Automation in Construction 20 (2011) 935–939

Contents lists available at ScienceDirect

Automation in Construction

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

Application of neural networks for detecting erroneous tax reports from construction companies

Jieh-Haur Chen [a,*], Mu-Chun Su [b], Chang-Yi Chen [a], Fu-Hau Hsu [b], Chin-Chao Wu [b]

[a] Institute of Construction Engineering and Management, National Central University, Jhongli, Taoyuan 32001, Taiwan
[b] Department of Computer Science and Information Engineering, National Central University, Jhongli, Taoyuan 32001, Taiwan

ARTICLE INFO

Article history:
Accepted 26 March 2011
Available online 14 April 2011

Keywords:
Neural networks
Pattern classification
Tax report
Construction company

ABSTRACT

In this study we develop an automatic detection model for discovering erroneous tax reports. The model uses a variety of neural network applications inclusive of the Multi-Layer Perceptrons (MLPs), Learning Vector Quantization (LVQ), decision tree, and Hyper-Rectangular Composite Neural Network (HRCNN) methods. Detailed taxation information from construction companies registered in the northern Taiwan region is sampled, giving a total of 5769 tax reports from 3172 construction companies which make up 35.98% of the top-three-class construction companies. The results confirm that the model yields a better recognition rate for distinguishing erroneous tax reports from the others. The automatic model is thus proven feasible for detecting erroneous tax reports. In addition, we note that the HRCNN yields a correction rate of 78% and, furthermore, generates 248 valuable rules, providing construction practitioners with criteria for preventing the submission of erroneous tax reports.

1. Introduction

The problem of submission by construction companies of erroneous tax reports to tax bureaus is critical. According to official information from the National Tax Administration of Northern Taiwan (NTANT), errors have been found in almost half to the total tax reports filed in recent years [1]. The government considers this behavior as tax evasion and it definitely affects government operations. It is also dangerous for the construction companies since those companies found guilty of tax evasion are subject to penalties, which can be serious enough to force them out of business. Detecting erroneous tax reports in the initial phase can be regarded as a two-class pattern recognition problem because of only two kinds of results which are "yes" for erroneous tax report and "no" for correct tax report; however, such detecting work in practice is normally dependent on experienced personnel. In their examination of tax reports, professional auditors look for the appearance of any suspicious values in the accounting, followed by comprehensive inspection to specify exact errors if any. Suspicious values vary widely, making such inspection an exhaustive task. The detection process is tedious and costs both companies and government a tremendous amount of money and manpower. In addition, auditing personnel may need the specification of solid upper and lower bounds for those accounts to facilitate the work. Developing an automatic model to solve the aforementioned problems is the motivation for this work. The study objective is to develop an automatic model for construction practitioners that will (1) save time and manpower from the wearisome task of manually reviewing massive tax reports, and (2) provide upper and lower bounds for each account to facilitate auditing. The research scope is limited to corporate tax information which is usually confidential and related to corporate privacy. Thus, there are two compulsory criteria to be met for the protection of the taxpayers: the selection of past taxation data that is at least 10 years old, and preservation of anonymity.

2. Applications of pattern classification in construction

One of the fundamental aspects of pattern recognition lies in determining which is featured should be employed for the best classification results. Next, an effective classifier is desirable. Construction practitioners and researchers have been able to solve realistic problems using the pattern classification concept for years. Chen utilized the k-mean Nearest Neighbor (kNN) approach to establish a knowledge-sharing model, which provides detailed resolutions as determined by the courts for construction disputes [2]. He also developed a hybrid model using the methods based on Artificial Neural Networks (ANN) and Case Based Reasoning (CBR) to predict the likelihood of litigation in cases of severe disputes by change orders [3]. The Ant Miner tool, integrated with rule-based classification extraction is introduced to predict the outcome of construction litigation [4]. Predicting performance for construction companies and projects is

* Corresponding author. Tel.: +886 3 4227151x34112; fax: +886 3 4257092.
E-mail address: jhchen@ncu.edu.tw (J.-H. Chen).

0926-5805/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.autcon.2011.03.011