Institute of Advanced Engineering and Science

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 2, No. 4, December 2013, pp. 151~158
ISSN: 2252-8938

An Expert System with Neural Network and Decision Tree for Predicting Audit Opinions

Seyed Mojtaba Saif, Mehdi Sarikhani, Fahime Ebrahimi
Department of Computer, Safashahr Branch, Islamic Azad University, Safashahr, Iran
Email: f_ebrahimi596@yahoo.com

Article Info

Article history:

Received Jul 19, 2013
Revised Oct 21, 2013
Accepted Nov 03, 2013

Keyword:

Audit opinions
Expert system
Artificial neural networks
Decision tree

ABSTRACT

Nowadays expert system, being used in various fields has received a great deal of attention. Auditing is one such field, along with determining the audit opinion type. An expert system consists of a knowledge database and an inference engine. The objective of this research is to make an expert system that will be of help to auditors in predicting and determining the different types of audit reports. The expert system receives data or knowledge from financial reports and determines the types of audit opinions by using an artificial neural network and a decision tree as an inference engine. An expert system should able to explain the solution, but presenting the reason for the results obtained with a neural network is difficult. This study attempts to provide a method that will present simple and understandable reasons for the results obtained with neural networks.

Corresponding Author:

Fahime Ebrahimi,
Department of Computer, Safashahr Branch,
Islamic Azad University, Safashahr, Iran
Email: f_ebrahimi596@yahoo.com

1. 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 required 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.

In recent year, auditing has been increasingly in demand with regard to the importance of quality and reliability of the audited financial reports for 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 modern technologies and knowledge such as expert system have caused new challenges in auditing methods [13]. Relevant research literature repeatedly recognizes the importance of the new technologies in auditing. For example, [15] 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.

Some researchers [3, 4] 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 [14].

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 defence against lawsuits [16].

One method used to predict the type of audit opinion is by use of an expert system, which must have expert knowledge. An expert system emulates the behavior of human experts in the domain of knowledge [9]. Generally, this system consists of a knowledge database and an inference engine. In this research, data mining techniques are used to build the knowledge base. Data mining is an umbrella term that includes methods used to extract human knowledge from data. There is a difference among the traditional way of data analysis and data mining. The former supposes that hypothesis are already constructed and validated against the data, whereas the latter supposes that the patterns and hypotheses are automatically extracted from the data [13]. In data processing methods, knowledge or hidden principles are extracted beyond the data to make different models for analyzing the data.

Data mining methods were used to predict the type of audit opinion. Doumpos et al. [5] 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. Gaganis et al. [7] 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.

Gaganis et al. [8] 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. 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. Kirkos et al. [13] 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%).

But the difference between this research and previous researches is in using neural network from data mining techniques in the framework of an expert system, which could also explain their solutions.

The objective of this research is to make an expert system that helps auditors to predict and determine types of the audit reports. Although a neural network can extract hidden knowledge from the available data, but explaining the neural network's solution is very difficult. On the other hand, although the decision tree, another data mining method, has little ability to extract knowledge from data, it can interpret the results very well. This paper describes an expert system that combines the capabilities of neural networks and decision trees.

2. EXPERT SYSTEM

An expert system is one of the most successful approximate solutions for problems of classical artificial intelligence. Feigenbaum [6], father of the expert system, defines it this way: "An intelligent computer program that uses the knowledge and inference methods for solving problems that need to have experience and skilled people." So the expert system is a computerized system that emulates the capability of deciding experts. This means that the expert system tries in all respects to act like an expert.

Knowledge in an expert system can be defined as experience or knowledge that is available through books, magazines, and scientists. In this research, the knowledge was obtained from audit reports that were coded by variables.

Figure 1 shows the basic concept of a knowledge-based expert system. The user feeds facts to the system and then obtains wise experience, expertise, and advice. The expert system consists of two main parts. The first part is the knowledge base. This base contains knowledge by which the second part, the inference engine, concludes. These results are expert system answers to user questions. Methods of inference in the expert system are very important because argument is a general technique by which the expert system is solving the problems. In this study, a neural network is used as the inference engine.

