European Journal of Operational Research 215 (2011) 498–501

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Short Communication

Some clarifications on the DEA clustering approach

Gholam R. Amin [a,*], Ali Emrouznejad [b], S. Rezaei [c]

[a] Department of Operations Management and Business Statistics, College of Commerce, Sultan Qaboos University, Oman
[b] Aston Business School, Aston University, Birmingham, UK
[c] Postgraduate Engineering Centre, Azad University, South Tehran Branch, Tehran, Iran

A R T I C L E I N F O

Article history:
Received 17 June 2010
Accepted 30 June 2011
Available online 13 July 2011

Keywords:
Data envelopment analysis
Clustering
Alternative optimal solutions

A B S T R A C T

This paper clarifies the role of alternative optimal solutions in the clustering of multidimensional observations using data envelopment analysis (DEA). The paper shows that alternative optimal solutions corresponding to several units produce different groups with different sizes and different decision making units (DMUs) at each class. This implies that a specific DMU may be grouped into different clusters when the corresponding DEA model has multiple optimal solutions.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Clustering is the problem of grouping similar observations of a data set into the same cluster and dissimilar units into different clusters. Many research studies in many different fields have investigated the problem of clustering high dimensional data (Thanassoulis, 1996; Yin et al., 2007; Leonard and Droege, 2008; Rege et al., 2008; Kim et al., 2009; Po et al., 2009). Recently, Po et al. (2009) suggested a data envelopment analysis (DEA) clustering approach for identifying different classes of high dimensional data. The proposed DEA clustering method (Po et al., 2009) uses the piecewise production functions obtained from DEA models for clustering the observed data with multiple inputs and multiple outputs. Besides of the identification of clusters it also provides the production functions of different classes. This note clarifies the role of alternative optimal solutions in the DEA clustering approach. We show that in the presence of multiple optimal solutions the proposed DEA clustering method (Po et al., 2009) concludes different clusters with different sizes and different units at each group. We demonstrate the role of alternative optimal solutions corresponding to a data set consisting of 23 car companies with three inputs and four outputs. We show that alternative optimal solutions corresponding to 23 car companies give different clusters with different sizes and different production functions.

* Corresponding author. Address: P.O. Box 20, PC 123, Muscat, Oman. Tel.: +968 24142910; fax: +968 24414043.
E-mail address: gamin@squ.edu.om (G.R. Amin).

0377-2217/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2011.06.043

2. DEA clustering with alternative solutions

As the proposed DEA clustering method (Po et al., 2009) is based on the optimal solutions of the corresponding DEA models this paper shows that multiple optimal solutions of DEA models generate different clusters for a given group of decision making units (DMUs). We use a data set containing 23 car companies each of which consumes three inputs and produces four outputs. The data is prepared from the Saipa car company in Iran. According to the opinions of experts we consider *research and development* (R&D) (in million Euros), *total assets* (in million Euros) and *average employees* (in 1000 persons) as inputs and *production* (in 1000 cars), *value added* (in million Euros), *net profit* (in million Euros), and *revenue* (in million Euros) as outputs of 23 car companies.

Some studies argued that using variables measured in terms of money, e.g. 'net profit', are not appropriate for measuring (pure) 'technical' efficiency (see Kortelainen and Kuosmanen, 2007; Kuosmanen et al., 2010; Banker et al., 2007). This is because the shadow prices (the variables of the multiplier form of CCR-DEA program) should value one; however this point is not important for the purposes of this work since the objective of this study is denoting the role of alternative optimal solutions in DEA clustering algorithm. Using the proposed method in this paper on any other datasets show the same results.

The data is shown in Table 1.

We solve the following multiplier form of the CCR input-oriented DEA model corresponding to the kth car company ($k = 1,\ldots,23$) given in Table 1.

Table 1
Data for 23 car companies.

DMUs	Inputs			Outputs			
	R&D	Total assets	Average employees	Production	Value added	Net profit	Revenue
DMU_1	2544	79057	103.7	1366.8	12429	2874	48999
DMU_2	5331	190022	365.8	4589.1	29706	3227	151589
DMU_3	1401	58303	170.1	2173.2	9260	1151	51832
DMU_4	317.1	9117.1	26.6	588.2	215	105.6	9982.3
DMU_5	315.8	6119.1	12.4	466.7	139	69.5	6600.1
DMU_6	3450.9	71479.5	141.3	3442	5588	4036.8	66992.1
DMU_7	489	10101.5	33	1140.7	2246	32.8	14551.4
DMU_8	367.1	5866.8	7.5	214	360	314.2	6206.3
DMU_9	372.6	7902	21.1	621	610	398.7	10695.6
DMU_{10}	647.3	12093.9	36.2	904.2	873	451.1	19742.1
DMU_{11}	540.9	9435.9	19.6	904	435	74.3	13740
DMU_{12}	3026.4	77630.7	183.5	3340.8	5644	3502.8	63748.6
DMU_{13}	2175	69050	210.2	3365.9	9518	63	56594
DMU_{14}	1963	68766	131	2385.1	3531	2943	41528
DMU_{15}	608	12506.7	40.1	2200	848	445.9	18569.9
DMU_{16}	82	3147	29.6	456.3	761	296	4108
DMU_{17}	323.2	6945.5	31.3	1142.5	400	226.7	9114.2
DMU_{18}	5494.7	194266.3	275.9	7711	14260	9277.9	142239.3
DMU_{19}	1982	18910	52.3	926.2	5514	1343	31142
DMU_{20}	157.9	3532.7	27.2	556.3	933	371.4	6838.3
DMU_{21}	4588	136603	328.6	5659.6	20065	2749	104875
DMU_{22}	4588	72085	310.6	5659.6	17803	1287	96004
DMU_{23}	30	2104.3	30	530.2	972	561.4	3683.6

$$\max \sum_{r=1}^{4} y_{rk}u_r$$
$$s.t.$$
$$\sum_{r=1}^{4} y_{rj}u_r - \sum_{i=1}^{3} x_{ij}v_i \leqslant 0 \quad j = 1,\ldots,23$$
$$\sum_{i=1}^{3} x_{ik}v_i = 1$$
$$u_r \geqslant 0 \quad r = 1,2,3,4, \quad v_i \geqslant 0 \quad i = 1,2,3$$

Table 2 gives the nonzero optimal solutions, or weights, corresponding to nine car companies.

As shown in Table 3, the optimal solutions of the remaining car companies have at least one zero weight.

The proposed DEA clustering algorithm (Po et al., 2009) has two steps. First, it initially sets

$$p = 0, \quad PF(p) = \phi, \quad C(p) = \phi$$
$$q = 0, \quad R(q) = \phi$$

Where, $PF(p)$ is the production function derived for the kth DMU and $C(p)$ denotes the DMU(s) belong to the pth cluster. This is

Table 2
The nonzero optimal solutions for nine car companies.

DMUs \Weights	v_1^*	v_2^*	v_3^*	u_1^*	u_2^*	u_3^*	u_4^*
DMU_1	0.00015928	0.0000036	0.00299344	0.00000991	0.00002746	0.00006284	0.00000948
DMU_2	0.00008091	0.0000009	0.0010856	0.00000074	0.0000107	0.0000199	0.00000405
DMU_6	0.00008257	0.00000433	0.00287225	0.00002396	0.00000919	0.00009422	0.0000063
DMU_8	0.00107017	0.00005607	0.03722835	0.0003106	0.0001191	0.00122126	0.00008168
DMU_9	0.00092236	0.0000321	0.01911649	0.00010405	0.00014614	0.00037694	0.00006507
DMU_{10}	0.00049306	0.00000984	0.01553981	0.00007422	0.00008074	0.00016112	0.00004
DMU_{11}	0.00074276	0.00001483	0.02340986	0.0001118	0.00012163	0.00024272	0.00006026
DMU_{18}	0.00009932	0.00000077	0.00110686	0.00000514	0.00000477	0.00002281	0.00000479
DMU_{19}	0.00022633	0.00000628	0.00827204	0.00004034	0.00003187	0.00010391	0.00002079

Table 3
DEA optimal solutions with at least one zero weight.

