We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



5

Self-Organizing Maps Infusion with Data Envelopment Analysis

Mithun J. Sharma[1] and Yu Song Jin[2]
[1]*Dibrugarh University*
[2]*Korea Maritime University*
[1]*India*
[2]*Republic of Korea*

1. Introduction

This chapter presents work on the use of an artificial intelligence technique to cluster stratified samples of container terminals derived from Data Envelopment Analysis (DEA). This technique is Kohonen's self-organizing map (SOM; (Kohonen, 1995)). Data envelopment analysis measures the relative efficiency of comparable entities called Decision Making Units (DMUs) essentially performing the same task using similar multiple inputs to produce similar multiple outputs ((Charnes et al., 1978)). The purpose of DEA is to empirically estimate the so-called efficient frontier based on the set of available DMUs. DEA provides the user with information about the efficient and inefficient units, as well as the efficiency scores and reference sets for inefficient units. The results of the DEA analysis, especially the efficiency scores, are used in practical applications as performance indicators.

There are many problems associated with applying the DEA in some applications. One problem is that the improvement projection for inefficient units in DEA analysis is concrete relative to its efficiency score. This means, in DEA, relative performance of any DMU can be contrasted only to the efficient DMUs that register unit efficiency score. There is no influence on the performance of efficient DMUs by presence or absence of inefficient DMUs. Therefore, the classical DEA does not actually provide a direct means to rank DMUs based on their relative degrees of efficiency or inefficiency ((Sharma & Yu, 2010)).

The second problem is that the DEA models assume that all DMUs are homogenous and identical in their operations ((Seiford, 1994)). Since various applications have heterogeneous DMUs and there is a high request to evaluate these applications under the DEA due to its acceptance as a performance measurement in different kind of business, we have to modify the DEA to work with these applications. If the heterogeneous DMUs are assessed by DEA without any modifications, the DEA yields a biased performance scores and inaccurate analyses. For example, the resources (land, equipment, and labor) of container terminals varies across the world, which requires to be evaluated in term of its common input characterstics. An essential requirement in analyzing these container terminals is to build a fair referencing system for each container terminal to manage and provide a solid plan that improves all inefficient terminals and supports all efficient terminals. This system can not be assessed under the standard DEA due to the non-homogenous nature of these container terminals in terms of their operations, different standards of equipments, infrastructure, and variety in quay length and area size. These factors will yield unfair benchmarking evaluation

if we apply the standard DEA. In order to conform to the homogeneity assumption, we use learning network clustering (SOM) procedure to minimize total dissimilarity. Nor is the SOM approach the only neural network approach which could be used, however, we find the Kohonen approach to be highly effective method.

2. Algorithm

The DEA tools which are necessary to appreciate the method are described in detail in ((Cooper et al., 2004);(Sharma & Yu, 2010)). There it is demonstrated how one can use these tools to measure efficiency and stratify samples for relative attractiveness to identify competitors with level wise target improvement. Here in this chapter, a new algorithm is provided to perform the DEA computation in non-homogenous DMUs by introducing the SOM-based DEA technique. In the proposed method, some of the non-homogenous DMUs are classified into seperate groups for appropriate benchmarking. Clustering stratified DMUs obtained from DEA using SOM is divided into two steps. The first step is to train the SOM against the DMUs as a training data set. The second one is to map input DMUs to output DMU clusters. The algorithm in the subsequent subsection achieves this objective.

2.1 SOM-based DEA

Assume there are n DMUs, each with m inputs and s outputs. We define the set of all DMUs as $J^1, J^1 = DMU_j, j = 1,...,n$ and the set of efficient DMUs in J^1 as E^1. Then the sequences of J^1 and $E^1$are defined interactively as $J^{l+1} = J^l - E^l$ where $E^l = DMU_p \in J^l | \phi_p^l = l$, and ϕ_p^l is the optimal value to the following linear programming problem:

$$max_{\lambda_i,\phi} \phi_p^l = \phi \tag{1}$$

$$s.t. \sum_{i \in F(J^l)} \lambda_i x_{ji} - x_{jp} \leq 0 \forall j \tag{2}$$

$$\sum_{i \in F(J^l)} \lambda_i y_{ki} - \phi y_{kp} \geq 0 \forall k \tag{3}$$

$$\lambda_i \geq 0, i \in F(J^l) \tag{4}$$

where $k = 1$ to s, $j = 1$ to m, $i = 1$ to n, y_{ki} = amount of output k produced by DMU_{i*}; x_{jp} = input vector of DMU_p, x_{ji} = amount of input j utilized by DMU_i; y_{kp} = output vector of DMU_p. $i \in F(J^l)$ in other words $DMU_i \in J^l$, i.e. $F(.)$ represents the correspondence from a DMU set to the corresponding subscript index set.