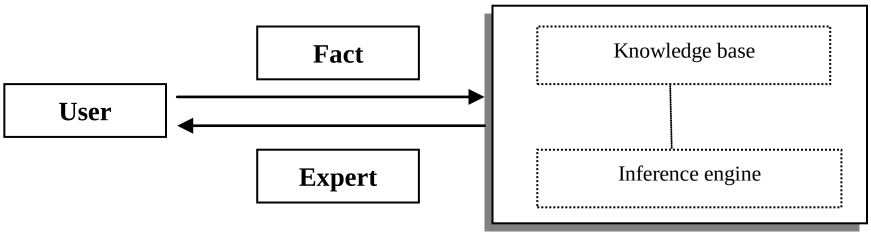

Figure 1. Schema of an expert system

Although neural networks are extremely useful, they also have limitations. Often these networks are trained by existing data, and their performances are such that when an input from training provides feed to the networks, they can reproduce output. In fact, these networks cannot reason or search information from the existent database and use it in their performances or explain about their output. They need a structured environment in which to act. However, new experiments show that if the expert system is used to create the performance environment, neural networks are a very useful act.

Perhaps the most direct way to combine these two branches of artificial intelligence is to use the neural network on the knowledge base of the expert system. This method adds learning ability from provided information to the expert system. Training may be done immediately or after an extended time [20].

3. NEURAL NETWORK

The neural network is human brain simulation. It is a nonlinear and parallel computer. Neural networks are made up of smaller units called neurons. The capability of a neural network is obtained from neurons in a constructive neural network. An artificial neuron is shown in Figure 2.

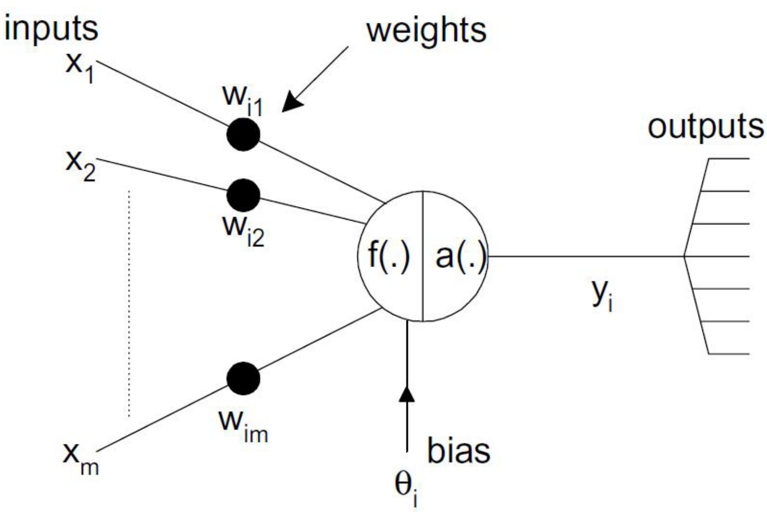

Figure 2. Artificial neuron model

Neurons are connected to each other by a link called a weight. Each neuron has an activation function to determine the output. In Figure 2, each X = (X_1; X_2; …; X_m) and represents the *m* input applied to the neuron; W_i represents the weight for input X_i; θ_i is a bias value; and (.) is the activation function.

Learning is used to find the relationship between input and output in the neural network. The data or samples are divided into a training data set and a testing data set. The training data set was previously used to determine connection weights and nodal thresholds that cause the neural network to estimate outputs that are sufficiently close to target values, and the other data set corresponds to testing of the model. A well-trained neural network is much faster than the mathematical models or simulation programs [12].

A back-propagation multilayer feed forward network (MLN) is a type of artificial neural network architecture and the most widely used for prediction. An MLN contains one input layer, one output layer, and one or more hidden layers [1, 19]. Learning in the MLN involves presenting the training samples, and then a learning method determines the weight of the connections to minimize errors between the network output and known outputs (target output). The learning process continues until the error is less than a value. The difference between these outputs and the target outputs is called error. The error is minimized by modifying the weights [11, 18, 19]. The back-propagation algorithm is a type of training method that implements a calculation of errors and adjusts weights of the hidden layer neurons.

An expert system should be possible to explain the solution. The possible explanation allows user questions from the system about how a particular conclusion is reached. Problem-using neural networks as an inference engine and part of the knowledge base cannot possibly explain the solution.

4. DECISION TREE

A decision tree is a classification algorithm. It sequentially divides the data set in subgroups so that all or almost all the elements in a subgroup belong to the same class. The model generated by a decision tree can be converted to IF-THEN rules, and the relationships between the output variables and input variables are explicit.

