Expert Systems with Applications 39 (2012) 7763–7775

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Analyzing the solutions of DEA through information visualization and data mining techniques: SmartDEA framework

Alp Eren Akçay [a,1], Gürdal Ertek [a,*], Gülçin Büyüközkan [b]

[a] Sabancı University, Faculty of Engineering and Natural Sciences, Orhanli, Tuzla, 34956 Istanbul, Turkey
[b] Department of Industrial Engineering, Galatasaray University, Çırağan Cad No: 36, 34257 Ortaköy, Istanbul, Turkey

ARTICLE INFO

Keywords:
Data envelopment analysis (DEA)
Information visualization
Data mining
Decision support system framework

ABSTRACT

Data envelopment analysis (DEA) has proven to be a useful tool for assessing efficiency or productivity of organizations, which is of vital practical importance in managerial decision making. DEA provides a significant amount of information from which analysts and managers derive insights and guidelines to promote their existing performances. Regarding to this fact, effective and methodologic analysis and interpretation of DEA results are very critical. The main objective of this study is then to develop a general decision support system (DSS) framework to analyze the results of basic DEA models. The paper formally shows how the results of DEA models should be structured so that these solutions can be examined and interpreted by analysts through information visualization and data mining techniques effectively. An innovative and convenient DEA solver, SmartDEA, is designed and developed in accordance with the proposed analysis framework. The developed software provides DEA results which are consistent with the framework and are ready-to-analyze with data mining tools, thanks to their specially designed table-based structures. The developed framework is tested and applied in a real world project for benchmarking the vendors of a leading Turkish automotive company. The results show the effectiveness and the efficacy of the proposed framework.

1. Introduction

Data envelopment analysis (DEA) is a widely used method in performance evaluation and benchmarking of a set of entities. The popularity of DEA can be easily confirmed in the article of Emrouznejad, Parker, and Tavares (2008) that has summarized previous DEA contributions during the past three decades. Its convenience in assessing the multiple input and output variables of these entities by not requiring congruity and an apriori relationship makes it a very popular management tool in many application areas. Another reason for its wide use is the managerial insights that come up with the solution of a DEA model. For instance, DEA assigns a peer group or reference set for an inefficient entity. The entity can take the entities in this reference set as role models in accordance with the assigned weights. Another important DEA result is the target values or projections for the input and output variables of an inefficient entity to achieve full efficiency. DEA thus provides significant amount of information from which analysts and managers derive insights and guidelines to enhance their existing performances. Regarding to this fact, effective and methodologic analysis and interpretation of DEA solutions is very critical (El-Mahgary & Lahdelma, 1995; Emrouznejad & De Witte, 2010; Yadav, Padhy, & Gupta, 2010).

The main contribution in this study is the development of a general framework that enables DEA analysts to extract the most important and interesting insights in a systematic manner. In order to do so, a computer science and data mining perspective is adopted for designing the structure of the DEA results. Various data mining and information visualization techniques can be appropriate for the analysis of different types of DEA models (Lin, Lin, Li, & Kuo, 2008; Seol, Lee, & Kim, 2011). The paper provides a fundamental basis for the implementation of these techniques in the DEA solutions. A convenient and general notation is proposed for the DEA data included in the model, the other data and the results data generated by DEA solvers. The ultimate goal of the study is then to build a structure framework for the analysis of DEA results, enabling researchers and practitioners to make analytical benchmarking and performance evaluations.

In accordance with the proposed framework, a user-friendly and convenient DEA software, SmartDEA, is designed and developed. The software generates DEA solutions with a structure consistent with the framework. The analysis steps performed by a DEA analyst are importing the model data, analyzing the data by solving the appropriate DEA model, and making analytical inquires on the generated solution data to evaluate and benchmark

* Corresponding author. Tel.: +90 216 483 9568; fax: +90 216 483 9550.
E-mail address: ertekg@sabanciuniv.edu (G. Ertek).

1 Present address: Carnegie Mellon University, Tepper School of Business, 5000 Forbes Avenue, Pittsburgh, 15213 PA, USA.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.059

the entities assessed in DEA model. The solution data generated by the software allows analysts to integrate the results and many of the data mining and information visualization techniques in a convenient and effective manner.

The rest of this paper is organized as follows: after a brief introduction to the DEA, the theory of the basic DEA models are explained in Section 2, providing a fundamental background for DEA. Since it is critical to know when and where DEA is an appropriate method, advantages and drawbacks of DEA are also explained in the same section. Proposed framework, which is based on the integration of DEA results with data mining and information visualization techniques, is presented in Section 3. An overview of existing DEA software is followed by the presentation of the developed DEA solver, SmartDEA, in Section 4. The testing of the framework and SmartDEA with the real world data of an automotive company is illustrated in Section 5. Concluding remarks are summarized in Section 6.

2. Data envelopment analysis (DEA)

2.1. Introduction to the DEA

DEA is a nonparametric performance evaluation technique with a wide range of application areas within various disciplines. In their breakthrough study, Charnes, Cooper, and Rhodes (1978) define DEA as "a mathematical programming model applied to observational data and a new way of obtaining empirical estimates of relations such as the production functions and/or efficient production possibility surfaces". Since its introduction in 1978, researchers in a multitude of disciplines distinguished and adopted DEA as a promising tool that is easily applicable to their fields for performance evaluation and benchmarking of entities and operational processes. A given set of entities, referred to as *decision making units* (DMU) in DEA terminology, can be conveniently compared in terms of multiple inputs and multiple outputs, assuming neither a specific form of relationship between inputs and outputs, nor fixed weights for the inputs and outputs of a DMU.

DEA firstly provides an *efficiency score* between 0 and 1 for each DMU involved in the analysis. The efficiency score for a DMU is determined by computing the ratio of total weighted outputs to total weighted inputs for it. DEA enables variable weights, which are calculated in such a way that the efficiency score for the DMU is maximized (Cooper, Seiford, & Tone, 2006). For a DEA model with *n* different DMUs, *n* different linear programming (LP) optimization models are solved to compute the efficiency scores of each of the DMUs.

A basic DEA model can provide important metrics and benchmarks for monitoring the comparative performances of entities in a group and take managerial actions to improve them. An *efficient frontier* or *envelopment surface*, which is drawn over the "best" DMUs, is the critical component of a DEA model. It is formed by the efficient DMUs which have efficiency scores of 1. The efficiency score of a DMU is basically the distance from each DMU to this efficient frontier. The efficiency scores of the inefficient DMUs are calculated in accordance with this distance represented as a Pareto ratio. Besides the efficiency scores, another result offered by DEA is the *reference sets*, or peer (sub) groups. For each inefficient unit, DEA identifies a set of corresponding efficient units that can serve as a benchmark peer group for the selected DMU. The solution of the LP formulation of the model results in the reference set for each DMU. Knowing that the DMUs in the reference set are efficient and have the same input and output structure, they can be regarded as "good" examples operating practices for the corresponding inefficient DMU (Boussofiane, Dyson, & Thanassoulis, 1991). The percentages of each reference set unit contributing to the composite unit (i.e. virtual producer – with respect to which the efficiency score of the inefficient DMU is found) are also computed by the DEA model.

One important shortcoming of DEA is the rank ordering of the efficiency scores for each DMU. Each inefficient DMU is labeled as such only in comparison to the given group, and this may be misleading: An inefficient DMU may have been labeled as efficient if it were part of an inferior group. Conversely, an efficient DMU may not necessarily be efficient when compared as a part of another group. Also, theoretically, only the DMUs with the same reference sets can be strictly rank ordered (Avkıran, 1999). Yet another important result DEA provides is the *target inputs* and *target outputs*, which are referred to as *projections*. They represent up to which value an input should be decreased while keeping the outputs at the same levels (i.e. input orientation) or how much an output should be increased while the input level remains unincreased (i.e. output orientation), respectively, so that the DMU becomes efficient.

2.2. Basic DEA models

Since its introduction by Charnes et al. (1978) in their seminal work *measuring efficiency of decision making units*, the CCR model served as the origin of many following ideas and models in DEA literature. Cooper et al. (2006) discusses the model in detail together with the classical alternative models. The theoretical formulations and summaries in this section are based on Cooper et al. (2006).

Let's suppose that there are n DMUs in the model: $\mathrm{DMU}_1, \mathrm{DMU}_2, \ldots, \mathrm{DMU}_n$. Suppose there are m inputs and s outputs for each one of them. For DMU_j the inputs and outputs are represented by (x_{1j}, x_{2j}, x_{mj}) and (y_{1j}, y_{2j}, y_{sj}), respectively. As stated previously, for each DMU_o, DEA tries to maximize the ratio

$$\frac{\text{Virtual output}}{\text{Virtual input}} \tag{1}$$

where

$$\text{Virtual input} = v_1 x_{1o} + \cdots + v_m x_{mo} \tag{2}$$

$$\text{Virtual output} = u_1 y_{1o} + \cdots + u_s y_{so} \tag{3}$$

and the weights v_i and u_r are not fixed in advance. Best weights are assigned according to the solution of the following fractional DEA model (**M1**):

$$\begin{aligned} (FP_o) \quad \max \quad & \theta = \frac{u_1 y_{1o} + \cdots + u_s y_{so}}{v_1 x_{1o} + \cdots + v_m x_{mo}} \\ \text{s.t.} \quad & \frac{u_1 y_{1o} + \cdots + u_s y_{so}}{v_1 x_{1o} + \ldots + v_m x_{mo}} \leqslant 1 \\ & v_1, v_2, \ldots, v_m \geqslant 0 \\ & u_1, u_2, \ldots, u_s \geqslant 0 \end{aligned}$$

The fractional model can be transformed into the linear model shown below (**M2**):

$$\begin{aligned} (LP_o) \quad \max \quad & \theta = u_1 y_{1o} + \cdots + u_s y_{so} \\ \text{s.t.} \quad & v_1 x_{1o} + \cdots + v_m x_{mo} = 1 \\ & u_1 y_{1j} + \cdots + u_s y_{sj} \leqslant v_1 x_{1j} + \cdots + v_m x_{mj} \\ & v_1, v_2, \ldots, v_m \geqslant 0 \\ & u_1, u_2, \ldots, u_s \geqslant 0 \end{aligned}$$

The fractional model (M1) is equivalent to linear model (M2), and *Unit Invariance Theorem* states that the optimal values of $max\,\theta = \theta^*$ in M1 and M2 are independent of the units in which the inputs and outputs are measured, under the requirement that these units are the same for every DMU.

The input and output data can be arranged in matrix notation X and Y, respectively: