Expert Systems with Applications 39 (2012) 12605–12617

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment

Stjepan Oreski [a,*], Dijana Oreski [b], Goran Oreski [a]

[a] Bank of Karlovac, I.G.Kovacica 1, 47000 Karlovac, Croatia
[b] Faculty of Organization and Informatics, University of Zagreb, Pavlinska 2, 42000 Varaždin, Croatia

ARTICLE INFO

Keywords:
Classification
Credit scoring
Neural network
Genetic algorithm
Feature selection

ABSTRACT

The databases of the banks around the world have accumulated large quantities of information about clients and their financial and payment history. These databases can be used for the credit risk assessment, but they are commonly high dimensional. Irrelevant features in a training dataset may produce less accurate results of classification analysis. Data preprocessing is required to prepare the data for classification to increase the predictive accuracy. Feature selection is a preprocessing technique commonly used on high dimensional data and its purposes include reducing dimensionality, removing irrelevant and redundant features, facilitating data understanding, reducing the amount of data needed for learning, improving predictive accuracy of algorithms, and increasing interpretability of models. In this paper we investigate the extent to which the total data, owned by a bank, can be a good basis for predicting the borrower's ability to repay the loan on time. We propose a feature selection technique for finding an optimum feature subset that enhances the classification accuracy of neural network classifiers. Experiments were conducted on the credit dataset collected at a Croatian bank to assess the accuracy of our technique. We found that the hybrid system with genetic algorithm is competitive and can be used as feature selection technique to discover the most significant features in determining risk of default.

1. Introduction

The credit crisis, which began in July 2007, has shaken financial markets, undermined consumer and investor confidence, raised serious concerns and fears of financial institutions about the stability of financial markets in general, and was threatening economies around the world. While this crisis had many causes, it is clear now that banks, governments and others institutions can do more to prevent many of these problems in the future.

In this context, Basel Committee's response to the crisis is a comprehensive set of reform measures, to strengthen the regulation, supervision and risk management of the banking sector. These measures form a new international regulatory framework for banks – "Basel III".[1] The reforms target (BIS, 2011):

- Bank-level, or microprudential, regulation, which will help raise the resilience of individual banking institutions to periods of stress.
- Macroprudential, system wide risks that can build up across the banking sector as well as the procyclical amplification of these risks over time.

Complementary measures are taken at a bank level and the system as a whole, because the greater resilience at the individual bank level reduces the risk of system wide shocks, and vice versa.

Regulation, on the one hand, competition even more so on the other, is forcing banks to apply advanced methods in risk management. Competition has reduced the interest margin to a level that the bank can be successful only if there are no unexpected losses. Similarly, the space for acquiring additional first-class collateral is increasingly narrow. The client is willing to provide additional first-class collateral only when the loan cannot be realized in other banks without such insurance. In these circumstances, bank management was forced to seek new solutions for their business, which will have, at the same time, more flexibility and sensitivity to risk. Therefore, we can observe the risk assessment is probably the most important and most difficult segment of banking operations, and bank management the most responsible in the prevention of the aforementioned problems.

Space for action at the level of banks has since Basel II allowed banks to measure credit risk using internal ratings based (IRB) approach in order to determine the capital level. In taking this step

* Corresponding author. Tel.: +385 98 246 328; fax: +385 47 614 306.
E-mail addresses: stjepan.oreski@kaba.hr (S. Oreski), dijana.oreski@foi.hr (D. Oreski), goran.oreski@kaba.hr (G. Oreski).

[1] The Basel Committee's oversight body – the Group of Central Bank Governors and Heads of Supervision (GHOS) – agreed on the broad framework of Basel III in September 2009 and the Committee set out concrete proposals in December 2009. These consultative documents formed the basis of the Committee's response to the financial crisis and are part of the global initiatives to strengthen the financial regulatory system that have been endorsed by the G20 Leaders. The GHOS subsequently agreed on key design elements of the reform package at its July 2010 meeting and on the calibration and transition to implement the measures at its September 2010 meeting (BIS, 2011). Basel III is part of the Committee's continuous effort to enhance the banking regulatory framework. It builds on the International Convergence of Capital Measurement and Capital Standards document (Basel II).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.05.023

(BIS, 2006), the Basel Committee is also putting forward a detailed set of minimum requirements designed to ensure the integrity of these internal risk assessments. In order for internal risk assessments systems to ensure the integrity, banks have to collect data from many sources on daily bases, and use it in the evaluation of loan applicants and in regular bases classification of its own clients. A regulatory requirement was made for banks to use sophisticated credit scoring models for enhancing the efficiency of capital allocation (Khashman, 2010). Consequently, a robust and systematic credit scoring model and a loan evaluation model is important in realizing the IRB approach. The use of a classification method depends on the complexity of the institution, the size and the type of the loan, and represents an important area of study credit scoring system.

