Advances in Economics and Business Management (AEBM)
Print ISSN: 2394-1545; Online ISSN: 2394-1553; Volume 1, Number 3; November, 2014 pp. 226-228

http://www.krishisanskriti.org/aebm.html

Application of Data Mining Techniques to Detect and Predict Accounting Fraud: A Comparison of Neural Networks and Discriminant Analysis

Mohammad Mehdi Nasrizar

Research Scholar of Banaras Hindu University Varanasi
E-mail: mehdi.nasrizar@yahoo.com

Abstract: *Accounting information systems enable the process of internal control and external auditing to provide a first line defense in detecting fraud (Turpen and Messina, 1997). There are few valid indicators at either the individual or the organizational level which are reliable indicators of fraud prevention (Groveman, 1995). Recent studies have shown that it is nearly impossible to predict fraud. In fact, many of the characteristics associated with white-collar criminals are precisely the traits which organizations look for when hiring employees (Lord, 1997). This paper proposes the use of information systems to deal with fraud through proactive information collection, data mining, and decision support activities.*

1. INTRODUCTION

ACCOUNTING FRAUD

Bookkeeping and record keeping methods were created during ancient and medieval times. the concept of double entry accounting began in the 14th century (Funk and Wagnall, 1994). While the concepts of accounting method rules and laws have changed significantly, one principle has remained constant. Accounting's primary purpose is to keep track of money and other assets (AICPA, 1997).

An accountant's first priority is to track all aspects of an organization's financial elements. Guidelines, rules, procedures and laws dictate the accounting profession. It is assumed that it is the duty of the accountant to insure that the financial statements provided are an accurate view of the firm. It is also assumed that it is the auditor's responsibility to detect fraudulent behavior(James A. Rodger).

Unfortunately, auditor's assumed that their responsibility was to detect material misstatements within their client's financial statements, not to detect fraud per se. This difference in opinion has been labeled the "expectation gap" and it is used to describe the difference between what auditors assume their responsibility to be and what the public perceives it to be (AICPA, 1998).

In an effort to reduce the "expectation gap" the Accounting Standards Board (ASB) issues Statements on auditing Standards (SAS) which are serially numbered pronouncements which interpret the auditing standards that accountants are mandated to follow. Specifically, SAS #53 and SAS #82 are the important statements regarding fraud detection (AICPA, 1998).

SAS #82 was issued in February of 1997 and is effective for audits of financial statements for periods ending on or after December 15, 1997. Prior to SAS #82, SAS #53 dealt with finding "errors and irregularities" in financial statements. SAS #53 defines errors simply as mistakes and says that irregularities include both fraudulent financial reporting and misappropriation of assets. However, SAS #82 provides an expanded description of fraud, and covers both fraudulent financial reporting and misappropriation of assets.

The ASB considers the detection responsibility in SAS #82 to be the same as in SAS #53. However, the detection responsibility in SAS #82 has been clarified to use the term "fraud" rather than the term "irregularities". In addition, SAS #82 also covers both audit planning and performance and provides auditors with additional operational guidance on the consideration and detection of material fraud in conducting financial statement audits.

2. DATA MINING TECHNIQUES

Most common data mining algorithms in use today , have broken the discussion into two sections, each with a specific theme:

2.1: Classical Techniques: Statistics, Neighborhoods and Clustering

2.2: Next Generation Techniques: Trees, Networks and Rules

Each section will describe a number of data mining algorithms at a high level, focusing on the "big picture" so that the reader will be able to understand how each algorithm fits into the

landscape of data mining techniques. Overall, six broad classes of data mining algorithms are covered. Although there are a number of other algorithms and many variations of the techniques described, one of the algorithms from this group of six is almost always used in real world deployments of data mining systems.

What is different between statistics and data mining?