A decision tree is a flowchart-like tree [10]. The top node is the root node; each internal node denotes an attribute test; each branch represents an outcome of the test; and each leaf node represents classes.

Classification and Regression Tree used in this research. Classification and Regression Tree reflects these two sides, covering the use of trees as a data analysis method, and in a more mathematical framework, proving some of their fundamental properties. In other word, Classification and regression trees are machine-learning methods for constructing prediction models from data. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition [3].

5. THE PROPOSED METHOD FOR CONSTRUCTING AN AUDITOR EXPERT SYSTEM

Different parts of the auditor expert system were prepared in several steps. The first step was to prepare the knowledge base. The knowledge based was built by observations that comprised 1018 audit reports for the years 2001 through 2007 containing 708 unqualified audit reports and 310 qualified ones. After excluding the financial companies (i.e., banks and investment companies), we selected the publicly listed firms that were qualified at least once over the 2001 to 2007 and that closed their fiscal year in mid-March (end of Persian calendar year).

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.

The audit reports were first coded by using 29 variables in Table 1, and a data set was created. The selection of these variables was based on past research that has been done [5, 13, 17]. This data set is the knowledge base of the expert system.

A perceptron neural network was prepared by part of the data. This data set is called training data set.

Table 1. List of variables

Y	Auditor's opinion
X_1	Z-Score*
X_2	Log net sales
X_3	Log total assets
X_4	Log number of employees
X_5	Current ratio
X_6	Quick ratio
X_7	Total debts to total assets
X_8	Working capital per employee
X_9	Total assets per employee
X_{10}	Net sales per employee
X_{11}	Profit per employee
X_{12}	Debtors turnover
X_{13}	Debtor collection period
X_{14}	Net assets turnover
X_{15}	Fixed assets turnover
X_{16}	EBIT margin
X_{17}	Earnings before tax margin
X_{18}	Cash from operating activities to net sales
X_{19}	Cash from investing activities to net sales
X_{20}	Return on equity
X_{21}	Equity to long-term debts
X_{22}	Return on total assets
X_{23}	Return on capital employed
X_{24}	Inventories turnover
X_{25}	Tax payables to sales
X_{26}	Provision for staff termination benefits per employee
X_{27}	Retained earnings to net sales
X_{28}	Litigation
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_2 = *Retained earnings / Total assets*
k_3 = *Working capital / Total assets*
k_4 = *Equity / Total debts*
k_5 = *EBIT / Net sales*
k_6 = *Current assets / Current liabilities*
k_7 = *Net profit / Net sales*
k_8 = *Total debts / Total assets*
k_9 = *Company size*

The neural network response is the expert system response, so the expert system should be able to explain it. The expert system uses a decision tree for the solution explanations. The decision tree should be able to explain the results of the neural network. Samples were classified correctly by the neural network, and in making and training the decision tree was used. Therefore the decision tree can simply explain the neural network performance. This decision tree can be converted to several simple rules, which explain the result of the neural network for each input.

6. CONDUCTED EXPERIMENTS TO TEST THE PROPOSED EXPERT SYSTEM

Each audit report was summarized and coded with 29 variables in Table 1 for entry into the expert system. Lastly, a table with 29 columns and 1018 rows were obtained as the audit report of a company in one year. From now on, each row is called a sample. The top and down percentiles of each column was calculated. Then the samples were deleted if they were smaller than the first percentile and larger than the 99th for reduction of the noise. Thus the number of samples is decreased to 780.

In this research, the sample was split into distinct groups, i.e., a training sample and a testing sample. The training sample covers the period of 2001-2005 and the testing sample the period of 2006-2007.

In this study, a multilayer perceptron neural network with two hidden layers was used. The number of neuron on the input layer and hidden layers was 29, 10 and 20 neuron sequentially that introduced by Poorheidari and Azami [17]. Two neurons were on the output layer. One is active for a qualified report, and the other for an unqualified report. Each input was normalized within the range of 0-1. The results of the neural network are present in Table 2, and each row represents the percentage of false negatives, false positives, and true positives for the qualified/unqualified audit reports.