The following algorithm accomplishes subsequent stratum.

Step 1: Set $l = 1$. Evaluate the entire set of DMUs, J^l, to obtain the set, E^1, of first-level frontier DMUs (which is equivalent to classical CCR DEA model), i.e. when $l = 1$, the procedure runs a complete envelopment model on all n DMUs and E^1 consists of all of the DMUs on the resulting overall best-practice efficient frontier.

Step 2: Exclude the frontier DMUs from future DEA runs and set $J^{l+1} = J^l - E^l$

Step 3: If $J^{l+1} = 3E^{l+l}$, then stop. Otherwise, evaluate the remaining subset of inefficient DMUs, J^{l+1}, to obtain the new best-practice frontier E^{l+1}.

Stopping Rule: The algorithm stops when $J^{l+1} = 3E^{l+l}$.

The training data is a set of all DMUs, without output variables but includes the class each DMU belongs to, $J = \{DMU_1, DUM_2, ..., DMU_n\}$ of already classified samples. Each sample

$DMUj = \{x_1, x_2, ..., x_m\}$ is a vector where $x_1, x_2, ..., x_m$ represent input variables of the sample. The training data is augmented with a vector $C = \{E_1, E_2, ..., E_{l+1}\}$ where $E_1, E_2, ..., E_{l+1}$ represent the class each sample belongs to. The SOM uses a set of neurons, often arranged in a 2D rectangular or hexagonal grid, to form a discrete topological mapping of an input space, $X \in \Re^n$. At the start of the learning, all the weights $\{w_{r1}, w_{r2}, ..., w_{rm}\}$ are initialised to small random numbers. w_{ri} is the weight vector associated to neuron i and is a vector of the same dimension, n, of the input. m is the total number of neurons. $\mathbf{r}i$ is the location vector of neuron i on the grid. Then the algorithm repeats the following steps.

– At each time t, present an input, $x(t)$, select the winner,

$$v(t) = \arg\min_{k \in \Omega} \|X(t) - W_k(t)\| \tag{5}$$

– Updating the weight of winners and its neighbours,

$$\Delta W_k(t) = \alpha(t\eta(v,k,t))\,[X(t) - W_v(t)] \tag{6}$$

– Repeat until the map converges,

where $\eta(v,k,t)$ is the neighborhood function and Ω is the set of neuron indexes.

3. Application

The proposed SOM-based DEA algorithm was applied to container port industry. Data was collected for 70 container terminals from relevant data sources like *Containerization International Year Book*, *The Drewry Annual Container Market Review and Forecast* and specific field studies of container port. The input and output data were selected from a comprehensive set of literature review ((Hayuth & Roll, 1993);(Park & De, 1993); (Tongzon, 2001);(Barros & Athanassiou, 2004)). The statistics related to the sample are shown in table 1.

Inputs and outputs used in DEA must be measurable, but they need not be measured in the same units. The number of variables used in the DEA formula has direct effect on any particular piece of data. (Szezepura et al., 1992) argue that the number of variables should be limited to the maximum extent. In general, the number of test units should be considerably greater than the total number of variables.

A number of container terminals constitute a container port. For one to one comparison (Song & Cullinane, 2003) we have investigated container terminals selected from underdeveloped to transition economies to developed economies that include large, medium and small container terminals. The following features/measures are chosen as inputs: (1) quay length (meters); (2) terminal area (sq. meters); (3) quay cranes (number); (4) transfer cranes (number); (5) reach stackers (number) and (6) straddle carriers (number). On the other hand, container throughput (TEU)[1] is the most appropriate and analytically tractable indicator of the effectiveness of the production of a port. Almost all previous studies treat it as an output variable, because it closely relates to the need for cargo-related facilities and services and is the primary basis upon which container ports are compared, especially in assessing their relative size, or activity levels. Therefore, throughput is chosen as an output variable.

