vol.1, No.2, 28-36 ISSN 1805-3602

Finding rules for audit opinions prediction through data mining methods

Seyed Mojtaba Saif[1], Mehdi Sarikhani[2], Fahime Ebrahimi[2]

[1]Department of Computer, Safashahr Branch, Islamic Azad University, Iran
[2]Department of Accounting, Safashahr Branch, Islamic Azad University, Iran

Received for publication August 31, 2012.
Accepted for publication September 16, 2012.

Abstract

Nowadays data mining, which is used in various accounting and financial applications, has received a great deal of attention. One of these applications is predicting and identifying the audit opinion type. The objective of research is to help auditors identify audit opinions by using a support vector machine from data mining methods. The system receives the data from financial reports and identifies the type of audit opinions. This approach combines support vector machine with a decision tree that can understand and interpret the obtained results. In this paper, a novel approach for rule extraction from support vector machine and decision tree is presented and its application is shown in the prediction of audit opinions. The research result is 30 rules that predict the audit opinions.

Keywords: Audit opinions, data mining, support vector machine, artificial neural networks, decision tree.

Introduction

In recent years, the qualitative growth and the increasing complexities of economic activities have caused financial information to play a significant role in evaluating entities insofar as the availability of reliable financial information is regarded as a requisite for the survival of society.

The investors, creditors, governments, and other users have been relying on financial information provided by company managers that will enable them to adopt reasonable decisions. In some circumstances, a conflict will arise between the purposes the providers of this information are following and those of their users. This can be regarded as a motive behind implementing auditing as an instrument for enhancing the ability of relying on financial reports presented by companies.

In Iran, all companies listed on the Tehran Stock Exchange are responsible to provide financial reports according to Iranian Generally Accepted Accounting Principles. These reports should be audited yearly by registered auditors of the Tehran Stock Exchange. These auditors must prepare reports that contain clear expressions of opinion about the financial statements.

An unqualified opinion is expressed when the financial statements «give a true and fair view» and have been prepared in accordance with the identified financial reporting framework. A qualified opinion is issued when either of the following two situations occurs. (A) There is a limitation to the scope of the auditors' examination that prevents them from obtaining sufficient evidence to express an unqualified opinion. (B) The auditors disagree with the treatment of the disclosure of a matter in the financial statements; in their judgment, the effect of the matter is or may be material to the financial statements, and therefore those statements may not or do not present a true and fair view of the matters on which the auditors are required to report or do not comply with relevant accounting or other requirements.

In recent years, auditing has been increasingly in demand with regard to the importance of quality and reliability of the audited financial reports for the purpose of optimized allocation of economic resources. The new development in the conceptual framework of auditing, a widespread use of information technology in commerce, and the creation of

Corresponding author: Mehdi Sarikhani, Department of Accounting, Safashahr Branch, Islamic Azad University, Safashahr, Iran. E-mail: mehdi_sarikhani@yahoo.com.

modern technologies and knowledge such as data mining have given rise to new challenges in auditing methods. Relevant research literature repeatedly recognizes the importance of the new technologies in auditing. For example, Koskivaara (2004) contends that the advanced tools of auditing can prevent the accounts from being manipulated by different companies, and it can help auditors to respond to today's demands by the business environment. Furthermore, an increasing number of frauds by managers have multiplied the use of modern auditing tools.

Bell and Tabor (1991) as well as Chen and Church (1992) note that auditors can use the output of such models to plan specific auditing procedures that can be applied to achieve an acceptable level of audit risk. These models can also be used as a quality control tool in the final or review stage of an engagement and for contingency analyses on how changes in specific variables could add or detract from the probability of obtaining a qualified opinion (Kleinman and Anandarajan, 1999).

Therefore these models of decision making for auditors contribute to presenting audit opinions. Auditors are able to screen many companies by making use of these models, and they can thus pay more attention to companies having a higher probability of receiving a qualified audit opinion leading to savings in time and money. Furthermore, the results of these models can play a significant role in evaluating potential clients, in peer reviews, in controlling quality, in predicting audit opinions under similar conditions, and in defense against lawsuits.