Table 2. The results of the neural network

Train sample				
Auditor's opinion	True positives	False positives	False negatives	Percent of samples misclassified
Qualified	%97.67	%6.46	%2.33	%6.20
Unqualified	%84.47	%5.59	%15.53	
Test sample				
Auditor's opinion	True positives	False positives	False negatives	Percent of samples misclassified
Qualified	%79.82	%40.35	%20.18	%35.12
Unqualified	%41.77	%29.11	%58.23	

Samples were classified correctly by the neural network; in making and training the decision tree was used. Figure 3 shows the decision tree, and it can be converted into a set of simple rules. These rules are presented in Table 3. They explain the results obtained by the neural network.

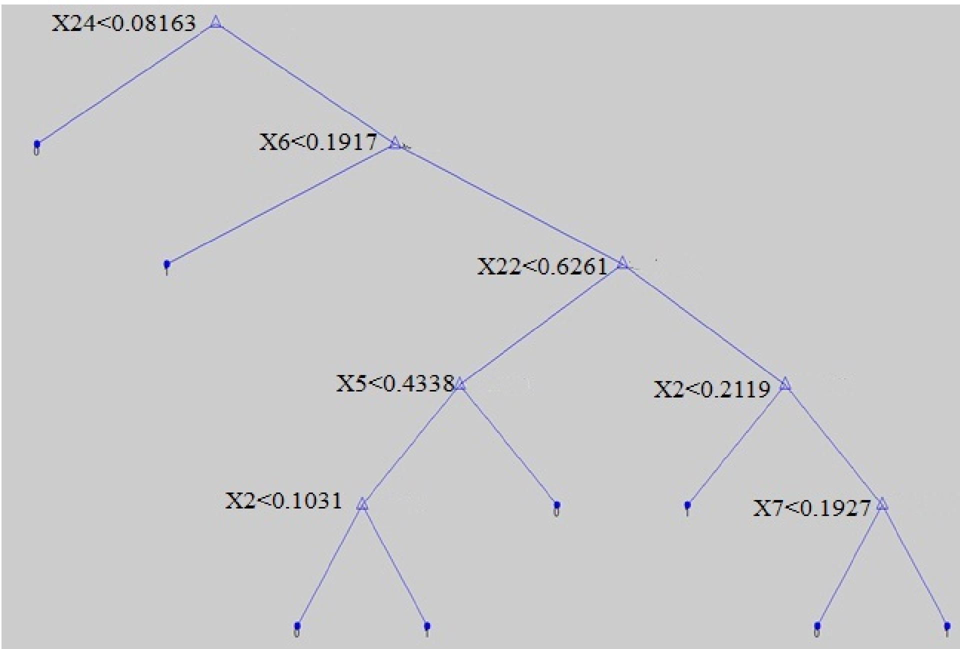

Figure 3. Decision tree model interprets the results

Table 3. The rules explain the results

Node1 :	if x_{24}<0.0816306 then node 2 else node 3
Node2 :	class = Qualified
Node3 :	if x_6<0.191773 then node 4 else node 5
Node4 :	class = Unqualified
Node5 :	if x_{22}<0.626199 then node 6 else node 7
Node6 :	if x_5<0.433883 then node 8 else node 9
Node7 :	if x_2<0.211922 then node 10 else node 11
Node8 :	if x_2<0.10314 then node 12 else node 13
Node9 :	class = Qualified
Node10:	class = Unqualified
Node11:	if x_7<0.192759 then node 14 else node 15
Node 12:	class = Qualified
Node13:	class = Unqualified
Node14:	class = Qualified
Node15:	class = Unqualified

The expert system is now ready for the auditor. We will explain how it works by means of two examples. A coded audit report of the Noosh Mazandaran Company by variable is importing to the expert system (see Table 4).

Example 1:

Input of expert system: Coded audit report of Noosh Mazandaran Company by variable

Answer of expert system: Unqualified

Explanation of expert system: The input audit report is unqualified because inventory turnover is greater than 0.08163 and quick ratio is less than 0.1977.

Table 4. The audit report of Noosh Mazandaran company

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	…	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}
Value	0.453789	0.251095	0.279794	0.310881	0.236915	0.158921	0.205532	0.569434	0.054615	…	0.825111	0.378247	0.473055	0.527583	0.430297	0.328883	0.001006	0.032071	0.661834	0.000000	0.150983

Coded audit report of Esfahan Petrochemical Company by variable is importing to the expert system (See Table 5).

Input of expert system: Coded audit report of Esfahan Petrochemical Company by variable

Answer of expert system: Qualified

Explanation of expert system: The input audit report is qualified because inventory turnover is greater than 0.08163, and quick ratio is greater than 0.1977; return on total assets is greater than 0.6261; log net sales is greater than 0.2119, and total debts to total assets is less than 0.1927.