There are several reasons. The first is that the classical data mining techniques such as CART, neural networks and nearest neighbor techniques tend to be more robust to both messier real world data and also more robust to being used by less expert users. But that is not the only reason. The other reason is that the time is right. Because of the use of computers for closed loop business data storage and generation there now exists large quantities of data that is available to users. IF there were no data - there would be no interest in mining it. Likewise the fact that computer hardware has dramatically upped the ante by several orders of magnitude in storing and processing the data makes some of the most powerful data mining techniques feasible today.

What is a Neural Network?

When data mining algorithms are talked about these days most of the time people are talking about either decision trees or neural networks. Of the two neural networks have probably been of greater interest through the formative stages of data mining technology.

To be more precise with the term "neural network" one might better speak of an "artificial neural network". True neural networks are biological systems (a k a brains) that detect patterns, make predictions and learn. The artificial ones are computer programs implementing sophisticated pattern detection and machine learning algorithms on a computer to build predictive models from large historical databases. Artificial neural networks derive their name from their historical development which started off with the premise that machines could be made to "think" if scientists found ways to mimic the structure and functioning of the human brain on the computer.

Don't Neural Networks Learn to make better predictions?

Because of the origins of the techniques and because of some of their early successes the techniques have enjoyed a great deal of interest. To understand how neural networks can detect patterns in a database an analogy is often made that they "learn" to detect these patterns and make better predictions in a similar way to the way that human beings do.

Are Neural Networks easy to use?

A common claim for neural networks is that they are automated to a degree where the user does not need to know that much about how they work, or predictive modeling or even the database in order to use them.

Just the opposite is often true. There are many important design decisions that need to be made in order to effectively use a neural network such as:

- How should the nodes in the network be connected?
- How many neuron like processing units should be used?
- When should "training" be stopped in order to avoid over fitting?

There are also many important steps required for preprocessing the data that goes into a neural network - most often there is a requirement to normalize numeric data between 0.0 and 1.0 and categorical predictors may need to be broken up into virtual predictors that are 0 or 1 for each value of the original categorical predictor.

Applying Neural Networks to Business

Neural networks are very powerful predictive modeling techniques but some of the power comes at the expense of ease of use and ease of deployment. As we will see in this section, neural networks, create very complex models that are almost always impossible to fully understand even by experts**.**

Predictive Discriminate Analysis (PDA):

PDA is a predictive classification technique deals with a set of multi-attributes and one classification variable, the latter being a grouping variable with two or more levels. Predictive discriminate analysis is similar to multiple regression analysis except that PDA is used when the criterion variable is categorical and nominally scaled. As in multiple regression, in PDA a set of rules is formulated which consists of as many linear combinations of predictors as there are categories, or groups.A PDA is commonly used for classifying observations to pre-defined groups based on knowledge of the quantitative attributes.

3. OBJECTIVES

Traditionally, audits have been a team effort in which new members learn from older members. Accounting information systems can be used to reverses these roles because the younger auditors are more at ease with using technology. Santhanan and Sein (1994) argue that the combination of effective hands-on training with theory helps end- users to develop good mental models of systems. DeSanctis and Jackson (1994) believe that end users, which get good user support services, should be more productive if the support services are planned around the user's (auditor) specific needs.

During the auditing process, auditors select audit procedures that are easiest to perform which take the least amount of time (McMillan and White, 1993). In reality, there are no auditing software applications for detecting and preventing fraud. Auditors may revert to word processors, spreadsheets, and calculators to decipher the firm's myriad of accounting information. However, more advanced information technology

methods are available for deterring fraud. For example, artificial neural networks (ANN) could be used for data mining and detecting key indicators of fraud, decision support systems (DSS) could be used for decision making, expert systems (ES) could be used for rule bases and discriminant analysis could be used to predict fraud. Rarely, if ever are these tools employed.

The objective of this research is to apply several of these advanced methods in an effort to detect and predict fraudulent behavior. Discriminant analysis will be applied to survey data in order to predict fraud according to key indicators such as poor internal controls, weak ethics policies, changes in employee lifestyles, working conditions, morale, and downturns in the economy (Turpen and Messina, 1997). ANN will then be used to data mine which factors are indicative of fraud. A comparison will then be made between the two methodologies.