One of the methods used to predict the type of audit opinion by utilizing the new models is data mining. It is an umbrella term that includes methods used to extract human knowledge from data. There is a difference between the traditional way of data analysis and data mining. The former supposes that hypotheses are already constructed and validated against the data, whereas the latter supposes that the patterns and hypotheses are automatically extracted from the data.

In data processing methods, knowledge or hidden principle is extracted beyond the data to make different models for analyzing the data. Data mining includes techniques such as decision trees, neural networks, support vector machines, genetic algorithm, and rough sets.

The purpose of this research is to find rules and an intelligent machine that will help auditors submit audit reports by supporting vector machines and decision tree.

Previous research

Keasey *et al.* (1988) studied the effective variables on the audit report type by means of a logistic model, making use of a sample obtained from some small companies in England within the years 1980–1982. They reached this conclusion: There is an increase in the probability of receiving qualified audit reports by the companies under study that utilize big audit firms. They concluded that there is an increase in the probability of receiving qualified audit reports by companies that have an increase in secured loans, a decrease in profitability, or a high percentage of management shareholders.

Dopuch *et al.* (1987) studied the effective variables on the type of audit opinion by making use of a logistic model. The result showed a negative relationship between the probability of receiving a qualified report and variables, such as profitability, growth, share of equity in a balance sheet, and number of employees. They also concluded that there is an increase in the probability of receiving a qualified audit report when there is an increase in the amount of debt.

Spathis (2003) tested the extent to which combinations of financial and nonfinancial information can be used to enhance the ability to discriminate between the choices of a qualified or unqualified audit report. The data are taken from a sample of 100 Greek companies. Logistic and ordinary least squares (OLS) regression models were estimated to assess the effect of company litigation and financial information on audit qualification opinions. He found that the qualification decision is associated by financial information, such as financial distress, and by nonfinancial information, such as company litigation. The model developed is accurate in classifying the total sample correctly with a rate of 78%.

Spathis *et al.* (2003) modeled the auditors qualification using a multicriteria decision aid classification method (UTADIS-UTilites' Additives Discriminates) and compared it with other multivariate statistical techniques, such as discriminant and logit analysis. The qualification decision is explained by financial ratios and by nonfinancial information, such as client litigation. The developed models are accurate in classifying the total sample correctly with rates of almost 80%.

Doumpos *et al.* (2005) used support vector machine for the development of linear and nonlinear models that explain qualifications in audit reports, based on a large sample of 5,189 unqualified audit reports and 859 qualified audit reports from 1,754 large UK companies over the period 1998–2003. The nonlinear models (Radial Basic Functions (RBF) and quadratic kernels) were not found to provide improved results compared with the simpler linear models. Nevertheless, in all cases the results of the support vector machine models were found robust to different sizes of the training sample, and they were analyzed to investigate the relative importance of financial variables as opposed to a credit-rating variable.

Pasiouras *et al.* (2006) studied the potentials of developing multicriteria decision aid models for reproducing the auditors' opinion on the financial statements of the companies. A sample consists of 823 private and public manufacturing companies over the period 1998–2003. Discriminant analysis and logit analysis are also used for comparison purposes. The out-of-time and out-of-sample testing results indicate that the two multicriteria decision aid techniques achieve almost equal classification accuracies and are both more efficient than discriminant and logit analyses.

Gaganis *et al.* (2007a) explored the potential of using Probabilistic Neural Networks (PNNs) in developing a model for explaining qualifications in audit reports. The analysis was based on a large sample of UK-listed companies for the period 1997-2004. The results demonstrate the high explanatory power of the PNN model in explaining qualifications in audit reports. The model is also found to outperform traditional artificial neural network (ANN) models, as well as logistic regression. Sensitivity analysis is used to assess the relative importance of the input variables and to analyze their role in the auditing process.