[1]TEU is the abbreviation for Twenty feet Equivalent Unit, referring to the most standard size for a container of 20 ft in length.

	Throughput	QC	TC	SC	RSC	QL	TA
Mean	882143.414	9	14.185	12.985	80.51	1105.042	517876.1
Std. error	98748.9083	0.666	2.084	2.951	0.852	82.505	48321.5
Median	573,049	8	9	0	7	927.5	350,000
Mode	N/A	6	0	0	2	600	300,000
Std. deviation	826192.642	5.579	17.442	24.692	7.172	690.286	404286.9
Sample variance	6.082*1011	31.130	304.24	609.72	51.441	476495.52	1.63*10
Kurtosis	4.269	1.304	5.114	2.222	2.426	4.064	0.931
Skewness	1.960	1.249	2.004	1.846	1.314	1.800	1.265
Range	3,901,632	24	90	94	36	3646	1,648,000
Minimum	98,368	2	0	0	0	300	20,000
Maximum	4,000,000	26	90	94	36	3946	1,668,000
Sum	61,750,039	630	993	909	596	77,353	36,251,334
Count	70	70	70	70	70	70	70
Confidence level (95%)	196998.679	1.330	4.159	5.887	1.710	164.59	96398.82

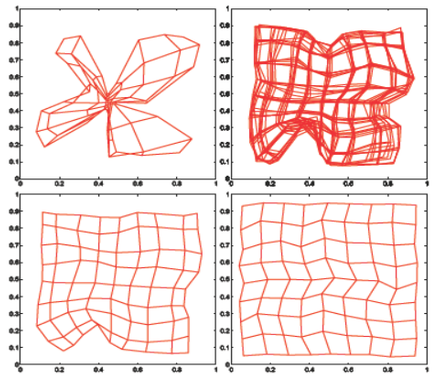
QC:Quay Cranes;TC:Transfer Cranes; SC:Straddle Carrier;RSC:Reach Stacker;QL:Quay Length;TA:Terminal Area

Table 1. Descriptive statistics of container terminal data

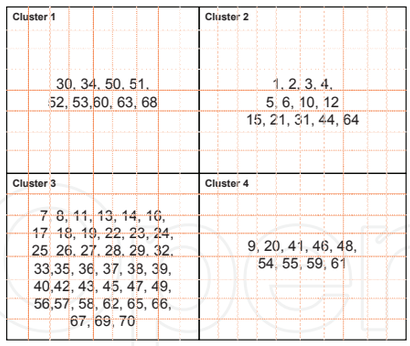
The input variable quay crane is a vital piece of equipment in the production process where it transfers the cargo from a container shipping line across a quay to the shore. This production process fundamentally decides the efficiency of a port, and is vital to its competitive position. As a storage area, the container terminalarea acts as a buffer between sea and inland transportation or transshipment. The size of a ship is very frequently thousands of times the size of the land vehicles that carry the cargo to and from the port. As such, the use of such storage space is normally inevitable. The main pieces of equipment used within a terminal area are the transfer cranes, reach stackers and straddle carriers. (Dowd & Leschine, 1978) argue that the production of a container terminal depends on the efficient use of labor, land and equipment. The measurement of terminal production, therefore, is a means of quantifying efficiency in the utilization of these three resources. Given the characteristics of container port production, the total quay length and the terminal area are the most suitable proxies for the land factor input and the number of quay cranes, the number of transfer cranes, the number of reach stackers and the number of straddle carriers are the most suitable proxies for the equipment factor input. Measures of these variables should be incorporated into the models as input variables

Usually traditional DEA method if employed to obtain efficiency measure for 70 container terminals, would give 18 efficient container terminals with unit score and 52 inefficient terminals with scores < 1. All these inefficient terminals are projected to the 18 efficient terminals for reference irrespective of their differences in efficiency scores. For example, a container terminal with a score of 0.07 is projected to the frontier. It is difficult for this inefficient container terminal to improve its performance by benchmarking a container terminal on the frontier due to huge performance gap along with the differences in their input characteristics. Therefore, it is important to have attainable benchmark target for improvement keeping in view the homogeneity assumption. The partitioning analysis is useful to provide an appropriate benchmark target for poor performers. By using the SOM-based DEA algorithm described in sub-section 3.1, we obtained five levels of efficient frontiers and four clusters. The efficient frontiers are as follows:

