Expert Systems with Applications 38 (2011) 5841–5850

Contents lists available at ScienceDirect

Expert Systems with Applications

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

An empirical analysis on auto corporation training program planning by data mining techniques

W.T. Lin [a], S.J. Wang [a], Y.C. Wu [b,*], T.C. Ye [a]

[a] Department of Industrial Engineering and Management, National Chin-Yi University of Technology, Taiwan
[b] Department of Industrial and System Engineering, Chung Yuan Christian University, Taiwan

ARTICLE INFO

Keywords:
TTQS (Taiwan TrainQuali System)
Human resources
Data mining
SOM (self-organization map)
K-means
Back-propagation artificial neural network

ABSTRACT

Under limited resources in corporation education training, to enhance human resources quality, making education training program planning more efficient is a significant issue in training future talents.

In accordance with Taiwan TrainQuali System (TTQS), the basic training structure is ton specify P (Plan) and D (Design). Ensuing results will be easier and successful. From TTQS database of Bureau of Employment and Vocational Training, corporations in Taoyuan, Hsinchu and Miaoli winning Gold Medals (Group B) have gaps outside control line in P and D. Enhancement is needed in the gap. The paper aims at a certain company winning Gold Medals in Taoyuan, Hsinchu and Miaoli to locate hidden or unobvious information with data mining, which will help future education training course planning and design.

The researchers use two-stage clustering (SOM and K-means) under data mining theory to collect personnel training data of Automobile Corporation A in Taiwan and China with data mining and analysis. The results under the two algorithms will serve as reference for future education training courses. In the end, in combination of back-propagation neural network to develop education training prediction model, the research offers reference for writing knowledge management system to enhance effects of personnel participation in training at corporations.

1. Study motivation and purpose

According to Lee and Wu (2007), in the knowledge economy age of the 21st century, *talents* are the major asset of an organization and the foundation to maintain competence: employees' knowledge and skill levels decide the operation performance of a company. Whether investment of a company in training truly helps enhance employees' competence and generates actual effects to reach the operation goal is the major evaluation of training results and the issue that enterprise owners and managerial persons care about.

Cascio (1991) pointed out that, in human resources accumulation, scholars believed education training was the most direct investment tool in human resources. Lee (1997) believed human resources development focused on talent development and training as well as organization growth. To avoid waste of resources or becoming superficial work, a lot of corporations were concerned about how to build an *education training mechanism* to improve organization performance and enhance productivity (Carlson, Bozeman, Kacmar, Wright, & McMahon, 2000).

The researcher wishes to conduct analysis and evaluation with effective use of the database. With manpower prediction evaluation model, current education training model is modified and education training system is established to obtain most suitable prediction model to offer suggestions for employee education training, enhance employee categorization accuracy and reduce waste of human resources while improving efficiency of training course planning.

The researcher analyzes variables of the case company education training database department category, employee ranks, and courses available. With data mining, the relation was understood. The researcher attempted to reach the following goals:

1. Analyzing past education training data with SOM and K-means to locate type of courses in the same nature as reference of future training course planning.
2. Establishing *education training prediction* model to explore training performance in light of practical viewpoints.
3. Verifying practical feasibility of *education training prediction* model in manufacture industry.

2. Literature review

2.1. Human resources

Chen (2007) mentioned that human resources management (HRM) was from labor management in Britain. Personnel manage-

* Corresponding author. Address: Department of Industrial and System Engineering, Chung Yuan Christian University, No. 200 Chung-Pei Rd., Chung Li City, Taoyuan County, 32023, Taiwan. Tel./fax: +886 4 23723808.
E-mail addresses: jason_wu1102@yahoo.com.tw, jason_wu@ms2.url.com.tw (Y.C. Wu).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.11.038

ment was used in the US. The term human resources management was not available until 1920's.

Hong (2007) held that, in the age of human capital, corporations held the key to enhanced competence if they had excellent employees. *Education* is the basic activity for corporations to seek sustainable development; *education training* is the major program for corporations to seek sustainable development. Thus, education training is indispensable to corporations. Education training is not only the key to maintain and train talents for better human resources. As we all know, competition among modern corporations is indeed competition of talents.

2.2. TTQS (Taiwan TrainQuali System)

In Service Industry Development Guidelines and Initiative (2004–2008) by Executive Yuan, talent training industry quality accreditation system was to be established in Talent Training Service Industry Development Measures. Council of Labor Affairs was in charge of planning introducing foreign training quality standard and submitted feasible strategies. Bureau of Employment and Vocational Training, in reference of ISO10015, European vocational training policies, Investors in People (IIP) in Britain, and aggressive vocational training policies in Australia and consideration of our challenges under global knowledge economic society in Taiwan, drafted our own Taiwan TrainQuali System (TTQS) in 2005 (see Fig. 1).

To ensure that management process model from training quality is planned, training process must be systemized to help the organization improve functions and satisfy the set training quality goal vision. Suggested revised TTQS has standard evaluation structure made of 17 training quality items in PDDRO. Each stage output is for the next stage output in a circular system.

2.3. Data mining

Data mining is to extract useful knowledge from information. Like digging minerals, we want to obtain interesting knowledge including unexplainable or undiscovered causal relation from chaotic, complicated database of great amount of information. In general, data mining has two major functions. The first one is to predict future trend and the other is to locate unknown patters (see Table 1).

From the preceding definitions, the purpose of data mining is a series of knowledge discovery analyses. As time changes data mining replaced knowledge discovery. The ultimate goal of data mining is to dig the rules that help decision making from the great amount of information.

Fig. 1. TTQS cycle. Source: Corporations training network, Taiwan.

2.4. Two-stage clustering

Sharma (1996) suggested using stratification and non-stratification if more precise categorization results were required. In the first stage stratification, Ward (1963) proposed Ward method for clustering to decide number of clusters. Cluster was made with K-means in the 2nd stage. In the first stage, Ward method was stratified clustering analysis. When two individuals were clustered together, they were always in the same cluster. The K-means in the 2nd stage offset the disadvantage to reach best number of clusters being homogeneous in the clusters and heterogeneous among clusters.

Using two-stage clustering for clustering reduces calculation cost. In direct use of K-means, users have to keep trying locating the most appropriate number of clusters. Two-stage method locates the initial cluster and saves calculation time cost.

2.4.1. SOM (self-organization Map)

SOM is a kind of non-supervision learning networks and was proposed by Kohonen (1990). The principle is that brains have the feature of *birds of a feather flock together*. The SOM imitates this feature. Output processing units affect one another. When network learning is completed, adjacent output processing units have similar functions. That is, they are similar for clustering.

SOM is in elicitation method structure. The learning termination condition is not optimization of models of input information processing. The final weight vectors are often related to training sample input learning order. Different initial conditions lead to heterogeneous output results. In different data collections, SOM algorithms such as learning rate and adjacent functions can only have best clustering effect through continuous learning cycle.

2.4.2. K-means

According to Buttrey and Karo (2002), K-means method was mostly frequently used in non-stratified clustering analysis. It required pre-determined number of clusters. Inappropriate number of clustering will lead to vague difference among clusters. In selecting clusters, it is recommended selecting different number of clusters for more algorithms to have reasonable explanation.

2.5. Neural network

Neural network was originated from 1950's. Scientists, in imitation of human brain organization and operation, designed neuron work math model (perception and association learning rules). Neural network could learn on its own (like human brains). Users did not need to design complicated programs to solve problems. With data, neural network could have self-learning.

Human brains have approx. 1011 nerve cells. Each cell has 104 synapses connected with other cells to form complicated nerve network. Ye (1999) mentioned that neural network was like

Table 1
Data mining definitions.

Scholars and experts	Definition
Peacock (1998)	Narrow sense mainly in machine learning methodology Broad sense emphasizing discovery of knowledge from database
Hand, Blunt, Kelly, and Adams (2000)	The process of locating interesting and valuable information from database
Berson, Smith, and Thearling (2001)	The attraction of data mining is it can predict, not review
Cheng, Chang, and Liu (2005)	Data mining helps increase understanding of customers' needs and behaviors to make customized service and enhance connection, communication and interaction with customers