

DEA based Hierarchical Structure Evaluation and Visualization Method

Kazushige Inoue

Department of Computer Science and
Intelligent Systems
Osaka Prefecture University
Osaka, Japan
inoue@mis.fcs.osakafu-u.ac.jp

Takeo Ichinotsubo

Department of Computer Science and
Intelligent Systems
Osaka Prefecture University
Osaka, Japan
ichinotsubo@mis.cs.osakafu-u.ac.jp

Shingo Aoki

Department of Computer Science and
Intelligent Systems
Osaka Prefecture University
Osaka, Japan
aoki@cs.osakafu-u.ac.jp

Abstract— Data Envelopment Analysis (DEA) is a method for evaluating the management efficiency of Decision Making Units (DMUs) relatively. However, when the number of inputs and outputs items is increased, the results of evaluation value are similar regrettably. In order to evaluate Japanese universities which have various activities, this study proposed the hierarchical structure DEA model which is divided into some activities (finance, achievement, attractiveness, research, and education). Hierarchical structure has constructed by some layers based on the theory that the efficiencies of lower layer can not exceed the ones of upper layer. In addition, this paper proposed the visualization method that the analyst can visually understand which field is strong or weak. Using proposed DEA model and visualizing method, the numerical study showed us to be able to understand the feature of each university.

Keywords—Decision Support System; Data Envelopment
Analysis; Visualization Method;

I. INTRODUCTION

Data Envelopment Analysis (DEA) is a nonparametric method for evaluating relative efficiency of Decision Making Unit (DMU) [1]. DEA assumes DMU's activity that multiple inputs to yield multiple outputs, and estimates this converting process as one evaluation value, so called "efficiency score". DEA has been applied to a variety of managerial and economic problem situations in both public and private sectors [2, 3, 4, 5, 6].

However, there are some problems in applying DEA to data set that consists of a lot of activities such as university [7]. In order to express a lot of activities, it is necessary to prepare a lot of inputs and outputs data. A use of larger data set causes following DEA problems: (1) evaluation does not make sense because the values of efficiency scores for each DMU may be similar, (2) understanding the strong points or weak points of DMU are difficult for analyst because complex relations among a lot of inputs and outputs data is expressed by tablet type result.

To deal with upper DEA problem, DEA analyst uses tacit knowledge that the number of DMUs should be satisfied with restriction of $n > \max \{m*s, 3*(m+s)\}$. Where, n indicates the number of DMUs, and m , s indicate the number of inputs and

outputs [2]. Therefore, if analyst can not collect an enough number of DMUs (n), analyst should decrease the number of inputs and outputs data (m , n). A difficult problem of extracting effective data set for a larger data set still remains.

To deal with lower DEA problem, traditional studies make an attempt to develop the visualization method [8, 9, 10]. These techniques help to understand the shape of efficiency frontier which is used as evaluation criteria. Using these methods, analyst can visually capture the existence of his competitors. However, these methods can not support to know the reason why specific DMU is his competitor.

In order to overcome above two DEA problems, this paper proposes the hierarchical structure DEA model which is divided into some activities and proposes the visualization method that analyst can understand which field is strong or weak.

II. DATA ENVELOPMENT ANALYSIS

A. Overview

DEA is a nonparametric method for identifying efficient production frontiers and evaluating the relative efficiency of DMUs, each of which is a company responsible for converting multiple inputs into multiple outputs. DEA assumes DMUs activity that uses multiple inputs to yield multiple outputs, and defines the process which changes multiple inputs into multiple outputs as efficiency. By comparison with Pareto optimal solution so called "efficient frontier", performance of DMUs is measured relatively.

B. Evaluation Value

In order to describe the mathematical structure of the evaluation value, this paper assumes that there are n DMUs ($DMU_1, DMU_2, \dots, DMU_k, \dots, DMU_n$), where each DMU is characterized by m inputs ($x_{1k}, x_{2k}, \dots, x_{ik}, \dots, x_{mk}$) and s outputs ($y_{1k}, y_{2k}, \dots, y_{rk}, y_{sk}$). Evaluation value of DMU_k is mathematically formulated by

$$\text{Evaluation value of DMU}_k = \frac{u_1 y_{1k} + u_2 y_{2k} + \dots + u_s y_{sk}}{v_1 x_{1k} + v_2 x_{2k} + \dots + v_m x_{mk}} \quad (1)$$

Here u_r is multiplier weight given to the r^{th} output, and v_i is multiplier weight given to the i^{th} input. In DEA, the optimal weights may vary from one DMU to another DMU. Each DMU is assigned a best set of weights with values. There are some calculation algorithms [2, 11] of input/output weights in formula (1). This paper uses the BCC model which is one of basic model. The optimal weights u_r and v_i are measured by following linear programming:

$$\begin{aligned} & \text{Maximize} && \sum_{r=1}^s u_r y_{rk} + \delta \\ & \text{subject to} && -\sum_{i=1}^m v_i x_{ij} + \sum_{r=1}^s u_r y_{rj} + \delta \leq 0 \\ & && (j = 1, 2, \dots, n) \\ & && \sum_{i=1}^m v_i x_{ik} = 1 \\ & && v_i \geq 0, u_r \geq 0, \delta : \text{free} \end{aligned} \quad (2)$$

Here, first restriction represents the condition that the productivity of all DMU becomes 100% or less. And the objective function represents the maximization of the virtual outputs of DMU_k, setting that the virtual inputs of DMU_k is equal to one which represented second restriction. Therefore, the optimal weights represent the convenient weight for DMU_k. Using these optimal weights, the evaluation value of DMU_k is called “efficiency score” in the manner that $\theta_k = 1(100\%)$ means the state of efficiency, while $\theta_k < 1$ means the state of inefficiency.

III. PROPOSED METHOD

A. Overview

This paper proposes the hierarchical structure DEA model and visualization method in order to overcome following two problem: (1) there are similar evaluation values among competitors due to a use of larger data set, (2) understanding the strong points or weak points of DMU are difficult.

B. Hierarchical structure DEA model

Hierarchical structure DEA model makes layered structure based on evaluation perspective and assigns input and output elements to each node that is perspective. Then input and output of upper layer include those of lower layer. It is possible to evaluate based on each perspective by calculating efficiency value of each node.

There are n evaluated DMUs. And they have m input elements and s output elements. Here input elements are $X = (X_1, \dots, X_m)$ and expressed as $X_i = (x_{i1}, \dots, x_{in})$. So X_i indicates i input group of all DMUs. Equally, $Y = (Y_1, \dots, Y_s)$,

$Y_r = (y_{r1}, \dots, y_{rn})$ are defined. The efficiency value of DMU_k shown as θ_k is calculated by formula (2). Moreover, efficiency value is calculated based on possible production set S formed by input and output.

Next it is necessary to consider the case that specific input and output are utilized for analysis. New input and output groups are selected and they are denoted as $X' = \{X' \in X, X' \neq \emptyset\}$, $Y' = \{Y' \in Y, Y' \neq \emptyset\}$. Then the efficiency value of DMU_k shown as θ'_k based on this data set (X', Y') is calculated by formula (2). It means this efficiency value is calculated based on possible production set S' . Thus $S' \in S$ is satisfied and two different efficiency values θ_k, θ'_k are followed inequality $\theta_k \geq \theta'_k$. The efficiency value of selected data is less than that of original data.

The proposed method hierarchizes by resolving the evaluation perspective and assigns input and output elements to each node. In this time, input and output of upper layer have to include those of lower layer. Then analysis is carried out in each node. Thanks to the proposition in previous section, the efficiency value of lower layer never exceeds that of upper layer. Therefore, it is possible to treat the efficiency value of each node evenly since efficiency signifies the state of each DMU by unified way. Thus inefficient DMU as a whole never be the state of efficient in some node. And the reason for decline about efficiency is revealed. For example, which parts of node affect to the efficiency badly or how degree the specific DMU influences the efficiency. It is possible to analyze these kinds of knowledge from all perspective analyst prepares.

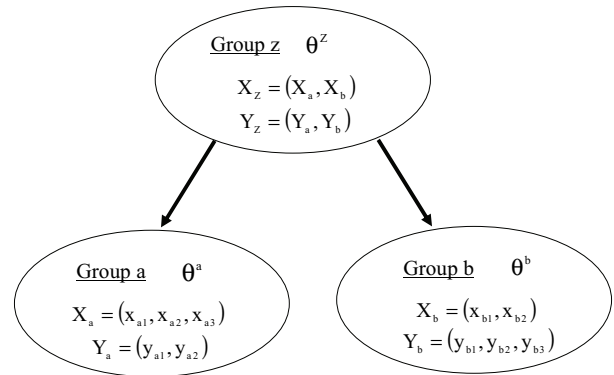


Figure 1. The example of hierarchical structure

Fig 1 shows the example of hierarchical structure. There are two layers (Group z and Group a plus Group b). Group z is divided into two another group (Group a, Group b). Lower nodes use a part of input and output elements in upper layer (Group z). So, proposed structure utilizes the DEA sensitivity analysis technique [2]. Therefore, the efficiency value of the lower nodes (θ^a, θ^b) cannot be exceeded the one's of upper node (θ^z).

Using this mathematical character, this paper overcomes first problem that there are similar evaluation values among

competitors due to a use of larger data set. So, this example assumed that Group z has a larger data set. If there is no difference of the efficiency value θ^z among competitors, the analyst can make a search for own character by digging down the layer.

C. Visualizaiton Method

This paper proposed the visualization method in order to overcome the problem that understanding the strong points or weak points of DMU is difficult.

This study puts the each node in 2D map. Coordinate is sets the efficiency value of each node to vertical axis, and sets the difference with average of each efficiency value to horizon axis.

Fig 2 shows the example of visualization. The node of Group z is set at high-middle position. This means that a similar evaluation among the competitors occurs because Group z has a larger data set. So, this evaluation criterion in Group z does not make a sense. However, second layer shows the difference with Group z. The node of Group a is set at high-right position. This mean that DMU has the strong points in the evaluation criterion Group a because a high efficiency value is maintained compared with competitors. Oppositely, the node of Group b is set at low-left position and analyst can understand the weak point in the evaluation criteria Group b.

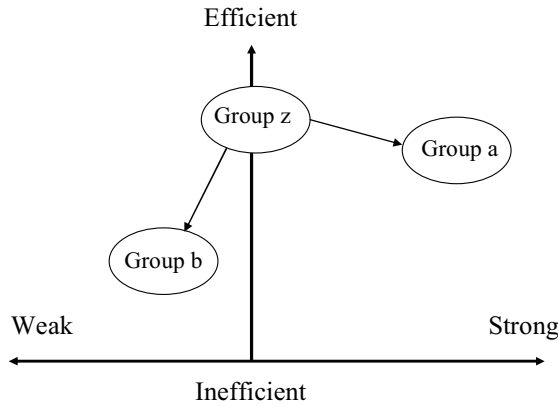


Figure 2. Example of visualization method

IV. NUMERICAL STUDY

A. Data Set

This paper carried out the numerical study using data of Japanese thirty-one universities [7]. There are following six input and eight output data:

Input:

- x_1 : number of faculty,
- x_2 : number of worker,
- x_3 : education and research expense,
- x_4 : grant-in aid for management,

x_5 : general and administrative expenses,

x_6 : profit of donation,

Output:

y_1 : maximum T-score,

y_2 : number of paper,

y_3 : number of graduate student,

y_4 : number of undergraduate student,

y_5 : number of books,

y_6 : grant-in-aid for scientific research,

y_7 : contract research funds,

y_8 : profit of business.

B. Hierarchical Structure

These data set is not satisfied the following tacit knowledge that the number of DMUs should be satisfied with restriction of $n > \max \{m*s, 3*(m+s)\}$. Where, n indicates the number of DMUs, and m, s indicate the number of inputs and outputs.

Then, this paper divided this data set into three layers which has eight nodes (University, Management, Research and Education, Finance, Achievement, Attractiveness, Research, and Education). Each layer and node has following input and output data.

<Third layer>

“Finance” uses three input and one output data (x_4, x_5, x_6, y_8). This efficiency value signifies how much university can get profit with possible funds that are input.

“Achievement” uses four input and one output data (x_2, x_4, x_5, x_6, y_1). This efficiency value signifies how high university has T-score with possible funds and human resources.

“Attractiveness” uses four input and two output data ($x_2, x_4, x_5, x_6, y_3, y_4$). This efficiency value signifies how many university gather students with possible funds and human resources.

“Research” uses two input and two output data (x_1, x_3, y_3, y_4). This efficiency value signifies how many university get external funds or accomplishments of research.

“Education” uses two input and three output data (x_1, x_3, y_2, y_6, y_7). This efficiency value signifies how much university invests funds and human resources to education for students.

<Second layer>

“Management” is combined by three lower nodes “Finance”, “Achievement”, and “Attractiveness”. Therefore, this node uses four input and four output data ($x_2, x_4, x_5, x_6, y_1, y_3, y_4, y_8$).

“Research and Education” is combined by two lower nodes “Research” and “Education”. This node uses two input and five output data ($x_1, x_3, y_2, y_3, y_4, y_6, y_7$).

<First layer>

“University” uses all input and output data ($x_1 \sim x_6, y_1 \sim y_8$). This node estimates the total efficiency of university.

C. Result of Hierarchical Sturcture

Figure 3 shows the result of hierarchical structure in one university. Score in box shows the each efficiency values. And the box also includes the following information: Range, Average, and Variance.

In the node “University”, we confirm that the high efficiency value “Score” is caused by a large data set. So, we can not estimate the character of DMU with only this node. However, we can make a search for DMU’s character by dipping down the layer. This DMU has a strong point in the evaluation criteria “Management”, “Achievement”, and “Attractiveness”. This means that proposed hierarchical structure overcome the following problem: (1) there are similar evaluation values among competitors due to a use of larger data set, (2) understanding the strong points or weak points of DMU are difficult.

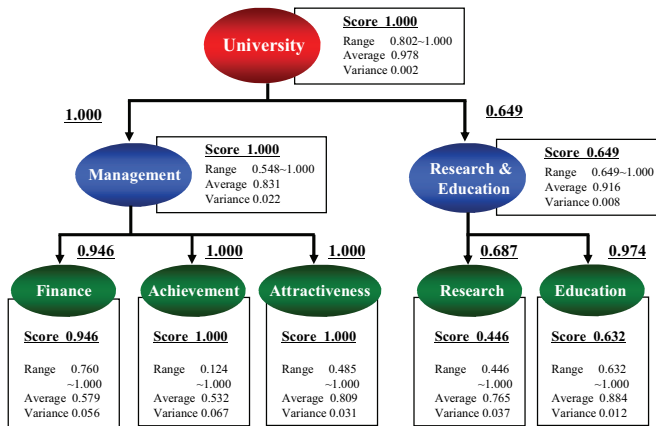


Figure 3. The result of hierarchical structure

D. Visualizaiton Method

Figure 4 shows the result of the visualization method in one university (University A). This figure has same volume of information as Figure 3. This method can intuitively understand the feature of DMU by seeing whether to be located right or left without seeing the numerical data of the score and average.

V. CONCLUSION

In order to evaluate Japanese universities which have various activities, this study proposed the hierarchical structure DEA model which is divided into some activities. Hierarchical structure has constructed by some layers based on the theory that the efficiencies of lower layer can not exceed the ones of upper layer. In addition, this paper proposed the visualization method that the analyst can visually understand

which field is strong or weak. Using proposed DEA model and visualizing method, the numerical study showed us to be able to understand the feature of university.

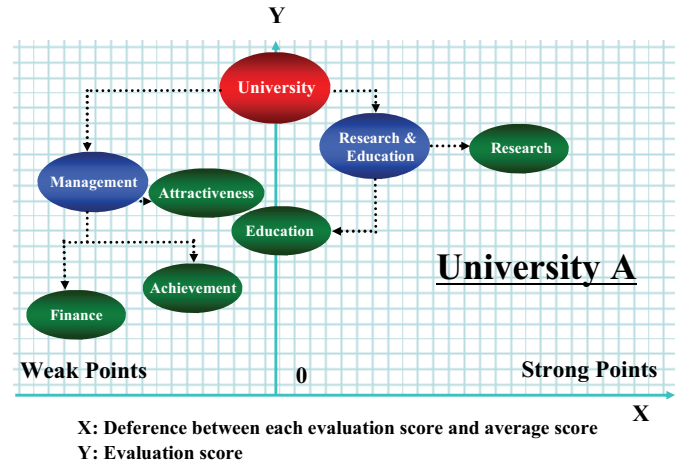


Figure 4. The result of visualization method

REFERENCES

- [1] A.Charnes, W.W.Cooper and E.Rhodes, “Measuring the efficiency of decision making units”, European Journal of Operational Research, Vol.2, pp429-444 (1978)
- [2] W.W.Cooper, L.M.Seiford and K.Tone, “Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-Solver software”, Kluwer Academic Publishers (2000)
- [3] S. Aoki, K. Mishima and H. Tsuji, “Two-Staged DEA model with Malmquist Index for Brand Value Estimation”, The 8th World Multiconference on Systemics, Cybernetics and Informatics, Vol. 10, pp.1-6 (2004)
- [4] E. W. N. Bernroider and V. Stix , “The Evaluation of ERP Systems Using Data Envelopment Analysis”, Information Technology and Organizations, Idea Group Pub, USA, pp.283-286 (2003)
- [5] Y. Zhou and Y. Chen, “DEA-based Performance Predictive Design of Complex Dynamic System Business Process Improvement”, Proceeding of Systems, Man and Cybernetics, IEEE International Conference, pp.3008-3013 (2003)
- [6] T. Sueyoshi, K. Onishi and Y. Kinase, “A Bench Mark Approach for Baseball Evaluation”, European Journal of Operational Research, Vol. 115, pp.429-438 (1999)
- [7] S. Aoki, K. Inoue and R. Gejima, “Data Envelopment Analysis for Evaluating Japanese Universities”, Journal of Artificial Life and Robotics, Vol. 15, No. 2, pp.165-170 (2010)
- [8] K. Honda, A. Notsu, H. Ichihashi, “PCA-guided k-Means with Variable Weighting and Its Application to Document Clustering,” Modeling Decisions for Artificial Intelligence (MDAI) 2009, Lecture Notes in Artificial Intelligence 5861, Springer, pp.282-292 (2009)
- [9] S. Aoki, K. Toyozumi, and H. Tsuji, “Visualizing Method for Data Envelopment Analysis,” Proc. of 2007 IEEE International Conference on Systems, Man and Cybernetics, pp. 474-479 (2007)
- [10] K. Honda, H. Ichihashi, F. Masulli, S. Rovetta, “Linear Fuzzy Clustering With Selection of Variables Using Graded Possibilistic Approach,” IEEE Transactions on Fuzzy Systems, Vol. 15, No. 5, pp.878-889 (2007)
- [11] H. Nakayama, M. Arakawa, Y. B. Yun, “Data Envelopment Analysis in Multicriteria Decision Making”,M. Ehrgott and X. Gandibleux (eds.) Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys, Kluwer Acadmic Publishers (2007)