$E^1 = \left\{ DMU_j | j = 19;20;29;34;36;39;41;42;45;46;53;54;57;59;60;63;67;69 \right\}$



(a) Two-dimensional Kohonen network



(b) Final clustering of container terminal data

Fig. 1. (a) shows the flattening of a two-dimensional Kohonen network in a quadratic input space. The four diagrams display the state of the network after 100, 1000, 5000, and 10000 iterations.(b) Four clusters of container terminal data at the final state of 1000 iterations.

$$E^2 = \left\{ DMU_j | j = 4;5;8;12;13;15;16;18;26;31;32;35;40;44;47;48;51;58;61;65;66;70 \right\}$$
$$E^3 = \left\{ DMU_j | j = 1;2;7;10;11;17;28;30;37;38;52 \right\}$$
$$E^4 = \left\{ DMU_j | j = 3;9;23;27;43;50;55;62 \right\}$$
$$E^5 = \left\{ DMU_j | j = 6;14;21;22;24;25;33;49;56;64;68 \right\}$$

The proposed SOM-based DEA algorithm produced five stratum of DMUs based on their efficiency level and four clusters as shown in figure 1(b) based on their input traits. Figure 1(a) shows the flattening of a two-dimensional Kohonen network in a quadratic input space. The four diagrams display the state of the network after 100, 1000, 5000, and 10000 iterations.After organizing the DMUs based on our proposed procedure, the projection of inefficient terminals were determined. The inefficient DMUs in the lowest stratum i.e. E^5 benchmarks their immediate upper stratum with similar input features. Same is the case with the DMUs in E^4, E^3, and E^2 belonging to separate clusters.

The application of the model reveals some interesting insight for improving poorly performing terminals. For example, let us consider DMUs 6, 21, 14, and 22 of E^5. DMUs 6 and 21 falls in cluster 2 whereas DMUs 14, and 22 are in cluster 3. Traditional DEA techinique will refer these inefficient terminals to the efficient frontier of 1 irrespective of difference in the efficiency scores. E.g., DMU 6 gets an efficiency score of 0.07 and for improvement, this particular DMU is referred to DMU 9 with efficiency score 1. However, DMU 6 varies from DMU 9 in various aspects of resource mix. Referring to DMU 9 for improvement is unrealistic due to the presence of heterogeneity in their input traits. Using SOM-based DEA technique, DMU 6 is referred to DMU 3 in efficient stratum E^4 with similar input mix (they belong to cluster 2). In the same way DMU 21 in E^5 is also referred to DMU 3 in E^4 as it falls in cluster 2. Whereas DMUs 14 and 22 of E^5 is referred to DMUs 23, 27, 43, and 62 of E^4 as they belong to cluster 3. Thus SOM-based DEA algorithm significantly enhances the capability of traditional DEA tool in prescribing realistic reference points for inefficient DMUs which otherwise is not possible with traditional DEA alone.

4. Conclusion

The benchmarking and improvement projection using the conventional DEA procedure is not desirable because the inefficient DMUs are projected to the efficient frontier ignoring the differences in the efficiency score and input traits. Upon analysis it was found that the efficiency score of DMUs ranged from 4.75% to 100% out of which 18 container terminals are found to be efficient with a score of 1. The 57 inefficient terminals had to refer these limited efficient terminals for improvement. In general, the benchmarking is done to improve the performance of DMUs. But a DMU with low score of 4.75% cannot make direct improvement projections to DMU of score 100%, it needs a DMU with reasonably equivalent characteristics and capacity for benchmarking and improvement.

In this chapter we remedied this limitation of conventional DEA by our proposed SOM-based DEA algorithm. Using the proposed algorithm we could organize the inefficient units into multiple efficient stratum ($E^2 = 22$, $E^3 = 11$, $E^4 = 8$, $E^5 = 11$) thereby providing a level-wise improvement path for poor performing DMUs and also context for evaluation with 4 clusters that conform to the homogeneity assumption, thereby minimizing total dissimilarity in the benchmarking procedure.