Current credit scoring models can be categorized into two major approaches (Li, Shiue, & Huang, 2006): specialized judgment and statistical modelling. The former relies on the expertise and tacit knowledge of specialists, which leads to financial experts' fatigue, misjudgment and slow response since the assessing process is usually time-consuming and laborious. The latter approach, on the other hand, can reduce such overheads due to its objectivity and consistence in nature.

Nowadays, financial institutions and researchers have developed many different quantitative credit scoring models. Šušteršic, Mramor, and Zupan (2009) have classified quantitative credit scoring models as follows: based on classical statistical methods, and based on artificial intelligence. Classical statistical methods are linear discriminant analysis, linear regression, logit, probit, tobit, binary tree and minimum method. The two most commonly used are discriminant analysis (DA) and logistic regression. Malhotra and Malhotra (2003) state that discriminant analysis suffers from bias of extreme data points, multivariate normality assumption, and equal group covariance assumptions. None of these restrictions apply to neural network models. Šušteršic et al. (2009) state that the weakness of the linear discriminant analysis is the assumption of a linear relationship between variables, which is usually nonlinear and the sensitivity to deviations from the multivariate normality assumption. Logistic regression does not require the multivariate normality assumption. Because of the linear relationship between variables both DA and logistical regression are reported to have a lack of accuracy.

There are also more sophisticated models known as artificial intelligence: expert systems, fuzzy systems, neural networks and genetic algorithms. Among these the neural networks are the possible alternative to the DA and logistic regression due to the possible complex nonlinear relationship between variables. In the literature in most cases of credit scoring problems the neural networks are more accurate than DA and logistic regression (Šušteršic et al., 2009). However, a large numbers of parameters, such as network topology, learning rate and training methods, have to be fine-tuned before the neural networks can be deployed successfully. Furthermore, drawbacks like trapping into local optimum, overfitting, and requiring huge time in learning computation tend to occur (Malhotra & Malhotra, 2003).

From the available literature, Khashman (2010) deduces that using neural networks for credit scoring and a loan evaluation has been effective over the past decade. The capability of neural networks, based on the back propagation learning algorithm, in such applications is due to the way the network operates, and the availability of training data. When feeding the information of a credit applicant to the neural network, variables are taken as input to the neural network and a linear combination of them is taken with arbitrary weights. The variables are linearly combined and subject to a non-linear transformation represented by a certain activation function (most frequently sigmoid function), then fed as inputs into the next layer for similar manipulation. The final function yields values which can be compared to a desired value. Each training case is submitted to the network, the final output compared with the observed value and the difference, the error, is propagated back through the network and the weights modified at each layer according to the contribution each weight makes to the error value. In essence, the network takes data, transforms it using the weights and activation functions into hidden value space and then possibly into further hidden value space; if further layers exist, and eventually into output layer space which is linearly separable.

Within the academic community there has been a growing body of literature on the application of many different methods for credit scoring and loan evaluation. There is no consensus as to which method a model developer should adopt for a given problem. Given this uncertainty, it is not unusual for a practitioner to construct several classifiers using different techniques, and then choose the one that yields the best solution for their problem. However, when comparing classifiers, it does not necessarily follow that the best classifier overall, outperforms all others throughout the regions of the problem domain. Consequently, error rates can often be reduced by combining the output of several classifiers. The research of classifier combination is rich in much of the relevant literature (Finlay, 2011; Twala, 2010), and represents another important area of the study of the credit scoring system.

In our opinion, the single classifier represents the first, combination of classifiers represents the second and the input data represents the third important area of credit scoring system study. Researchers did not regard the selection of variables as a crucial step of model development, possibly due to the problem of data availability. Hence, the issue of variable selection is a crucial and a challenging problem to solve before different credit scoring techniques are used to develop the best performing model (Šušteršic et al., 2009). As it is known, different variable selection techniques give different results on the same dataset. In this paper, we aim to design a hybrid system with genetic algorithm and artificial neural networks (GA-NN) for finding an optimum feature subset at retail credit risk assessment that enhances the classification accuracy of neural network classifier. We examine various combinations of the input data in terms of their contribution to correct classification of the credit applicant from the aspect of credit risks.

The remaining sections of this paper are organized as follows. Section 2 describes the problem of consumer loans to be studied in the paper and reviews the previous literature related to the problem. A brief overview of techniques and concepts used in the research is given in the third section. Section 4 describes the experimental design for data collection, feature selection, classification, performance evaluation and comparison. Section 5 discusses the experimental analysis and results that focus on prediction accuracy and misclassification costs. Section 6 concludes this paper and gives some guidelines for future work.

2. Problem statement and literature review

According to BIS (2006), credit risk is most simply defined as the potential that a bank borrower or the counterparty will fail to meet their obligations in accordance with the agreed terms.

The accuracy of the forecasts of a good or bad customer in terms of credit risk can be improved by: a good selection of input data, using the best methods of classification and combining the results of different classification methods. Until a few years, the body of research on consumer credit risk measurement was quite sparse. Quantitative consumer credit scoring models were developed much later than those for business credit, mainly, due to the problem of availability of data. Data were limited to the own databases of financial institutions. Nowadays, some data are publicly avail-