Available online at http://ijdea.srbiau.ac.ir

Int. J. Data Envelopment Analysis (ISSN 2345-458X)

Vol.3, No.3, Year 2015 Article ID IJDEA-00331, 9 pages

Research Article

International Journal of Data Envelopment Analysis Science and Research Branch (IAU)

Ranking DMUs with Interval Data Using DEA and CA Approaches

Kianoosh Kianfar [a*], Mahnaz Ahadzadeh Namin [b], Akbar Alam Tabriz [c], Esmaeil Najafi [d], Farhad Hosseinzadeh Lotfi [e]

(a) *Department of Industrial Engineering, Science and Research branch, Islamic Azad University, Tehran, Iran.*
(b) *Department of Mathematics, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran.*
(c) *Department of Management, Shahid Beheshti University, Tehran, Iran*
(d) *Department of Industrial Engineering, Science and Research branch, Islamic Azad University, Tehran, Iran.*
(e) *Department of Mathematics, Science and Research branch, Islamic Azad University, Tehran, Iran.*

Received 10 November 2015, Revised 14 February 2016, Accepted 21 May 2016

Abstract

Data envelope analysis (DEA) is an approach to estimate the relative efficiency of decision making units (DMUs). Several studies were conducted in order to prioritize efficient units and some useful models such as cross-efficiency matrix (CEM) were presented. Besides, a number of DEA models with interval data have been developed and ranking DMUs with such data was solved. However, presenting an obtained crisp data derived interval data is a critical problem, so that many researches were implemented so as to compute weights and averaging the interval data. In this paper we propose the new algorithm to find more suitable weight applying a data mining approach of DMU's data. For this purpose, we employed clustering and pair-wise comparison matrix on given relative efficiency from CEM. Results indicate there is meaningful different between efficiency of DMUs with lower bound and that of DMUs with upper bound.

***Keywords**:* Data envelope analysis, Cross-efficiency matrix, Cluster analysis

[*] Corresponding Author: kianoush.kianfar@gmail.com

1. Introduction

Data envelope analysis (DEA) is a linear programming model and non-parametric approach that evaluates relative technical efficiencies of decision making units (DMUs) on the basis of multiple inputs and outputs by computing the ratio of weighted sum of their outputs to their inputs (Arieh and Gullipalli, 2012; Jahanshahloo et al, 2009; Smirlis et al, 2006). This technique has been used in many fields successfully with crisp values, however in real application there are inaccurate data similar to probabilistic, interval, ordinal, qualitative, or fuzzy. Hence, some researchers conducted several theoretical development of DEA model with data such as interval (Despotis and Smirlis, 2002; Jahanshahloo et al, 2004; Jahanshahloo et al, 2009). Nevertheless, there are many models and techniques to solve this problem, but there is a new problem for ranking the efficient DMUs with interval data, so that in some researches *DMU*s were ranked by these ideal points (Jahanshahloo et al, 2011; Wu et al, 2013). There are several models to rank DMUs with crisp data (Hashimoto and Wu, 2004). However, in all researches, ranking DMUs with interval data has been solved by using ranking approaches such as AHP or TOPSIS or hybrid algorithm to find suitable weight in order to calculate crisp efficiency basis of interval inputs and outputs. Therefore, we conduct new approach using data mining techniques similar to clustering to obtain these weights as new model.

2. Overreview of the research techniques

2.1. DEA models

In the previous section DEA technique was defined completely. In this part we describe existing models related to DEA. There are three commonly orientations for DEA model, which can be formulated as below:

1. Input oriented model is related to the minimizing level of the inputs in order to achieve a given level of the outputs.

$$\min \ \theta_p = \sum_{i=1}^{m} v_i \, x_{ip} / \sum_{r=1}^{s} u_r \, y_{rp}$$