4. RESULTS

In order to more closely approximate the training and testing methodology utilized by neural networks, the jackknife method of discriminant analysis was utilized. This "leave- one-out" principle is a sophisticated method based on estimation with multiple subsets of the sample. In other words, the discriminant function is fitted to repeatedly drawn samples of the original sample. The jackknife method yielded 50.4% of original grouped cases correctly classified.

In comparison, the neural network method classified 75.9 % good parts. The training set consisted of 200 respondents. The data was normalized, and ran for one hour and twenty minutes before it resulted in 100 good parts. A weight was obtained, and the network was saved. The testing data set (composed of 211 respondents) was then placed into the saved network and run. After twenty-four hours, the neural network had achieved 75. 9% good parts.

5. CONCLUSIONS AND RECOMMENDATIONS

Discriminant analysis yielded 50.4% of original grouped cases correctly classified. No significant relationship was found (.149) between attitude, morale, internal controls, increases in expenditures and whether or not fraud was actually committed. Cronbach's Alpha of reliability was .6626 and offered somewhat reliable results in this exploratory research.

Neural networks did a much better job of predicting fraud (75.9%) good parts than discriminant analysis (50.4%). Neural networks were able to find patterns in the training set and then correctly identify more than three fourths of similar patterns in the testing set. Therefore, it can be concluded that neural networks outperform discriminant analysis by 25.5% in this data set.

6. LIMITATIONS AND FUTURE DIRECTIONS

The survey asks the respondent to play two roles. One role is that of an auditor from outside the organization. The second role asks the respondent to assess the environment within the company. The result is respondent confusion. There may also be a degree of social desirability in which the accountant filling out the survey may not wish to divulge information because they may be afraid of managerial repercussions.

In the future, it may be worthwhile to create different surveys for different types of accountants (internal, external, auditors, etc.) and then to tailor the questions in a manner that better identifies their perceptions of fraud. The present survey confuses the responding accountants as to whether the survey should be answered from an internal or external perspective.

7. ACKNOWLEDGEMENTS

This research was made possible by a grant from the Small Grants Program, Central Research Development fund at the University of Pittsburgh. We would like to extend a special thanks to Tracy Kaiser and Michelle Callihan for assistance with the data entry and statistical analysis. Acknowledgement also is in order to the members of the Pennsylvania Institute of Certified Public Accountants (PICPA) for their cooperation in gathering the data for this research.

REFERENCES

[1] George C. J. Fernandez Discriminant Analysis, A Powerful Classification Technique in Data Mining George C. J. Fernandez

[2] E.W.T. Ngai a The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature

[3] A. Agresti, Categorical Data Analysis, Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1990.

[4] M. Agyemang, K. Barker, R. Alhajj, A comprehensive survey of numeric and symbolic outlier mining techniques, Intelligent Data Analysis 10 (6) (2006) 521–538.

[5] S.R. Ahmed, Applications of data mining in retail business, International Conference on Information Technology: Coding and Computing 2 (2) (2004) 455–459.

[6] M. Artís, M. Ayuso, M. Guillén, Modeling different types of automobile insurance fraud behaviour in the Spanish market, insurance, Mathematics and Economics 24 (1) (1999) 67–81.

[7] M. Artís, M. Ayuso, M. Guillén, Detection of automobile insurance fraud with discrete choice models and misclassified claims, The Journal of Risk and Insurance 69 (3) (2002) 325–340.

[8] J.A. Atwood, J.F. Robinson-Cox, S. Shaik, Estimating the prevalence and cost of yield-switching fraud in the federal crop insurance program, American Journal of Agricultural Economics 88 (2) (2006) 365–381.

[9] James A. Rodger: Utilization Of Data Mining Techniques to Detect and Predict Accounting fraud

[10] Alex Berson, Stephen Smith, and Kurt Thearling: Building Data Mining Applications for CRM