Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran

Eghbal Rahimikia,*, Shapour Mohammadi, Teymur Rahmani, Mehdi Ghazanfaria

a School of Industrial Engineering, Iran University of Science & Technology, Tehran 16846-13114, Iran
b Faculty of Management, University of Tehran, Tehran 14155-6211, Iran
c Faculty of Economics, University of Tehran, Tehran 14155-6445, Iran

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form 25 November 2016
Accepted 18 December 2016
Available online xxxx

Keywords:
Corporate tax evasion detection
Data mining
Hybrid intelligent system
Support vector machine
Neural network
Harmony search

ABSTRACT

This paper concentrates on the effectiveness of using a hybrid intelligent system that combines multilayer perceptron (MLP) neural network, support vector machine (SVM), and logistic regression (LR) classification models with harmony search (HS) optimization algorithm to detect corporate tax evasion for the Iranian National Tax Administration (INTA). In this research, the role of optimization algorithm is to search and find the optimal classification model parameters and financial variables combination. Our proposed system finds optimal structure of the classification model based on the characteristics of the imported dataset. This system has been tested on the data from the food and textile sectors using an iterative structure of 10-fold cross-validation involving 2451 and 2053 test set samples from the tax returns of a two-year period and 1118 and 906 samples as out-of-sample using the tax returns of the consequent year. The results from out-of-sample data show that MLP neural network in combination with HS optimization algorithm outperforms other combinations with 90.07% and 82.45% accuracy, 85.48% and 84.85% sensitivity, and 90.34% and 82.26% specificity, respectively in the food and textile sectors. In addition, there is also a difference between the selected models and obtained accuracies based on the test data and out-of-sample data in both sectors and selected financial variables of every sector.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Tax returns contain useful information for detecting tax evasion. However, much of this information is voluminous and complex. Therefore, it is necessary to use mathematical and statistical models to analyze this information. Because of the costs associated with auditing individual companies’ tax returns, finding models that accurately identify misstated returns is important. Consequently, the goal for this data analysis is to help tax auditors and tax authorities to detect companies with a high probability of misstated activities in order to give these suspect companies a more detailed audit. By using these models, tax authorities can significantly increase tax income and decrease the human resource costs associated with manually auditing tax returns.

* Corresponding author.
E-mail address: erahimi@ut.ac.ir (E. Rahimikia).
Due to voluminous and complex data sets, tax authorities in many countries have started adopting new systems based on data mining (DM) and artificial intelligence (AI) to help them to detect misstatements in financial statements. Because of this, academic research in this area (specially published research reports for tax evasion detection systems) is lagging behind and the need for more attention to this topic by the academia is especially felt more than ever. Generally, tax evasion detection, used by the tax authorities to detect tax returns suspected of tax evasions, is somewhat similar to financial statement fraud detection.

Tax administrations have to deal with a variety of risks, such as the risk of non-compliance, the risk of tax evasion and the risk of insolvency by the taxpayer. Compliance risk management allows us to deal with these risks by looking at the behavior of taxpayers. For these reasons, intelligent predictive models to identify tax returns for additional examination can be used as a tool to increase the accuracy and efficiency of auditing.

The Iranian National Tax Administration (INTA) is a governmental institution and is established under the supervision of Iran’s ministry of economic affairs and finance. INTA provides full support for administering tax reforming plans and executing tax collection legal procedures in an efficient manner. In the future, its responsibilities will also extend to the monitoring of tax law enforcements and regulations and the creation of a proper basis to achieve tax objectives.

In this paper, we present a new hybrid intelligent system to detect corporate tax evasion in INTA. Our focus is to compare the effectiveness of using a system that combines the multilayer perceptron (MLP) neural network, support vector machine (SVM), and logistic regression (LR) classification models with the harmony search (HS) optimization algorithm to detect corporate tax evasion separately in two sectors. The HS optimization algorithm optimizes the selection of financial variables in addition to model parameters; so an optimized model with acceptable accuracy in terms of a combination of optimal financial variables and model parameters can be developed.

The hybrid model was developed and tested with data from the food and textile sectors. This paper is among the first research based on actual data with a large sample set of tax returns from different sectors operating in Iran. To gain a good level of confidence in our results, the output from our model was evaluated against out-of-sample data (in addition to test data and total data). Furthermore, cross-validation was repeated 8 times using parallel computing to improve the reliability of the results. Most of the articles published in this or related areas mainly focus on binary results instead of probability results. However, the binary results are not really helpful in practice. Therefore, all classification models presented in this proposed system generate the probability of membership in each of the two classes and for reporting purposes, the outputs are sorted based on the obtained probabilities. This allows auditing of companies with higher probabilities to be done with more scrutiny. Furthermore, comparing the selected financial variables based on a number of repetitions in 10 system runs for the food and textile sectors is helpful to answer the question of whether tax evasion behavior patterns are different across sectors.

This paper is structured as follows: First intelligent systems for fraud detection is presented in Section 2, then tax evasion detection algorithms that were used are described in Section 3, the hybrid intelligent system structure for the corporate tax evasion detection is explained next in Section 4, method is presented in Section 5 followed by the results and discussion in Section 6 and finally Section 7 provides conclusions and future research directions of this work.

2. Intelligent systems for fraud detection

The current published research mainly focuses on financial statement fraud detection. Relatively little published academic work is available on tax evasion detection. This is due to the limited access to datasets at the national level. It should be noted that in terms of modeling frameworks and applied methods, tax evasion detection and financial statement fraud detection have a lot in common.

Krieger (1996) mentions an example from the USA’s Internal Revenue Service (IRS) which uses neural networks and polynomial regression to pinpoint potential tax non-compliance cases. Their tools focused on feature classification and extraction, descriptive statistics, polynomial networks and clustering algorithms. Through these tools, the database variables were reduced from 150 to 8. The accuracy of the polynomial network was confirmed by other neural networks (backpropagation neural network (BPN), learning vector quantization (LVQ) and self-organizing map (SOM)). Fischthal (1998) registered a patent in the USA for detecting fraud using a neural network. The architecture of this system involves first employing a conceptual clustering technique to generate a collection of classes from historical data. Neural networks are provided for each class created by the clustering step and the networks are trained using the same historical data. Yu et al. (2003) present a system architecture for the fraudulent tax declaration detection problem of Chinese commercial enterprises that includes communication with domain experts, choice of the core data mining algorithm, design guidelines of the data mining system architecture and incorporation of domain experts’ knowledge. Their decision tree implementation results achieved an 85–90 % accuracy rate.

Gupta and Nagadevara (2007) implement 8 models based on different combinations of DT, Discriminant model and LR in India Value-Added Tax (VAT) from 2003 to 2004 and a sample containing 402 dealers. They found that all models developed through data mining techniques were better than random selection. Wu et al. (2012) design a VAT evasion detection model in Taiwan by association rule mining (IBM DBMiner 2.0 tool) from 2003 to 2004. The results show that the designed model enhanced the detection of tax evasion and therefore can be employed to effectively reduce or minimize losses from VAT evasion. Hsu et al. (2015) present a case study of a pilot project that was developed to evaluate the use of data mining in audit selection for the Minnesota Department of Revenue (DOR) for Sales and Use Tax. Researchers used a combination of C4.5 DT, Naive Bayes, MLP neural network, SVM and other techniques to build classification models using 10943 samples from 2004–2006 as training
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه‌شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات