Journal of the Korean Institute of Industrial Engineers
Vol. 40, No. 6, pp. 628-641, December 2014.
ISSN 1225-0988 | EISSN 2234-6457
http://dx.doi.org/10.7232/JKIIE.2014.40.6.628

<Original Research Paper>

An Optimization Approach to the Construction of a Sequence of Benchmark Targets in DEA-Based Benchmarking

Jaehun Park[1] · Sungmook Lim[2] · Hyerim Bae[1†]

[1]Department of Industrial Engineering, Pusan National University
[2]Dongguk Business School, Dongguk University-Seoul

DEA 기반 벤치마킹에서의 효율성 개선 경로 선정을 위한 최적화 접근법에 관한 연구

박재훈[1] · 임성묵[2] · 배혜림[1]

[1]부산대학교 산업공학과 / [2]동국대학교-서울캠퍼스 경영대학 경영학부

Stepwise efficiency improvement in data envelopment analysis (DEA)-based benchmarking is a realistic and effective method by which inefficient decision making units (DMUs) can choose benchmarks in a stepwise manner and, thereby, effect gradual performance improvement. Most of the previous research relevant to stepwise efficiency improvement has focused primarily on how to stratify DMUs into multiple layers and how to select immediate benchmark targets in leading levels for lagging-level DMUs. It can be said that the sequence of benchmark targets was constructed in a myopic way, which can limit its effectiveness. To address this issue, this paper proposes an optimization approach to the construction of a sequence of benchmarks in DEA-based benchmarking, wherein two optimization criteria are employed : similarity of input-output use patterns, and proximity of input-output use levels between DMUs. To illustrate the proposed method, we applied it to the benchmarking of 23 national universities in South Korea.

Keywords: Data Envelopment Analysis(DEA), Multicriteria, Benchmarking Path, Optimization

1. Introduction

Benchmarking has been defined as "a continuous, systematic process for evaluating the products, services, and work processes of organizations that are recognized as representing best practices for the purpose of organizational improvement" (Spendolini, 1992). Generally, a benchmarking process consists of three steps. The first step is to identify a unit (or a group of units) that is acknowledged as the best performer, the second is to set a benchmarking goal and to recognize the activities required in order to catch up with the best performer, and the third is to implement the best practices and, thus, achieve the benchmarking objective (Donthu *et al*., 2005). For benchmarking, an effective methodology for best-performer identification is essential, for which purpose, data envelopment analysis (DEA) has been widely used (Ross and Droge, 2002). DEA is a methodology for measuring the relative efficiencies of a set of homogeneous decision making units (DMUs) and providing integrated benchmarking information. It identifies an efficient frontier (trade-off curve) that comprises Pareto optimal DMUs along with their respective efficiency scores. DMUs on the efficient frontier can serve as empirical benchmark targets for inefficient DMUs. DEA has

This work was supported by a 2-Year Research Grant of Pusan National University.
† Corresponding author : Professor Hyerim Bae, Department of Industrial Engineering, Pusan National University, Busan 609-735, Korea, Tel : +82-51-510-2733, Fax : +82-51-512-7603, E-mail : hrbae@pusan.ac.kr
Received October 7, 2014; Revision Received November 25, 2014; Accepted December 1, 2014.

been applied to relative-efficiency evaluation of DMUs and benchmarking in various fields, such as banking, health care, agriculture, farming, transportation, education, manufacturing, and others. For more details on DEA application, please refer to Liu *et al*. (2013).

DEA, whereas it is a useful tool for identification of best performers (efficient units), its use has some limitations, in that it requires an inefficient DMU to achieve its target's efficiency in a single move, which might not be feasible in practice. This is true especially when an inefficient DMU under evaluation is far from its benchmark target DMU on the efficient frontier (Cooper *et al*., 2006).

To overcome this problem, various alternative DEA-based benchmarking methods, by which inefficient DMUs can choose benchmark targets in a stepwise manner and, thereby, achieve gradual performance improvement, have been proposed. Talluri (2000) proposed a performance evaluation and benchmarking method that utilizes a combination of DEA, game theory, and clustering methods in identifying effective stepwise benchmarks for poorly performing business processes. Alirezaee and Afsharian (2007) developed a layered efficiency evaluation model for resolving DEA difficulties in the presence of outlier data, based on which a stepwise improvement strategy for inefficient DMUs can be formulated. Estrada *et al*. (2009) proposed a proximity-based stepwise benchmark target selection method whereby DEA is used to determine efficiency scores, a self-organizing map (SOM) is utilized for clustering DMUs according to their input levels, and a reinforcement learning algorithm is adopted to determine the optimal path to the frontier. However, optimal benchmarking paths obtained by this method can vary significantly according to the selected SOM parameter values (e.g., minimum increase rate, discount factor, and map size). Suzuki and Nijkamp (2011) developed an integrated DEA-based technique by combining the distance friction minimization (DFM) and context-dependent models. Although this methodology can provide a stepwise efficiency-improving projection, its practical use is limited, since it might, in the end, provide only hypothetical (not actually observed) benchmark targets for inefficient DMUs. Sharma and Yu (2010) introduced decision tree (DT)-based context-dependent DEA to enhance the capability and flexibility of general DEA. This model proceeds by construction of multiple efficient frontiers for the level-wise reference set and diagnosing the factors that differentiate inefficient DMU performance. Sharma and Yu (2009) proposes a combined data mining/DEA model as a diagnostic tool that can effectively measure the efficiencies of inefficient terminals and prescribe a stepwise projection to reach the frontier in accordance with their maximum capacity and similar input properties. Lim *et al*. (2011) advocated, for stepwise benchmark target selection, the use of the attractiveness and progress measures of context-dependent DEA along with the consideration of feasibility. While their approach employs some optimization criteria for choosing targets, it seeks to select only locally optimal targets, and thus does not assure a globally optimal target sequence. Park *et al*. (2012) proposed a stepwise benchmark target selection method based on preference, direction and similarity criteria. This method integrates the three criteria to construct a more practical and feasible sequence of benchmarks. Park *et al*. (2012) introduced, for the purposes of for port-efficiency improvement, a DEA-based stepwise benchmarking method that considers a minimization-improving performance measure.

As is perhaps apparent, most of the previous research relevant to DEA-based stepwise benchmarking can be considered to have followed a myopic (or local) optimization approach in focusing primarily on how to stratify DMUs into multiple layers and how to select immediate benchmark targets in leading levels for lagging-level DMUs. To address this local optimization issue, we propose a global optimization approach to the construction of a sequence of stepwise benchmark targets in DEA-based benchmarking, in which two optimization criteria are employed: similarity of input-output use patterns, and proximity of input-output use levels between DMUs. The proposed approach suggests a reduced benchmarking network, which is a network structure consisting of an alternative sequence of benchmark targets considering the first optimization criterion, similarity of input-output use patterns. Subsequently it provides a method for selection of the optimal stepwise benchmarking path based on the second optimization criterion, the proximity of the input-output use levels between DMUs. In order to illustrate the effectiveness and demonstrate the advantages of this method, it was applied to 23 national universities in South Korea for determination of optimal stepwise benchmarking paths. In general, two kinds of benchmark reference target for an inefficient DMU can be considered in DEA-based benchmarking: existing units and hypothetical units. The former is an actually existing DMU that is located on the efficient frontier. The latter is not an existing DMU but rather a projection point representing a convex combination of DMUs on the efficient frontier. In setting benchmarking goals or strategies, selecting existing units as the benchmark target can be more practical, because the inefficient DMU that wants to improve its efficiency can utilize actual information. On the other hand, selecting hypothetical units as the benchmark target can lead to unrealistic benchmarking in setting benchmarking strategies or implementing the best practice, because learning additional knowledge from hypothetical units is problematic (Lim *et al*., 2011). For this reason, in the present study, only existing units were considered as benchmark targets.

This paper is organized as follows. Section 2 discusses the framework of the proposed approach, and Section 3 defines the two optimization criteria. Section 4 defines a benchmarking network and a reduced benchmarking network, and Section 5 explains the process of determining the optimal benchmarking path in the reduced benchmarking network. Section 6 details our empirical study, in which the proposed method was applied to 23 national universities in South Korea. Finally, Section 7 summarizes our work and discusses future research.

2. Framework of proposed method

The framework of the proposed method consists of two parts, as shown in <Figure 1> : formulation of a benchmarking network, and choice of the optimal benchmarking path. The procedure for the formulation of a benchmarking network starts by stratifying all DMUs into several layers according to their efficiency scores. After thus stratifying the DMUs, a benchmarking network consisting of every alternative sequence of intermediate benchmark targets (IBTs) is constructed for the use of an inefficient DMU that wants to improve its efficiency score according to an ultimate benchmark target (UBT). Finally, a reduced benchmarking network is constructed by refining the benchmarking network based on the similarity of input-output use patterns. To achieve the reduced benchmarking network, all of the DMUs are classified into several clusters by utilizing a combined cross-efficiency DEA method/K-means clustering algorithm. Based on the reduced benchmarking network, the choice of the optimal benchmarking path can be the optimal sequence of stepwise benchmark targets that maximizes, by application of a network optimization model, the proximity of the input-output use levels between DMUs. To determine the proximity of the input-output use levels between DMUs, the similarity coefficients for the reduced benchmarking network are measured.

The procedure of the proposed method can be regarded as rather complex, since it is composed of several steps including stratification, K-means clustering algorithm, and network optimization. Similarly, we can appreciate that most of the previous methodologies reviewed in section 1 can also be regarded as rather complex. A short summary of the previously presented methodologies is provided in <Table 1>. Nevertheless, the previous methods have been considered practical benchmarking approaches in that they enable an inefficient DMU to improve its efficiency gradationally. Notably, our approach deals with a global optimization issue in order to overcome a drawback of the previous methods, in that they selected immediate benchmark targets only in the leading level for lagging-level DMUs, and suggests two practical optimization criteria for selection of benchmark targets. Thus, the contribution of our approach is the facilitation of inefficient units' efficiency-level increases in terms of global optimization.

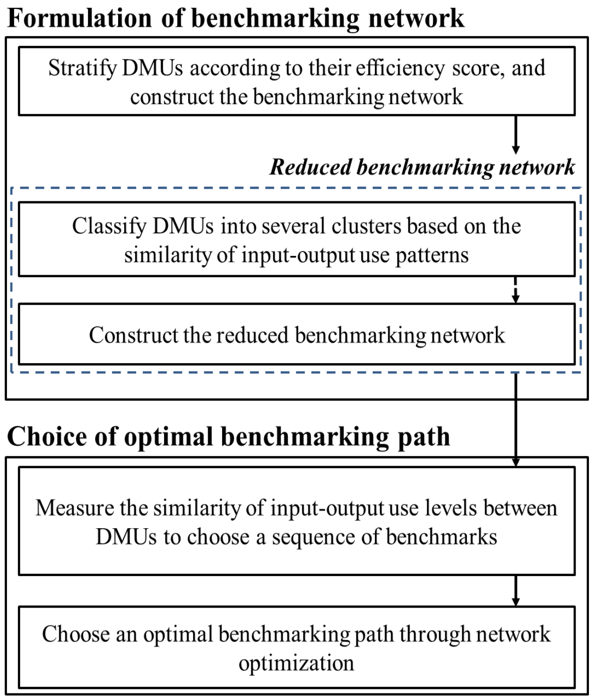


Figure 1. Framework of proposed method

Table 1. Methodologies applied in relevant research

Ref	Methodology	Characteristic
Talluri (2000)	Multifactor productivity analysis, Game theory, Clustering method	
Alirezaee and Afsharian (2007)	Classification, Improvement approach	
Estrada *et al*. (2009)	DEA, Self-organizing map, Reinforcement learning algorithm	
Suzuki and Nijkamp (2011)	Distance friction minimization, Context-dependent model	
Sharma and Yu (2010)	Decision tree, Context-dependent model (with attractiveness, progress, prioritize)	Local optimization approach
Sharma and Yu (2009)	Self-organizing map, Context-dependent model	
Lim *et al*. (2011)	Context-dependent model (with attractiveness, progress, feasibility)	
Park *et al*. (2012)	Self-organizing map, Context-dependent model, Distance measure	
Park *et al*. (2012)	Context-dependent model, Distance-minimization, Sensitive analysis	
This paper	Context-dependent model, Clustering, Cross-evaluation, Network optimization	Global optimization approach

3. Optimization Criteria : Similarity of Input-Output Use Patterns and Proximity of Input-Output Use Levels between DMUs

In the real-world situations, many companies compete with other companies having a similar input or output sizes. For example, a small-sized company would set as a benchmark target not a major company but a small or medium-sized company. González and Álvarez (2001) suggested that when a firm is informed that it is inefficient, a reasonable strategy for its benchmark target selection would be to select and benchmark the efficient firm that is most similar in its input use. In other words, considering the similarity of resource size makes the benchmark target selection more practical. It was for this reason that the present research's two resource-size criteria considered for construction of the optimal sequence of stepwise benchmark targets were similarity of input-output use patterns and proximity of input-output use levels between DMUs.

3.1 Similarity of Input-Output Use Patterns

The first criterion, similarity of input-output use patterns, is employed to reduce the alternative sequence of benchmark targets from the benchmarking network by clustering IBTs, which have the similar input-output use pattern. To illustrate the economical utility of this criterion, consider the simple numerical supermarket example introduced in (Cooper *et al.*, 2006) but with more DMUs added, as shown in <Table 2>. There are twelve DMUs, each consuming two inputs (employee and floor area) and yielding one output (sales). The data and efficiency scores are plotted on a 2-dimensional plane in <Figure 2>.

In <Figure 2>, we suppose that J can choose an IBT between D and H (both are the same efficiency score: 0.667), and that it can then also choose an UBT between A and B (both are the same efficiency score: 1). Given these assumptions, J has four alternative stepwise benchmarking paths: $J \rightarrow D \rightarrow A$, $J \rightarrow D \rightarrow B$, $J \rightarrow H \rightarrow A$, and $J \rightarrow H \rightarrow B$. Let's compare two of them, $J \rightarrow D \rightarrow A$ and $J \rightarrow D \rightarrow B$. In general, an inefficient DMU that wants to improve its efficiency score (hereafter called evaluated DMU) improves by reducing its input usage or increasing its output yield in the DEA. Therefore, when J is regarded as the evaluated DMU, for the benchmarking path $J \rightarrow D \rightarrow A$, J has to reduce inputs ($x1$, $x2$) by 3 and 3, respectively, to benchmark D, and then it has to reduce inputs by 1 and 2, respectively, to benchmark A. On the other hand, for the benchmarking path $J \rightarrow D \rightarrow B$, J has to reduce inputs ($x1$, $x2$) by 3 and 3, respectively, to benchmark D, and then it has to reduce inputs ($x1$, $x2$) by -1 (not reduce $x1$, but inversely increase by 1) and 4 to benchmark B. In other words, for stepwise benchmarking in case of the benchmarking path $J \rightarrow D \rightarrow B$, J has to reduce $x1$ by 3 first and then increase it by 1 inversely. In general, it increasing some inputs or decreasing some outputs should not be considered an unreasonable strategy for a DMU's efficiency score improvement. However, the above-noted benchmarking case can be an unnecessary and ineffective efficiency-improvement approach, because J has to reduce $x1$ and then increase it inversely (hereafter referred to as zigzagging). This zigzagging can be regarded as an unreasonable efficiency-improvement activity.

