DOI: 10.6220/joq.2015.22(5).02

Applying Decision Tree and Neural Network to Raise the Performance of Human Training Quality

Yung-Tsan Jou[1], Yih-Chuan Wu[2*], Wen-Tsann Lin[2]

[1]Department of Industrial and Systems Engineering, Chung Yuan Christian University

[2]Department of Industrial Engineering and Management, National Chin-Yi University of Technology

Abstract

Human resources are the most important asset of organization. In the trend of globalization, human resources turned to be the core department of a company. The fundamental job is to establish an excellent human training system and enhance training quality. Taiwan TrainQuali System, TTQS provides businesses and training institutions with the tools to carry out internal human training by offering a sound system of evaluating the education and training. The assessment of how good the processes and outcomes of all the training programs in an enterprise by operating the TTQS and following its PDDRO (Plan, Design, Do, Review, Outcomes) procedural standards can integrate the executive results and the enterprise's performance into a systematic integral plan, making the education and training better meet the business needs. This study applies data mining techniques to explore human training quality of 2012 TTQS new version assessment review database and to find TTQS critical assessment indicators. Back-propagation neural networks first apply to evaluate TTQS database classification accuracy and performance. Training patterns and testing patterns prediction accuracy are greater than 95%. Then the study analyzes and compares decision tree algorithms (C5.0, CART, CHAID), and chooses a higher accuracy rate algorithm to discuss TTQS critical indicators. C5.0 has better accuracy rate under any partition proportions and the highest testing accuracy is 89.41%. K-Means clustering analysis is to identify the critical indicators chosen by the decision tree. Through cross comparison with decision tree and K-Means results, this study identifies 9 important critical indicators of TTQS assessment to help enterprise in Taiwan to introduce TTQS, to grasp TTQS assessment critical indicators, and to enhance the quality of human training and performance.

Keywords: back-propagation neural network, decision tree, IIP, K-Means, TTQS

Received: 06/2014; Revised: 08/2014; Accepted: 12/2014

[*]Correspondence: Department of Industrial Engineering and Management, National Chin-Yi University of Technology
57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung City 41170, Taiwan, R.O.C.
E-mail: jason_wu1102@yahoo.com.tw

1. INTRODUCTION

Global competition turns to knowledge economy-based form from industry. Advanced industry relies on innovation, intelligent property and capital to stay competitive. This results in an increased demand for skilled and knowledge talents. Businesses have to keep investing in talents to ensure upgraded function to meet the changing markets and stay competitive in practice. Currently in Taiwan, to encourage business talents continue learning and studying, Small and Medium Enterprise Administration, Ministry of Economic Affairs, promoted Life Long Learning Passport (National Association of Small & Medium Enterprises, R.O.C., 2012). This shows that education training is hightly regarded in Taiwan and explains that talents are the most significant asset for businesses (Hsu, 2011). Any technologies, resources and strategies can be imitated; only the abilities of talents can not be taken away. Developing potential resources of talents is one of the most critical issues for businesses.

If an organization understands the learning motives of its employees and combines them with the concept, vision and mission of its operation and development further to enhance their willingness to participate in the education and training, it should be able to effectively increase the training effects. The main structure in this research is Taiwan TrainQuali System, TTQS to explore correlation between key success factors and performance of businesses in the hope to provide businesses with major reference in future introduction of human development or quality training management TTQS system, while enhancing professional ability of training talents, increasing labor productivity and international competence.

This study is based on the TTQS evaluation database of the central Taiwan in 2012. First, the data mining technique of Back-Propagation Neural Networks categorization is employed to assess the best network framework and performance of TTQS database, and verify that the correct level of TTQS categories. Then through the results of the decision tree analysis, this study compares the predictive accuracy of the decision tree algorithm analysis by CHAID, CART, and C5.0. Identifying TTQS critical indicators and providing them to training units as a reference, enterprises can grasp the critical indicators of the assessment. Finally, confirmative comparison is made by clustering in K-Means to verify the decision tree critical indicators. Further, this study verifies whether this training quality system can effectively enhance the human training quality in organizations, such that the organizations that introduce and carry out TTQS are all able to construct the human training system that satisfies themselves.

2. LITERATURE REVIEW

2.1 Investors in People, IIP

Amid the international competition that has become fiercer all the time, challenging the industrial transformation, the U.K. began to work on the startup of the IIP (Investors in People) system in 1990 by leading enterprises, experts and laborers organizations who jointly drew up the IIP standards. Such project was officially launched in 1991 with the aims as follows:

(1) To uplift the levels of education and training with this system when facing the integration of European market and the pressure from unions, to assist the enterprises in developing their goals.
(2) To encourage the organizations to link the employee development with the business operation goals in a systematic way to improve the organizations' performance.
(3) To further enhance the national competitive edge through the upgrading and development of employee abilities as well as the connection to business goals.

IIP is the world's first "standards for human quality assurance" and able to adapt to organizations of various trades and different scales. Since its formulation in 1990 and on, the British IIP has come a long way. A set of standards for human quality assurance, IIP is also an important object of reference for TTQS. But, unlike TTQS that uses scorecards, IIP uses both compliance evaluation and maturity evaluation. IIP operates in three principles, ten indicators and 196 evidence requirements, of which 196 requirements some should be known or executed by top management, some others by officers and senior officers and further others by all from employees to senior officers. The three principles are Plan, Do and Review (Investors in People, 2012).

2.2 Taiwan TrainQuali System, TTQS

Human capital is one of the most important elements of productivity; human training is to elevate human quality and build up human capital. It is fundamental to establish a sound education and training system for fostering the necessary human power for the industries. Being an open economy with limited natural resources, Taiwan can rely only on the accumulation of human capital. Taiwan TrainQuali System (TTQS) provides businesses and training institutions a sound system of evaluating education and training to carry out internal human training. The assessment of how good the processes and outcomes of all the training programs in an enterprise by introducing and operating the TTQS and following its PDDRO (Plan, Design, Do, Review, Outcomes) procedural standards can integrate the executive results and the enterprise's performance into a systematic integral plan, making the education and training better meet the business and training institution's needs.