Gaganis *et al.* (2007b) investigated the efficiency of k-nearest neighbors (k-NN) in developing models for estimating auditors' opinions, as opposed to models developed with discriminant and logit analyses. The sample consists of 5,276 financial statements of UK companies. They developed two industry- specific models and a general one using data from the period 1998-2001, which are then tested over the period 2002-2003. The comparison of methods revealed that the k-NN models can be more efficient, in terms of average classification accuracy, than the discriminant and logit models. Lastly, the results were mixed concerning the development of industry specific models, as opposed to general models.

Kirkos *et al.* (2007) employed three data mining classification techniques to develop models capable of identifying cases of qualified audit opinions. The three models have been proven capable of distinguishing the qualified cases. The decision tree model achieves the highest accuracy rate against the training set. They estimated the true predictive power of the models by using tenfold cross validation. According to these results, the Bayesian Belief Network achieves the highest classification accuracy (82.22%) of total observations). The multilayer perceptron model achieves a marginally lower performance (81.11%). The decision tree model achieved the lowest performance (77.69%).

Research Methodology

Sample

The support vector machine was trained and evaluated by training and training samples. Samples comprised 1018 observations for the years 2001 through 2007 that contained 708 qualified observations and 310 unqualified observations. After excluding the financial companies (i.e., banks and insurance companies), we selected the publicly listed firms that were qualified at least once over the 2001 to 2007 and closed their fiscal year on mid-March (end of Persian calendar).

The required data for examination were extracted from the information of market and financial statements. For this purpose, a large section of the information was extracted from the Tadbir Pardaz and RahAvard Novin softwares (two Iranian Softwares) and the rest of the information was extracted through the information database of the Islamic Studies and Research Management Center of the Tehran Stock Exchange.

Variables

To determine the effective variables on the identification and prediction of audit reports based on the results of previous research and position of Iran (Setayesh and Jamalianpoor, 2010; Poorheidari and Azami, 2011), some effective variables will be taken into consideration, including the index related to liquidity, profitability, efficiency of companies leverage, growth, company size, productivity of employees, financial distress, and cash flow. Each of these indicators has been measured by one or more variables.

In the present study, the researchers will implement current ratio and quick ratio to determine the liquidity position. The liquidity ratios are used as the main criteria for financial health. On one hand the high amount of liquidity because of the exaggerated amount of assets leads to an increase in the probability of a qualified report (Ireland, 2003). On the other, there will be an increase in the probability of receiving a qualified report because of a decrease in the financial health of a company (Spathis, 2003). Ireland (2003) came to this conclusion: In comparison with other companies, those having low liquidity and high risk that report material contingent liabilities or which do not pay dividends, receive a qualified report with higher probability. Laitinen and Laitinen (1998) concluded that there is no significant difference between liquidity of Finnish companies that receive qualified and unqualified reports.

The profitability ratios used in this research include earnings before interest and taxes margin, earnings before taxes margin, return on assets, return on equity, and return on capital employed. There have been many research projects whose results indicate that those companies receiving qualified reports or that have falsified financial statements have less amount of profitability in comparison to others. Moreover, as already mentioned, there will be an increase in the probability of a qualified audit report because of a decrease in the financial health of companies (Doumpos *et al.,* 2005; Pasiouras *et al.,* 2006). Spathis (2003) also believes that profitability of a company may be influenced by the utility of the management. Exaggerating the assets and revenue is considered one of the fraud methods in financial statements. Furthermore, some researchers contend that some managers might manipulate the inventory in a company (Doumpos *et al.,* 2005).

Furthermore, to study the efficiency and activities are managed in a company, we must recognize that some variables can play a significant role, including the debtors turnover, debtor collection period, net assets turnover, fixed assets turnover and inventory turnover. Spathis (2003) concluded that the activity ratios can be regarded as helpful variables in the prediction of the type of opinion expressed by the auditors.

A company's ability to respond to its obligations is evaluated by two criteria, the ratio of total debts to total assets and the ratio of the equity to long-term debts. A company's reliance on financing assets by means of liability may lead to its bankruptcy in comparison to other methods used for financing. Numerous studies indicate that the number of qualified reports received by companies with high percentages of bankruptcy risk is higher than those received by other companies (Spathis, 2003). Ireland (2003) concluded that British companies with high leverage receive a high probability of the going-concern and nongoing-concern modified audit reports in comparison to other companies. Similarly, Laitinen and Laitinen (1998) drew the conclusion that the greater amounts of shares of equity in a balance sheet, the higher the probability of receiving an unqualified report.

To measure the size of a company, the researcher makes use of the logarithms of the book value of assets, net sales, and number of employees. Beasley *et al.* (1999) concluded that companies committing fraud in the United States are generally small. In another study, Palmrose (1986) showed that the greater the size of a company, the greater the number of supervisory contracts and corporate governance. Therefore, auditors pay more attention to submitting the audit report. Ireland (2003) concluded that big companies have good accounting systems and internal controls with more probability, leading to a decrease in disagreements and limitations on scope.

The growth of a company is measured by considering the variable «change in total assets.» Laitinen and Laitinen (1998) reached the conclusion that the less the growth of a company, the more the probability of receiving a qualified report. According to Gaganis *et al.* (2007a), the productivity of employees is evaluated by considering the four criteria in this study, that is, working capital per employee, total assets per employee, net sales per employee, and profit per employee. Moreover, the dummy variable is used to study the effects of litigation. According to Spathis (2003), litigation can have an influence on receiving a qualified report.

Also, Z-score calculated by Kupaei (2008) is implemented to determine the financial distress of Iranian companies. The previous studies related to this area indicate that these companies, having a high probability of bankruptcy, are more probable to receive qualified reports because there is more doubt in their abilities to continue their activities (Doumpos *et al.*, 2005; Pasiouras *et al.*, 2007).

To study the cash flow, researchers are going to study the ratios of cash from operating activities to the sales and the cash from investing activities to the sales. Furthermore, to determine other effective variables, the researchers make use provision for staff termination benefits per employee and the ratios of tax payables and retained earnings to sales.

The list of the explanatory variables used to predict audit qualification is presented in Table 1.

Table 1. List of variables

y	Auditor's opinion	x_{15}	Fixed assets turnover
x_1	Z-Score*	x_{16}	EBIT margin
x_2	Log net sales	x_{17}	Earnings before tax margin
x_3	Log total assets	x_{18}	Cash from operating activities to net sales
x_4	Log number of employees	x_{19}	Cash from investing activities to net sales
x_5	Current ratio	x_{20}	Return on equity
x_6	Quick ratio	x_{21}	Equity to long-term debts
x_7	Total debts to total assets	x_{22}	Return on total assets
x_8	Working capital per employee	x_{23}	Return on capital employed
x_9	Total assets per employee	x_{24}	Inventories turnover
x_{10}	Net sales per employee	x_{25}	Tax payables to sales
x_{11}	Profit per employee	x_{26}	Provision for staff termination benefits per employee
x_{12}	Debtors turnover	x_{27}	Retained earnings to net sales
x_{13}	Debtor collection period	x_{28}	Litigation
x_{14}	Net assets turnover	x_{29}	Growth

*$Z\text{–}Score = 3.20784k_1 + 1.80384k_2 + 1.61363k_3 + 0.50094k_4 + 0.16903k_5 – 0.39709k_6 – 0.12505k_7 + 0.33849k_8 + 1.42363k_9$

k_1=EBIT / Total assets	*k_6=Current assets / Current liabilities*
k_2=Retained earnings / Total assets	*k_7=Net profit / Net sales*
k_3=Working capital / Total assets	*k_8=Total debts / Total assets*
k_4=Equity / Total debts	*k_9=Company size*
k_5=EBIT / Net sales	

Methods

This section describes an innovative method that used to for finding and extracting the existing laws in data. After using the method, we can get logical rules set. This method combines support vector machine and decision tree from data mining methods.

Support vector machines presented by Vapnik (1995) have been studied increasingly because of their success in overcoming many problems such as bioinformatics (Bonneville *et al.,* 1998), image processing (Huang *et al.,* 1998), text classification (Joachims, 1998), and financial markets (Chen *et al.,* 2006). There are some advantages to using support vector machine, including effective avoidance of over fitting, information condensing of the given data set, and the ability to handle large feature spaces, etc.