Subject to

$$\sum_{r=1}^{m} v_i \, x_{ij} / \sum_{r=1}^{s} u_r \, y_{rj} \le 1$$

j=1,…,n (1)

$$u_r, v_i \ge 0$$

2. Output oriented model is concerned with the maximizing level of the outputs per given level of the inputs (Zohrehbandian & Sadeghi, 2013; Samoilenko et al, 2008; Caklovic & Hunjak, 2012).

$$\max \quad \theta_p = \sum_{r=1}^{s} u_r \, y_{rp} / \sum_{i=1}^{m} v_i \, x_{ip}$$

Subject to

$$\sum_{r=1}^{s} u_r \, y_{rj} / \sum_{r=1}^{m} v_i \, x_{ij} \le 1$$

j=1,…,n (2)

$$u_r, v_i \ge 0$$

3. Base oriented model unlike the others, is pertains to the optimal combination of the inputs and outputs. Consequently, this model has control over inputs as well as outputs, concluding the efficiency of input utilization

and efficiency of output production (Samoilenko et al, 2008).

Instead of exact data, we will apply models with interval data in order to rank DMUs. Input oriented with interval data for upper bound efficiency and lower efficiency is formulated, respectively as below:

Upper bound efficiency:

$$\theta^U = Max \sum_{r=1}^{s} u_r\, y_{rp}^U$$

Subjected to

$$\sum_{i=1}^{m} v_i\, x_{ip}^L = 1$$

$$\sum_{r=1}^{s} u_r\, y_{rj}^L - \sum_{i=1}^{m} v_i\, x_{ij}^U \leq 0 \quad j = 1, \dots, n, \qquad j \neq p$$

$$\sum_{r=1}^{s} u_r\, y_{rp}^U - \sum_{i=1}^{m} v_i\, x_{ip}^L \leq 0$$

$$v_i, u_r \geq 0 \qquad (3)$$

Lower bound efficiency:

$$\theta^L = Max \sum_{r=1}^{s} u_r\, y_{rp}^L$$

Subjected to

$$\sum_{i=1}^{m} v_i\, x_{ip}^U = 1$$

$$\sum_{r=1}^{s} u_r\, y_{rj}^U - \sum_{i=1}^{m} v_i\, x_{ij}^L \leq 0 \qquad j = 1, \dots, n, \;\; j \neq p$$

$$\sum_{r=1}^{s} u_r\, y_{rp}^L - \sum_{i=1}^{m} v_i\, x_{ip}^U \leq 0$$

$$v_i, u_r \geq 0 \qquad (4)$$

2.2 Cross-efficiency matrix

The cross-efficiency matrix was introduced by Sexton et al in 1986. This approach aids us to evaluate efficiency of one DMU considering the optimal input and optimal output weights of another DMU (Caklovic & Hunjak, 2012). The matrix element θ_{ij} of the cross-efficiency matrix (CEM) in *i*-th column and *j*-th row of CEM represents the efficiency of DMU_i when evaluated with the optimal weights of DMU_j, according to below relationship:

$$\theta_{DMU_i}^{DMU_j} = \theta_{ij} = eff_i(u^j, v^j) = \sum_{r=1}^{s} u_r\, y_{rp} \qquad (5)$$

It is expected that 'good' DMU has several high values in its row.

2.3 Cluster analysis

Clustering is a popular data mining approach that deal with the separating of a set of objects into a useful set of mutually exclusive clusters in order that the similarity between the observations from the different clusters (i.e., subset) is low, whereas the similarity between the observations within each cluster is high (Samoilenko et al., 2010; 2008). Unlike decision trees which assign a class to an instance (supervised method), clustering procedures are applied when instances are divided into natural groups (unsupervised method). There are different ways to produce these clusters. The groups may be exclusive i.e. any instance belongs to only one group probabilistic or fuzzy i.e. an instance belongs to each group to a certain probability or degree (membership value) hierarchical i.e. there is a crude division of instances into groups at the top level and each of these groups are refined further up to individual instances (Thomassey et al., 2006). In other literature, overview of two general approaches to clustering was provided: hierarchical clustering, partitional clustering (e.g., k-means, k-median). Hierarchical clustering could make clusters by one of the two methods, agglomerative or divisive. Agglomerative method assumes that each data point is its own cluster, and with each step of the clustering process, these clusters are combined to form larger clusters, which are eventually combined to combine a single cluster. Divisive method of the hierarchical clustering, on the contrary, commences with the single cluster including all data points within the sample and proceeds to divide it into the smaller dissimilar clusters. Unlike hierarchical clustering, k-means clustering requires the number of resulting cluster, k, to be specified prior to analysis. Thus, k-means clustering will produce k different clusters of greatest possible distinction (Samoilenko et al., 2008).

3. Methodology

In this section, we introduce multi steps algorithm so as to compute weights for combination of lower bound and upper bound efficiency of DMUs so that we can obtain the crisp efficiency instead of the interval efficiency.

In proposed algorithm, there are 5 stages as below:

1. *Evaluating efficiency of DMUs*: the DMU's performance is measured using DEA (θ^L, θ^U) based on equations (3) and (4). It is underlined that, we applied input oriented model, because in the conducted research by Samoilenko and et al in 2008, the most natural grouping of DMUs was provided by results of that model using constant return to scale (CRS) criterion (i.e. CCR model)

2. *Applying cross-efficiency matrix:* the efficient DMUs according to CEF and equation (5) are prioritized. The matrixes are implemented as table1 and table 2.

Table 1. CEM for the lower efficiency of the DMU

	1	2	3	…	n	Avg.
1	θ^L_{11}	θ^L_{12}	θ^L_{13}	…	θ^L_{1n}	$(\sum_{j=1}^{n} \theta^L_{1j})/n$
2	θ^L_{21}	θ^L_{22}	θ^L_{23}	…	θ^L_{2n}	$(\sum_{j=1}^{n} \theta^L_{2j})/n$
…	…	…	…	…	…	…
n	θ^L_{n1}	θ^L_{n2}	θ^L_{n3}	…	θ^L_{nn}	$(\sum_{j=1}^{n} \theta^L_{nj})/n$

Table 2. CEM for the upper efficiency of the DMU

	1	2	3	…	n	Avg.
1	θ^U_{11}	θ^U_{12}	θ^U_{13}	…	θ^U_{1n}	$(\sum_{j=1}^{n} \theta^U_{1j})/n$
2	θ^U_{21}	θ^U_{22}	θ^U_{23}	…	θ^U_{2n}	$(\sum_{j=1}^{n} \theta^U_{2j})/n$
…	…	…	…	…	…	…
n	θ^U_{n1}	θ^U_{n2}	θ^U_{n3}	…	θ^U_{nn}	$(\sum_{j=1}^{n} \theta^U_{nj})/n$

3. *Cluster analysis of upper and lower:* the DMUs is clustered using K-mean approach (indicators including outputs and inputs as the attributes):

a. Clustering of DMUs with lower efficiency is done according to data points of table 3.

b. Clustering of DMUs with upper efficiency is applied by data points of table

4. *Obtaining the score for each cluster:* we compute the average relative efficiency of some clusters identified in previous step according to the research by semoilinko et al in 2008. It is noted that averaging for clusters in lower bound is different from that of clusters in upper bound.

5. *Assigning the relative weight to each DMU:* Approximation a numerical scale derived graphic scale in AHP (analytic hierarchy process) and assign that to each cluster according to given score for that cluster.

Table 3. Clustering with lower efficiency

	Indicators						Cluster.No.
Attributes	x_{11}^L	x_{i1}^L	x_{m1}^L	y_{11}^L	y_{r1}^L	y_{s1}^L	
	…	…	…	…	…	…	
		x_{ij}^L			y_{rj}^L		
	x_{1n}^L	x_{in}^L	x_{mn}^L	y_{1n}^L	y_{rj}^L	y_{sn}^L	

Table 4. Clustering with upper efficiency

	Indicators						Cluster.No.
Attributes	x_{11}^U	x_{i1}^U	x_{m1}^U	y_{11}^U	y_{r1}^U	y_{s1}^U	
	…	…	…	…	…	…	
		x_{ij}^U			y_{rj}^U		
	x_{1n}^U	x_{in}^U	x_{mn}^U	y_{1n}^U	y_{rj}^U	y_{sn}^U	

We use pair-wise comparison matrix in order to obtain relative importance for indices (e.g. clusters) and ranking them. In fact, results indicate relative importance of DMU within each cluster considering obtained relative importance of its cluster.