DMUs\Weights	v_1^*	v_2^*	v_3^*	u_1^*	u_2^*	u_3^*	u_4^*
DMU_3	0.00046973	0	0.00201035	0	0.00001062	0	0.0000174
DMU_4	0.00224027	0	0.0109104	0.00004803	0	0	0.00009123
DMU_5	0.00111734	0	0.05240209	0.00040918	0	0.0005162	0.00008894
DMU_7	0	0.00007596	0.00705090	0.00027974	0.00012564	0	0.00002740
DMU_{12}	0.00008811	0.00000392	0.00233677	0	0.00000697	0.00008666	0.00000694
DMU_{13}	0.00029546	0	0.00170057	0.00001506	0.00001354	0	0.00001052
DMU_{14}	0.00012395	0.00000514	0.00307735	0	0	0.00012882	0.00000938
DMU_{15}	0	0.00007315	0.0021228	0.00020515	0	0	0.00002955
DMU_{16}	0.0047747	0	0.02061281	0.00023171	0	0	0.00017341
DMU_{17}	0	0.00011982	0.0053624	0.00077898	0	0	0
DMU_{20}	0	0.00028307	0	0	0	0	0.00014624
DMU_{21}	0.00011688	0	0.00141127	0.00001231	0.0000147	0.00001932	0.00000385
DMU_{22}	0.00007154	0.00000386	0.00126676	0	0.000047	0	0.0000007
DMU_{23}	0	0.00047522	0	0	0.0002038	0	0.0002177

Table 4
23 car companies clustered in nine groups.

Clusters	1	2	3	4	5	6	7
1	DMU_1	DMU_{22}					
2	DMU_2	DMU_3	DMU_{13}	DMU_{21}			
3	DMU_6						
4	DMU_8	DMU_{12}	DMU_{14}				
5	DMU_9						
6	DMU_{10}						
7	DMU_{11}	DMU_7					
8	DMU_{18}	DMU_4	DMU_{15}	DMU_{16}	DMU_{17}	DMU_{20}	DMU_{23}
9	DMU_{19}	DMU_5					

related to a DMU with the optimal solution $(\boldsymbol{u}^*, \boldsymbol{v}^*)$ for which $u_r^* > 0$ and $v_i^* > 0$ for each ($r = 1,\ldots,s$ & $i = 1,\ldots,m$) (case 1 in step 1). Furthermore, if the optimal solution of the kth DMU has at least one zero component then this DMU is assigned to $R(q)$, case 2 in step 1. These sets are initially assumed to be empty. Also, the algorithm assigns the kth DMU to one of the already constructed clusters $C(1),\ldots,C(p)$ if

$$f(x_1,\ldots,x_m,y_1,\ldots,y_s) = \sum_{r=1}^{s} u_r^* y_r - \sum_{i=1}^{m} v_i^* x_i = 0$$

exist in one of the already observed production functions $PF(1),\ldots,PF(p)$. The second step of DEA clustering algorithm (Po et al., 2009) re-evaluates the DMUs belong to $R(1),\ldots,R(q)$ and reclassify them to one of the clusters $C(1),\ldots,C(p)$. To do this, the algorithm multiplies the input items of $R(j)$th DMU by t and substitutes the $R(j)$th DMU by $(tx_{R(j)1},\ldots,tx_{R(j)m},y_{R(j)1},\ldots,y_{R(j)s})$. Then it denotes the solutions of the following equations

$$f_z(tx_{R(j)1},\ldots,tx_{R(j)m},y_{R(j)1},\ldots,y_{R(j)s}) = 0 \quad z = 1,\ldots,p$$

By $t(1),\ldots,t(p)$, respectively, where $f_z(.)$ is the production function of the zth cluster ($z = 1,\ldots,p$) obtained in step 1. Therefore, this DMU is classified into cluster $C(k^*)$ where

$$t(k^*) = \max\{t(1),\ldots,t(p)\}$$

We employed this algorithm and obtained the following nine clusters for 23 companies as shown in Table 4.

The first row of Table 4 shows the first cluster containing DMU_1 and DMU_{22}. Also, the last row shows the last group consisting of two units, DMU_{19} and DMU_5. Note that car companies given in Table 3, contain at least one zero element in their optimal solutions and therefore they are assigned to the clusters according to the second step of DEA clustering algorithm (Po et al., 2009). For example, consider car company 14, DMU_{14}. It is assigned into the fourth group where DMU_8 has all nonzero weights and the maximum index k^* is obtained as follows.

$$t(k^*) = \max\{t(1),\ldots,t(9)\} = t(8) = 0.7521$$

where,

$$\begin{aligned} PF(8) = &-0.00009932 \times 1963t(8) - 0.00000077 \times 68766t(8) \\ &- 0.00110686 \times 131t(8) + 0.00000514 \times 23851t(8) \\ &+ 0.00000477 \times 3531t(8) + 0.00002281 \times 2943t(8) \\ &+ 0.00000479 \times 41528t(8) \\ = &\ 0 \end{aligned}$$

So, DMU_{14} is assigned to the fourth group. The other DMUs are assigned to the clusters similarly. Now we consider alternative optimal solutions for the car companies and obtain different clusters both in the size and the observations content at each group. Note that one can obtain some alternative optimal solutions of a multiplier DEA model from the optimal table of simplex algorithm by using some pivoting steps (Bazarra et al., 2010). It is also possible to obtain an optimal adjacent extreme point, or an alternative optimal solution, corresponding to the DEA model from the current optimal extreme point by a pivoting scheme. It should be noted that the result of this paper can be achieved by finding some alternative optimal solutions, hence we do not need to obtain all alternative

Table 5
Alternative nonzero DEA weights.

DMUs\Weights	v_1^*	v_2^*	v_3^*	u_1^*	u_2^*	u_3^*	u_4^*
DMU_6	0.00008257	0.00000433	0.00287225	0.00002396	0.00000919	0.00009422	0.0000063
DMU_9	0.00109709	0.0000236	0.01921395	0.0000047	0.00017314	0.00038876	0.00006886

Table 6
Alternative DEA solutions.

DMUs\Weights	v_1^*	v_2^*	v_3^*	u_1^*	u_2^*	u_3^*	u_4^*
DMU_1	0.00016206	0.00000384	0.00274215	0	0.00003017	0.00005469	0.00000955
DMU_2	0.00007916	0.00000096	0.00108216	0	0.00001051	0.00001993	0.00000411
DMU_3	0.00046973	0	0.00201035	0	0.00001062	0	0.0000174
DMU_4	0.00224027	0	0.0109104	0.00004803	0	0	0.00009123
DMU_5	0.00111734	0	0.05240209	0.00040918	0	0.0005162	0.00008894
DMU_7	0	0.00007596	0.00705090	0.00027974	0.00012564	0	0.00002740
DMU_8	0.00071118	0.00008528	0.03193033	0.00025493	0	0.00123332	0.00008990
DMU_{10}	0.00002872	0.00007124	0.00331331	0	0	0	0.00005065
DMU_{11}	0.00030621	0.00006161	0.01292383	0.00010871	0.00006699	0	0.00006351
DMU_{12}	0.00008811	0.00000392	0.00233677	0	0.00000697	0.00008666	0.00000694
DMU_{13}	0.00029546	0	0.00170057	0.00001506	0.00001354	0	0.00001052
DMU_{14}	0.00012395	0.00000514	0.00307735	0	0	0.00012882	0.00000938
DMU_{15}	0	0.00007315	0.0021228	0.00020515	0	0	0.00002955
DMU_{16}	0.0047747	0	0.02061281	0.00023171	0	0	0.00017341
DMU_{17}	0	0.00011982	0.0053624	0.00077898	0	0	0
DMU_{18}	0.00013337	0	0.00096847	0	0.00000257	0.00001746	0.00000563
DMU_{19}	0	0.00005009	0.00101052	0	0.00005451	0	0.00002246
DMU_{20}	0	0.00028307	0	0	0	0	0.00014624
DMU_{21}	0.00011688	0	0.00141127	0.00001231	0.0000147	0.00001932	0.00000385
DMU_{22}	0.00007154	0.00000386	0.00126676	0	0.000047	0	0.0000007
DMU_{23}	0	0.00047522	0	0	0.0002038	0	0.0002177

Table 7
Alternative clustering with two classes.

Clusters	
1	DMU_6, DMU_{14}, DMU_{15}, DMU_{17}, DMU_{23}
2	DMU_9, DMU_1, DMU_2, DMU_3, DMU_4, DMU_5, DMU_7, DMU_8, DMU_{10}, DMU_{11}, DMU_{12}, DMU_{13}, DMU_{16}, DMU_{18}, DMU_{19}, DMU_{20}, DMU_{21}, DMU_{22}

optimal solutions. This is important because as it is mentioned in Cooper et al. (2007) different software may give different optimal solutions.

Table 5 gives alternative optimal solutions.

According to Table 5 only two car companies, DMU_6 and DMU_9, have positive weights for all inputs and outputs. Therefore, the remaining 21 DMUs contain at least one zero weight in the corresponding alternative optimal solutions as shown in the following Table 6.

Therefore, according to the DEA clustering algorithm (Po et al., 2009) they are assigned into two clusters as shown below.

This means that the DEA clustering algorithm proposed by Po et al. (2009) gives different clusters both in terms of the number of clusters and the units within in each class, see Table 4 and Table 7, when we consider alternative optimal DEA weights. For the discussed car companies we consider only two sets of alternative optimal solutions and conclude two different clusters. As it is shown here the first set of optimal DEA weights concluded nine clusters meanwhile in the second time we clustered the companies into two classes. For example, in the first clustering DMU_{14} is assigned into the fourth cluster with DMU_8 and DMU_{12}, as it is shown in Table 4. On the other hand this unit is moved into the first group in the second running of the algorithm with units DMU_6, DMU_{14}, DMU_{15}, DMU_{17}, and DMU_{23}. So, DMU_{14} is clustered into two different groups with different members and also with different production functions. The number of different groups may be increased if we consider more alternative optimal solutions. The DEA clustering algorithm proposed by Po et al. (2009) does not explain this important case. The results of different running of the DEA clustering algorithm can be different for high dimensional data set and this may confuse the user of the algorithm. Now consider the non-zero DEA weights corresponding to the production functions of the obtained clusters. The production function of DMU_{14} in the first clustering, given in the fourth row of Table 2, and the production function of DMU_{14} in the second clustering, given in the first row of Table 5, are independent vectors. This means that DMU_{14} can be clustered into two different groups and this reveals the role of alternative optimal solutions in the DEA clustering algorithm.

3. Conclusion

This paper clarified the role of alternative optimal solutions for the DEA clustering approach introduced in Po et al. (2009). It is shown that different optimal solutions may conclude different clusters with different sizes and different production functions.

Acknowledgements

The authors thank the anonymous reviewers and the area editor of EJOR for their insightful comments and suggestions.

References

Banker, R.D., Chang, H., Natarajan, R., 2007. Estimating DEA technical and allocative inefficiency using aggregate cost or revenue data. Journal of Productivity Analysis 27, 115–121.

Bazarra, M., Jarvis, J.J., Sherali, H.D., 2010. Linear Programming and Network Flows, fourth ed. John Wiley & Sons, USA.

Cooper, W.W., Ruiz, J.L., Sirvent, I., 2007. Choosing weights from alternative optimal solutions of dual multiplier models in DEA. European Journal of Operational Research 180, 443–458.

Kim, J., Yang, J., O'lafsson, S., 2009. An optimization approach to partitional data clustering. Journal of the Operational Research Society 60 (8), 1069–1084.

Kortelainen, M., Kuosmanen, T., 2007. Eco-efficiency analysis of consumer durables using absolute shadow prices. Journal of Productivity Analysis 28, 57–69.

Kuosmanen, T., Kortelainen, M., Sipiläinen, T., Cherchye, L., 2010. Firm and industry level profit efficiency analysis using absolute and uniform shadow prices. European Journal of Operational Research 202, 584–594.

Leonard, S.T., Droege, M., 2008. The uses and benefits of cluster analysis in pharmacy research. Research in Social and Administrative Pharmacy 4, 1–11.

Po, R.W., Guh, Y.Y., Yang, M.S., 2009. A new clustering approach using data envelopment analysis. European Journal of Operational Research 199, 276–284.

Rege, M., Dong, M., Fotouhi, F., 2008. Bipartite isoperimetric graph partitioning for data co-clustering. Data Mining and Knowledge Discovery 16 (3), 276–312.

Thanassoulis, E., 1996. A data envelopment analysis approach to clustering operating units for resource allocation purposes. Omega 24, 463–476.

Yin, X., Han, J., Yu, P.S., 2007. Crossclus: User-guided multi-relational clustering. Data Mining and Knowledge Discovery 15 (3), 321–348.