In support vector machines, firstly N-dimensional input data is transformed into higher dimensional feature space via a nonlinear mapping function. Then, it constructs the optimal separating hyperplane with maximum distance from the closest points of the training data (Burges, 1998).

If we take a two classification task with the training data $(\overline{x_i}\overline{x_i}, y_i)$, i= 1,...,N, $\vec{x_i}\,\vec{x_i}$ ε R_m, y_i={+1,-1}, then the hyperplane decision function can be written as follows:

$$f(\vec{x}) = \text{sign}\left(\sum_{i=1}^{sv} \alpha_i\, y_i K\langle \vec{x}, \vec{x}_i \rangle + b\right) \quad (1)$$

In the Eqs. (1), α_i is a Lagrange multiplier. There are many functions to be used as a kernel. However, typical examples of kernels successfully used in support vector machines are linear, polynomials, RBF and hyperbolic tangent:

Linear kernel: $$K\langle \vec{x}, \vec{x}_i \rangle = x * x' \quad (2)$$

Polynomial kernel: $$K\langle \vec{x}, \vec{x}_i \rangle = (x * x' + c)^d \quad (3)$$

RBF kernel: $$K\langle \vec{x}, \vec{x}_i \rangle = \exp\left(-\frac{\gamma (x-x\prime)^2}{\delta}\right) \quad (4)$$

Sigmoid kernel: $$K\langle \vec{x}, \vec{x}_i \rangle = \tanh(\gamma x * x' + c) \quad (5)$$

Decision trees have several advantages. They are highly interpretable since they can easily be transformed to a set of meaningful IF-THEN rules. Decision trees are also nonparametric, they make no assumptions about the distribution of the data, and they incorporate a built in feature selection method which makes them immune to the presence of irrelevant independents. Decision trees also have a fast learning mechanism and they can handle categorical values.

For the rules extraction, first, support vector machine is trained with the training samples. Then the data are fed in the support vector machine for data classification. A new data set is created from samples that are classified correctly by support vector machine. A decision tree is made on the new data set finally; decision tree is converted to IF-THEN rules set.

Experiments and Results Analysis

In this research, we focused on the rule extraction for understanding prediction of audit options. We should get the rules which have biological meaning.

In the experiments, the data set was spilt into distinct subset, i.e. training set and testing set. The training set covers the period of 2001-2005 and the testing set the period of 2006-2007.

The Samples of data set were deleted if they were smaller than the first percentile and larger than the 99th percentile, for reduces the noise; the samples were reduced to 780.

First, the support vector machine is trained by using a training sample. The sample is organized in tuples (rows) and attributes (columns).The RBF kernel was used in the support vector machine. Experiments were performed with a different kernel and the RBF kernel was selected for the support vector machine. The training data set fed into support vector machine to get the model and used training data as validation. The results obtained are reported in Table 2. Second, we selected the instance into a new data set which was used later for building rules by comparing the prediction result from support vector machine on data set. The training samples that were correctly classified with support vector machine as training data to train decision tree of rule. Finally we convert decision tree to the IF-THEN rules (Table 3) and obtain the logical rules with accounting meaning. The decision tree is shown in fig 1. In fig 1, the zero number shows the qualified class and the one number shows the unqualified class.

Table 2. The results of the support vector machine for predicting audit opinions.

Results obtained using the test data							
Predicted by the model / Auditor's opinion	Qualified	Unqualified	Accuracy of model predictions	Total Data	The overall prediction accuracy of model	Type I error	Type II error
Qualified	17	62	22%	193	64.25%	0.78	0.06
Unqualified	7	107	94%				
Results obtained using the training data							
Predicted by the model / Auditor's opinion	Qualified	Unqualified	Accuracy of model predictions	Total Data	The overall prediction accuracy of model	Type I error	Type II error
Qualified	144	17	89%	521	96.53%	0.11	0.01
Unqualified	2	358	99%				

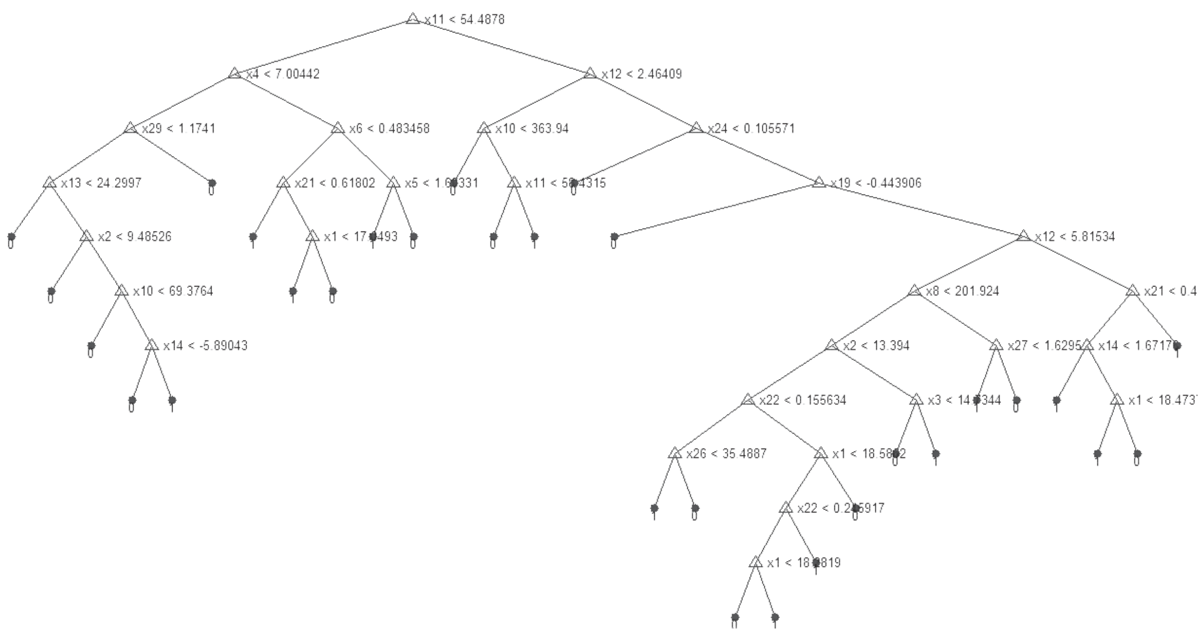

Figure 1. Decision tree model interprets the support vector machine model.

Table 3. The rules for audit opinions identification.