Workforce Development Agency (WDA) by incorporating ISO 10015, European vocational training policies, the British Investor in People (IIP) and the Australian proactive vocational training policies as well as Taiwan's current condition, a set of "training quality system," TTQS for short, that is unique to Taiwan was developed. Starting in 2005, the TTQS specifications began to take shape with the efforts of many experts, aiming at the training of reviewers, developing grading criteria with TTQS score cards. With the modification by the government, the present changes in foci on and objectives of TTQS development are apparent. Because of TTQS older version indicators content improvement, WDA announced TTQS new version assessment indicators in 2012. TTQS training circle is shown as in Figure 1.

Figure 1. TTQS training circle

Source: This study.

2.3 Data mining theory and technology

Data mining had two major functions -- predicting future trend from built model to offer decision makers dependable information and locating unknown forms of information (Berry and Linoff, 1997). The former is represented by statistics and analysis, including all descriptive statistics, probability and category analysis in statistics. Data mining mainly focused on information with a lot of variables and great number of pieces of data with factor analysis for generalizing variables in multivariate analysis, discriminate analysis for classification and cluster analysis for dividing groups (Han and Kamber, 2000). The improved technology widely used artificial intelligence with decision tree theory, gene algorithm, neuron network, and rule induction.

2.4 Back-Propagation Neural Networks, BPN

Studies and application of Artificial Neural Networks (ANN) have been widely available in different fields (Rumelhart *et al.*, 1994). ANN tolerates incomplete, lost and noise information (Chin, 2008; Lin *et al.*, 2010). It does not need parameter or prior hypothesis on data distribution. It can establish any complicated nonlinear and continuous functions in correspondence to questionnaire (Vellido *et al.*, 1999). Between 1992 and 1998, 78% of ANN application in commerce is supervision Backpropagation Neural Network (BPN). According to Chen (2010), as a whole, the basic framework for neural network can be divided in the levels of “processing element,” “layer” and “network.”

2.4.1 Processing Element, PE

Otherwise called artificial neurons, the process elements, PEs, are the basic components of neural network. They are divided into different functions; Summation, Activity, and Transfer. Transfer function is the most frequently used and discussed processing element, which affects directly the network's learning structure. With different transfer functions, the networks built have different characteristics.

(1) Summation function

Summation functions play a summing role in network; it sums up all calculation results sent by the other PEs through the network; i.e.,

$$I = f_1(W, X) \tag{1}$$

Where W is the weight of the network, X is the output of other PEs.

Generally, commonly used summation functions are a. weighted product sums and b. sum of Euclidean distances, each is expressed by below equations, respectively.

a. Weighted product sums

$$I_{ij} = \Sigma_i w_{ij} X_I \tag{2}$$

b. Sum of Euclidean distances

$$I_j = \Sigma_i X_I - w_{ij} \tag{3}$$

(2) Activity function

$$net_j = f_2(I^n{}_j, I_j^{n-1}, net_j^{n-1}) \tag{4}$$

Its purpose is to combine the value of summation function and the network status of the PE. There are three types of commonly used activity functions: a. direct use of summation function output; b. incorporation with previous summation function output; and c. incorporation with previous activity function output, described in below expressions, respectively:

a. Direct use of summation function output

$$net_j^n = I_j^n \tag{5}$$

b. Incorporation with previous activity function output

$$net_j^n = I_j^n + C \bullet I_j^{n-1} \tag{6}$$

c. Incorporation with previous activity function output

$$net_j^n = I_j^n + C \bullet net_j^{n-1} \tag{7}$$

Activity functions of neural network use function outputs directly, while in BPN models, a previous activity function output is incorporated.

(3) Transfer function

$$Y = f_3(net) \tag{8}$$

2.4.2 Layer

Several PEs with the same function form into a layer, which is characterized by competitive output and competitive learning. In a basic neural network, the structure comprises the input lay, the hidden layer and the output layer. The input layer receives the input signals from outside and resends them to all the neurons in the hidden layer; it, however, does not process the inputted data as it actually does not contain calculation neurons. The hidden layer provides the neural network to present the interaction between PEs; it includes an inner structure capable of processing problems. The number of layers is usually determined by the test mode as optimal; that is, to be decided by the extent to which the problems are complex. The output layer receives signals from the hidden layer and outputs them to exterior environment (Yeh, 2009).

2.4.3 Network

In a nutshell, a complete neural network consists of several structural layers with different functions, where the relationship between layers is defined by the transfer function. Each structural layer contains a great number of PEs by which, along with the relationships, it is possible to construct an intelligent system that is capable of learning. In general, the operation of the neural network can be divided in learning and recalling; the former is a process of learning from examples via learning algorithm to adjust the weight of network connection; the latter is a process of determining network output data from input data by the recalling algorithm and includes the algorithms of supervised learning, unsupervised learning and associated learning.

A neural network comprises of many Artificial Neurons, which are also called processing elements, PEs. A PE makes fan-shaped output as input for many other elements. Such action of biological neurons can be simulated by a simple structure and algorithmic rules, as shown in Equation (9) (Yeh, 2010):

$$Y_j = F(\Sigma_i W_{ij} X_i - Z_j) \tag{9}$$

Where Y_j = output signal from the mimicked biological neuron model

F = Transfer function of the mimicked biological neuron model; a mathematic equation that converts the weighted product of the input received by other PEs into PE output

W_{ij} = the node intensity of the mimicked biological neuron model; also called connection weight

X_i = input signal to the mimicked biological neuron model

Z_j = threshold of the mimicked biological neuron model.

Neural networks are widely applied, to failure diagnosis and process monitoring at chemical factories, for example, in industry. Commerce-wise, examples are analysis of stock investment and decision-making in futures transaction; management-wise, an example is selection of scheduling strategy. Among the above, supervised learning networks are the most applied and most successful network model, accounting for more than 95% of the present applications (Yeh, 2003).

Back-Propagation Neural Networks is a supervised learning network among neural networks and currently one of the widest applied techniques and most typical of neural network, suiting in the aspects like diagnosis and prediction(Su *et al.*, 2002). It is connected by multiple single-layer networks, each of which is formed by a number of neurons (Chen and Lin, 2010).

This study employs BPN to identify human training quality, because BPN has high learning precision and can deal with the complex problem of sample identification and that of synthesis of highly-nonlinear functions (Yeh, 1999, 2003). The structure of BPN includes input layer, hidden layer, and output layer. (as shown in Figure 2)

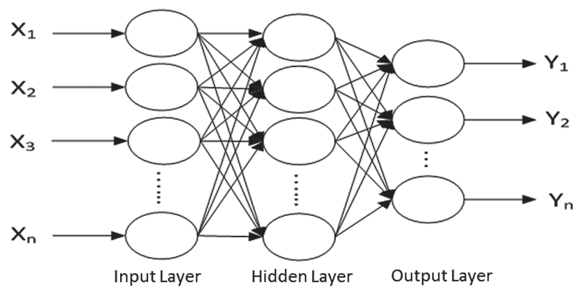

Figure 2. Backpropagation neural network framework

Source: This study.

2.5 Decision tree theory

Decision tree algorithm is one of the classification model methods in data mining to automatically classify data based on split conditions, locate source rules with induction method, establish expert system and predict unknown data. Decision tree data structure is like a tree with internal node and leaf. In decision tree, each internal node represents test of a certain property. Each node is a judgment formula to judge on a variable to see if the data were larger or smaller than a value. Nodes could be divided into different classes per input information. Leaf means the corresponding decision value. Branch of each internal node represented possible value of the property. Hastie *et al.* (2001) mentioned that the most widely used decision tree algorithm includs CART (classification and regression trees), C5.0 and CHAID (chi-square automatic interaction detector).

(1) CHAID: Kass (1980) combined Chi-Square Test and Automatic Interaction Detector as CHAID. CHAID aimed at chi-square test in each split to calculate node *p*-value, which decided whether the decision tree should continue growing. (Biggs *et al.*, 1991). No tree pruning was needed. CHAID and C5 normally were applied in categorical information.

(2) CART: CART was porposed by Breiman in 1984; it was a binary split method applied in database of continuous information. Each split divided the information into 2 sub-set based on gini rule for split condition selection. C5 algorithm also first built a complete tree and then had pruned tree. What was different was CART was based on entire error rate for tree pruning (Breiman, 1984).

(3) C5.0: C5.0 algorithm was proposed by Quinlan in 1996 (Chen, 2011). C5.0 algorithm, also known as rule-based reasoning model for handling large amounts of data. C5.0 algorithm was for category information such as, categorical split in high, medium and low in percentage of education training budget. C5.0 selected property with greatest addition in information as split property. C5.0 first built a complete tree and pruned tree with user defined predicted error rate of each node. The comparison of decision tree algorithms is shown in Table 1.

2.6 K-Means cluster algorithm

Clustering is the process of partitioning or grouping a given set of patterns into disjoint clusters. Cluster analysis is a logical procedure of objective classification based on similarity and difference, having the purpose of facilitating the identification of the similarity between certain things and the division of them into a number of clusters by such feature such that those in a same cluster have high homogeneity and the samples among different clusters have high heterogeneity. Abidi and Ong (2000) proposed the technique of two-stage clustering (SOM + K-Means) as the clustering strategy for conducting data mining.

K-Means cluster algorithm was proposed by J. B. MacQueen in 1967, which is used to deal with the problem of data clustering, the algorithm is relatively simple, so generate a wide influence in the scientific field research and industrial applications. The K-Means objective function is usually taken squared error criterion and shown as in Equation (10).

Table 1. Comparison of decision tree algorithms

Algorithm	Data property	Split rules	Tree pruning rules
C5.0	Categorical information	Gain ratio	Predicted error rate
CHAID	Categorical information	Chi-square test	No pruning
CART	Continuous information	Gini index	Entire error rate

Source: This study.

$$E = \sum_{i=1}^{K} \sum_{p \in C_i} | p - m_i |^2 \quad (10)$$

Where, E is total square error of all the objects in the data cluster, p bellows to data object set, m_i is mean value of cluster C_i.

In terms of economy, simplicity and effectiveness, K-Means is a method worth applying. It is a non-layer-type clustering, not subject to the effects of outliers, errors in distance measurement and selection of method of distance calculation. The results of cluster are better if the initial point of the cluster is known. The knowledge of number of clusters ahead of making clusters enables effective process of large amount of numeric data; this method is suitable for dealing with the characteristics of convex clusters, while ending only fits local optimization.

3. RESEARCH METHODOLOGY

The exploration of this study targets at the "Taiwan TrainQuali System, TTQS," based on the WDA-provided database of businesses applications for TTQS review. The purpose is chiefly to inspect whether the TTQS evaluation system can offer effective and appropriate benefits to the human capitals of the industries in Taiwan. It is expected to observe, from the Review database provided by WDA, whether the indicators and specifications in the businesses who have introduced TTQS have all been effectively internalized and carried out, and to turn the database contents into information with referential values. The research methodology is described below.

This research first explores TTQS performance effects at enterprises with back-propagation neural networks from 2012 TTQS review database. Then the decision tree is discussed to compare the predictive accuracy of the decision tree algorithm analysis by CHAID, CART, and C5.0. Identifying TTQS critical indicators and provided to training units as references, enterprises can handle the critical indicators of the assessment. Finally, confirmative comparison is made by K-Means cluster alrigotthm to verify the decision tree critical indicators. This research hopes to grasp the TTQS critical indicators to enhance enterprise performance and understand training strategies of various human resources.

3.1 TTQS indicators

This research uses the 2012 TTQS central Taiwan review database for analysis. In order to facilitate subsequent employment of related software to make analysis, this section will number the TTQS indicators. The indicators, bases for classification assessment and code list are as shown in Table 2 below.

Based on TTQS assessment indicator scores and summarized, TTQS performance and grade distribution are arranged and categorized as in Table 3 below. Threshold is of 53.5 points to 63 points (Grade E); Bronze medal is 63.5 points to 74 points (Grade D); Silver medal is 74.5 points to 85

Table 2. TTQS indicators and code

Construct	Code	Indicators
Plan	B1	Disclosure of institutional vision/mission/strategy. Setting targets and requirements.
Plan	B2	Specific training policy and target. Top managers for training commitment and participation.
Plan	B3	Specific PDDRO training system and specific core training categories.
Plan	B4	Training quality management system and documents.
Plan	B5	Linking the training planning and the operation objects.
Plan	B6	Administrative management of the training facility. Train-related functions.
Design	B7	Competency-related analysis and application of training requirements.
Design	B8	System design of training programs.
Design	B9	Process involvement by stakeholders.
Design	B10	Purchasing procedure and selection criteria of training products and services.
Design	B11	Combination of training and target requirements.
Do	B12	12a. Planning of learner selection and compliance. 12b. Planning of selection and compliance of teaching materials. 12c. Planning of teacher selection and compliance. 12d. Planning of selection and compliance of teaching methods. 12e. Planning of selection for teaching environment and equipment.
Do	B13	Learning results transfer and use.
Do	B14	Classification and filing of training materials. Systemization of management information.
Review	B15	Assessment report and regular summarized analysis.
Review	B16	Monitoring and handing for abnormal correction.
Outcomes	B17	Diversity and completeness of training results evaluation.
Outcomes	B18	Top managers for training cognition and feeling.
Outcomes	B19	Training results.

Source: This study.

Table 3. TTQS classification standard

Levels	Performance	Scores	Guidance
Platinum (Grade A)	Benchmarking	96.5 and above	As a model of success.
Gold (Grade B)	Excellent	85.5 ~ 96	Duly execution in continuous improvement.
Silver (Grade C)	Fair	74.5 ~ 85	Appropriate execution, requiring key improvement.
Bronze (Grade D)	Slightly bad	63.5 ~ 74	Partial execution, requiring major improvement.
Threshold (Grade E)	Bad	53.5 ~ 63	Without distinctive evidence of execution performance.

Source: This study.

points (Grade C); Gold medal is 85.5 points to 96 points (Grade B); Platinum medal is 96.5 points or above (Grade A) (Lin *et al.*, 2009).

3.2 Back-Propagation Neural Networks

To conduct analysis, this study applies the "Back-Propagation Neural Networks" of the neural network; hopes to verify the goodness of fit of the classification by means of "training (learning)" and "testing (verification)" of the Back-Propagation Neural Networks.

3.2.1 Entry data scaling

Variable value fields have to be unified to avoid entire network learning controlled by larger value fields from great changes of variable value fields in units. ANN training can be more efficient in data processing network before input and goals (Wu and Liu, 2011). They are between 0 and 1 (Kulatilake *et al.*, 2010). All data are scaled before use (Lin *et al.*, 2010). Scaling equation is as follows Equation (11):

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{11}$$

Where, y_i: scaling data, x: original data, xmax: maximum value in original data, $x_{\min}$: minimum value in original data.

3.2.2 Entry of variable definition and goal output

BPN often divides data into two patterns; training and testing (Yeh and Wu, 2009). The 350 pieces effective sample are divided into training patterns and testing patterns; 70%, 80%, 90% effective samples are for network training and balance 30%, 20%, 10% are from network tests and compare its accuracy rate under different circumstances division. Number of layers is 3: input, hidden and output. Setup variable of input layer is based on 19 indicators of TTQS (as in Table 2); that of goal output is based on five levels of TTQS assessment, Platinum, Gold, Silver, Bronze and Threshold. Those of hidden layer include number of hidden layers, number of neurons, and transfer function. Yeh (2003) suggests one to two layers to have best convergence; Ye also emphasizes unhidden layer structure must be first used for analysis prior to solving problems with BPN. Provided the prediction ability is better than that with hidden layer, the problem is free from BPN questions, as the result of BPN without hidden structure is close to linear regression analysis.

3.3 Decision tree algorithm

Decision tree algorithm automatically split and classified based on pre-set distinctive level and then established a tree structure from the classified events. With the tree structure, certain rules in the events were generalized to locate the relation and predict these events. Simply put, decision tree algorithm help transfer complicated information into relatively simple and easily understood information in quick judgment. CHAID is applied in categorical information in categorical split. CHAID uses Chi-square test to determine split conditions. The information could be categorized.

Chi-square test calculates p-value in nodes. Value size of p-value determined whether split continued and there was no need of pruning tree. CART is a binary split method applied in database of continuous information. Each split divide the information into 2 sub-set based on Gini rule for split condition selection (Chien *et al.*, 2008). C5 algorithm is for category information and select property with greatest addition in information as split property. Let p $(j|D)$ denotes the fraction of records belonging to class j at a given node D. It sometimes omits the reference to node D and expresses the fraction as p_j. The Gini rule is described in Equation (12) (Wang *et al.*, 2010) as follows:

$$Gini(D) = 1 - \sum_{j=1}^{n} P_j^2 \tag{12}$$

Where, p_j is the probability of class j vectors in the node.

3.4 K-Means clustering

The K-Means clustering algorithm, which is also called the GLA, is widely used for data clustering. The K-Means clustering algorithm partitions data points into K clusters Si (i = 1, 2, ..., K) and the cluster Si is associated with a representative (cluster center) **C**i (Lai and Liaw, 2008; Lin *et al.*, 2011).

The K-Means clustering algorithm is briefly described as follows:

(1) Begin with an initial set of cluster centers SC_0. Set $p = 0$.

(2) Given the set of cluster centers SC_p, perform the Lloyd iteration to generate the improved set of cluster representatives SC_{p+1}.

(3) Compute the average distortion for SC_{p+1}. If it is changed by a small enough amount since the last iteration, then stop. Otherwise set $p + 1 \rightarrow p$ and go to Step (2).

The nearest cluster center is determined by computing the Euclidean distance between each cluster center and a data point. The Euclidean distance between a data point $\mathbf{X} = (x_1, x_2, ..., x_d)\ t$ and a cluster center $\mathbf{C} = (c_1, c_2, ..., c_d)\ t$ is defined as below Equation (13):

$$d(X, C) = \left[\sum_{i=1}^{d} |X_i - C_i|^2\right]^{0.5} \tag{13}$$

4. RESULTS

4.1 Back-propagation neural networks learning results

The exploration of this study targets at the "Taiwan TrainQuali System, TTQS," based on the businesses applications for TTQS review. To conduct analysis, this study first applies the "Back-Propagation Neural Networks" of the neural network. Training and testing (verification) samples

Table 4. Neuron output and accuracy

Partition (Train/Test)	Training		Testing	
	Samples	Accuracy	Samples	Accuracy
70/30	242	96.75%	108	95.58%
80/20	286	99.32%	64	98.46%
90/10	318	97.25%	32	100.00%

Source: This study.

prediction accuracy was greater than 95%. Table 4 below indicates that the rates of correctness of the output in neural network analysis are 97.25% and 100% in training group and verification group, respectively, with "90% effective samples are for training patterns and balance 10% are from verification patterns," whereas those are 99.32% and 98.46% with "80% effective samples are for training patterns and balance 20% are from verification patterns" and 96.75% and 95.58% with "70% effective samples are for training patterns and balance 30% are from verification patterns." It thus reveals that the neural network learning of the 2012 TTQS Review database retained good review mechanism; that is, for the businesses who introduced TTQS review to evaluate their execution of TTQS, the grades can verify TTQS review system as a good one.

The model effects shown by assessing the nodes in the graph are used, hoping to use the diagonal lines of the effects assessment graph to represent all the desired target data that this study can only locate by using 100% of the data as the basis, and, the built neural network model is further used to explore the model effects. When neural network learning is conducted on its learning group, it takes about 6% of its original data to achieve the same effect; again, when analyzing in the verification group, it takes about 7% to achieve the same effect. As such, it is clear that this kind of neural network analysis has high consistence in review (as in Figure 3).

4.2 Decision Tree Prediction Model

4.2.1 Goal variable and prediction variables

Only one goal variable in the study, the five performance levels of Platinum Medal (Grade A), Gold Medal (Grade B), Silver Medal (Grade C), Bronze Medal (Grade D) and the threshold (Grade E) categorized by TTQS review database. Prediction variables are 19 indicators of TTQS assessment. This study analyzed and compared for a decision tree algorithm, and chooses a higher accuracy rate algorithm as to discuss TTQS critical indicators. C5.0 shows better accuracy rate under any partition proportions. In which "80% for the training patterns and 20% for testing patterns" is the highest accuracy ratio (as in Table 5).

4.2.2 Decision tree results

Decision tree map is generated by IBM SPSS Modeler 14 software to find TTQS critical

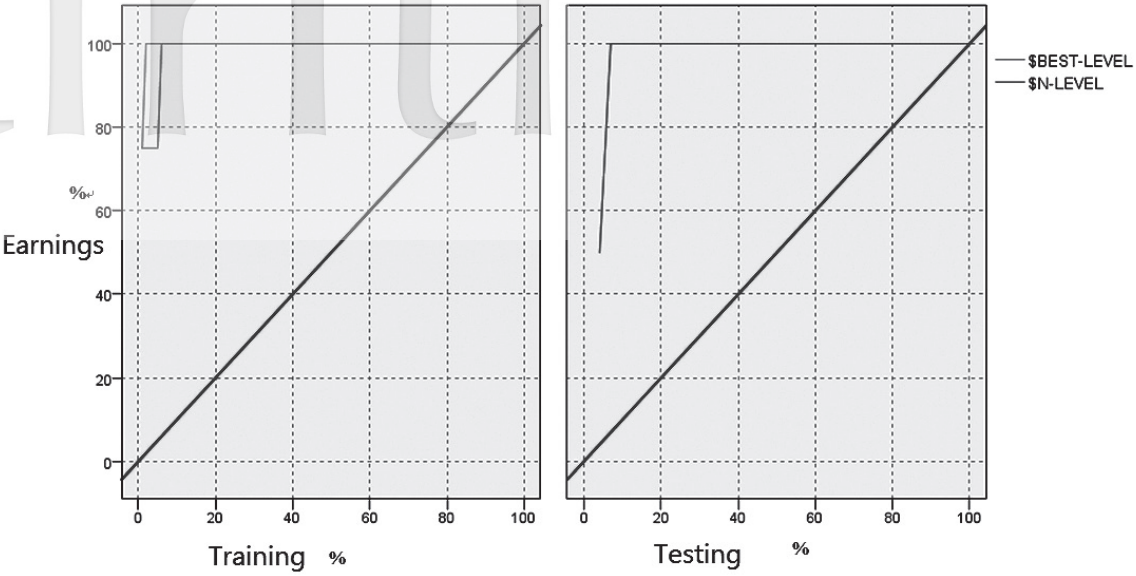

Figure 3. Training (learning) group and testing (verification) group model effects

Source: This study.

Table 5. Comparison of decision tree accuracy

Partition	70/30		80/20		90/10	
	Training accuracy	Testing accuracy	Training accuracy	Testing accuracy	Training accuracy	Testing accuracy
CHAID	93.74%	86.84%	92.70%	81.18%	94.44%	81.12%
C5.0	94.29%	87.72%	95.99%	89.41%	94.77%	88.68%
CART	90.61%	84.21%	89.78%	81.18%	87.91%	83.02%

Source: This study.

indicators of assessment. In Figure 4 decision tree association map, all assessment ratings are predicted by C5.0 algorithms.

Node 0 is the “B15-Assessment report and regular summarized analysis” and split to nodes 1 and 4. Node 1 includes 173 threshold and 29 bronze medals; its main decision-making rule is downward to predict the “threshold” and “bronze” level of assessment. Node 4 includes all the assessment levels; most of them are in bronze medals followed by silver medals. Node 4 main decision-making rule is downward to predict the “silver” and “gold” levels of assessment. Through the decision tree cross analysis, TTQS critical indicators include B1-Disclosure of institutional vision/mission/strategy and setting targets and requirements; B2-Specific training policy and target and top managers for training commitment and participation; B9-Process involvement by stakeholders; B10-Purchasing procedure and selection criteria of training products and services; B15-Assessment report and regular summarized analysis. These five critical indicators are attributed to Plan, Design, and Review construct (as in Table 6).

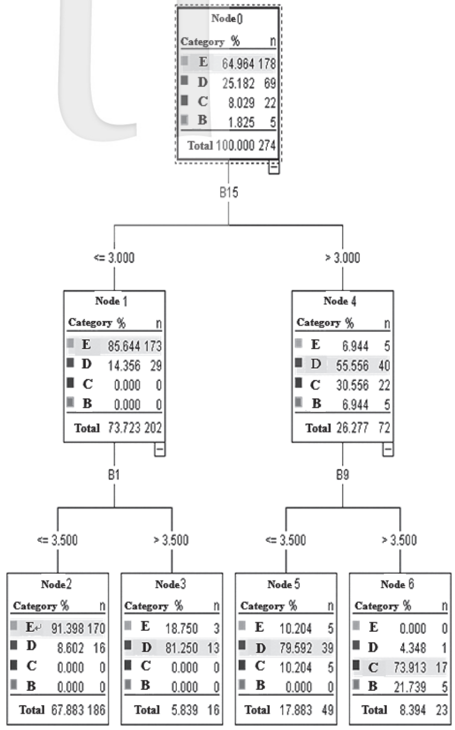

Figure 4. Decision tree C5.0 map

Source: This study.

Table 6. TTQS critical indicators by decision tree

Construct	Critical indicators
Plan	B1-Disclosure of institutional vision/mission/strategy. Setting targets and requirements.
Plan	B2-Specific training policy and target. Top managers for training commitment and participation.
Design	B9-Process involvement by stakeholders.
Design	B10-Purchasing procedure and selection criteria of training products and services.
Review	B15-Assessment report and regular summarized analysis.

Source: This study.

4.3 K-Means clustering results

K-Means clustering is generated by IBM SPSS Modeler 14 software based on the 4 assessment levels gold, silver, bronze, and threshold. Input data is 2012 TTQS assessment records.

In Figure 5, “Cluster-1” is generated from “gold, silver and bronze” medals, where “gold and silver” precise classification in the “Cluster-1.” Few of the “bronze” enterprises are classified in “Cluster-1,” the main reason is the enterprises in the “Outcome construct” indicators cannot get a higher score; they have to intensify “Outcome construct” indicators to get a better or higher assessment level. “Cluster-2” is generated from “Threshold;” it represents the enterprises of Cluster-2 without distinctive evidence of TTQS execution performance and they cannot grasp the critical assessment indicators.

"Cluster-3" and "Cluster-4" are composed of "bronze and threshold." These two clusters are between "lower bronze" and "higher threshold." The study found "threshold" enterprises of "Cluster-3" and "Cluster-4" are higher than the classification of the "Cluster-2." Thus, "bronze" enterprises still have to conduct a major improvement to get the higher assessment scores; "threshold" enterprises should grasp the critical indicators of TTQS assessment and continue forward to bronze medal or higher levels.

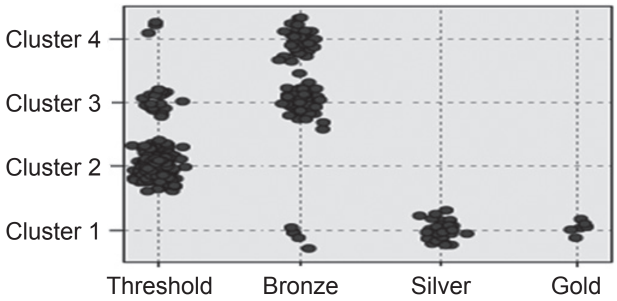

Figure 5. K-Means distribution charts

Source: This study.

In K-Means prediction results, TTQS database prediction variables important analysis represent indicators "B2-Specific training policy and target and top managers for training commitment and participation, B3-Specific PDDRO training system and specific core training categories and B4-Training quality management system and documents" in the construct of "Plan," indicators "B8-System design of training programs, B10-Purchasing procedure and selection criteria of training products and services and B11-Combination of training and target requirements" in the construct of "Design," indicators "B12- 12a. Planning of learner selection and compliance; 12b. Planning of selection and compliance of teaching materials; 12c. Planning of teacher selection and compliance; 12d. Planning of selection and compliance of teaching methods; 12e. Planning of selection for teaching environment and equipment and B14-Classification and filing of training materials and Systemization of management information" in the construct of "Do," and indicators "B15-Assessment report and regular summarized analysis and B16-Monitoring and handing for abnormal correction" in the construct of "Review" TTQS 10 critical assessment indicators .

4.4 Comparison K-Means clustering and decision tree algorithm results

Comparison TTQS 10 critical indicators by K-Means clustering and 5 critical indicators chosen by decision tree algorithm, this study arranged the TTQS critical indicators as in Table 7. In Table 7 indicates B2-Specific training policy and target and top managers for training commitment and participation, B10-Purchasing procedure and selection criteria of training products and services, B15-Assessment report and regular summarized analysis the absolutely important indicators of TTQS because these three indicators represent both in K-Means and decision tree algorithms. B1-Disclosure of institutional vision/mission/strategy and setting targets and requirements and B9-Process involvement by stakeholders are two important indicators because they represent in decision tree algorithm. B8-System design of training programs, B11-Combination of training and target requirements, B12- 12a. Planning of learner selection and compliance; 12b. Planning of selection and compliance of teaching materials; 12c. Planning of teacher selection and compliance; 12d. Planning of selection and compliance of teaching methods; 12e. Planning of selection for teaching environment and equipment, B14-Classification and filing of training materials and Systemization of management information are equally important indicators because these four indicators represent in K-Means clustering algorithm and have prediction variables higher than 90%.

Cross comparison with decision tree and K-Means, this study thus identify TTQS 9 critical indicators (as in Table 7) to help enterprise to introduce TTQS, grasp TTQS assessment critical indicators and enhance the quality of human training.

5. CONCLUSIONS

This study applies the back-propagation neural networks of the neural network to analyze 2012 TTQS new version review database. The forecasted precisions of the neural network are 96.75% and 95.58% for training group and testing group, respectively. It thus reveals that the neural network learning of the 2012 TTQS review database retained good review mechanism. When neural network learning is conducted on its training group, it takes about 6% of its original data to achieve the same effect; again, when analyzing in the testing group, it takes about 7% to achieve the same effect. As such, it is clear that this kind of neural network analysis has high consistence in review.

Comparison of the decision tree algorithms (CHAID, C5.0, CART) classification accuracy, in

Table 7. Comparison of decision tree and K-Means indicators

Algorithms	Decision tree	K-Means	Critical indicators
Assessment indicators	B1	B1	◆
	B2	B2	●
	B8	B8	▲
	B9	B9	◆
	B10	B10	●
	B11	B11	▲
	B12	B12	▲
	B14	B14	▲
	B15	B15	●

Source: This study.
Note: ● : Absolutely important; ◆ : Important; ▲: Equally important.

different partitions proportion, C5.0 algorithm has maintained better accuracy; the best accuracy rate is of 89.41%. Applications of C5.0 algorithm obtain B1, B2, B9, B10 and B15 five critical indicators. The K-Means clustering analysis is to identify 10 critical indicators of TTQS, B14, B8, B12, B11, B15, B10, B3, B16, B2, and B4, respectively. Cross comparison with decision tree and K-Means, this study identify the TTQS absolutely important indicators include B2, B10, and B15, important indicators include B1 and B9, and equally important indicators include B8, B11, B12, and B14. The study results obtained TTQS 9 critical assessment indicators include B1-Disclosure of institutional vision/mission/strategy; Setting targets and requirements, B2-Specific training policy and target; Top managers for training commitment and participation, B8-System design of training programs, B9-Process involvement by stakeholders, B10-Purchasing procedure and selection criteria of training products and services, B11-Combination of training and target requirements, B12- 12a. Planning of learner selection and compliance; 12b. Planning of selection and compliance of teaching materials; 12c. Planning of teacher selection and compliance; 12d. Planning of selection and compliance of teaching methods; 12e. Planning of selection for teaching environment and equipment, B14-Classification and filing of training materials and Systemization of management information, and B15-Assessment report and regular summarized analysis.

It is hoped that the assessment model provided in this study can assist in the promotion and execution of TTQS, providing better insight into the quality of human training inside the organization to effectively decrease costs and increase the probability of success in introduction. When serving as the strategy and exercise for the organizations to upgrade the values of human capital, it can, by referencing the measures this study offers as solutions, boost the competitive edge of the human capital for the entire organization.

Acknowledgements

The authors would like to thank Mr. Ya-lin Shieh for his comments and expresses the gratitude for the budget support from National Science Council (NSC 100-2410-H-167-010-MY3) to make this research possible.

References

Abidi, S. S. and Ong, J. A., 2000, A data mining strategy for inductive data clustering: a synergy between self-organising neural networks and K-Means clustering techniques, *Proceedings of TENCON 2000, Vol. 2*, 568-573.

Berry, M. J. A. and Linoff, G. S., 1997, *Data Mining Technique: For Marketing, Sale, and Customer Support*, John Wiley, New York.

Biggs, D., De Ville, B., and Suen, E., 1991, A method of choosing multiway partitions for classification and decision trees, *Journal of Applied Statistics*, 18(1), 49-62.

Breiman, L., 1984, *Classification and Regression Trees,* Chapman & Hall/CRC, Boca Raton, FL.

Chen, C.-J., 2010, *Prediction and Evaluation of Fitness for Shoe Insert with Artificial Neural Network*, Master Thesis, Tunghai University, Taichung, Taiwan.

Chen, R.-C. and Lin, Y.-C., 2010, Using back-propagation neural network to improve the RFID 3D indoor location system, *International Journal of Advanced Information Technologies*, 4(2), 44-55.

Chen, Y.-T., 2011, *Using Data Mining to Investigate the Risk Factors for Arterio-Venous Fistula Occlusion in Hemodialysis Patients*, Master Thesis, National Yunlin University of Science & Technology, Yunlin, Taiwan.

Chien, C.-F., Lin, Y.-S., and Cheng, J.-C., 2008, Construct fuzzy decision tree for mining interrelated semiconductor manufacturing data for yield enhancement, *Journal of Quality*, 15(3), 193-210.

Chin, J.-S., 2008, *To Construct the Performance Evaluation Mode for Domestic Enterprises to Carry out ISO 10015 Certification System*, Master Thesis, National Chin-Yi University of Technology, Taichung, Taiwan.

Han, J. and Kamber, M., 2000, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco, CA.

Hastie, T., Tibshirani, R., and Friedman, J., 2001, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York.

Hsu, S.-W., 2011, *Application of Neural Network on the Taiwan Training Quali System Impetus of Performance Evaluation*, Master Thesis, National Chin-Yi University of Technology, Taichung, Taiwan.

Investors in People, (accessed September, 2012), Investors in people. <https://www.investorsinpeople.com/>.

Kass, G.-V., 1980, An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, 29(2), 119-127.

Kulatilake, P. H. S. W., Wu, Q., Hudaverdi, T., and Kuzu, C., 2010, Mean particle size prediction in rock blast fragmentation using neural networks, *Engineering Geology*, 114(3-4), 298-311.

Lai, J. Z. C. and Liaw, Y.-C., 2008, Improvement of the K-Means clustering filtering algorithm, *Pattern Recognition*, 41(12), 3677-3681.

Lin, W.-T., Kung, C.-Y., and Lin, L.-L., 2009, Comparing IIP, ISO10015 and TTQS, *Quality Magazine*, 45(4), 52-56.

Lin, W.-T., Wang, S.-J., Wu, Y.-C., and Ye, T.-C., 2011, An empirical analysis on auto corporation training program planning by Data Mining Techniques, *Expert Systems with Applications*, 38(5), 5841-5850.

Lin, W.-T., Wu, Y.-C., Tung, C.-L., Huang, M.-R., and Qin, R.-S., 2010, Establishing ISO 10015 accreditation system performance model for domestic enterprises, *Expert Systems with Applications*, 37(6), 4119-4127.

National Association of Small & Medium Enterprises, R.O.C., 2012, *Analytic Report on Visit to IIP, an International Human Resources Quality System*, Bureau of Employment and Vocational Training, Council of Labor Affairs, Executive Yuan, Taipei.

Rumelhart, D. E., Widrow, B., and Lehr, M., 1994, The basic ideas in neural networks, *Communications of the ACM*, 37(3), 87-92.

Su, C. T., Yang, T., and Ke, C. M., 2002, A neural-network approach for semiconductor wafer post-sawing inspection, *IEEE Transactions on Semiconductor Manufacturing*, 15(2), 260-266.

Vellido, A., Lisboa, P. J. G., and Vaughan, J., 1999, Neural networks in business: a survey of applications (1992-1998), *Expert Systems with Applications*, 17(1), 51-70.

Wang, X. and Niu, R., 2010, Landslide intelligent prediction using object-oriented method, *Soil Dynamics and Earthquake Engineering*, 30(12), 1478-1486.

Wu, J.-D. and Liu, J.-C., 2011, Development of a predictive system for car fuel consumption using an artificial neural network, *Expert Systems with Applications,* 38(5), 4967-4971.

Yeh, T.-H., 2010, *The Application of Neural Network to Simulate Design Thinking Strategy*, Master Thesis, National Yunlin University of Science & Technology, Yunlin, Taiwan.

Yeh, I.-C., 1999, *Application and Practice of Neural Network Models*, Scholars Books, Taipei, Taiwan.

Yeh, I.-C., 2003, *Application and Practice of Neural Network Models*, Scholars Books, Taipei, Taiwan.

Yeh, I.-C., 2009, *Application and Practice of Neural Network Models*, Scholars Books, Taipei, Taiwan.

Yeh, I.-C. and Wu, P.-J., 2009, Taguchi method based on neural networks and cross validation methodology, *Journal of Quality*, 16(4), 261-279.

應用決策樹與類神經網路以提升人力訓練品質之績效

周永燦[1] 吳益銓[2*] 林文燦[2]
[1] 中原大學工業與系統工程學系
[2] 國立勤益科技大學工業工程與管理學系

摘 要

人力資源是組織中最重要的資產。在全球化的競爭趨勢下，人力資源成爲公司主要的核心部門，最根本的工作任務就是建立優秀的人力培訓系統以提高訓練品質。台灣訓練品質系統 TTQS 爲協助國內事業單位、訓練機構針對內部人力教育訓練的執行，提供一套完善的教育訓練品質系統，藉由 TTQS 系統的導入及運作，並依照 TTQS 制度之 PDDRO (Plan、Design、Do、Review、Outcomes) 五環程序標準及 19 項評核指標，評估企業所有人力訓練方案過程及結果之優劣，使執行後的結果與企業績效成爲一個系統性的整體規劃，讓人力教育訓練能更符合事業單位的需要。本研究應用資料探勘技術探討人力訓練品質與績效，以 2012 年 TTQS 新版評核資料庫爲探勘基礎，尋找 TTQS 之關鍵評核指標項目。首先應用倒傳遞類神經網路評估 TTQS 資料庫分類之準確率與績效，其訓練樣本與測試樣本預估準確率都高達 95% 以上。接著比較決策樹演算法 (C5.0, CART, CHAID)，選擇準確率較高之演算法來探討 TTQS 關鍵指標，發現 C5.0 演算法無論在任何資料分割比例下，皆保有較佳之學習準確率，其測試準確率最高達 89.41%。K-Means 集群分析法用來驗證比較決策樹演算法的結果。經交叉比對 K-Means 集群分析與決策樹 C5.0 演算之結果，本研究找出 TTQS 之重要關鍵評核指標九項。研究結果可幫助台灣之公司企業導入台灣訓練品質系統，掌握 TTQS 評核關鍵指標內容 , 強化人力訓練品質與績效。

關鍵詞：倒傳遞類神經網路、決策樹、IIP、K-Means、TTQS

收件日：103/06；修改日：103/08；接受日：103/12
[*] 聯絡作者：國立勤益科技大學工業工程與管理學系，41170 臺中市太平區坪林里中山路二段 57 號。
E-mail: jason_wu1102@yahoo.com.tw