Engineering Applications of Artificial Intelligence 63 (2017) 54–68

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction

Mingjing Wang [a], Huiling Chen [a,*], Huaizhong Li [b], Zhennao Cai [a], Xuehua Zhao [c], Changfei Tong [a], Jun Li [a], Xin Xu [d]

[a] College of Physics and Electronic Information Engineering, Wenzhou University, 325035 Wenzhou, China
[b] Department of Computing, Lishui University, Lishui 323000, Zhejiang, China
[c] School of Digital Media, Shenzhen Institute of Information Technology, Shenzhen 518172, China
[d] Electric Power Research Institute, State Grid Jilin Electric Power Company Limited, Changchun 130021, China

ARTICLE INFO

Article history:
Received 13 May 2016
Received in revised form
17 February 2017
Accepted 9 May 2017

Keywords:
Kernel extreme learning machine
Parameter tuning
Grey wolf optimization
Bankruptcy prediction

ABSTRACT

This study proposes a new kernel extreme learning machine (KELM) parameter tuning strategy using a novel swarm intelligence algorithm called grey wolf optimization (GWO). GWO, which simulates the social hierarchy and hunting behavior of grey wolves in nature, is adopted to construct an effective KELM model for bankruptcy prediction. The derived model GWO-KELM is rigorously compared with three competitive KELM methods, which are typical in a comprehensive set of methods including particle swarm optimization-based KELM, genetic algorithm-based KELM, grid-search technique-based KELM, extreme learning machine, improved extreme learning machine, support vector machines and random forest, on two real-life datasets via 10-fold cross validation analysis. Results obtained clearly confirm the superiority of the developed model in terms of classification accuracy (training, validation, test), Type I error, Type II error, area under the receiver operating characteristic curve (AUC) criterion as well as computational time. Therefore, the proposed GWO-KELM prediction model is promising to serve as a powerful early warning tool with excellent performance for bankruptcy prediction.

1. Introduction

Due to financial crisis all over the world, company bankruptcy predication attracts significant attention for financial institutions. It is important for enterprises to build a trustworthy and accurate early warning system to predicate potential risk of company's bankruptcy beforehand.

Bankruptcy predication generally forms a binary classification that needs to be resolved in a rational approach. The output result generated from the classification models has two types, namely, type 1 represents a company with bankruptcy and type 0 otherwise. Input values of the classification models are often financial statistic ratios derived from credible financial statements in the real enterprises. So far, considerable amount of classification models based on different domain knowledge has been proposed for bankruptcy prediction. In general, the proposed predication models can be classified as statistical approaches or artificial intelligence methods (AI).

A great deal of typical statistical approaches that are constructed for bankruptcy prediction models apply simple univariate analysis (Beaver, 1966), multivariate discriminant analysis (Altman, 1968), logistic regression (Ohlson, 1980) and factor analysis (West, 1985). Recently, AI methods are drawing more attention for failure prediction. Approaches that are based on the AI means, such as artificial neural networks (ANN) (Atiya, 2001a), support vector machines (SVM) (Min and Lee, 2005; Shin et al., 2005), k-nearest neighbor (KNN) approach (Chen et al., 2011c), Bayesian network models (Sarkar and Sriram, 2001; Sun and Shenoy, 2007), extreme learning machine and ensemble methods (Fedorova et al., 2013; Abellán and Mantas, 2014), as well as different hybrid approaches, have been widely used in financial area. Reddy and Ravi (2013) constructed two novel kernels based soft computing techniques for classification task. The experimental results indicated that the proposed approaches could perform well for bankruptcy prediction. Sharma et al. (2013) successfully proposed a hybrid algorithm based on ant colony optimization and Nelder-Mead simplex for training neural networks with an application to bankruptcy prediction. Paramjeet and Ravi (2011) modified bacterial foraging technique to train wavelet neural network in order to predict bankruptcy in banks. A hybrid approach based on differential evolution and radial basis function network (DERBF) proposed by Naveen et al. (2010) was applied to

* Corresponding author.
E-mail address: chenhuiling.jlu@gmail.com (H. Chen).

http://dx.doi.org/10.1016/j.engappai.2017.05.003
0952-1976/

bankruptcy prediction. The results showed that DERBF had a good performance of generalization on bank bankruptcy datasets. Chauhan et al. (2009) employed differential evolution algorithm to train wavelet neural network (DEWNN), predicting the bankruptcy in banks. The results on the four bankruptcy datasets revealed that the DEWNN was obviously superior to other existed methods. Ravi and Pramodh (2008) proposed a new architecture called principal component neural network (PCNN) applied to bankruptcy prediction problem in commercial banks. It is inferred that the proposed PCNN hybrids outperformed other classifiers on the bankruptcy dataset. A new neural network architecture kernel principal component neural network (KPCNN) trained by threshold accepting was presented in Ravisankar and Ravi (2009). Its application to bankruptcy prediction in banks reveled that KPCNN yields comparable results with all the techniques. Vasu and Ravi (2011) proposed new principal component analysis-wavelet neural network hybrid (PCATAWNN) architecture trained by threshold accepting algorithm to predict bankruptcy in banks. The experimental results showed that the PCATAWNN could convincingly outperformed other techniques in terms of area under ROC curve (AUC) in 10-fold cross-validation. In all of the employed methods, ANN (Tsai and Wu, 2008; Atiya, 2001b; Zhang et al., 1999) has become more and more popular for financial prediction, thanks to its prominent ability to capture the nonlinearity relationship that exists between different features in real data set. Nevertheless, it is worth to point out that traditional ANN learning methods, such as the back-propagation approach, are based on the gradient descent strategy which may result in local optimum. Furthermore, it is generally required that a fair amount of network parameters be tuned.

In order to avoid ANN's drawbacks, Huang et al. proposed a new machine learning paradigm named extreme learning machine (ELM) (Huang et al., 2006). ELM is a representative learning model of neural network named after single hidden layer feedforward neural networks (SLFNs). The hidden biases and input weights in this method can be randomly generated, and the output weights are mathematically determined using Moore-Penrose (MP) generalized inverse. It is well-known that the universal approximation can reflect the approximation capabilities of the neural networks. The approximation capabilities of multilayer feedforward networks were proved by Hornik (1991), namely, noconstant bounded continuous activation functions and continuous mappings could be approximated in measure by neural networks. Leshno et al. (1993) advocated that continuous functions could be approximated by feedforward networks with a non-polynomial activation function. Guang-Bin and Babri (1998) proposed that SLFNs with N hidden nodes and almost nonlinear activation function could exactly learn N distinct observations. Due to its classification performance, ELM has been adopted in fields such as image classification (Cao et al., 2016a; Jun et al., 2011), disease diagnosis (Chen et al., 2015; Zhang et al., 2007), and engineering application (Cao et al., 2016b, 2015). In addition, methods based on ELM have also been widely applied in financial areas such as bankruptcy prediction (Yu et al., 2014), corporate life cycle prediction (Lin et al., 2013) and corporate credit ratings (Zhong et al., 2014). One limitation of ELM, nevertheless, is that the randomly assigned input weights can increase the variations of accuracies obtained by classifiers in multiple trials. In order to overcome this limitation, Huang et al. (2012) proposes an extension version of ELM, namely, kernel extreme learning machine (KELM), whose connection weights between hidden layers and input are not necessary. Compared with ELM, KELM can achieve comparative or more excellent property with faster training speed and much easier implementation in applications such as hyperspectral remotesensing image classification (Pal et al., 2013; Chen et al., 2014), activity recognition (Deng et al., 2014), 2-D profiles reconstruction (Liu et al., 2014), disease diagnosis (Chen et al., 2016) and fault diagnosis (Jiang et al., 2014).

We recently applied the KELM to bankruptcy prediction's issue (Zhao et al., 2017), and obtained better performance than other five competitive approaches including SVM, ELM, random forest (RF), particle swarm optimization boosted fuzzy KNN, and Logit model on the same real data set. Nevertheless, it should be noticed that the two significant parameters in KELM with RBF kernel are kernel penalty parameter C and bandwidth γ. C controls the tradeoff between the model complexity and the fitting error minimization, while γ defines the non-linear mapping from the input space to some high-dimensional feature space. Several studies have illustrated that these two parameters have an important effect on KELM's performance, similar to that in SVM. Thus, these two key parameters must be properly set prior to its application to realistic problems. These parameters are traditionally obtained using the grid-search method whose main drawback, however, is that it is easy to be trapped in a local optimum. Presently, it has been shown that biologically-inspired methods (such as the genetic algorithm (Liu et al., 2014), particle swarm optimization (PSO) (Zhang and Yuan, 2015), and artificial bee colony (Ma et al., 2016) are more likely to find the global-best solution than the grid-search method. As a new member in the nature-inspired methods, Grey Wolf Optimizer (GWO) (Mirjalili et al., 2014) mimics the social hierarchy and hunting behavior of grey wolves in nature. The main traits of GWO are social hierarchy, encircling prey, hunting, attacking prey (exploitation), and search for prey (exploration).

Due to its good search ability, GWO has been applied in a various fields. Muangkote et al. (2014) used the GWO with improvements to training q-Gaussian Radial Basis Functional-link nets neural networks. The experimental result indicated that the proposed algorithm obtained competitive performance comparing with other meta-heuristic methods. Komaki and Kayvanfar (2015) successfully applied GWO for the two-stage assembly flow shop scheduling problem with release time to greatly improve the efficiency. Sulaiman et al. (2015) used GWO to solve optimal reactive power dispatch problem. Mirjalili (2015) employed GWO to train multi-layer perceptron and eight standard datasets including five classification and three function-approximation datasets ware evaluated. The results demonstrated that a high level of accuracy in classification and approximation of the proposed trainer could be obtained. However, to the best of our knowledge, the potential of GWO has not been explored to fine tune the optimal parameters appeared in KELM. Therefore, this study aims at exploring the GWO technique's ability to address KELM's model selection problem for classification, and further applying the resulted model GWO-KELM to successfully and effectively predict company bankruptcy. For verification purpose, the effectiveness and efficiency of the proposed GWO-KELM is compared against the common methods such as grid-search optimized KELM (GS-KELM), genetic algorithm optimized KELM (GA-KELM), particle swarm optimization optimized KELM (PSO-KELM) and other four advanced machine learning methods including original ELM, self-adaptive evolutionary extreme learning machine proposed by Cao et al. (2012) (SaE-ELM), SVM and RF on the real-life financial dataset. All methods are compared in terms of the training accuracy, the validation accuracy and test accuracy, Type I error, Type II error and the area under the receive operating characteristic curve (AUC) criterion. For the stability of the results, the cross validation (CV) strategy is also adopted including external 10-fold CV and the inner 5-fold CV. The experimental results show that our proposed methodology, GWO-KELM, performs better when compared with some other well-known common methods. The main contribution of this study can be summarized as follows: