Data envelopment analysis classification machine

Hong Yan\textsuperscript{a,*}, Quanling Wei\textsuperscript{b}

\textsuperscript{a}Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong
\textsuperscript{b}Institute of Operations Research and Mathematical Economics, Renmin University of China, China

\textbf{A R T I C L E   I N F O}

Article history:
Received 4 December 2009
Received in revised form 1 June 2011
Accepted 2 July 2011
Available online 12 July 2011

Keywords:
Data envelopment analysis
Classification machine
Preference cone

\textbf{A B S T R A C T}

This paper establishes the equivalent relationship between the data classification machine and the data envelopment analysis (DEA) model, and thus build up a DEA based classification machine. A data is characterized by a set of values. Without loss of the generality, it is assumed that the data with a set of smaller values is preferred. The classification is to label if a particular data belongs to a specified group according to a set of predetermined characteristic or attribute values. We treat such a data as a decision making unit (DMU) with these given attribute values as input and an artificial output of identical value 1. Then classifying a data is equivalent to testing if the DMU is in the production possibility set, called acceptance domain, constructed by a sample training data set. The proposed DEA classification machine consists of an acceptance domain and a classification function. The acceptance domain is given by an explicit system of linear inequalities. This makes the classification process computationally convenient. We then discuss the preference cone restricted classification process. The method can be applied to classifying large amount of data. Furthermore, the research finds that DEA classification machines based on different DEA models have the same format. Input-oriented and output-oriented DEA classification machines have similar properties. The method developed has great potential in practice with its computational efficiency.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

In recent years, the focus of global economics has increasingly shifted towards knowledge management. According to the forecast of Merrill Lynch (Merrill Lynch Report, 2003), the market value of data mining related to customer relation management analysis will soon reach $3.5 billion. Large scale data processing, including classification, mining, clustering, association, forecasting, estimation, assessment and others, plays a more and more important role in the time.

A classification is to judge if a data belongs to a particular group by evaluating a set of attribute values. In general, a classification machine contains the following four fundamental components: (1) a set of attribute or characteristic values according with some assumptions or a postulate system, (2) a sample training data set, (3) an acceptance domain and (4) a classification function. The classification process is to first construct the acceptance domain from the sample training data set, according to the given attribute values, and then to judge if a new data is in the acceptance or be rejected.

In practice, a health care decision making system determines if a patient has an illness according to some given symptom indexes and physical conditions. Likewise an oil production company makes decision if it is suitable to drill at a potential oil well based on past experience and observed information. In a classification machine, the sample training data set and attribute indicators are selected by experts according to passed experience or by senior management following corporate...
strategy. The techniques and methods for the conventional classification machine can be found in the works of Han and Kamber [8] and Burges [2]. Yi et al. [29] proposed a classifier in the decision tree algorithm for multi valued data based on the analysis for multi-valued attribute and multi labeled data. Small and Roth [16] described margin-based active learning techniques for structure analysis of learning machine. Tang et al. [17] gave a local learning based algorithm for image processing. A wide range of techniques are utilized for classification, including decision tree [6], support vector machine [19] and neural network [10]. Furthermore, Peng et al. [13] discussed a data gravitation based classification method and Han et al. [9] proposed a customized classification learning method based on query projections. All these methods aim to constructing a classification function according to a training set given by expert team, although the principles and procedures are totally different.

This paper extends the data envelopment analysis (DEA) method for the large data classification problem. We treat a data as a decision making unit (DMU) which has the attribute values as input and has a single output of value 1. Without loss of the generality, we assume that the general classification to select a data with smaller attribute values. Therefore, under the DEA framework, we first construct the acceptance domain which is equivalent to the production possibility set. And the further classification of a data is to check if the data is in the acceptance domain. In this sense, we call our process developed here as the DEA classification machine.

The DEA theory, model and method are used to evaluate the relative efficiency among a given number of decision making units (DMUs) with multiple input and multiple output, by solving linear programming problems consecutively for each DMU according to the observation of data. The most representative DEA models include the CCR model by Charnes et al. [3], the BCC model by Banker et al. [1], the FG model by Färe and Grosskopf [7], and the ST model by Seiford and Thrall [15]. The economic meaning of DEA efficiency varies under the different models above. Yu, Wei and Brockett unified the above models and proposed a generalized DEA model [25,30]. Research on DEA models has several main aspects, such as the production possibility set and its surfaces, production frontier, Pareto solutions of the corresponding multi-objective programming problem, and properties of each DMU including technical efficiency, returns to scale, as well as evidence of congestion. Generally speaking, conventional DEA models are not designed for evaluating or processing large amount of DMUs due to the following two limitations. First, the related data for the DMUs are usually not observed simultaneously. Second, even if we have all the data observed, it would not be practically possible to run a linear programming based DEA analysis with such a large number of DMUs and large size of the problem formed.

Trott et al. [18] developed a DEA model CCR based method for credit applicant acceptance systems. In their work, a sample based decision system is proposed to make a decision on whether or not to accept or reject a credit risk, based on samples predetermined by experts. This is a piece of pioneer work extending the DEA model to large number data processing. However, a main issue left is that the description of the acceptance boundary is given by the intersection of the some supporting planes instead of the production possibility set surface. Therefore, acceptance errors always occur. Pendharkar et al. [12] applied the method of Truett et al. et al. to discuss the potential use of a DEA approach for discovering patterns in breast cancer. Their work indicates that DEA outperforms statistical linear discriminate analysis. Pendharkar [11] further extended his work on developing a DEA neural network for classification. Yeh et al. [28] combined DEA, rough set and support machines and used DEA efficiency as predictive variables to predict the business failure. In addition, Wu [26] used DEA result to train decision tree and neural network for further evaluation. This idea is close to that a “evaluation machine” for large number of data processing.

We can then see that the critical step for the DEA classification method is how to construct the acceptance domain, similar to the idea of support vector machine [19]. The key point in our method is to find the linear inequality representation of each surface of the DEA production possibility set. A DEA production possibility set is given in sum-form in conventional DEA literature. In this paper, we transform the conventional sum-form of the production possibility set into an intersection form, which is a linear inequality system, by the procedures we provided earlier [22,27]. We applied the intersection form production possibility set to the DEA efficiency analysis on a large number DMU [23] and the evaluation of returns to scale and evidence of congestion of a large number of DMUs [24].

In this work, as discussed above, we treat each data as a DMU which has the original data values as its input and has a single output with value 1. A set of DMUs is selected to form a “sample training data set”. Based on this training data set, we construct “acceptance domain” under a postulate system. In addition, we define a classification function according to the intersection form acceptance domain. The framework established in this way is thus called a “DEA Classification Machine”. The acceptance boundaries are given by the surface of acceptance domain in linear inequalities. These linear inequality representation does not contain acceptance errors. The second stage of classification is to check if a new DMU satisfies the linear inequality system. The DEA classification machine therefore is computationally efficient and can be applied to large amount data processing.

We also discuss the preference cone restricted DEA classification machine. The preference cone reflects the different importance of the attribute or characteristic values of the DMUs. In particular, we discuss in detail when the preference structure is given by a polyhedral cone. Furthermore, we discuss the DEA classification machine based on other classic DEA models and compare the input-oriented and the output-oriented DEA model. It is found that the DEA classification machine defined has fundamentally the same form under different models. It thus has great potential in practical application and is a complementary approach for data mining.

In the last, we need to point that our method has a high generalization capability. Generalization capability measures the ability of the algorithm to classify a general data, which could be out of the given classification data set [14]. The techniques for
improving the generalization of data mining algorithm are widely discussed in the information and computer sciences areas [20,21]. Our method is a DEA based framework which is a non-parametric evaluation process considering precise data. Therefore, as long as the data to be evaluated has the same data structure, or the same attributes, the algorithm would work properly.

The rest of the paper is arranged as the follows. Section 2 describes a theoretical background for the classification machine, including the DEA postulate system, concept of classification function, and the relationship between the boundary of the acceptance domain and its corresponding Pareto solution. Section 3 gives the classification machine in the intersection form production possibility. Section 4 considers the preference cone added to the DEA classification machine. Section 5 concludes the research.

2. Postulates For DEA classification machine

The data classification is to judge if a data belong to a specified group, or set. For example, a medical doctor need to judge if a patient gets a specified disease based on some observed symptoms. In other words, the data classification is first to determine a patient gets a specified disease based on some observed symptoms. In other words, the data classification is first to determine if a data is said to be of a DMU, or is to be selected. In this sense, we can consider each data as a decision making unit (DMU) to the acceptance domain.

A DEA classification machine developed in this paper is to use the DEA model to construct such an acceptance domain. All data to be classified are described by some numerical indexes. Without loss of the generality, we can assume that the data with smaller value is preferred, or is to be selected. In this sense, we can consider each data as a decision making unit (DMU) with these numerical values as the input value and an artificial single output with the identical value 1. To classify if a data is preferred with smaller values is then equivalent to test if the DMU is in the production possibility domain. In the following, we discuss how to construct the acceptance domain from the set of sample data.

Consider a sample training data set of size \( n \), and output \( y_j \).

(i) (Ordinary postulate) The observed \( x_j \) \( \in T \) for all \( j = 1,2,\ldots,n \).

(ii) (Convexity postulate) If \( x \in T \) and \( x \in T \), then \( \lambda x + (1 - \lambda) x \in T \), for \( \lambda \in [0,1] \).

(iii) (Inefficiency postulate) If \( x \in T \), and \( x > x \), then \( x \in T \).

(iv) (Expansion postulate) If \( x \in T \), then \( x \lambda \in T \), for all \( \lambda \geq 1 \).

(v) (Minimum extrapolation postulate) \( T \) is the intersection set of all \( T \) satisfying postulates (i) to (iv).

It can be shown that such an acceptance domain \( T \) must be in the form of

\[
T = \left\{ x : \sum_{j=1}^{n} x_j \lambda_j \leq x, \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0, j = 1,\ldots,n \right\}.
\]

It is clear that this problem can be described by DEA model CCR [3] with the DMUs, \( j = 1,\ldots,n \), of the input given by \( x_j \), \( j = 1,\ldots,n \), and output \( y_j = 1, j = 1,\ldots,n \). That is, the sample training data set in the DEA model is given by \((x_j, y_j), j = 1,\ldots,n\).

Example 1. Consider a sample training data set of \( x_1, x_2, x_3 \) and \( x_4 \) and their characteristic values are as follows:

\[
x_1 = (1,4)^T, \quad x_2 = (2,2)^T, \quad x_3 = (4,1)^T, \quad x_4 = (4,4)^T.
\]

Giving \( y_j = 1, j = 1,\ldots,n \), the corresponding DMUs, \( DMU_1, DMU_2, DMU_3 \) and \( DMU_4 \) have input and output data as below.

\[
(x_1, 1) = (1,4,1)^T, \quad (x_2, 1) = (2,2,1)^T, \quad (x_3, 1) = (4,1,1)^T, \quad (x_4, 1) = (4,4,1)^T.
\]

Following the postulates given above, the acceptance domain \( T \) is given as follows (see Fig. 1):

\[
T = \left\{ x : \sum_{j=1}^{4} \lambda_j \leq x, \sum_{j=1}^{4} \lambda_j = 1, \lambda_j \geq 0, j = 1,\ldots,4, x \in \mathbb{E}^2 \right\}
\]

It is clear that this problem can be described by DEA model CCR [3] with \( x_j, j = 1,2,3,4 \), as input and \( y_j = 1, j = 1,2,3,4 \), as output. That is, the input and output data of \( DMU_j, j = 1,2,3,4 \), are given as

\[
(x_1, 1) = (1,4,1)^T, \quad (x_2, 1) = (2,2,1)^T, \quad (x_3, 1) = (4,1,1)^T, \quad (x_4, 1) = (4,4,1)^T.
\]
Consider a classification data set \( \mathbb{T} \) which contains a large number of DMUs to be classified. Take \( x \in \mathbb{T} \), the problem of classification is to construct a classification function as follows:

\[
d(x) = \begin{cases} 
1, & x \in \mathbb{T}, \\
-1, & x \notin \mathbb{T},
\end{cases}
\]

where \( x \in \mathbb{E}^n \) and \( x > 0 \). If \( d(x) = 1 \), we say that \( x \) is classified in \( \mathbb{T} \), which means that \( x \) is accepted. That is, it is equivalent to say that a DMU is efficient if its input have relatively smaller values.

Then DEA model CCR with input–output value \((x_j, 1)\) for DMU \( j = 1, \ldots, n \), to evaluate DMU \( j_0 \) is given by

\[
\min \theta \\
\text{s.t.} \sum_{j=1}^{n} x_{ij} \lambda_j \leq \theta x_0 \\
\sum_{j=1}^{n} 1 \cdot \lambda_j \geq 1 \\
\lambda_j \geq 0, \quad j = 1, \ldots, n, \theta \in \mathbb{E}^1.
\]

where \( x_0 = x_{j_0}, 1 \leq j_0 \leq n \). Then (\( P \)) can be rewritten into the following

\[
\min \theta \\
\text{s.t.} \theta x_0 \in \mathbb{T}, \quad \theta \in \mathbb{E}^1.
\]

The acceptance domain \( \mathbb{T} \) satisfying the above postulates (i)–(v) corresponds to the production possibility set of the DEA model CCR. It is not difficult to see that the dual programming problem of (\( P \)) is given by

\[
\max \mu_0 \\
\text{s.t.} \omega x_j - \mu_0 \geq 0, \quad j = 1, \ldots, n, \\
\omega x_0 = 1, \\
\omega \geq 0, \quad \mu_0 \geq 0.
\]

In (\( D \)), variables \( \omega \in \mathbb{E}^n \) and \( \mu_0 \in \mathbb{E}^1 \).

In the following, we discuss properties of the acceptance domain \( \mathbb{T} \) and its boundaries, by DEA model CCR, (\( P \)) and (\( D \)), and related DEA terminologies. Similar to the definition of weak efficiency given in [5]

**Definition 1.** If the optimal values of (\( P \)) and (\( D \)) are 1, then DMU \( j_0 \) is called weakly DEA efficient.

If DMU \( j_0 \) is weakly DEA efficient, then (\( D \)) has an optimal solution \((\omega, \mu_0^0)\), with \( \omega^c \geq 0 \) and \( \omega^e \neq 0 \), such that

\[
\begin{align*}
\omega x_j - \mu_0^0 & \geq 0, \quad j = 1, \ldots, n, \\
\omega x_0 & = 1, \\
\omega^c & \geq 0, \quad \mu_0^e \geq 0, \\
\mu_0^c & = 1.
\end{align*}
\]

Therefore, \( \forall x \in \mathbb{T} \), we have
Theorem 1. \( DMU_j \) of multi-objective programming is weakly DEA efficient if and only if \( x_0 \) is a weak Pareto solution to \((VP)\). 