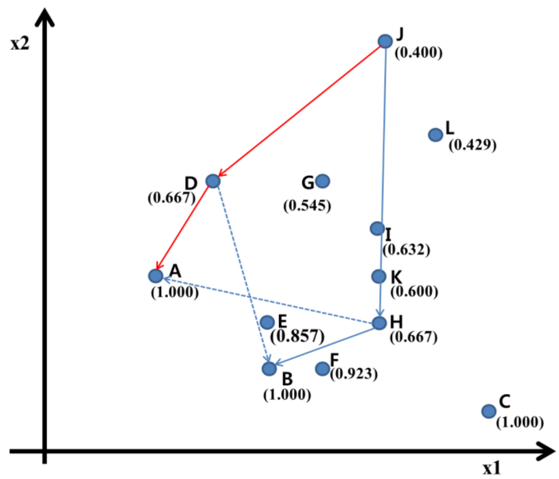


Figure 2. Sample data on 2-dimensional plane

Let's now consider the two other alternative benchmarking paths: $J \rightarrow H \rightarrow A$ and $J \rightarrow H \rightarrow B$. For the benchmarking path $J \rightarrow H \rightarrow B$, J has to reduce inputs ($x1$, $x2$) by 0 and 3, respectively, to benchmark H, and then reduce inputs by 2 and 1, respectively, to benchmark B. On the other hand, for the benchmarking path $J \rightarrow H \rightarrow A$, J has to reduce $x1$ by 4 and increase $x2$ by 1 inversely to benchmark A after it benchmarks H. In the benchmarking path $J \rightarrow H \rightarrow A$, H can be an inadequate IBT, since it leads to an ineffective resource-improvement activity. In other words, if J benchmarks A as its

Table 2. Supermarket example

Store		*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Employees (unit : 10)	x^1	2	4	8	3	4	5	5	6	6	6	6	7
Floor area (unit : 100m^2)	x^2	4	2	1	6	3	2	6	3	5	9	4	7
Sales (unit : \$100,000)	y	1	1	1	1	1	1	1	1	1	1	1	1

UBT, D can be a more effective and proper IBT than H in terms of resource improvement.

In order to minimize the probability of selection of H as the IBT when J benchmarks A as the UBT (zigzagging activity), we suggest a DMU resource use pattern. Let's consider the following case. We assume that an absolute UBT of all DMUs is a hypothetical DMU on zero point ($x1 = 0$, $x2 = 0$). If J, D and A want to benchmark this UBT, they have to improve their resources as (6, 9), (3, 6) and (2, 4), respectively; their resource pattern ratio can be represented as (1:1.5), (1:2) and (1:2), respectively. Further, that of H and B can be represented as both (1:0.5), respectively. Note that the resource pattern ratio in this example can seem commensurable, but it is represented under the condition that units of input 1 and 2 are (10 persons and $100m^2$), respectively. Here, we can see that the resource pattern ratios of A and D are similar to that of J, whereas those of H and B are dissimilar to that of J. Additionally, we can consider the benchmarking direction from the evaluated DMU to the UBT for the resource-improvement pattern. As shown in <Figure 2>, if J wants to benchmark A as its UBT, and if it can select D and H as its IBTs, D is located closer to A than H. Otherwise, if J wants to benchmark B as its UBT, H is closer to B than D. Consequently, although considering the resource use pattern cannot completely avoid the selection of inadequate DMUs as the IBT, it can reduce the probability of doing so.

Furthermore, considering the similarity of the input-output use pattern in selecting IBTs implies the consideration of the similarity of benchmarking strategies or implementation plans. As noted above, an inefficient organization, after selecting its benchmark targets, needs to establish benchmarking strategies and implementation plans. Stepwise benchmarking, however, entails the carrying out of several benchmarking activity steps, each of which requires its own benchmarking strategy and implementation plan. This onerous task can diminish benchmarking efficiencies. Alternatively, a more effective and efficient benchmarking activity for an inefficient DMU might be to select IBTs gradationally, according to similar benchmarking strategies and implementation plans.

3.2 Proximity of Input-Output Use Levels Between DMUs

The proximity of the input-output use levels is employed to select a benchmarking path consisting of a benchmark sequence that minimizes the input reduction or output increase necessary for improvement. In other words, by considering the proximity of the input-output use levels between DMUs, the evaluated DMU can select its IBTs according to a minimum reduction/expansion of the input and output resources. In this regard, Park *et al*. (2012), Baek and Lee (2009) and González and Álvarez (2001) made the point that the minimal effort required to become efficient is closely related to the similarity between DMUs. Especially, Yi *et al*. (2003) emphasized the similarity of benchmarks, and computed the Euclidean distance between the two vectors to determine the degree of similarity between two benchmark DMUs. Amirteimoori and Kordrostami (2010) also proposed a Euclidean-distance-based efficiency measure for DEA. Baek and Lee (2009) proposed a least distance measure (LDM) method to obtain the shortest projection from the evaluated DMU to the strongly efficient production frontier. They checked the validity of the LDM for providing a reasonable measure of efficiency considering conditions such as strongly monotonic, translation invariant and unit invariant. Similarly, Park *et al*. (2012) proposed a benchmarking method that can select benchmark targets with minimum reduction/expansion of the inputs and outputs of the evaluated DMU.

To determine the proximity of the input-output use levels, the LDM method is applied to measure the similarity coefficient with respect to the inputs and outputs. In the general distance concept, a shorter distance denotes less difference between two physical points. In the present context, the difference between two points can be regarded as the necessary extent of resource reduction or expansion between two DMUs. Therefore, selecting a benchmark target by minimizing the reduced inputs or expanded outputs is the same as selecting a benchmark target at the shortest distance from the evaluated DMU.

Let's consider the supermarket example in <Table 2> for better understanding of the concept of the proximity of input and output levels. Assume that L can select its IBT between D and I, which have a relatively similar efficiency. In order for L to benchmark D and I, it has to reduce its inputs by 4 and 1 and 1 and 2, respectively. Measuring the distances from L to D and I by the Euclidean distance method, the distances obtained are 4.12 and 2.24, respectively. In other words, for L, I as a benchmark target minimize the input reduction more than D.

In summary, the first criterion, namely the similarity of input-output use pattern, is used to refine the benchmarking network by grouping IBTs having similar input-output use patterns; the second criterion, the proximity of the input-output levels between DMUs, is used to choose IBTs that minimize the input reduction or output increase necessary for improving the evaluated DMU. The following sections detail the procedure of the proposed method as well as how the proposed two optimization criteria are utilized to construct a sequence of stepwise benchmark targets.

4. Benchmarking Network

4.1 Stratified Benchmarking Paths and Benchmarking Network

According to the framework of the proposed method, all of the DMUs are stratified into several layers according to their efficiency scores so that the evaluated DMU can gradually select IBTs. The context-dependent DEA method proposed by Seiford and Zhu (2003) is utilized to stratify DMUs. By this method, we signify that J^l is the DMU set in the l-th layer and that E^l is the efficient DMU set in the i-th layer. When $l = 1$, the DMUs in set E^1 define the first-level efficient frontier, which might be the most efficient layer. For gradual selection of benchmark targets based on the stratified layers, we specify that the evaluated DMU can sequentially select IBTs in each layer. For example, the evaluated DMU in E^4 can select its IBT in E^3, which is relatively more efficient than that in E^4, and the evaluated DMU in E^3 can select its IBT in E^2.

When the above DEA stratification method is applied to the supermarket example data in <Table 2>, the following stratification result is obtained : E^1 (1-st layer) = $\{A, B, C\}$, E^2 (2-nd layer) = $\{D, E, F\}$, E^3 (3-rd layer) = $\{G, H\}$, E^4 (4-th layer) = $\{K, I\}$, and E^5 (5-th layer) = $\{J, L\}$. Here, the DMUs in the 5-th layer can choose IBTs by selecting DMUs sequentially in layers 4, 3, 2, and 1. Because many DMUs can be included in each layer, an evaluated DMU can produce multiple stratified benchmarking paths.

To represent these multiple stratified benchmarking paths from the evaluated DMU, we define the benchmarking network as follows.

Definition 1 : **Benchmarking Network**
The benchmarking network of the evaluated DMU is defined based on the directed graph, as $p = (D, L)$.

- $D = \bigcup_{i=1}^{L} J^l$ is a DMU set of all J^l.
- $L \subseteq \{l_{jh} = (d_j, d_h) \mid d_j \in E^l, d_h \in E^{l-1}, j \neq h, l = 2, \cdots, L\}$ is a links set, where d_j and d_h represent the index of the j-th DMU and the h-th DMU in D, respectively, and l_{jh} represents the link between the two DMUs d_j and d_h. The element (d_j, d_h) represents the fact that d_j immediately precedes d_h, which means that d_j can benchmark d_h.
- d_h in l_{jh} indicates that the DMU exists in a layer that is more efficient than the layer containing d_j.
- d_j can be linked to multiple d_h, and d_h can be linked to multiple d_j.

Using definition 1 stated above, the benchmarking network comprising multiple stratified benchmarking paths, each of which consists of a sequence of IBTs selected from each layer. If we treat L as the evaluated DMU in the supermarket example, the benchmarking network can be illustrated as in <Figure 3>.

Each circle indicates a node that represents the DMUs in the layers, and each arc between nodes indicates that the previous DMU of an arc can benchmark the next DMU of an arc. This full-sized benchmarking network includes all of the 36 possible alternative benchmark target sequences from the evaluated DMU (L) to the UBTs (A, B, C). This proposed method refines the full-sized benchmarking network by applying similar input-output use patterns.

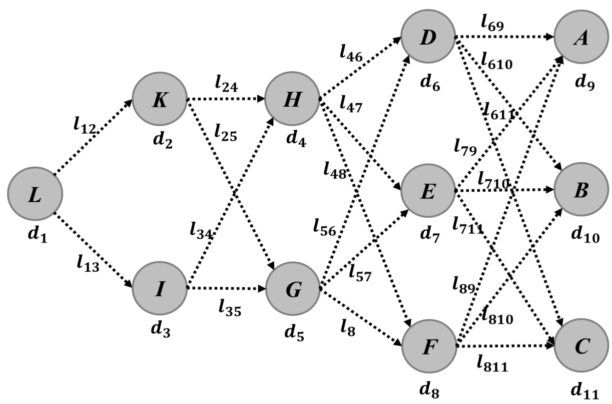


Figure 3. Benchmarking network of supermarket example data of <Table 2>

4.2 Reduced Benchmarking Networks

For application of the similarity of input-output use patterns, all of the DMUs in the full-sized benchmarking network are classified into several clusters. To that end, note that we utilize a method combining a cross-efficiency DEA method with a K-means clustering algorithm. The classification process is performed in two steps : composition of a DMU cross-efficiency matrix using the cross-efficiency DEA method proposed by Sexton *et al*. (1986), and classification of those DMUs by the K-means clustering algorithm proposed by MacQueen (1967). This combined method has been applied by Doyle and Green (1994) and Talluri (2000). However, whereas they classified DMUs to construct layers for the benchmark target selection, we apply this combined method to consider the similarity of input-output use patterns. Additionally, we define a new protocol whereby the DMUs that are similar in terms of input-output use patterns are classified into the same cluster.

In the first step, the cross-efficiency indicates the efficiency score of the DMU under evaluation (hereafter called the target DMU), according to favorable weights assigned by pairwise comparison with other DMUs (hereafter called competitor DMUs) (see Sexton *et al*. (1986)). A target DMU selects optimal weights that maximize its efficiency score and at the same time minimize or maximize the competitor DMUs in turn. Therefore, a single run of the model with a target DMU yields a set of multiple optimal scores; one is for the target DMU itself, and the others are for the competitor DMUs. Generally, the cross-efficiency DEA method can be separated into two models. The first is an aggressive model that aims to maximize the efficiencies of all DMU under evaluation as well as minimize the cross-efficiencies of the other DMUs. The second is a benevolent model that aims to simultaneously maximize both the efficiencies of all of the DMUs under evaluation and the cross-efficiencies of the other DMUs. Because the benevolent model has a drawback, which is its requirement that the efficiency score be raised abnormally, we apply one of the aggressive models, specifically the PEG (Pairwise Efficiency Game) model proposed by Talluri (2000). The PEG model is represented as model (1), where p is the target and Θ_{pp} is the efficiency score of the p-th target DMU evaluated by model (2), which is the general DEA model.

$$\min \sum_{r=1}^{s} v_r y_{rj} \tag{1}$$

$$
\begin{aligned}
s.t. & \sum_{i=1}^{m} u_i x_{ij} = 1 \\
& \sum_{r=1}^{s} v_r y_{rj} - \sum_{i=1}^{m} u_i x_{ij} \leq 0, \ \forall\, i \neq k \\
& \sum_{r=1}^{s} v_r y_{rp} - \theta_{pp} \sum_{i=1}^{m} u_i x_{ip} = 0 \\
& u_i, v_r \geq 0, \ i = 1, \cdots, m, \ r = 1, \cdots, s, \\
& \qquad j = 1, \cdots, J, \ p = 1, \cdots, J
\end{aligned}
$$

$$\theta_{pp} = \max \sum_{r=1}^{s} u_r y_{rk} \tag{2}$$

$$
\begin{aligned}
s.t. & \sum_{i=1}^{m} v_i x_{ik} = 1 \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0; \ j = 1, \cdots, n \\
& u_r, v_i \geq 0; \ r = 1, \cdots, s; \ i = 1, \cdots, m
\end{aligned}
$$

In model (2), u_r is the weight given to the r-th output, v_i is the weight given to the i-th input, n is the number of DMUs, s is the number of outputs, m is the number of inputs, k is the DMU being measured, y_{rj} is the amount of the r-th output produced by DMU j, and x_{ij} is the amount of the i-th input produced by DMU j, respectively.

The PEG model is repeatedly solved by altering the target DMU, resulting in n-1 optimal weights. In the end, each DMU will have n-1 optimal cross-efficiency scores given by n-1 target DMUs along with its own optimal efficiency score. Thus, each DMU constructs a cross-efficiency matrix (J×J). Applying the PEG model to the supermarket example results in the cross-efficiency matrix (12×12) listed shown in <Table 3>.

The cross-efficiency scores of the competitor DMUs vary according to the weights of the target DMU. More specifically, the competitor DMUs that have similar input-output use patterns will have similar cross-efficiency scores under the same target DMU. For example, the cross-efficiency scores of competitor DMUs A and D in <Table 3> are very similar, because their input use patterns are the same as noted in section 3.1 (the input use pattern of A and B is 1:2, respectively). In another case, we can identify that the cross-efficiency scores of competitor DMUs B and H, which have the same input use pattern, 1:0.5, are also very similar.

Based on the cross-efficiency matrix, we classify the DMUs by applying the K-means clustering algorithm, wherein the competitor DMU represents objects to be clustered and the cross-efficiency scores are attributes describing the objects. In the K-means clustering algorithm, the number of clusters is determined by the number of k-centroids. In setting the number of k-centroids, we satisfy two conditions: the UBTs are distributed evenly in each cluster, and there are no inefficient DMUs that have any UBTs in their clusters. Because the efficient DMUs in the 1-st layer can be regarded as the UBTs of the evaluated DMU, and since these UBTs have to be distributed evenly in each cluster and all clusters have to include at least one UBT, the number of k-centroids cannot be greater than the number of efficient DMUs in the

Table 3. Cross-efficiency matrix of supermarket example

		Target DMUs											
		A	***B***	***C***	***D***	***E***	***F***	***G***	***H***	***I***	***J***	***K***	***L***
Competitor DMUs	***A***	1.000	0.500	0.250	0.667	0.500	0.400	0.400	0.333	0.286	0.333	0.333	0.286
	B	0.500	1.000	0.500	0.333	0.667	0.800	0.333	0.667	0.571	0.222	0.500	0.286
	C	0.250	0.500	1.000	0.167	0.333	0.500	0.167	0.333	0.333	0.111	0.250	0.143
	D	1.000	0.500	0.250	0.667	0.500	0.400	0.400	0.334	0.286	0.334	0.334	0.286
	E	0.643	0.857	0.429	0.429	0.857	0.686	0.429	0.571	0.490	0.286	0.571	0.367
	F	0.462	0.923	0.577	0.308	0.615	0.923	0.308	0.615	0.615	0.205	0.462	0.264
	G	0.818	0.681	0.341	0.545	0.681	0.545	0.545	0.454	0.389	0.363	0.454	0.389
	H	0.500	1.000	0.500	0.334	0.667	0.800	0.334	0.667	0.572	0.222	0.500	0.286
	I	0.474	0.948	0.553	0.316	0.632	0.885	0.316	0.632	0.632	0.211	0.474	0.271
	J	0.900	0.600	0.300	0.600	0.600	0.480	0.480	0.400	0.343	0.400	0.400	0.343
	K	0.600	0.900	0.450	0.400	0.800	0.720	0.400	0.600	0.514	0.267	0.600	0.343
	L	0.751	0.751	0.375	0.501	0.751	0.601	0.501	0.501	0.429	0.334	0.501	0.429

1-st layer. Let *kn* be the number of k-centroids and *en* the number of efficient DMUs in the 1-st layer; thus, the procedure for setting the number of k-centroids is as follows.

Procedure : Setting the number of k-centroids

```
k = 1;
Do {
  DMU clustering with the number of k-centroids as k;
  k++;
}
While ((k-en <= 1} or (UBTs cannot be included in any
        cluster));
kn = k-1;
```

For example, if the number of efficient DMUs in the 1-st layer is 4, the number of k-centroids cannot be greater than 4; and further, if any cluster does not include any UBTs under the 4 k-centroids condition, the number of k-centroids is 3. <Figure 4> shows the classified result with the number of k-centroids being 3, that is, the same number as the number of efficient DMUs in the 1-st layer. *A* in the 1-st layer is regarded as the UBT of *J*, *B* is regarded as the UBT of *L*, and the DMUs in each cluster can be considered to be similar in their input-output use patterns. As shown in <Figure 4>, we can identify that *D* and *G* in Cluster 1 are located close to the *J* to *A* benchmarking direction, and that *I*, *K*, *J*, *E* and *F* in Cluster 2 are located close to the *L* to *B* benchmarking direction.

As emphasized in Section 3.1, consideration of the input-output use pattern in selecting IBTs is for the purpose of at least limiting (if not perfectly avoiding) the occurrence of the zigzagging activity and selecting IBTs suitable for maintaining the established gradational benchmarking strategies and implementation plans.

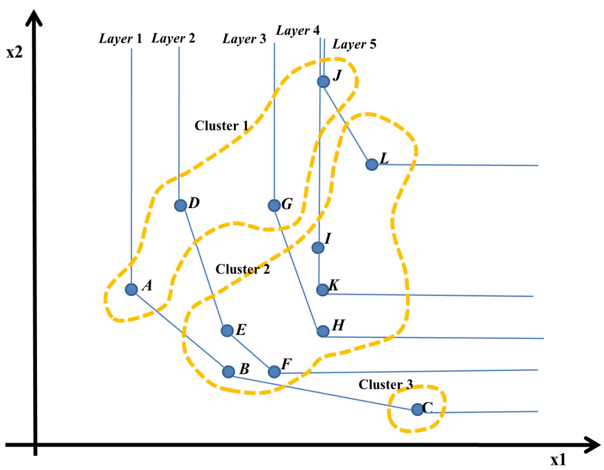


Figure 4. Classification result with number of k-centroids set at 3

Based on the DMU classification, the benchmarking network can be reduced by first defining the benchmarking candidate set, as follows.

Definition 2 : **Benchmarking candidate set** (R^l)
The benchmarking candidate set of the l-th layer is defined as $R^l = \{DMU_{j \in}(E^l \cap C^e) \mid l = 1, \cdots, L\text{-}1\}$, where E^l is the DMU set in the l-th layer, and C^e is the DMU set in the cluster that includes the evaluated DMU. In other words, the benchmarking candidate set is the DMU set in both C^e and E^l.

For example, in <Figure 4>, the benchmarking candidate sets of *L* are $R^1 = \{B\}$, $R^2 = \{E, F\}$, $R^3 = \{H\}$, $R^4 = \{K, I\}$, and $R^5 = \{L\}$. Here, we define that the DMUs in R^l can be regarded as the IBTs of the DMUs in R^{l+1}. In other words, *K* or *I* in R^4 can be regarded as the first IBT of *L* in R^5, and then *H* in R^3 can be regarded as the second IBT of both *K* and *I*. Additionally, *E* or *F* in R^2 is regarded as the third IBT of *H*, and finally, *B* in R^1 can be regarded as the UBT of both *E* and *F*. Here, the number of benchmarking steps for an evaluated DMU depends on not only the number of layers but also whether the benchmarking candidate set exists in the l-th layer. For example, if there are five layers, and each layer has a benchmarking candidate set as shown in <Figure 5>, *L* has four benchmarking steps, whereas if we assume that the 2-nd layer (R^2) has no benchmarking candidate set, *L*, by skipping the 2-nd layer, has just three benchmarking steps.

The reduced benchmarking network is constructed by linking the IBTs between R^l and R^{l+1} in the benchmarking candidate set. In the supermarket example, the reduced benchmarking network of *L* is illustrated in <Figure 5>. Here, we can see that the reduced benchmarking network is a more simplified structure than the benchmarking network.

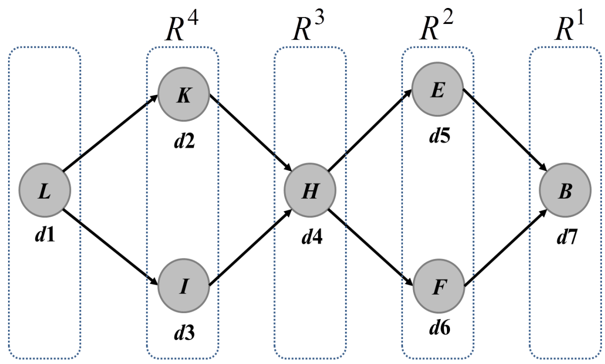


Figure 5. Reduced benchmarking network of DMU L

5. Optimal Choice of Benchmarking Path

5.1 Alternative Criterion for Choosing Sequence of Benchmark Targets

In the previous section, we suggested a benchmarking network comprising multiple stratified benchmarking paths, as well as a reduced benchmarking network refined by consideration of the similarity of input-output use patterns. This section, then, will discuss how we choose the optimal benchmarking path in the reduced benchmarking network. The proximity of the input-output levels between DMUs is considered as the second optimization criterion, and a network optimization method is applied to determine the optimal benchmarking path.

As we mentioned earlier, the distance measure is used for the proximity of the input-output levels measure by applying LDM. The proximity of the input-output use levels between DMUs can be calculated by model (3). Compared with the LDM, the main difference is that model (3) calculates the distance for a real existing unit, not a hypothetical unit.

$$SC_{jh} = \frac{1}{m+s}\sqrt{\sum_{i=1}^{m}\left(\frac{(x_{ij}-x_{ih})}{R_i^-}\right)^2 + \sum_{r=1}^{s}\left(\frac{(y_{rj}-y_{rh})}{R_r^+}\right)^2}$$

where

$$R_i^- = \max_{j=1,\cdots,n}\{x_{ij}\} - \min_{j=1,\cdots,n}\{x_{ij}\},\ i=1,\cdots,m$$

$$R_r^+ = \max_{j=1,\cdots,n}\{y_{rj}\} - \min_{j=1,\cdots,n}\{x_{rj}\},\ r=1,\cdots,s \tag{3}$$

where SC_{jh} represents the similarity coefficient value between d_j and d_h, n is the number of DMUs, s is the number of outputs, m is the number of inputs, x_{ij} and x_{ih} are the amounts of the i-th input produced by d_j and d_h, and y_{rj} and y_{rh} are the amounts of the r-th output produced by d_j and d_h. SC_{jh} has a normalized value between 0 and 1. Note that the lower the similarity coefficient value, the closer the DMUs are in terms of the input and output levels. <Table 4> lists the results of SC_{jh} for all l_{jh} in the supermarket example's reduced benchmarking network. Based on the SC_{jh} for all l_{jh}, the optimal benchmarking path, which maximizes the proximity of the input and output levels between DMUs, is selected through network optimization.

5.2 Choice of Optimal Benchmarking Path Through Network Optimization

To choose the optimal benchmarking path, we apply the method for solving the Shortest Path Problem (SPP), which is one of the common network optimization approaches. In the SPP approach, we define variable x_{jh} : if d_h and d_j in l_{jh} are able to benchmark each other, then x_{jh} is assigned 1, otherwise x_{jh} is assigned 0. When SC_{jh} is given to the reduced benchmarking network, the construction of the optimal sequence of benchmarks can generally be obtained by 0-1 integer programming using model (4).

$$min\ Sp = \sum_{j=1}^{J}\sum_{h=1}^{J} SC_{jh}x_{jh} \tag{4}$$

$$s.t.\ \sum_{h=1}^{J} x_{jh} - \sum_{k=1}^{J} x_{kj} = \left\{\begin{array}{ll} 1 & (j=1) \\ 0 & (j=2,3,\cdots,J-1) \\ -1 & (j=J) \end{array}\right\}$$

$$x_{jh} = 0\ \ or\ \ 1\ (h,j=1,2,\cdots,J)$$

Because, as already noted, a smaller SC_{jh} indicates closer values of input and output level, the objective function minimizes the sum of SC_{jh} assigned to x_{jh}. If this model is applied to the reduced benchmarking network in the supermarket example, the feasible solution is $x_{12}=1$, $x_{13}=0$, $x_{24}=1$, $x_{34}=0$, $x_{45}=0$, $x_{46}=1$, $x_{57}=0$, and $x_{67}=1$. Thus, the optimal sequence of benchmark targets of L can be selected as $d1 \rightarrow d2 \rightarrow d4 \rightarrow d6 \rightarrow d7$, as shown in <Figure 6>. More specifically, L can benchmark K as a first stepwise benchmark target, which can then lead to the sequential benchmarking of H, E and B.

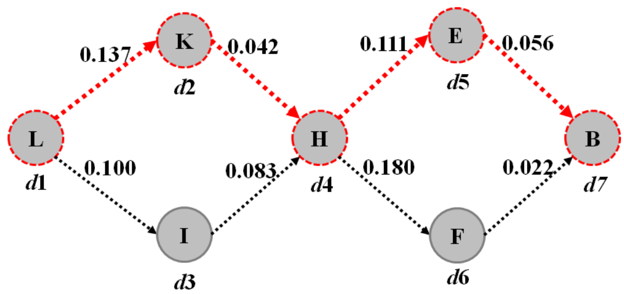


Figure 6. Optimal sequence of DMU L benchmarks

As stressed earlier, most of the previous research relating to the stepwise benchmarking method constructed a sequence of benchmarks in a myopic (local) way; by contrast, the sequence of benchmarks by the proposed method, in <Figure 6>, is optimal in terms of the overall benchmarking path. For better understanding of the difference between these two approaches, consider the <Figure 6> example. If L selects the first IBT between K and I in the partial benchmarking path aspect, I will be selected, because the proximity of its input-output uses levels is greater than that of K (the similarity coefficient score of I is smaller than that of K). However, when L selects H as the second IBT, K will be a more proper target than I, because the sum of the similarity coefficients for $L \rightarrow K \rightarrow H$ is smaller than that for $L \rightarrow I \rightarrow H$ (the sum of SC_{jh} for $L \rightarrow K \rightarrow H$ is 0.179, while for $L \rightarrow I \rightarrow H$ it is 0.183). This result indicates that I is the optimal target of L in the partial benchmarking path aspect, but this DMU might

Table 4. Similarity coefficients for each link in reduced benchmarking network

l_{jh}	SC_{jh}	l_{jh}	SC_{jh}	l_{jh}	SC_{jh}	l_{jh}	SC_{jh}
d_1-d_2 (l_{12})	0.137	d_2-d_4 (l_{24})	0.042	d_4-d_5 (l_{45})	0.111	d_5-d_7 (l_{57})	0.056
$d1$-d_3 (l_{13})	0.100	d_3-d_4 (l_{34})	0.083	d_4-d_6 (l_{46})	0.180	d_6-d_7 (l_{67})	0.022

not be guaranteed to be the optimal target in terms of the overall benchmarking path. The choice of I might, in the long-term efficiency-improvement perspective, restrict the accuracy of stepwise benchmark target selection. To show a more definitive difference, we compare the overall benchmarking paths derived using the myopic approach and the proposed optimized approach, respectively. The benchmarking path by the myopic approach and the proposed method are constructed as $L \rightarrow I \rightarrow H \rightarrow E \rightarrow B$ and $L \rightarrow K \rightarrow H \rightarrow E \rightarrow B$, respectively. As a result, the total sum of the similarity coefficient scores of the benchmarking path by the proposed method (0.349) is smaller than that by the myopic approach (0.339). This result also indicates that the proposed method can construct a more accurate sequence of benchmarks than can the myopic approach.

Since one or more UBTs can be included in a cluster, the evaluated DMU can benchmark one or multiple UBTs. Let's look at another example, this one shown in <Figure 7>. In <Figure 7>(a), inefficient DMU a has two UBTs (g and h). The reduced benchmarking network can be divided into two different reduced benchmarking networks according to each UBT, and the optimal benchmarking network for each reduced benchmarking network can be calculated. The divided reduced benchmarking networks according to UBTs g and h, and each optimal sequence of benchmarks, $a \rightarrow b \rightarrow d \rightarrow e \rightarrow h$ and $a \rightarrow b \rightarrow d \rightarrow f \rightarrow g$, respectively, as shown in <Figure 7>(b) and <Figure 7>(c). Both benchmark sequences can be regarded as optimal benchmarking paths of DMU a. However, if evaluated DMU a wants to select a more favorable sequence of benchmarks in terms of the proximity of input-output uses levels, benchmarking path $a \rightarrow b \rightarrow d \rightarrow e \rightarrow h$ can be selected, because its SC_{jh} sum is less than that of $a \rightarrow b \rightarrow d \rightarrow f \rightarrow g$.

As is already widely known, the SPP can be solved very efficiently. So efficiently in fact, that, from a computational point of view, there would be no need to reduce the full-sized benchmarking network by using the similarity of input-output use pattern. However, note that, as emphasized in section 3.2 above, the reason for reducing the full-sized benchmarking network is not to reduce the computation time but rather to minimize the occurrence of zigzagging activity in selecting IBTs.

6. Application

In a case study, we applied our proposed method to national universities in South Korea in order to select the optimal stepwise benchmarking path of an inefficient university under evaluation. We based it on 2012～2013 data for 23 national and incorporation universities, excluding educational universities, provided by the Ministry of Higher Education in Korea (http://www.academyinfo.go.kr). The data reflected the following selected performance measures.

Inputs :

(i) Tuition : Average tuition for all departments in university

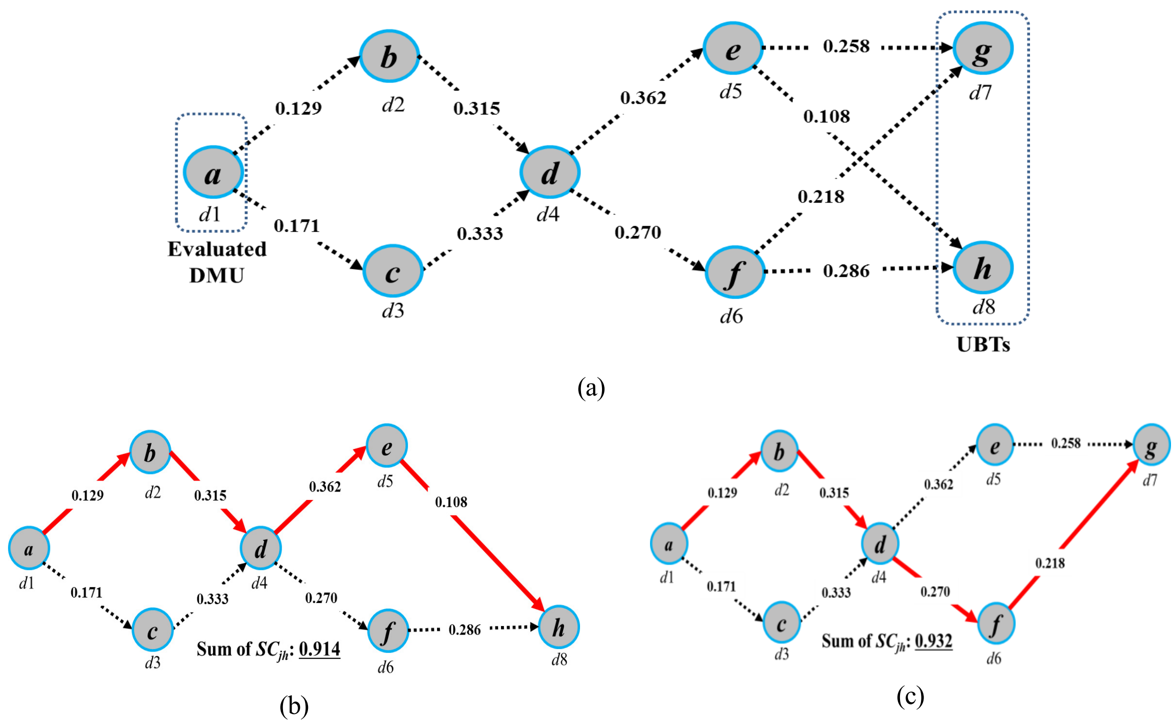


Figure 7. Optimal benchmarking paths when two UBTs are included in cluster

(ii) Student-to-Faculty : Ratio of full-time-equivalent students to full-time-equivalent faculty

Outputs :

(iii) Scholarship-to-Student : Ratio of scholarship money to full-time-students

(iv) Fund-to-Faculty : Ratio of secured funding to full-time faculty

(v) Publication-to-Faculty : Number of publications to full-time faculty

The descriptive statistics of the input and output data and the relative efficiency scores for the 23 universities are listed in <Table 5> and <Table 6>, respectively.

Assessing the relative efficiency, four universities were determined to be the most efficient (efficiency score = 1), while the remaining nineteen universities were determined to be inefficient. From the stratification, six layers of the efficient frontier were determined. We selected university 21 as the evaluated DMU, and set the K-centroids number as 3, according to the procedure noted in section 4.2 above. The stratification and classification are presented in <Figure 8>, and the reduced benchmarking network and similarity coefficient results for the DMUs' proximity of input-output use levels are shown in <Figure 9>, respectively. Note that although the universities were stratified into six layers, there were only four stepwise benchmarking steps, not five, because layer 2 did not have any benchmarking candidate set for university 21.

Because the evaluated DMU had two UBTs (universities 22 and 23), the reduced benchmarking network was divided into two different benchmarking networks, and the optimal benchmarking paths according to the two UBTs were obtained as shown in <Figure 10>. The benchmarking path in <Figure 10>(a) was more favorable in terms of the proximity of input-output uses levels, because its SC_{jh} sum was less than that of the path shown in <Figure 10>(b).

Next, we conducted a comparative experiment with the method proposed by Lim *et al*. (2011) (hereafter called the L-method) to demonstrate the effectiveness of the proposed method. The L-method is regarded as a similar approach to our method in that it was designed to select the optimal benchmarking path considering multiple criteria. However, in the strict sense, the L-method is a partial optimal benchmarking path method because it seeks to select only locally optimal targets, not an overall optimal benchmarking path method. Accordingly, we compared the sequence of benchmarks between the L-method and the proposed method in terms of both the partial and the overall benchmarking path.

Table 5. Descriptive statistics for inputs and outputs used

	Resource	Unit	Max.	Min.	Avg.	Std.Dev.
Inputs	Tuition	₩1,000	6,177.4	3,024.0	4,000.7	708.5
	Student-to-Faculty	Person	41.9	15.6	29.1	6.4
Outputs	Scholarship-to-Student	₩1,000	6,592.4	1,271.1	2,227.1	1,262.9
	Fund-to-Faculty	₩1,000	294,444.5	1,316.0	81,595.7	76,087.8
	Publication-to-Faculty	%	100.0	5.0	38.5	29.2

Table 6. Relative efficiency scores for 23 national universities in South Korea

Univ. No.	University Name	θ	Univ. No.	University Name	θ
19	Seoul National University	**1.00**	2	Kangwon National University	**0.50**
20	Ulsan National Institute of Science and Technology	**1.00**	9	Pukyong National University	**0.48**
22	Korea Advanced Institute of Science and Technology	**1.00**	5	Kongju National University	**0.46**
23	Gwangju Institute of Science and Technology	**1.00**	8	Mokpo National University	**0.44**
10	Pusan National University	**0.88**	14	Cheju National University	**0.42**
4	Gyeongsang National University	**0.65**	18	Korea Maritime University	**0.42**
12	Chonnam National University	**0.63**	11	Andong National University	**0.42**
3	Kyungpook National University	**0.63**	6	Kunsan National University	**0.36**
7	Kumoh National Institute of Technology	**0.61**	15	Changwon National University	**0.32**
13	Chonbuk National University	**0.60**	1	Gangneung-Wonju National University	**0.30**
16	Chungnam National University	**0.58**	21	Incheon National University	**0.24**
17	Chungbuk National University	**0.58**			

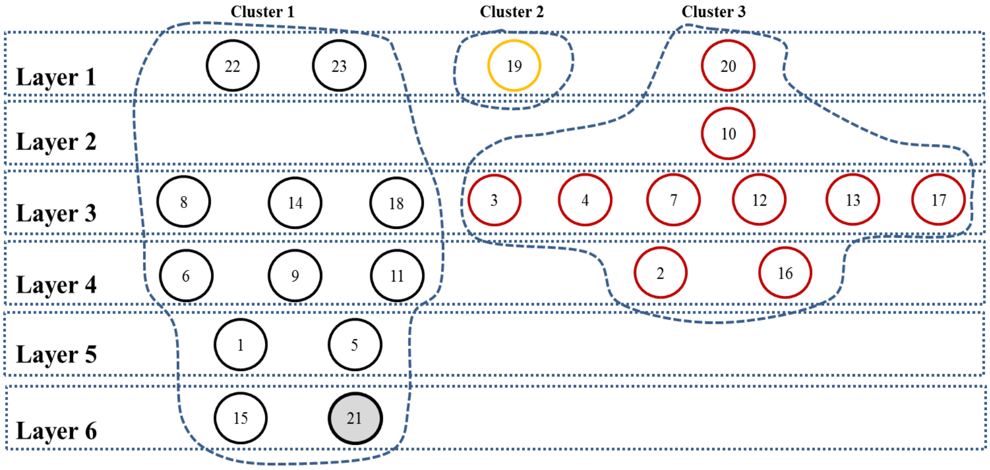
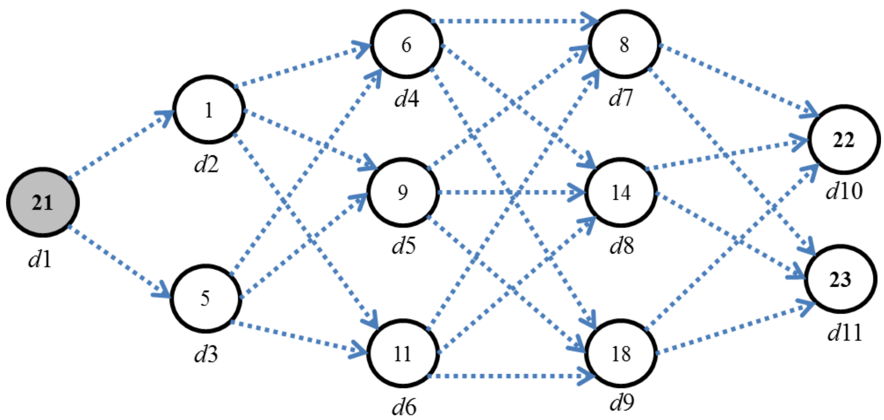
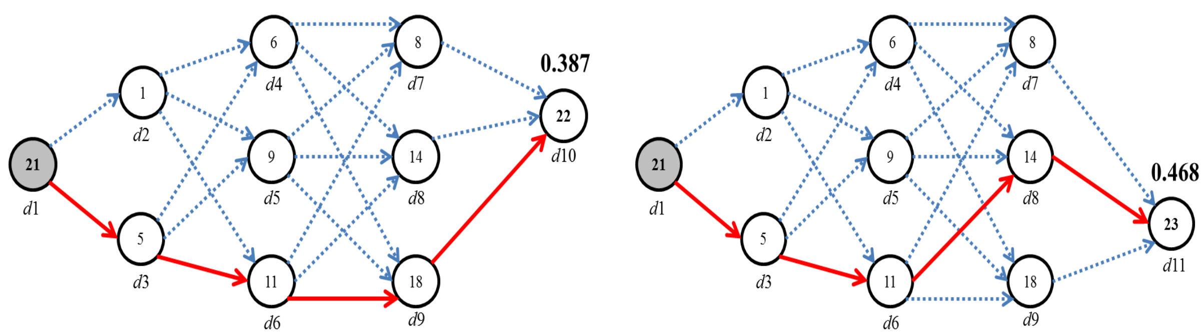


Figure 8. Stratification and clustering results for 23 national universities



l_{jh}	SC_{jh}	l_{jh}	SC_{jh}	l_{jh}	SC_{jh}	l_{jh}	SC_{jh}	l_{jh}	SC_{jh}
*d*1-*d* 2	0.14	*d*3-*d*4	0.05	*d*4-*d*9	0.05	*d*6-*d*8	0.05	*d*7-*d*11	0.27
*d*1-*d* 3	0.12	*d*3-*d*5	0.04	*d*5-*d*7	0.06	*d*6-*d*9	0.04	*d*8-*d*11	0.26
*d*2-*d* 4	0.04	*d*3-*d*6	0.04	*d*5-*d*8	0.05	*d*7-*d*10	0.32	*d*9-*d*11	0.29
*d*2-*d*5	0.10	*d*4-*d*7	0.06	*d*5-*d*9	0.06	*d*8-*d*10	0.32		
*d*2-*d*6	0.06	*d*4-*d*8	0.04	*d*6-*d*7	0.05	*d*9-*d*10	0.19		

Figure 9. Reduced benchmarking network and similarity coefficient results



(a) (b)

Figure 10. Two different benchmarking networks and their optimal benchmarking paths

We first constructed the reduced benchmarking networks considering the similarity of input-output use patterns, and selected the optimal stepwise benchmarking paths that maximize the weighted sum of *Attractiveness*, *Progress* and *Feasibility* (APF) scores used in the L-method instead of the proximity of the input-output use levels. We selected university 21 as the evaluated DMU, and we assigned the same weights as in the case study in (Lim *et al*., 2011) : 0.45, 0.45 and 0.1 to *Attractiveness*, *Progress* and *Feasibility*, respectively. The stepwise benchmarking paths by the L-method, indicated by the green dotted line, and by the proposed method, indicated by the red line, are shown in <Figure 11>. The weighted sums of the APF scores are indicated in parentheses on each arc. In the results of the comparative experiment, the two sequences of universities generated by the L-method and the proposed method differed. The L-method determined university 1 as the first IBT, which has the maximum weighted sum of APF score from university 21, while the proposed method determined university 5 as the first IBT, which is one part of a sequence of benchmarks maximizing the total weighted sum of the APF scores from university 21 to university 22. The total weighted sum of the APF scores from university 21 and university 22 by the L-method and the proposed method were determined to be 0.480 and 0.634, respectively. This result, showing a larger total weighted sum of APF scores by the proposed method, indicates that the L-method could not guarantee the optimal benchmarking path in the overall benchmarking path aspect, and that the benchmarking path by the proposed method was closer to the optimal path.

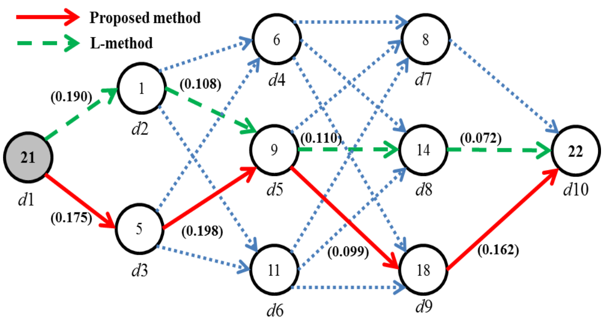


Figure 11. Optimal benchmarking paths by proposed method and L-method

7. Concluding Remarks

In this paper, we proposed an optimization approach to the construction of a sequence of benchmarks in DEA-based benchmarking in terms of an overall stepwise benchmarking path, in which two optimization criteria are employed: the similarity of input-output use patterns, and the proximity of input-output use levels between DMUs. We first suggested a benchmarking network and a reduced benchmarking network considering the similarity of input-output use patterns. Next, we suggested a method for selection, by network optimization, of the optimal stepwise benchmarking path in the reduced benchmarking network, as based on the proximity of the input-output use levels between DMUs. The proposed method was applied to the determination of the optimal stepwise benchmarking path for 23 national universities in South Korea, and the result was compared with that for the L-method. We expect that an effective and reliable benchmark target selection and schedule could be established, because, by considering the two optimization criteria, the optimal stepwise benchmarking path could be selected in terms of the overall benchmarking path.

The proposed methodology, notwithstanding its great utility, does not consider the number of benchmarking steps necessary for an inefficient DMU to reach the UBT. In an actual inefficient organization, the number of benchmarking steps can be an important decision factor : if there are too many stepwise benchmark targets, the benchmarking task might incur significant practical difficulty for the DMU. Therefore, the issue of the number of benchmarking steps as it affects the practical utility of stepwise benchmarking will be a focus of future research.

References

Alirezaee, M. R. and Afsharian, M. (2007), Model improvement for computational difficulties of DEA technique in the presence of special DMUs, *Applied Mathematics and Computation*, **186**(2), 1600-1611.

Amirteimoori, A. and Kordrostami, S. (2010), A Euclidean distance-based measure of efficiency in data envelopment analysis, *Optimization : A Journal of Mathematical Programming and Operations Research*, **59**(7), 985-996.

Baek, C. and Lee, J. (2009), The relevance of DEA benchmarking information and the Least-Distance Measure, *Mathematical and Computer Modeling*, **49**, 265-275.

Cooper, W. W., Seiford, L. M., and Tone, K. (2006), Introduction to data envelopment analysis and its uses : with DEA solver software and references. Interface.

Donthu, N., Hershberger, E. K. and Osmonbekov, T. (2005), Benchmarking marketing productivity using data envelopment analysis, *Journal of Business Research*, **58**(11), 1474-1482.

Doyle, J. and Green, R. (1994), Efficiency and cross-efficiency in DEA : derivations, meanings and uses, *Journal of Operational Research Society*, **45**(5), 567-578.

Estrada, S. A., Song, H. S., Kim, Y. A., Namn, S., H. and Kang, S. C. (2009), A method of stepwise benchmarking for inefficient DMUs based on the proximity-based target selection, *Expert Systems with Applications*, **36**(9), 11595-11604.

González, A. and Álvarez, A. (2001), From efficiency measurement to efficiency improvement : the choice of a relevant benchmark, *European Journal of Operational Research*, **133**(3), 512-520.

Lim, S., Bae, H., and Lee, L. H. (2011), A study on the selection of benchmarking paths in DEA, *Expert System with Applications*, **38**(6), 7665-7673.

Liu, J. S., Lu, L. Y. Y., Lu, W. M., and Lin, B. J. Y. (2013), A survey of DEA application, *OMEGA International Journal of Management Science*, **41**(5), 893-902.

MacQueen, J. (1967), *Some methods for classification and analysis of multivariate observations*, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA : University of California Press, **1**, 281-297.

Park, J., Bae, H., and Lim. S. (2012), A DEA-based method of stepwise benchmark target selection with preference, direction and similarity criteria, *International Journal of Innovative Computing*, *Information and Control*, **8**(8), 10-19.

Park, J., Lim, S., and Bae, H. (2012), DEA-based port efficiency improvement and stepwise benchmarking target selection, *Information-An International Interdisciplinary Journal*, **15**(12c), 6155-6172.

Ross, A. and Droge, C. (2002), An integrated benchmarking approach to distribution center performance using DEA modeling, *Journal of Operations Management*, **20**(1), 19-32.

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

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

Seiford, L. M. and Zhu, J. (2003), Context-dependent data envelopment analysis-measuring attractiveness and progress, *OMEGA International Journal of Management Science*, **31**(5), 397-408.

Sexton, T. R., Silkman, R. H. and Hogan, A. J. (1986), Data envelopment analysis : critique and extensions, *New Directions for Program Evaluation*, **1986**(32), 73-105.

Spendolini, M. J. (1992), The benchmarking book. New York : Amacom.

Suzuki, S. and Nijkamp, P. (2011), A stepwise-projection data envelopment analysis for public transport operations in Japan, *Letters in Spatial and Resource Sciences*, **4**(2), 139-156.

Talluri, S. (2000), A benchmarking method for business-process reengineering and improvement, *The International Journal of Flexible Manufacturing System*, **12**(4), 291-304.

Yi, J. J., Lilja, D. J., and Hawkins, D. M. (2003), *A statistically rigorous approach for improving simulation methodology*, Ninth International Symposium on High Performance Computer Architecture, 281-291.