As a result, we can calculate final crisp efficiency for each DMU by computing weighted average (W.A.) of the $(\sum_{j=1}^{n} \theta_{1j}^{L})/n$ and $(\sum_{j=1}^{n} \theta_{1j}^{U})/n$. Therefore, more W.A. indicates more rank for each DMU.

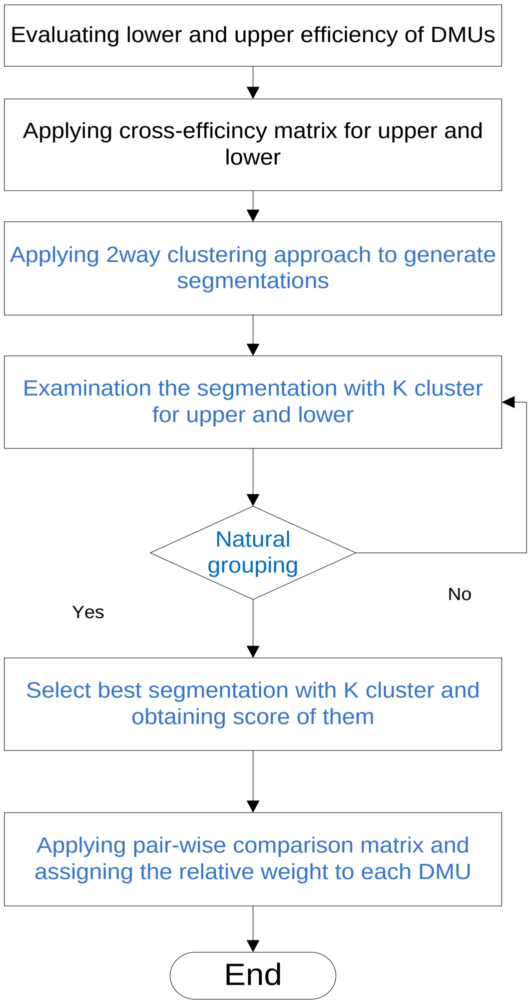

Figure 1. Proposed algorithm

4. Numeric example and results

4.1 Data

The numerical example is taken into account and we apply this approach to selected commercial bank branch in Iran , which have been used as interval data in conducted research by jahanshahloo et al in 2009 (reference of data). Each branch utilizes three inputs to generate five outputs and there are 20 branches as DMU.

4.2 Implementation, results and discussion

In the first place, the cluster analysis is applied on whole data including 20 bank branches. For this purpose, we employ k-means approach according to the introduced approach of the conducted research by Samoilenko et al in 2008, so that they defined K_{max} and τ_{Outlier} as a parameters, then K_{max} clusters was generated and after the examination of the segmentation with K clusters, they were able to indicate that current segmentation with K clusters does not provide the natural grouping of DMUs, if K > 1 and there is at least one cluster is that includes less than τ_{Outlier} percent of DMUs. Therefore, by decreasing the number of clusters, previous examination was being repeated in order to evaluate new segmentation. Otherwise, they could access to best segmentation with K clusters.

Noted that τ_{Outlier} is the index to identify natural clustering Moreover, we consider silhouette index introduced by Peter J. Rousseeuw in 1987 so as to find natural and qualified clustering. Average silhouette width is inside interval [-1,1] so that, the value near to 1 indicates natural grouping and near to -1 indicates incorrect clustering.

Parameters K_{max} and τ_{Outlier} are set 5 and 10%, thus with average silhouette more than 0.5 we could come up with two solutions that disaggregate upper data and lower data into two and three clusters respectively. Results for clustering are shown in table 5.

As it was discussed in steps of algorithm, we applied CCR (constant return to scale) and input-oriented model in order to measure relative efficiency of DMUs with interval data. Details are indicated in table 6.

Table 5. Results of clustering

	Number of clusters	**Number of DMUs in each cluster**	**Average Silhouette width**
Upper	Five clusters	1, 7, 1, 6, 5	0.2481
	Four clusters	11, 5, 3, 1	0.4734
	Three clusters	7, 1, 12	0.5098
	Two clusters	12, 8	0.6037
lower	Five clusters	3, 5, 8, 3, 1	0.2335
	Four clusters	7, 1, 5, 7	0.2039
	Three clusters	3, 14, 3	0.5735

Table 6. Interval efficiency of DMU

DMU	(θ^L, θ^U)	**DMU**	(θ^L, θ^U)
1	(1.0,0.29)	11	(1.0,1.0)
2	(0.21,0.77)	12	(0.32,0.49)
3	(0.52,1.0)	13	(0.44,0.70)
4	(1.0,1.0)	14	(0.25,0.72)
5	(0.63,0.38)	15	(0.41,1.0)
6	(0.90,1.0)	16	(0.22,1.0)
7	(0.73,1.0)	17	(1.0,1.0)
8	(1.0,1.0)	18	(0.26,0.95)
9	(1.0,1.0)	19	(0.99,1.0)
10	(1.0,1.0)	20	(0.18,0.97)

We also calculated the average relative efficiencies of the two and three clusters identified of upper and lower data separately that results are shown in table 7.

In this stage, we approximate numerical scale considering given score of each cluster (table 8, 9). Therefore, the relative importance as the weight of each DMU for upper efficiency and lower efficiency is obtained (Table 10).

Table 7. Score of clusters

	Cluster No	Score
Upper	C1	0.9113
	C2	0.8317
Lower	C1	0.9967
	C2	0.6433
	C3	0.5814

Table 8. Numerical scale of lower data

Index	C1	C2	C3
C1	1	2	7
C2	0.5	1	6
C3	0.143	0.167	1
Sum	1.643	3.147	14

Table 9. Numerical scale of upper data

Index	C1	C2
C1	1	2
C2	0.5	1
Sum	1.5	3

Table 10. Relative importance of clusters

	Cluster No	Score
Upper	C1	0.67
	C2	0.33
Lower	C1	0.58
	C2	0.35
	C3	0.07

Finally, the crisp efficiency of each DMU is computed using obtained relative weight and score concerning to upper and lower data in interval data (table 11). Ranking process is implemented for below parts separately:

1. DMU with interval efficient (θL=1, θU=1)

2.DMU with interval semi-efficient (θL<1, θU=1) or (θL=1, θU<1)

3.DMU with interval inefficient (θL<1, θU<1)

It is noted that efficient value of DMUs is set according to results of cross-efficiency matrix.

5. Conclussion

In this paper, we studied on ranking methodology of DMUs with interval data. There were several approaches to prioritize DMUs using combination of DEA and ranking techniques such as AHP or TOPSIS. In contrast, we applied data mining (DM) techniques similar to cluster analysis in order to partition data (DMUs) based on their attributes. Assigning the relative weights to DMUs with interval data (lower and upper) helps us compute weighted average of lower and upper data, however the approximation of weights and suitable methodology to obtain these would be an important problem.