Proof. From (1), \( \forall x \subseteq T \),
\[
\omega^\prime x - \mu_0^* \geq 0,
\]
where \((\omega^\prime, \mu_0^*)\) is an optimal solution to \((D)\), \(\omega^\prime \geq 0, \omega^\prime \neq 0\). Since \( DMU_j \) is weakly DEA efficient, then
\[
\omega^\prime x_0 = 1, \quad \mu_0^* = 1.
\]
Thus, \( \forall x \subseteq T \), we have
\[
\omega^\prime x - \mu_0^* \geq 0 = \omega^\prime x_0 - \mu_0^*.
\]
This implies
\[
\omega^\prime x \geq \omega^\prime x_0.
\]
That is, \( x_0 \) is an optimal solution to the linear weighted sum problem, \( \min_{x \subseteq T} \omega^\prime x \). Since weight \( \omega^\prime \geq 0 \) and \( \omega^\prime \neq 0 \), then \( x_0 \) is a weak Pareto solution to \((VP)\).

On the other hand, let \( x_0 \) be a weak Pareto solution to \((VP)\). Then
\[
\sum_{j=1}^{n} x_j \lambda_j \leq x < x_0,
\]
\[
\sum_{j=1}^{n} 1 \cdot \lambda_j \geq 1,
\]
\[
\lambda_j \geq 0, \quad j = 1, \ldots, n, x \subseteq E^m
\]
is infeasible. Thus, \((P)\) must have the optimal value 1, which implies that \( DMU_j \) is weakly DEA efficient. This completes the proof. \(\square\)

From the definitions and theorem above, a point inside \( T \) (an interior point of \( T \)) is not weak Pareto solution, thus is not weakly DEA efficient. Those weakly DEA efficient points form the boundary of the acceptance domain \( T \), and are weak Pareto solutions to \((VP)\). That is, assuming that \( x \) is a boundary point of \( T \), there is no other point in \( T \) with all characteristic value components strictly smaller than those of \( x \). The face constructed by weak Pareto solutions of \( T \) is called a surfacing plane of \( T \).

3. Classification by intersection acceptance domain

In Example 1, \( DMU_1 \), \( DMU_2 \) and \( DMU_3 \) are weakly DEA efficient. At each of these \( DMUs \), the optimal solution to \((D)\) gives the normal direction (by \( \omega^\prime \)) and the intercept (by \( \mu_0^* \)) of the supporting plane, \( k = 1, 2, 3 \). Let
\[
(\omega^1, \mu_0^*) = (1, 0, 1), \quad (\omega^2, \mu_0^*) = (1, 1, 4), \quad (\omega^3, \mu_0^*) = (0, 1, 1)
\]
be the optimal solutions to \((D)\) for \( j_0 = 1 \), \( j_0 = 2 \) and \( j_0 = 3 \) respectively. Then we can construct a set \( \bar{T} \) by these supporting planes,
\[
\bar{T} = \{ x | x_1 - 1 > 0, x_1 + x_2 - 4 \geq 0, x_2 - 1 \geq 0 \}.
\]
It is clear that if we use optimal solutions obtained by weakly DEA efficient DMUs to construct the "acceptance domain" $T$, then we have $T \subset \bar{T}$, and there would be an error area, as shown in shadow of Fig. 2. It is easy to show that such shadow areas always exist no matter what normal directions (given by the optimal solution to (D)) are used, since the supporting planes cannot exactly give the surfaces of the acceptance domain $T$. The errors are generated due to the fact that the optimal solution to (D) is not unique. Therefore, we use the method of transferring a polyhedral cone from its intersection form to its sum form, rewriting $T$ from its sum form into the intersection form. The surfacing planes of $T$ give the exact boundaries for classification.

For the sample training data set $T = \{x_{ij} | j = 1, \ldots, n\}$, denote
$$Q = \{(\omega, \mu_0) | \omega x_j - \mu_0 \geq 0, \ j = 1, \ldots, n; \ \omega \geq 0, \ \mu_0 \geq 0\}.$$  

(3)

$Q$ is an intersection form polyhedral cone. Using the method transferring a polyhedral cone from its intersection form to its sum form given in [22,27], we can obtain the extreme rays of the polyhedral cone $Q$ as

$$(\omega^k, \mu^k_0), \ k = 1, \ldots, l.$$  

Then, the sum form of $Q$ is given by

$$Q = \left\{ \sum_{k=1}^l (\omega^k, \mu^k_0) z_k | z_k \geq 0, \ k = 1, \ldots, l \right\}. \tag{4}$$

The intersection form of $T$ is given by Theorem 2, as the following

$$T = \{x | \omega^k x - \mu^k_0 \geq 0, \ k = 1, \ldots, l\}.$$  

Before giving Theorem 2, we point out the following property which is to be used in the proof of the theorem. From the intersection form and sum form of $Q$, there must be

$$\omega^k x_j - \mu^k_0 \geq 0, \ j = 1, \ldots, n; \ k = 1, \ldots, l.$$  

(5)

**Theorem 2.** Denote

$$S = \{x | \omega^k x - \mu^k_0 \geq 0, \ k = 1, \ldots, l\},$$

where $(\omega^k, \mu^k_0)$ are extreme rays of $Q$ in intersection form, $k = 1, \ldots, l.$ Then

$$T = S.$$

**Proof.** We first show that $T \subset S$. Take $x \in T$, then there exist $\lambda_j, j = 1, \ldots, n$, satisfying

$$\sum_{j=1}^n x_j \lambda_j \leq x,$$

$$\sum_{j=1}^n \lambda_j \geq 1,$$

$$\lambda_j \geq 0, \ j = 1, \ldots, n.$$
Then, for \( k = 1, \ldots, l \), we have
\[
\omega^k x - \mu_0^k \geq \omega^k \sum_{j=1}^n x_j \lambda_j - \mu_0^k = \sum_{j=1}^n (\omega^j x_j - \mu_0^j) \lambda_j + \mu_0^k \left( \sum_{j=1}^n \lambda_j - 1 \right) \geq 0 \text{ (from (5))}
\]
That is, \( x \in S \).

On the other hand, if \( S \not\subset T \), that is, there exists \( x \in S \) but \( x \not\in T \). Since \( T \) is a closed convex set, from the separation theorem of closed convex set, there is \( \omega \neq 0 \) and \( \mu_0 \in E_1 \), such that
\[
\omega^k x < \mu_0^k \leq \omega x, \quad \forall x \in T
\]
and there exists \( x \in T \) such that
\[
\omega^k x = \mu_0^k.
\]
In the following, we show that \((\omega, \mu_0) \in Q\). That is
\[
\omega x_j - \mu_0 \geq 0, \quad j = 1, \ldots, n.
\]
Therefore, (8) holds. Since
\[
T = \left\{ x \left| \sum_{j=1}^n x_j \lambda_j \leq x_j, \sum_{j=1}^n \lambda_j \geq 1, \lambda_j \geq 0, j = 1, \ldots, n \right. \right\}
\]
when every component of \( x \) is large enough, we must have \( x \in T \). Therefore, from (6), we know that \( \omega \geq 0 \), this is (9). Further more, from (7),
\[
\mu_0 = \omega x \geq 0.
\]
Thus (10) holds. From (8)–(10), \((\omega, \mu_0) \in Q\). From (4), there exist \( x_k \geq 0, k = 1, \ldots, l \), such that
\[
(\omega, \mu_0) = \sum_{k=1}^l (\omega^k x_k, \mu_0^k) x_k.
\]
That is,
\[
\omega = \sum_{k=1}^l \omega^k x_k, \quad \mu_0 = \sum_{k=1}^l \mu_0^k x_k.
\]
Therefore,
\[
\omega x - \mu_0 = \left( \sum_{k=1}^l \omega^k x_k \right) x - \sum_{k=1}^l \mu_0^k x_k = \sum_{k=1}^l (\omega^k x - \mu_0^k) x_k \geq 0 \text{ (since } x \in S)\]
This is a contradiction to (6). Therefore, \( S \subset T \). The proof is completed. \( \square \)

From Theorem 2, we know that the acceptance domain \( T \) can be given in its intersection form
\[
T = \left\{ x | \omega^k x - \mu_0^k \geq 0, k = 1, \ldots, l \right\}.
\]
Then, if \( x \in T \), the classification function is given by
\[
d(x) = \text{sign} \left( \min_{1 \leq k \leq l} (\omega^k x - \mu_0^k) \right).
\]
To judge if a given data \( x \) is acceptable using the classification function \( d(x) \) is equivalent to identify if \( x \in T \), the acceptance domain. With the intersection form \( T \), it only needs to compute the values of \( \omega^k x - \mu_0^k \), for \( k = 1, \ldots, l \) to test if data \( x \in T \). We conduct a numerical experiment based on a preselected sample training data set containing 100 DMU’s. Each DMU has 10 characteristic values. On the notebook computer lenovo-X200, it takes 289 s to construct the intersection form acceptance domain. Then, it takes 665 s to complete the classification of 10,000 data randomly generated. Note that the classification process is independent of the construction of intersection form acceptance domain. The computation time on classification is linearly increasing as the number of data to be processed. Therefore, the DEA classification machine provides an efficient tool for data classification.
Example 2. Consider Example 1 again. The intersection form of $Q$ is given as
\[ Q = \left\{ (\omega, \mu_0) \right\} \]
\[ \begin{align*}
 & \omega_1 + 4\omega_2 - \mu_0 \geq 0, \\
 & 2\omega_1 + 2\omega_2 - \mu_0 \geq 0, \\
 & 4\omega_1 + \omega_2 - \mu_0 \geq 0, \\
 & 4\omega_1 + 4\omega_2 - \mu_0 \geq 0.
\end{align*} \]
Its sum form is
\[ Q = \left\{ (1, 0, 1)x_1 + (2, 1, 6)x_2 + (1, 2, 6)x_3 + (0, 1, 1)x_4 | x_i \geq 0, \right\}
\[ \left\{ i = 1, 2, 3, 4 \right\}. \]
From Theorem 2, $T$ has its intersection form (see Fig. 3.)
\[ T = \left\{ x | x_1 - 1 \geq 0, 2x_1 + x_2 - 6 \geq 0, x_1 + 2x_2 - 6 \geq 0, x_2 - 1 \geq 0 \right\}. \]
Let $\tilde{T} = \{ \tilde{x}_1, \tilde{x}_2, \tilde{x}_3 \} = \{ (0.5, 5)^T, (1, 3)^T, (5, 0.5)^T \}$. Then,
\[ d(\tilde{x}_1) = \text{sign} \left( \min_{1 \leq k \leq 4} (\omega^k \tilde{x}_1 - \mu_0^k) \right) = \text{sign}(-0.5) = -1; \]
\[ d(\tilde{x}_2) = \text{sign} \left( \min_{1 \leq k \leq 4} (\omega^k \tilde{x}_2 - \mu_0^k) \right) = \text{sign}(-1) = -1; \]
\[ d(\tilde{x}_3) = \text{sign} \left( \min_{1 \leq k \leq 4} (\omega^k \tilde{x}_3 - \mu_0^k) \right) = \text{sign}(-1) = -1. \]

4. DEA classification with preference cone

In practice, the given $m$ characteristic values of DMUs often have different importance to the decision maker. For example, it is clear that in the case described by Pendharkar et al. [12], the “input factors” used for breast cancer status detection should have different weights. Those physiological indicators, such as menopausal status, are more important than those census indicators, such as age. In the DEA model, such a different importance is described by a preference cone (see [4,25,30]). Consider a DEA model with preference cone $W$
\[ \begin{align*}
 & \max \quad \mu_0 \\
 & \text{s.t.} \quad \omega x_j - \mu_0 \geq 0, \quad j = 1, \ldots, n, \\
 & \omega x_0 = 1, \\
 & \omega \in W, \quad \mu_0 \geq 0
\end{align*} \]
and its dual programming problem
\[ \begin{align*}
 & \min \quad \theta \\
 & \text{s.t.} \quad \sum_{j=1}^n (x_j \lambda_j - \theta x_0)^T \in W^+, \\
 & \sum_{j=1}^n \lambda_j \geq 1, \\
 & \lambda_j \geq 0, \quad j = 1, \ldots, n, \theta \in E^1.
\end{align*} \]
In the above, \( W \subset E_+^m = \{ \omega^T \in E^m | \omega \geq 0 \} \). \( W \) is a closed convex cone, \( W^* \) is the negative polar cone of \( W \), and \( x_j \in \text{Int} W^* \), \( j = 1, \ldots, n \).

For the DEA model with preference cone \( W \), we have the similar definition of weak DEA efficiency for a DMU. That is, when both \( (\mathcal{P}) \) and \( (\mathcal{D}) \) have the optimal value 1, the DMU is weakly DEA efficient. Since the constraint set of \( (\mathcal{D}) \) is a subset of the constraint set of \( (D) \) (since \( W \subset E_+^m \)), then if a DMU is not weakly DEA efficient under \( (D) \), it cannot be weakly DEA efficient under \( (\mathcal{D}) \).

Usually, a polyhedral cone \( W \) is given in the intersection form. For example, we may have \( W = \{ \omega | 0 < x_1 \leq \frac{x_2}{x_3} \leq \beta, \omega_i \geq 0, i = 2, \ldots, m \} \) or \( W = \{ \omega | 0 < \omega_1 \leq \omega_2 \leq \cdots \leq \omega_m \} \). We can thus transform it into the sum form by the method given in [22,27]. That is, we can find a \( m' \times m \) parameter matrix \( A \), extreme directions of cone \( W \), such that the sum form of \( W \) can be represented by,

\[
W = \{ \omega A | \omega^T \geq 0 \},
\]

where

\[
A = (a_1^T, \ldots, a_{m'}^T)^T, \quad a_i^T \in E^m, \quad a_i \geq 0, \quad i = 1, \ldots, m'.
\]

Then, it is not difficult to see that

\[
\min \theta
\]

\[
(\mathcal{P}) \quad \text{s.t.} \quad \sum_{j=1}^{n} (Ax_j) \lambda_j \leq \theta(Ax_0),
\]

\[
\sum_{j=1}^{n} \lambda_j \geq 1,
\]

\[
\lambda_j \geq 0, \quad j = 1, \ldots, n; \quad \theta \in E^1.
\]

And the dual problem is (for \( \omega^T \in E'^m \))

\[
\max \mu_0
\]

\[
(\mathcal{D}) \quad \text{s.t.} \quad \omega^T(Ax_j) - \mu_0 \geq 0,
\]

\[
\omega^T(Ax_0) = 1,
\]

\[
\omega^T \geq 0, \quad \mu_0 \geq 0.
\]

Then, when a preference cone is considered, it is equivalent to change the sample training data set

\[
\mathcal{T} = \{ x_j | j = 1, \ldots, n \}
\]

into

\[
\mathcal{T}^W = \{ Ax_j | j = 1, \ldots, n \}
\]

for classification. And correspondingly, the classification data set \( \hat{\mathcal{T}} \) becomes \( \hat{\mathcal{T}}^W = \{ Ax | x \in \hat{\mathcal{T}} \} \). Denote

\[
Q^W = \{ (\omega^T, \lambda_0^T) | \omega^T(Ax_j) - \mu_0^T \geq 0, \quad j = 1, \ldots, n, \quad \omega^T \geq 0, \quad \mu_0^T \geq 0 \}.
\]

