Expert Systems with Applications 38 (2011) 1364–1373

Contents lists available at ScienceDirect

Expert Systems with Applications

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

An integrated Data Envelopment Analysis–Artificial Neural Network–Rough Set Algorithm for assessment of personnel efficiency

Ali Azadeh [a,*], Morteza Saberi [b,c], Reza Tavakkoli Moghaddam [a], Leili Javanmardi [a]

[a] Department of Industrial Engineering and Center of Excellence for Intelligent-Based Experimental Mechanics, College of Engineering, University of Tehran, Iran
[b] Department of Industrial Engineering, University of Tafresh, Iran
[c] Institute for Digital Ecosystems & Business Intelligence, Curtin University of Technology, Perth, Australia

ARTICLE INFO

Keywords:
Personnel
Efficiency
Data Envelopment Analysis (DEA)
Artificial Neural Network (ANN)
Rough Set Theory (RST)
Cross Validation Test Technique (CVTT)

ABSTRACT

Personnel specifications have greatest impact on total efficiency. They can help us to design work environment and enhance total efficiency. Determination of critical personnel attributes is a useful procedure to overcome complication associated with multiple inputs and outputs. The proposed algorithm assesses the impact of personnel efficiency attributes on total efficiency through Data Envelopment Analysis (DEA), Artificial Neural Network (ANN) and Rough Set Theory (RST). DEA has two roles in the proposed integrated algorithm of this study. It provides data ANN and finally it selects the best reduct through ANN result. Reduct is described as a minimum subset of attributes, completely discriminating all objects in a data set. The reduct selection is achieved by RST. ANN has two roles in the integrated algorithm. ANN results are basis for selecting the best reduct and it is also used for forecasting total efficiency. The proposed integrated approach is applied to an actual banking system and its superiorities and advantages are discussed.

0. Significance

This is the first study which proposes an integrated algorithm for assessment of the impact of personnel efficiency attributes on total efficiency through Data Envelopment Analysis (DEA), Artificial Neural Network (ANN) and Rough Set Theory (RST). The outcome helps managers to construct helpful system to forecast efficiencies by selected attributes. The integrated algorithm is successfully applied to 102 branches of a large private bank, evaluating personnel attributes impact on bank branches efficiencies. The results of this study show that four subsets of conditional attributes with total number of nine attributes from 28 attributes have a critical impact on the accuracy of the optimal solution. This reduction in attributes number decrease the time of decision-making and consequently reduces the cost of efficiency evaluation.

1. Introduction

Efficiency is a key concept for organizations. Too many immeasurable influences and complex relationships among attributes influence efficiency in organizations. Efficiency relevant to human attributes is a goal that is rarely questioned in contemporary organizations. As personnel specifications have greatest impact on efficiency, they can help us to design work environments and enhance total efficiency. As providing information on multiple input and output factors are a complicated and time-consuming procedure, determining critical personnel attributes is useful. The purpose of proposed integrated algorithm in present study is to alert management to the important attributes that should be considered if an effective decision to enhance efficiency is to be formulated. This is because there is a great desire to identify the critical attributes for sensitivity analysis of inefficient decision-making units (DMUs) regarding efficiency attributes. The algorithm proposes an analytic function that predicts these attributes exactly. This model is applicable for all problems associated with decision-making in organizations composed of decision-making units (DMUs) and will be valuable for executives and senior managers. The outcome helps managers to construct helpful system to forecast DMUs efficiency by selected attributes. Furthermore, reduction in attributes number decrease the time of decision-making and consequently reduces the cost of efficiency analysis.

The integrated algorithm uses Data Envelopment Analysis (DEA), ANN and Rough Set Theory (RST). DEA has two roles in the proposed integrated algorithm of this study. It provides data ANN and finally it selects the best reduct. Reduct is described as a minimum subset of attributes, completely discriminating all objects in a data set. The reduct selection is achieved by RST. ANN has two roles in the integrated algorithm. ANN results are basis for

* Corresponding author.
E-mail address: aazadeh@ut.ac.ir (A. Azadeh).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.07.033

selecting the best reduct and it is used for forecasting total efficiency. The proposed algorithm is applied to an actual banking system.

Weakness of DEA in forecasting is the reason to use ANN (Wu, Yang, & Liang, 2006). ANN has been viewed as a good tool to approximate nonlinear problems and is useful for manager in predicting system (Cybenko, 1989; Patuwo, Hu, & Hung, 1993). In addition, RST and ANN are combined to construct the forecasting feature of the proposed algorithm. The best reduct is selected when ANN performance is at its best with the selected reduct. In order to apply necessity available error measurement of ANN in best reduct determination procedure, DEA is used.

1.1. ANN and efficiency

Athanassopoulos and Curram (1996) first introduced the idea of combination of neural networks and DEA for classification and/or prediction. They treated DEA as a preprocessing methodology to screen training cases in a study. Their application is bank with multi-output: 4 inputs, 3 outputs. After selecting samples, the ANNs are then trained as tools to learn a nonlinear forecasting model. They assume that inefficiency distributions are semi-normal and exponential and conclude that DEA is superior to ANN for measurement purpose. Their study indicates that ANN results are more similar with the constant returns to scale and less with the variable returns to scale results. The latter, is a consequence of the implicit assumption of constant returns to scale adopted by the ANN models.

Costa and Markellos (1997) analysed the London underground efficiency with time series data for 1970–1994 where there are 2 inputs – fleet and workers – and 1 output – km s. They explain how the ANNs results are similar to COLS and DEA. They proposed two procedures: (a) similar way to COLS after neural training; (b) by an oversized network until some signal to noise ratio is reached. Then, inefficiency is determined as observation-frontier distance. However, ANNs offer advantages in the decision-making, the impact of constant versus variable returns to scale or congestion areas (Costa & Markellos, 1997). Santin and Valino (2000) study on education efficiency by a two-level model: student – production function is estimated by ANNs – and school. They infer that ANN is superior to econometric approach at frontier estimation. Pendharkar and Rodger (2003) used DEA as a data screening approach to create a sub sample training data set that is 'approximately' monotonic, which is a key property assumed in certain forecasting problems. Their results indicate that the predictive power of an ANN trained on the 'efficient' training data subset is stronger than the predictive performance of an ANN trained on the 'inefficient' training data subset. Santin, Delgado, and Valino (2004) used a neural network approach for a simulated nonlinear production function and compared its performance with conventional alternatives such as stochastic frontier and DEA in different observations and noise scenarios. The results suggested that ANNs are a promising alternative to conventional approaches, to fit production functions and measure efficiency under nonlinear contexts. Wu et al. (2006) presented a DEA–NN study for performance assessment of branches of a large Canadian bank. The results are operable to the normal DEA results overall. They concluded that the DEA–NN approach produces a more robust frontier and identifies more units that are efficient because better performance patterns are explored. Furthermore, for worse performers, it provides the guidance on how to improve their performance to different efficiency ratings. Ultimately, they concluded the neural network approach requires no assumptions about the production function (the major drawback of the parametric approach) and it is highly flexible. ANNs have been viewed as a good tool to approximate numerous nonparametric and nonlinear problems.

Azadeh, Ghaderi, Anvari, and Saberi (2006a), Azadeh, Ghaderi, Anvari, Saberi, and Izadbakhsh (2007) and Azadeh, Ghaderi, Anvari, and Saberi (2007) proposed a highly unique flexible ANN algorithm to measure and rank the decision-making unit's (DMUs) efficiency. Their algorithm calculated efficiency score of Iran steam power plants in 2004. Results indicate that the proposed algorithm estimates the values of efficiency scores closer to the ideal efficiency. They concluded that the propose algorithm estimates more robust results and more efficient units than the conventional approach because better performance patterns are explored. In addition, they proposed a method to integrate their pervious algorithm (Azadeh, Anvari, & Saberi, 2007, 2008).

1.2. Rough Set Theory

Several immeasurable influences and complex relationships among attributes impact efficiency in organizations. Rough Set Theory (RST) proposed by Pawlak, is one of the techniques for the identification and recognition of common patterns in data Pawlak (1982, 1991). This technique has found applications in knowledge discovery from data bases, data mining, fault diagnosis, machine learning, knowledge acquisition, expert systems and decision support systems (Błaszczyński, Greco, & Słowiński, 2007; Fan, Liu, & Tzeng, 2007; Inuiguchi & Miyajima, 2007). It is also used to study uncertainty (Beynon & Peel, 2001; Lili & Zhi, 2001; Ziarko, 1993), prediction (Becerra-Fernandez, Zanakis, & Walczak, 2002; Kusiak, Kern, Kernstine, & Tseng, 2000; Sanchis, Segovia, Gil, Heras, & Vilar, 2007), service organizations (Chou, Cheng, & Chang, 2007; Hassanien, 2007; Kowalczyk & Slisser, 1997; Sikder & Gangopadhyay, 2007; Tsumoto, 1997), financial firms (Ravi Kumar & Ravi, 2007; Ruhe, 1996; Shyng, Wang, Tzeng, & Wu, 2007), and scheduling problems (Liu, Chen, Wu, & Li, 2006; Triantaphyllou, Liao, & Iyengar, 2002). Stefanowski and Slowinski have studied rough sets as a tool for feature selection by studying attribute dependencies (Stefanowski & Slowinski, 1997). Kusiak and Tseng (2000) have proposed two independent algorithms for accurate feature selection in medical, industrial and engineering case studies. Others like Xia and Wu discusses feature extraction technique of Rough Set Theory for supplier selection to select best suppliers according to different tangible and intangible attributes (Xia & Wu, 2007). Moreover, there are some other application of Rough Set Theory to feature selection in customer relationship management (Tseng & Huang, 2007), product quality evaluation (Zhai, Khoo, & Fok, 2002) and healthcare (Wang, Yang, Jensen, & Liu, 2006). However, existing heuristic rough set approaches to feature selection are insufficient at finding optimal reductions. On the other hand, it is not feasible to search for optimal in even in average sized datasets. Therefore, the combination of this method by other robust data mining tools may help practitioners to go further into feature selection to obtain more accurate results.

1.3. Data Envelopment Analysis

DEA is a non-parametric method that uses linear programming to calculate the efficiency in a given set of decision-making units (DMUs). The DMUs that make up a frontier envelop, the less efficient firms and the relative efficiency of the firms is calculated in terms of scores on a scale of 0 to 1, with the frontier firms receiving a score of 1. DEA models can be input or output oriented and can be specified as constant returns to scale (CRS) or variable returns to scale (VRS).

1.4. Artificial Neural Networks

An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous sys-