly aims merely to minimize the practical prediction error, SVM simultaneously minimize practical prediction error and maximize the geometric margin in classification (Öğüt et al., 2012; Yeo et al., 2009). This is done by generating a discriminating hyperplane which optimally separates instances form two classes with maximum margin; hence it is also known as maximum margin classifier (Boser et al., 1992). The basic illustration of SVM is expressed as follows. Given a training set of instance-label pair (x_i, y_i), $i = 1,\ldots,h$ where $x_i \in R^q$ and $y_i \in \{\pm 1\}$. The SVM determines an optimal discriminating hyperplane to separate the two different classes, so that to maximize the margin. The margin is viewed as the distance between the discriminating hyperplane and nearest point. The process is equivalent to handle the following optimization task.

$$\begin{aligned} &Minimize\ \frac{1}{2}(w \cdot w) + C\sum_{i=1}^{h}\xi_i \\ &subject \quad to \\ &y_i((w \cdot x_i) + e) \geq 1 - \xi_i, \quad i = 1,\ldots,h \\ &\xi_i \geq 0, \quad i = 1,\ldots,h \end{aligned} \tag{1}$$

where the slack variable yields a soft discriminating boundary was expressed as ξ_i, weight vector was represented as w, the error was indicated as e and C denotes as a constant corresponding to the value w^2. The solution to dual problem is to maximize the quadratic from Eq. (2) under the constraint of Eq. (3).

$$\underset{\gamma}{Maximum} \sum_{i=1}^{h}\gamma_i - \frac{1}{2}\sum_{i,j=1}^{h} y_i y_j \gamma_i \gamma_j \Theta\left(x_i, x_j\right) \tag{2}$$

subject to

$$\begin{aligned} &0 \leq \gamma_i \leq C, \quad i = 1,\ldots,h \\ &\sum_{i=1}^{h}\gamma_i y_i = 0 \end{aligned} \tag{3}$$

where Lagrange multipliers from the quadratic programming problem were expressed as γ_i. Support vectors (SVs) were utilized to decide the surface of decision, and correspond to the subset of non-zero γ_i, the vectors can be seemed as the most essential training vectors. The kernel function $\Theta(x_i, x_j)$ embedded into SVM is used to transform the