\( Q^W \) is in intersection form. The sum form of \( Q^W \) is given by

\[
Q^W = \left\{ k \omega^k \mu_0^k | x_k \geq 0, \quad k = 1, \ldots, l' \right\}.
\]

From Theorem 2, the intersection of \( T^W \) is given by

\[
T^W = \left\{ y \in E'^m | \omega^k y - \mu_0^k \geq 0, \quad k = 1, \ldots, l' \right\}.
\]

The classification function is given by

\[
d(x) = \text{sign} \left( \min_{k \leq k \leq l} (\omega^k(Ax) - \mu_0^k) \right).
\]

where \( x \in \hat{\mathcal{T}} \).

**Example 3.** Consider Example 1 again, and a preference cone (see Fig. 4)

\[
W = \{ (\omega_1, \omega_2) | \omega_2 \geq 4 \omega_1, \quad \omega_1 \geq 0, \quad \omega_2 \geq 0 \},
\]
which says that the second characteristics value is at least four time more important than the first characteristics value. It is clear that \( W \) is given in the intersection form, a set of linear inequalities. With the method give in [22,27], we can transform \( W \) into the sum form as below. Or, directly from Fig. 4, we can see that \( W \) has two extreme directions: \((0,1)\) and \((0.1,0.4)\). Thus the preference cone \( W \) in sum form is then given as follows:
\[
W = \{(\omega_1, \omega_2)A|\omega_1 \geq 0, \omega_2 \geq 0\},
\]
where
\[
A = \begin{pmatrix}
0 \\
0.1 \\
1 \\
0.4
\end{pmatrix} ; m = m' = 2.
\]
It is shown as Fig. 4.

In Example 1 and Example 2,
\[
T = \{x_1, x_2, x_3, x_4\} = \{(1,4)^T, (2,2)^T, (4,1)^T, (4,4)^T\},
\]
\[
\hat{T} = \{x_1, x_2, x_3\} = \{(0.5,5)^T, (1,3)^T, (5,0.5)^T\}.
\]
Then
\[
Ax_1 = (4,1.7)^T, \ Ax_2 = (2.1)^T, \ Ax_3 = (1,0.8), \ Ax_4 = (4,2)^T
\]
and
\[
\hat{A}x_1 = (5,2.05)^T, \ \hat{A}x_2 = (3,1.3)^T, \ \hat{A}x_3 = (0.5,2.5)^T.
\]
It is then computed that \((\omega^1, \mu_0^1) = (1,0.1)\) and \((\omega^2, \mu_0^2) = (0,1,0.8)\). The intersection form of \(T^W\) is given by (see Fig. 5.)
\[
T^W = \{y \in E^2|y_1 - 1 \geq 0, \ y_2 - 0.8 \geq 0\}.
\]
Then,
\[
d(\hat{x}_1) = \text{sign} \left( \min_{1 \leq k \leq 2} (\omega^k(A\hat{x}_1) - \mu_0^k) \right) = \text{sign}(1.25) = 1;
\]
\[
d(\hat{x}_2) = \text{sign} \left( \min_{1 \leq k \leq 2} (\omega^k(A\hat{x}_2) - \mu_0^k) \right) = \text{sign}(0.5) = 1;
\]
\[
d(\hat{x}_3) = \text{sign} \left( \min_{1 \leq k \leq 2} (\omega^k(A\hat{x}_3) - \mu_0^k) \right) = \text{sign}(-0.5) = -1.
\]

Note that in Example 2 where the preference cone is not considered, two characteristic values have the same importance and the classification function values are \(-1\) for \(DMU_{s_1}, DMU_{s_2}\) and \(DMU_{s_3}\). The preference cone given in this example indicates that the second characteristic is at least four times more important than the first characteristics value. The classification function value for both \(DMU_{s_1}\) and \(DMU_{s_2}\) are \(1\). However, classification function value for \(DMU_{s_3}\) is \(-1\) when the preference cone is considered, since the second characteristic value of \(DMU_{s_1}\) is relatively small.

5. Remarks and conclusion

**Remark 1.** The DEA classification machine discussed in this paper is based on the DEA model CCR. The other classic DEA models include BCC [1], FG [7] and ST [15] models. The following Property 1 and (i)–(iii) point out that the DEA classification machine constructed based on DEA model CCR and the DEA classification machine based other classical DEA models such as BCC, FG and ST are in the same format. All these DEA models have the same acceptance domain \(T\) in DEA classification machine.
Property 1. Let
\[ T = \left\{ x \left| \sum_{j=1}^{n} x_j \leq x, \sum_{j=1}^{n} \lambda_j \geq 1, \lambda_j \geq 0, j = 1, \ldots, n \right. \right\}, \]
\[ T' = \left\{ x \left| \sum_{j=1}^{n} x_j \leq x, \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0, j = 1, \ldots, n \right. \right\}. \]

Then, we must have \( T = T' \).

