Expert Systems with Applications 36 (2009) 9105–9112

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Supplier selection: A hybrid model using DEA, decision tree and neural network

Desheng Wu *

Reykjavik University, Kringlunni 1, IS-103 Reykjavík, Iceland and RiskLab, University of Toronto, 1 Spadina Crescent, Toronto, Canada ON M5S 3G3

ARTICLE INFO

Keywords:
Supplier selection
Data envelopment analysis (DEA)
Decision tree (DT)
Neural networks (NNs)
Data mining (DM)
Classification
Prediction

ABSTRACT

As the most important responsibility of purchasing management, the problem of vendor evaluation and selection has always received a great deal of attention from practitioners and researchers. This management decision is a challenge due to the complexity and various criteria involved. This paper presents a hybrid model using data envelopment analysis (DEA), decision trees (DT) and neural networks (NNs) to assess supplier performance. The model consists of two modules: Module 1 applies DEA and classifies suppliers into efficient and inefficient clusters based on the resulting efficiency scores. Module 2 utilizes firm performance-related data to train DT, NNs model and apply the trained decision tree model to new suppliers. Our results yield a favorable classification and prediction accuracy rate.

1. Introduction

Supply chain vendor evaluation is a very important operational decision, involving not only selection of vendors, but other decisions with respect to quantities to order from each vendor. Globalization has led to the opportunities for many to utilize sources from around the world. This, of course, introduces additional decision-making considerations. Vendor selection decisions are complicated by the fact that various criteria must be considered in the decision-making process. Dickson (1966), in one of the early works on supplier selection, identified over 20 supplier attributes which managers trade off when choosing a supplier. The criteria may have quantitative as well as qualitative dimensions. A strategic approach towards purchasing may further emphasize the need to consider multiple criteria. In the case of strategic supplier selection, Wu (in press) stressed the need not only to consider traditional criteria such as price and quality but also longer term and qualitative criteria such as "strategic fit" and "assessment of future manufacturing capabilities". Nassimbeni (2006) surveyed 78 Italian enterprises concerning their international sourcing, finding that quality and technological content were the highest ranked criteria for vendor selection, with cost ranked only fifth. Purchasing portfolio models have been extensively studied and findings indicate that there is no simple, standardized blue print for the application of the portfolio analysis (Olsen & Ellram, 1997; Gelderman & van Weele, 2003).

On the other hand, supplier selection requires the information about potential suppliers' credit history, performance history and other personal information, which are often not available to the public. Therefore, data available to supplier selection often incur problems such as small dataset available to the public, missing values, inconsistent values, errors, etc. In addition, companies conducting supplier performance evaluation always have a great deal of data but lack the knowledge of the data. That is to say, these data are not fully and effectively explored and used and they cannot provide predictive functions for the future decision-making.

Many quantitative models have been proposed for vendor selection in supply chains (De Boer, Labro, & Molrlacchi, 2001). Fuzzy programming was proposed by Kumar, Vrat, and Shankar (2006) to allow consideration of various levels of uncertainty. Among various quantitative methods, both DEA analysis (Wu and Olson, 2008) and data mining (DM) techniques have been presented in vendor assessment. DEA method aids the buyer in classifying the suppliers (or their initial bids) into two categories: the efficient suppliers and the inefficient suppliers. Weber has primarily discussed the application of DEA in supplier selection in several publications; see Weber and Ellram (1993), and Weber and Desai (1996). Apart from simply categorizing suppliers, Weber demonstrated how DEA can be used as a tool for negotiating with inefficient suppliers. However, classical DEA often fails to work effectively since DEA calls for restrictions on data such as the requirement of rule of thumb, no outliers and statistical noise (Li & Reeves, 1999; Wu, 2006).

DM approaches use historical data to train a sui model and make prediction of new supplier performance with the trained model. The application cases of DM approaches to supplier selection include neural networks and expert systems (Albino & Garavelli, 1998; Khoo, Tor, & Lee, 1998). Due to the recent development of some methods, relatively few reports of such applications can be found in the literature. Thus it is important to investigate these methods and examine their potential. DM

* Tel.: +1 416 8805219.
E-mail address: dash@ru.is.

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2008.12.039

approaches such as decision trees (DT) and neural networks (NNs) provide good tools to approximate numerous nonparametric and nonlinear problems as well as qualitative data. The main advantage of DM algorithms involves: handling of qualitative attributes; flexibility in dealing with missing information; exploitation of large data sets for model development goal by efficient computation procedures and derivation of easily understandable classification models (classification rules or trees). The main disadvantage of using DM is that most DM techniques requires the known value of the target. In the supplier selection problem, we do not know the actual class of the supplier. This can be accomplished by DEA where efficient and inefficient signals are provided to classify all suppliers.

This paper presents a hybrid model to assess supplier performance using data envelopment analysis (DEA), decision trees (DT) and neural networks (NNs). The model consists of two modules: Module 1 applies DEA and classifies suppliers into efficient and inefficient clusters based on the resulting efficiency scores. Module 2 utilizes firm performance-related data to train DT, NNs model and apply the trained decision tree model to new suppliers. Our results yield a favorable classification and prediction accuracy rate. Moreover, to our knowledge, there is no work to analyze the supplier selection problem by jointly using DEA and DM approaches. It is very attractive to use DEA and DM approaches to develop an integrated model, which exhibits the advantages of both DEA and DM approaches. This route has been proved effective in our previous work in the field of efficiency analysis of financial entities (Wu, Yang, & Liang, 2006; Wu, 2006). De Boer et al. (2001) argues that not all methods are equally useful in every possible purchasing situation. Our demonstration shows a purchasing situation where purchasing information is characterized with small dataset, qualitative data, and missing values.

We begin in the following section with the proposed methodology. Section 2 provides the models and methodology utilized in this paper. Section 3 gives the DEA/DM results and further discussion. Section 4 discusses handling of qualitative data and missing data. Finally, our conclusions are presented in Section 5.

2. Models and methodology

The model can function as a classification model or a regression model, based on the problem taken into account. For either the classification model or the regression model, it generally consists of two modules. Module 1 applies DEA to calculate the DEA score given to each supplier. If classification of suppliers is of interest, the calculated DEA scores are used to derive the class for each supplier, typically classified as efficient and inefficient clusters. Module 2 utilizes supplier performance-related data to train decision tree or numeral network model and apply the trained models to new suppliers.

2.1. DEA supplier selection model

Data envelopment analysis (DEA) is a nonparametric programming approach for evaluating the relative efficiency of DMUs with performance characterized by multiple attributes. DEA supplier selection models have been presented by Kleinsorge, Schary, and Tanner (1992), Weber and Desai (1996), Liu, Ding, and Lall (2000), and Wu and Olson (2008). Considered game theoretic models for negoations in vendor selection, citing the closeness of his approach to DEA. Listed no fewer than 14 different methodologies over the period 1969 through 2003 (including multiple criteria methods and DEA). Wu and Olson (2008) consider uncertainty in vendor selection and compares stochastic DEA and stochastic dominance.

DEA is based on the ‘efficiency’ analysis of alternative suppliers. The suppliers are evaluated on benefit criteria (outputs) and cost criteria (inputs). Supplier efficiency is defined as the ratio of the weighted sum of its outputs (i.e. the performance of the supplier) to the weighted sum of its inputs (i.e. the costs of using the supplier). We present DEA basics in the Appendix. Assume that there are n suppliers indexed by $_j(j = 1, 2, \dots, n)$ to be evaluated. The jth supplier, i.e., S_j has m different inputs x_{ij} and s different outputs y_{rj}. Let the observed input and output vectors of S_j be $X_j = (x_{1j}, x_{2j}, \dots, x_{mj})^T > 0$, $j = 1, 2, \dots, n$, and $Y_j = (y_{1j}, y_{2j}, \dots, y_{sj})^T > 0$, respectively. The relative efficiency of S_j is calculated as

$$S_j = \frac{\sum_{r=1}^{s} u_r y_{kj}}{\sum_{i=1}^{m} v_i x_{ij}} = \frac{U^T Y_j}{V^T X_j},\ j = 1, 2, \dots, n,$$

where $V = (v_1, v_2, \dots, v_m)^T$ and $U = (u_1, u_2, \dots, u_s)^T$ are input and output weight vectors, respectively. The calculation of the relative efficiency of S_j can be accomplished by solving a set of linear programming (see the Appendix) Wu (2009).

2.2. Decision tree and neural network models

A decision tree is a graphical representation of a procedure for classifying or evaluating an alternative of interest. By graphical representation, they clearly show how to reach a decision, and they are able to construct automatically from labeled instances. Two well-known programs for constructing decision trees are C4.5 (Quinlan, 1993) and CART (classification and regression tree) (Breiman, Friedman, Olshen, & Stone, 1984).

A decision tree is typically constructed recursively in a top-down manner (Friedman, 1977). If a set of labeled instances is sufficiently pure, then the tree is a leaf, with the assigned label being that of the most frequently occurring class in that set. Otherwise, a test is constructed and placed into an internal node that constitutes the tree so far. The test defines a partition of the instances according to the outcome of the test as applied to each instance. A branch is created for each block of the partition, and for each block, a decision tree is constructed recursively.

Neural networks provide a new way for feature extraction (using hidden layers) and classification (e.g. multilayer perceptrons). In addition, existing feature extraction and classification algorithms can also be mapped into neural network architectures for efficient (hardware) implementation. For classification or prediction, backpropagation neural network (BPNN) is the most widely used neural network technique (Liang and Wu, 2005). This algorithm is adopted in this study.

2.3. Hybrid conceptual model

Fig. 1 depicts the *conceptual* model for supplier selection using DEA, DT and NNs. As mentioned before, the hybrid model can function as both a classification model and a regression model. For either the classification model or the regression model, it generally consists of two modules. Module 1 applies two-stage DEA and classifies suppliers into efficient and inefficient clusters based on the computed efficiency scores. Module 2 is a classification or regression module based on the decision tree or the neural network, which utilizes supplier performance-related data to train decision tree or the neural network model and apply the trained classifier or predictor to new suppliers. Qualitative variables and missing values can be introduced and handled by the tree or the neural network model, as argued in previous sub-section. In Module 2 in the figure, the goal is to address the classification or the regression problem, which involves the development of a relationship (e.g. a function or a rule) between the classes and the criteria. To develop such a functional relationship, it is usually required that the classification error rate or prediction errors between the priori efficiency