Table 11. Ranking DMU with interval data

(θL=1, θU=1)						
DMU	Score	Weight	Score	Weight	Crisp	Rank
4	0.466	0.07	0.578	0.33	3.520857	4
8	0.7	0.58	1.195	0.67	2.750431	6
9	0.525	0.07	1.432	0.67	14.90129	2
10	0.959	0.07	4.097	0.33	20.60343	1
11	0.376	0.07	0.599	0.67	6.779286	3
17	0.892	0.58	1.257	0.67	3.014052	5
(θL<1, θU=1) or (θL=1, θU<1)						
DMU	Score	Weight	Score	Weight	Crisp	Rank
1	1.166	0.35	0.29	0.67	2.391143	13
3	0.52	0.35	1.256	0.67	3.594343	9
6	0.9	0.07	0.659	0.33	4.336714	8
7	0.73	0.07	0.436	0.33	3.115429	11
15	0.41	0.35	1.136	0.67	3.254629	10
16	0.22	0.07	0.951	0.33	5.033286	7
19	0.99	0.58	1.128	0.67	2.963034	12
(θL<1, θU<1)						
DMU	Score	Weight	Score	Weight	Crisp	Rank
2	0.21	0.07	0.77	0.33	4.17	16
5	0.63	0.07	0.38	0.33	2.751429	20
12	0.32	0.07	0.49	0.33	2.96	19
13	0.44	0.07	0.7	0.33	4.07	17
14	0.25	0.07	0.72	0.33	3.974286	18
18	0.26	0.07	0.95	0.33	5.068571	15
20	0.18	0.07	0.97	0.33	5.082857	14

Therefore, clustering as a DM approach has ability to explore appropriate relative importance for all DMUs that are similar to each other. On the other hand, efficiency of DMUs is evaluated applying DEA and is ranked by using CEM approach. The proposed algorithm was employed on 20 bank branches. Results show crisp efficiency obtained for each DMU by using combined cluster analysis and CEM has been computed based upon correct weighting. Because this weighting originates from natural similarity of DMUs to each other's so that their inputs and outputs have been considered as attributes. For future researches, it is suggested that proposed algorithm is applied in order to rank DMUs with fuzzy data.

References

[1] Caklovic, L. & Hunjak T. (2012). Measuring DMU-efficiency by modified cross-efficiency approach. Mathematical.

Communication, 17, 559–573.

[2] Hashimoto, A., & Wu, D.A. (2004). A DEA- COMPROMISE PROGRAMMING MODEL FOR COMPREHENSIVE RANKING. Journal of the Operations Research Society of Japan, 47, 73-81.

[3] Jahanshahloo, G.R. Hosseinzadeh Lotfi, F., Rostamy Malkhalifeh, M., & Ahadzadeh Namin M., (2009). A generalized model for data envelopment analysis with interval data. Applied Mathematical Modelling, 33, 3237–3244.

[4] Jahanshahlooa, G.R., Hosseinzadeh Lotfi, F., Rezaieb, V., & Khanmohammadia M. (2011). Ranking DMUs by ideal points with interval data in DEA. Applied Mathematical Modelling, 35, 218–229.

[5] Rousseeuw, P. J. (1987). Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics, 20, 53–65.

[6] Samoilenko, S., Osei-Bryson, K. M. (2008). Increasing the discriminatory power of DEA in the presence of the sample heterogeneity with cluster analysis and decision trees. Expert Systems with Applications, 34, 1568–1581.

[7] Samoilenko, S., & Osei-Bryson, K. M. (2010). Determining sources of relative inefficiency in heterogeneous samples: Methodology using Cluster Analysis. DEA and Neural. European Journal of Operational Research, 206, 479–487.

[8] Sexton, T. R., Silkman, R. H., & Hogan, A. J. (1986). “Data Envelopment Analysis: Critique and Extensions.” in Silkman RH (eds.) Measuring Efficiency: An Assessment of Data Envelopment Analysis Jossey-Bass: 73-105.

[9] Thomassey, S., & Fiordaliso, A. (2006). A hybrid sales forecasting system based on clustering and decision trees. Decision Support Systems, 42, 408-421.

[10] Wu, J., Sun, J., Song, M., & Liang, L. (2013). A ranking method for DMUs with interval data based on dea cross-efficiency evaluation and TOPSIS. Journal of Systems Science and Systems Engineering, 22, 191-20.

[11] Zohrehbandian M. & Sadeghi Gavgani, S. (2013). Cross-efficiency Evaluation under the Principle of Rank Priority of DMUs. World Applied Sciences Journal, 21, 46-49.