Table 5. The audit report of Esfahan petrochemical

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	…	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}
Value	0.729110	0.542743	0.518485	0.611006	0.223485	0.336133	0.170172	0.564832	0.068250	…	0.869278	0.587066	0.396983	0.891100	0.634981	0.058933	0.135352	0.280184	0.280184	0.000000	0.251903

7. CONCLUSION

This research is an attempt to introduce an approach to building an auditor expert system that is of help to auditors. In this approach, data mining methods used for building inference engines, and also the problem presenting the reason for the results obtained with neural network was solved by a decision tree. The neural network is from the data mining method that was used to construct the inference engine. The results of the research present an approach to the building of an auditor expert system with the accuracy and ability to explain solutions for users.

REFERENCES

[1] Akcayol MA, Cnar C, Bulbul HI, Klcarslan A. ANN based modeling of injection pressure in diesel engines. *World Scientific and Engineering Academy and Society (WSEAS) Transactions on Computers*. 2004; 3: 1538-1544.
[2] Bell T, Tabor R. Empirical analysis of audit uncertainty qualifications. *J. Account. Res*. 1991; 29: 350-370.
[3] Breiman L, Friedman J, Olshen R, Stone C. *Classification and Regression Trees*. Chapman and Hall/CRC Press, Boca Raton, Florida. 1984.
[4] Chen K, Church B. Default on debt obligations and the issuance of going concern opinions. *Auditing-J. Pract. Th.* 1992: 30-49.
[5] Doumpos M, Gaganis C, Pasiouras F. Explaining qualifications in audit reports using a SVM methodology. *Intell Syst. account. Financ. Manage*. 2005; 13: 197-215.
[6] Figenbaum EA, Cohen PR. *Handbook of artificial Intelligence*. William Kaufmann & HeurisTech Press. 1982.
[7] Gaganis C, Pasiouras F, Doumpos M. Probabilistic neural networks for the identification of qualified audit opinions. *Expert. Syst. Appl*. 2007; 32: 114-124.
[8] Gaganis C, Pasiouras F, Spathis C, Zopounidis C. A comparison of nearest neighbours, discriminant and logit models for auditing decisions. *Intell. Syst. Account. Financ. Manage*. 2007; 15: 23-40.
[9] Giarratano JC, Riley G. *Expert Systems: Principles and Programming*. Pwskent Pub. Co. 1989.
[10] Han J, Kamber M. *Data Mining Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 2005.
[11] Hosoz M, Ertunc HM. ANN analysis of an automobile air conditioning system. *Energ. Convers. Manage*. 2006; 47: 1574-1587.
[12] Kafkas F, Karatas C, Sozen A, Arcaklioglu E, Saritas S. Determination of residual stresses based on heat treatment conditions and densities on a hybrid powder metallurgy steel using ANN. *Mater. Design*. 2007; 28: 2431-2442.
[13] Kirkos E, Spathis C, Nanopoulos A, Manolopoulos Y. Identifying qualified auditors opinions: a data mining approach. *J. Emerg. Technol. Account*. 2007; 4: 183-197.
[14] Kleinman G, Anandarajan A. The usefulness of off-balance sheet variables as predictors of auditors going concern opinions: an empirical analysis. *Manage. Audit. J*. 1999; 14(6): 286-293.
[15] Koskivaara E. Artificial neural networks in analytical review procedures. *Manage. Audit. J*. 2004; 19(2): 191-223.
[16] Laitinen EK, Laitinen T. Qualified audit reports in Finland: evidence from large companies. *Eur. Account. Rev*. 1998; 7(4): 639–653.
[17] Poorheidari O, Azami Z. Identifying audit opinion type using neural network. *Account Knowl*. 2011; 3: 77-97.
[18] Rakesh K, Kaushikb SC, Gargb SN. Heating and cooling potential of an earth-to-air heat exchanger using ANN. *Renew. Energ*. 2006; 31: 1139-1155.
[19] Sencan A. Artificial intelligent methods for thermodynamic evaluation of ammonia water refrigeration systems. *Energ. Convers. Manage*. 2006; 47: 3319-3332.
[20] Wang S. Neural networks in generalizing export knowledge. *Comput. Ind. Eng*. 1997; 32: 67-76.