Proof. It is obvious that \( T \supseteq T' \). On other hand, let \( x \in T \), then there are \( \lambda_j, j = 1, \ldots, n \), such that
\[ \sum_{j=1}^{n} x_j \lambda_j \leq x, \sum_{j=1}^{n} \lambda_j \geq 1, \lambda_j \geq 0, j = 1, \ldots, n. \]

And without loss of the generality, assume that \( \sum_{j=1}^{n} \lambda_j > 1 \). Let \( x = \left( \sum_{j=1}^{n} \lambda_j \right)^{-1} \), it is clear that \( 0 < x < 1 \). Then
\[ \sum_{j=1}^{n} (x \lambda_j) = 1, \]
\[ x \lambda_j \geq 0, \quad j = 1, \ldots, n, \]
\[ \sum_{j=1}^{n} x_j (x \lambda_j) \leq 2x < x. \]

Therefore, \( x \in T \). \( \square \)

The same as before, consider the sample training data set \( \mathcal{T} \),
\[ \mathcal{T} = \{ x_j | j = 1, 2, \ldots, n \}. \]

It is clear that \( \mathcal{T} \) is corresponding to the reference set for the DEA model, with the identical output \( (x_j, 1) | j = 1, 2, \ldots, n \).

Then we have the following results. (The postulate systems corresponding to DEA models BCC, FG and ST can be found in [25].)

(i) The acceptance domain corresponding to the DEA model BCC is (from Property 1)
\[ T_{BCC} = \left\{ x \left| \sum_{j=1}^{n} x_j \lambda_j \leq x, \sum_{j=1}^{n} \lambda_j \geq 1, \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0, j = 1, \ldots, n \right. \right\} = T. \]

(ii) The acceptance domain corresponding to the DEA model FG is
\[ T_{FG} = \left\{ x \left| \sum_{j=1}^{n} x_j \lambda_j \leq x, \sum_{j=1}^{n} \lambda_j \geq 1, \sum_{j=1}^{n} \lambda_j \leq 1, \lambda_j \geq 0, j = 1, \ldots, n \right. \right\} = T_{BCC} = T. \]

(iii) The acceptance domain corresponding to the DEA model ST is
Therefore, all DEA models have the same acceptance domain in DEA classification machine.

Remark 2. The model discussed above is to translate the characteristic values into input data and consider an input-oriented DEA model with single identical output. Therefore, it could be regarded as an input-oriented DEA classification machine. Similarly, an output-oriented DEA model can be used. If the sample training data set is $T$

$$T = \{y_j| j = 1, 2, \ldots, n\},$$

where $y_j \in E^1, j = 1, \ldots, n$, are decision making units (DMUs) with a certain characteristics. For DMU$_j$, the characteristic value is given by $y_j = (y_{j1}, y_{j2}, \ldots, y_{jm})^T$. Similarly, with appropriate data processing, we can require that the larger value of $y_{ij}$ is preferred. Then the acceptance domain is,

$$T = \left\{y \sum_{j=1}^{n} y_{ij} \lambda_j \geq y \sum_{j=1}^{n} \lambda_j \leq 1, \lambda_j \geq 0, j = 1, \ldots, n\right\}.$$ 

The corresponding reference set in DEA model is $(1, y_j) - j = 1, 2, \ldots, n$. The DEA model CCR is given by

$$\max z$$

$$\text{(P)} \quad \text{s.t.} \quad \sum_{j=1}^{n} y_{ij} \lambda_j \geq zy_0,$$

$$\sum_{j=1}^{n} \lambda_j \leq 1,$$

$$\lambda_j \geq 0, \quad j = 1, \ldots, n, z \in E^1.$$ 

The dual of (P) is given by

$$\min \mu_0$$

$$\text{(D)} \quad \text{s.t.} \quad \mu_0 - \mu y_{ij} \geq 0, \quad j = 1, \ldots, n,$$

$$\mu y_0 = 1,$$

$$\mu \geq 0, \quad \mu_0 \geq 0.$$

6. Conclusion

In line with the idea of Troutt et al. [18], we discussed and defined a DEA classification machine, in DEA terminologies. We treat the data under evaluation to be a decision making unit with the given attribute values as the input and a single output of value 1. We then used a set of decision making units (DMUs) to form a sample training data set and construct the acceptance domain based on the sample set for classification. By the method of transferring a polyhedral cone between intersection-form and sum-form, we proposed and proved that the acceptance domain can be given by a linear inequality system without acceptance error. We also discussed DEA classification with a preference cone, reflecting different importance weights on different characteristic values. We discussed the classification machine under other typical DEA models and revealed that the DEA classification machine is in the same format. The theoretical work and the computational results show that the method developed has great potential in practice.

Acknowledgement

The first author is partially supported by the National Natural Science Foundation of China, NNSF 70531040 and NNSF 70871114 and the 985 Research Grant of Renmin University of China. The research is partially supported the Hong Kong CERG Research Fund PolyU 5485/09H and 5515/10H.

References