	Conditions (IF)	Results (THEN)
1	$X_{11}<54.48$, $X_4<7.00$, $X_{29}<1.17$, $X_{13}<24.29$	Qualified
2	$X_{11}<54.48$, $X_4<7.00$, $X_{29}<1.17$, $X_{13}>24.29$, $X_2<9.48$	
3	$X_{11}<54.48$, $X_4<7.00$, $X_{29}<1.17$, $X_{13}>24.29$, $X_2>9.48$, $X_{10}<69.37$	
4	$X_{11}<54.48$, $X_4<7.00$, $X_{29}<1.17$, $X_{13}>24.29$, $X_2>9.48$, $X_{10}>69.37$, $X_{14}<-5.89$	
5	$X_{11}<54.48$, $X_4<7.00$, $X_{29}>1.17$	
6	$X_{11}<54.48$, $X_4>7.00$, $X_6<0.48$, $X_{21}<0.61$, $X_1>17.49$	
7	$X_{11}<54.48$, $X_4>7.00$, $X_6<0.48$, $X_5>1.63$	
8	$X_{11}>54.48$, $X_{12}<2.46$, $X_{10}<363.94$	
9	$X_{11}>54.48$, $X_{12}<2.46$, $X_{10}>363.94$, $X_{11}<58.43$	
10	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}<0.10$	
11	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}<-0.44$	
12	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_2<13.39$, $X_{22}<0.15$, $X_{26}>35.48$	
13	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_2<13.39$, $X_{22}>0.15$, $X_1<18.58$, $X_{22}<0.24$, $X_1<18.28$	
14	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_2<13.39$, $X_{22}>0.15$, $X_1>18.58$	
15	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_2>13.39$, $X_3<14.73$	
16	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_{27}>1.62$	
17	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_{21}<0.46$, $X_{14}>1.67$, $X_1>18.47$	
18	$X_{11}<54.48$, $X_4<7.00$, $X_{29}<1.17$, $X_{13}>24.29$, $X_2>9.48$, $X_{10}>69.37$, $X_{14}>-5.89$	Unqualified
19	$X_{11}<54.48$, $X_4>7.00$, $X_6<0.48$, $X_{21}<0.61$	
20	$X_{11}<54.48$, $X_4>7.00$, $X_6<0.48$, $X_{21}<0.61$, $X_1<17.49$	
21	$X_{11}<54.48$, $X_4>7.00$, $X_6<0.48$, $X_5<1.63$	
22	$X_{11}>54.48$, $X_{12}<2.46$, $X_{10}>363.94$, $X_{11}>58.43$	
23	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_2<13.39$, $X_{22}<0.15$, $X_{26}<35.48$	
24	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_2<13.39$, $X_{22}>0.15$, $X_1<18.58$, $X_{22}<0.24$, $X_1>18.28$	
25	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_2<13.39$, $X_{22}>0.15$, $X_1<18.58$, $X_{22}>0.24$	
26	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_2>13.39$, $X_3>14.73$	
27	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_8<201.92$, $X_{27}<1.62$	
28	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_{21}<0.46$, $X_{14}<1.67$	
29	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_{21}<0.46$, $X_{14}>1.67$, $X_1<18.47$	
30	$X_{11}>54.48$, $X_{12}>2.46$, $X_{24}>0.10$, $X_{19}>-0.44$, $X_{12}<5.81$, $X_{21}>0.46$	

Table 4. The results of the support vector machine for predicting audit opinions.

Results obtained using the training data							
Auditor's opinion \ Predicted by the model	Qualified	Unqualified	Accuracy of model predictions	Total Data	The overall prediction accuracy of model	Type I error	Type II error
Qualified	12	5	71%	124	91.94%	0.29	0.05
Unqualified	5	102	95%				
Results obtained using the training data							
Auditor's opinion \ Predicted by the model	Qualified	Unqualified	Accuracy of model predictions	Total Data	The overall prediction accuracy of model	Type I error	Type II error
Qualified	99	45	69%	529	82.99%	0.31	0.12
Unqualified	45	340	88%				

In assessing the performance of a model, an important consideration is the Type I and Type II error rates. A Type I error is committed when a qualified company is classified as unqualified. A Type II error is committed when an unqualified company is classified as qualified.

The auditor can predict audit opinions by the rules exist in Table 3. In Table 3, the rules also reduce the number of variables. You can see the name of all variables in Table 1. These rules clearly explain the intelligent machine performance that made by support vector machine and auditors can easily understand it. Table 4 shows the results of decision tree.

Conclusions

There have been many studies that focused on the prediction of audit opinions using data mining methods. However, these studies were not able to explain the process by which a learning result was reached and why a decision was being made.

The support vector machine is a classification algorithm that provides performance in a wide variety of domains. It has shown generalization ability. However, the success of support vector machine comes at a cost poor comprehensibility which may hinder the wide acceptance of this technique in many areas, especially in accounting and auditing. Reasonable interpretation is useful to guide, and the extracted rules render possible the integration of computational intelligence with symbolic artificial intelligence systems for advanced deduction. In this paper, a novel approach for rule extraction from support vector machine and decision tree is presented. This approach combines support vector machine with decision tree into a new algorithm. We applied the method to the prediction of audit opinions. Using support vector machine as a preprocessor for decision tree aids in the selection of strong instances to generate rules.

The experimental results for audit opinions prediction using our new algorithm show that the comprehensibility of new algorithm is better than that of support vector machine and meaningful and high quality rules can be generated. The most important thing is that the explanation of the rules is very useful in accounting and auditing. The auditor using logical rules obtained to an audit opinions prediction. Finally, this research obtains the 30 rules with 20 variables that help auditor for the audit opinion prediction.

References

Beasley S.M., Carcello J.V., Hermanson D.R., 1999. Fraudulent financial reporting: 1987–1997: an analysis of US public companies. Research Report, COSO.

Bell T., Tabor R., 1991. Empirical analysis of audit uncertainty qualifications. Journal of Accounting Research 29: 350-370.

Bonneville M., Meunier J., Bengio Y., Soucy J.P., 1998. Support vector machines for improving the classification of brain PET images. In Proceedings of the SPIE medical imaging symposium (Vol. 3338, pp. 264-273).San Diego, CA.

Burges C.J.C., 1998. SA tutorial on support vector machines for pattern recognition. Data Mining and knowledge 2(2):121-167.

Chen K., Church B., 1992. Default on debt obligations and the issuance of going concern opinions. Auditing: A Journal of Practice and Theory (Fall): 30-49.

Chen W., Shih J., 2006. A study of Taiwan's issuer credit rating systems using support vector machines. Expert Systems with Applications 30: 427-435.

Dopuch N., Holthausen R., Leftwich R., 1987. Predicting audit qualifications with financial and market variables. The Accounting Review 62 (3): 431-454.

Doumpos M., Gaganis Ch., Pasiouras F., 2005. Explaining qualifications in audit reports using a SVM methodology. Intelligent Systems in Accounting, Finance and Management 13: 197-215.

Gaganis Ch., Pasiouras F., Doumpos M., 2007. Probabilistic neural networks for the identification of qualified audit opinions. Expert Systems with Applications 32:114-124.

Gaganis Ch., Pasiouras F., Spathis Ch., Zopounidis C., 2007. A comparison of nearest neighbours, discriminant and logit models for auditing decisions. Intelligent Systems in Accounting, Finance and Management 15: 23-40.

Han J., Kamber M., 2005. Data Mining Concepts and Techniques. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Ireland J., 2003. An empirical investigation of determinants of audit reports in the UK. Journal of Business Finance and Accounting 30 (78): 975-1015.

Joachims T., 1998. Text categorization with support vector machines: learning with many relevant features. In Proceedings of the 10 European conference on machine learning, (Vol.1, pp. 137-142). Chemnitz, Germany.

Keasey K., Watson R., Wynarczyk P., 1988. The small company audit qualification: a preliminary investigation. Accounting and Business Research 18 (72): 323-333.

Kirkos E., Spathis Ch., Nanopoulos A., Manolopoulos Y., 2007. Identifying qualified auditors opinions: a data mining approach. Journal of Emerging Technologies in Accounting 4:183-197.

Kleinman G., Anandarajan A., 1999. The usefulness of off-balance sheet variables as predictors of auditors going concern opinions: an empirical analysis. Managerial Auditing Journal 14 (6): 286-293.

Koskivaara E., 2004. Artificial neural networks in analytical review procedures. Managerial Auditing Journal 19(2): 191-223.

Laitinen E.K., Laitinen T., 1998. Qualified audit reports in Finland: evidence from large companies. European Accounting Review 7 (4): 639-653.

Palmrose Z.V., 1986. Audit fees and auditor size: further evidence. Journal of Accounting Research 24: 97–110.

Pasiouras F., Gaganis Ch., Zopounidis C., 2006. Multicriteria decision support methodologies for auditing decisions: the case of qualified audit reports in the UK. European Journal of Operational Research 180 (3): 1317-30.

Poorheidari O., Azami Z., 2011 .Identifying audit opinion type using neural network, Accounting Knowledge 3: 77-97.

Setayesh M., Jamalianpoor M., 2010. Investigating the relationship between financial and nonfinancial variables and audit opinion. Accounting Researches 2: 130-157.

Spathis Ch., 2003. Audit qualification, firm litigation, and financial information: an empirical analysis in Greece. International Journal of Auditing 7: 71-85.

Spathis Ch., Doumpos M., Zopounidis C., 2003. Using client performance measures to identify pre-engagement factors associated with qualified audit reports in Greece. The International Journal of Accounting 38:267-284.

Vapnik V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York, NY.