5. References

Barros, C.P., Athanassiou, M. (2004). Efficiency in European seaports with DEA: Evidence from Greece and Portugal. *Maritime Economics and Logistics*, Vol. 6, 122 140.

Cooper, W. W., Seiford, L. M., & Zhu, Joe (2004). Name of paper, In: *Handbook on data envelopment analysis*, Kluwer.

Charnes, A., Cooper, W.W., Rhodes, E. (1978). Measuring the efficiency of decision making unit. *European Journal of Operational Research*, Vol. 2, 429 444.

Dowd, T.J., Leschine, T.M. (1990). Container terminal productivity: A perspective. Maritime Policy and Management 17, 107 111.

Hayuth, Y., & Roll, Y.(1993). Port performance comparison applying data envelopment analysis (DEA). *Maritime Policy and Management*, Vol. 20, 153 161.

Kohonen T., 1995. *Self Organizing Maps*, Springer-Verlag, Berlin.

Park, R-K.,& De, P.(2004). An alternative approach to efficiency measurement of seaports. *Maritime Economics and Logistics*, Vol. 6, 53 69.

Seiford, L. (1994). A DEA bibliography in data envelopment analysis (19781992). DEA bibliography. Merest: University of Massachusetts.

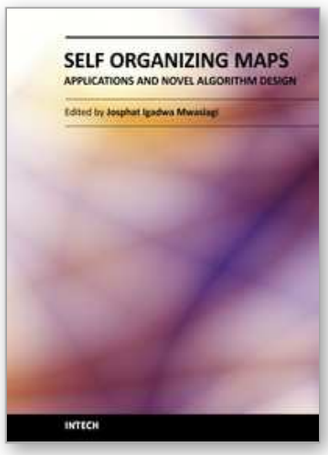
Song, D., Cullinane, K. (2003). A stochastic frontier model of the productive efficiency of Korean container terminals. *Applied Economics*, Vol. 35, 251 267.

Sharma, M.J., Yu, S.J. (2009). Performance based stratification and clustering for benchmarking of container terminals. *Expert Systems with Applications*, Vol. 36, 5016 5022.

Sharma, M.J., Yu, S.J. (2010). Benchmark optimization and attribute identification for improvement of container terminals. *European Journal of Operational Research*, Vol. 201, 568 580.

Szezepura, A., Davis, C., Fletcher, J., Bousoffiane, A.(1992). Applied data envelopment analysis. In: Health Care: The Relative Efficiency of NHS General Practices. Warwick Business School Research Bureau, Coventry.

Tongzon, J.(2001). Efficiency measurement of selected Australian and other international ports using data envelopment analysis. *Transportation Research Part A*, Vol. 35, 113 128.



Self Organizing Maps - Applications and Novel Algorithm Design
Edited by Dr Josphat Igadwa Mwasiagi

ISBN 978-953-307-546-4
Hard cover, 702 pages
Publisher InTech
Published online 21, January, 2011
Published in print edition January, 2011

Kohonen Self Organizing Maps (SOM) has found application in practical all fields, especially those which tend to handle high dimensional data. SOM can be used for the clustering of genes in the medical field, the study of multi-media and web based contents and in the transportation industry, just to name a few. Apart from the aforementioned areas this book also covers the study of complex data found in meteorological and remotely sensed images acquired using satellite sensing. Data management and envelopment analysis has also been covered. The application of SOM in mechanical and manufacturing engineering forms another important area of this book. The final section of this book, addresses the design and application of novel variants of SOM algorithms.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Mithun J. Sharma and Yu Song Jin (2011). Self-Organizing Maps Infusion with Data Envelopment Analysis, Self Organizing Maps - Applications and Novel Algorithm Design, Dr Josphat Igadwa Mwasiagi (Ed.), ISBN: 978-953-307-546-4, InTech, Available from: http://www.intechopen.com/books/self-organizing-maps-applications-and-novel-algorithm-design/self-organizing-maps-infusion-with-data-envelopment-analysis

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen