https://doi.org/10.5392/JKCA.2019.19.07.099

DEA와 의사결정 나무(C5.0)의 하이브리드 모델을 사용한 항만의 효율성 평가

Using a Hybrid Model of DEA and Decision Tree Algorithm C5.0 to Evaluate the Efficiency of Ports

홍한국*, 임병학, 김삼문*****

동의대학교 경영정보학부*, 부산외국어대학교 경영학부**, 동의대학교 응용소프트웨어공학과***

Han-Kook Hong(honghk@deu.ac.kr)*, Byung-hak Leem(bhleem@pufs.ac.kr),**
Sam-Moon Kim(sammoon@deu.ac.kr)***

요약

비모수 생산성 분석기법인 Data Envelopment Analysis (DEA)는 여러 분야의 효율성 평가에 적용되고 있다. DEA 방법론이 다양한 분야의 문제에 대한 현실적 적용에 있어 단점이 있다. 예를 들어 DEA는 각 의사결정단위의 상대적인 효율성 평가에 적합하다. 그러나 이론적인 최대치와의 비교가 아닌 벤치마킹해야 할 참조그룹과 얼마만큼 개선해야 할지를 단지 알려 줄 뿐이다. 즉, 새로운 의사결정단위의 효율성을 측정하기 위해 우리는 과거에 사용된 의사결정단위 데이터와 함께 완전히 새로운 DEA를 적용해야만 한다. 또한 우리는 다시 DEA를 적용하지 않고서 새로운 의사결정단위의 효율성 수준을 예상할 수 없다. 우리는 이러한 DEA의 단점을 보완하기 위해 C5.0과 결합한 하이브리드 분석방법론을 제안한다. 35개의 항만의 효율성 평가를 통해 새로운 의사결정단위는 기존의 의사결정단위와 함께 다시 DEA를 실행할 필요 없이 제안한 방법론을 적용하여 어느 등급에 속하는지 예상할 수 있다.

■ 중심어 : | DEA | C5.0 | Tier Analysis | CCR | 인공지능 |

Abstract

Data Envelopment Analysis (DEA), a non-parametric productivity analysis tool, has become an accepted approach for assessing efficiency in a wide range of fields. Despite of its extensive applications, some features of DEA remain bothersome. For example DEA is good at estimating "relative" efficiency of a DMU(Decision Making Unit), it only tells us how well we are doing compared with our peers but not compared with a "theoretical maximum." Thus, in order to measure efficiency of a new DMU, we have to develop entirely new DEA with the data of previously used DMUs. Also we cannot predict the efficiency level of the new DMU without another DEA analysis. We aim to show that DEA can be used to evaluate the efficiency of ports and suggest the methodology which overcomes the limitation of DEA through hybrid analysis utilizing DEA along with C5.0. We can generate classification rules C5.0 in order to classify any new Port without perturbing previously existing evaluation structures by proposed methodology.

■ keyword : | DEA | C5.0 | Tier Analysis | CCR | AI |

* This Work was supported by Dong-eui University Foundation Grant 2018.

접수일자 : 2019년 06월 17일
수정일자 : 2019년 07월 11일
심사완료일 : 2019년 07월 11일
교신저자 : 홍한국, e-mail : honghk@deu.ac.kr

I. INTRODUCTION

The DEA model is a fractional linear program that aims to assess the comparative efficiency of Decision-Making Units (DMUs) where there are multiple possibly incommensurate inputs and outputs. DEA was developed by Charnes et al.[1] as a generalization of the framework of Farrell [2] on the measurement of productive efficiency. They generalized Farrell's model and allowed it to cast in the form of a fractional expression or ratio. Numerous researches on efficiency measurement of real life problems using DEA have been conducted. DEA has been tested empirically in many settings including schools[3], criminal superior courts[4], fast food restaurants[5], university departments[6], and branch network of a bank[7].

As the earlier list of applications suggests, DEA can be a powerful tool when used wisely. A few of characteristics that make it powerful are as following: First, It doesn't require an assumption of a functional form relating inputs to outputs; Second, it allows managers to consider simultaneously multiple inputs and multiple outputs of a DMU; Third, it provides managers with a procedure to differentiate efficient DMUs from the inefficient ones; Fourth, it pinpoints the sources and the amount of deficiency for each of the inefficient DMUs; Finally, it can be used to detect specific inefficiencies that may not be detectable through other techniques such as linear regression or ratio analyses.

Despite of its extensive applications and merits, some features of DEA remain bothersome. For example DEA is good at estimating "relative" efficiency of a DMU, it only tells us how well we are doing compared with our peers but not compared with a "theoretical maximum." Thus, in order to measure efficiency of a new DMU, we have to develop entirely new DEA with the data of previously used DMUs. Also we cannot predict the efficiency level of the new DMU without another DEA analysis. Second, for DMUs directly compared with a peer or combination of peers, DEA offers no guidelines where relatively inefficient DMUs improve. Finally, it does not provide stepwise paths for improving the efficiency of each inefficient DMU.

In this paper, we aim to show that DEA can be used to evaluate the efficiency of Ports and to suggest the methodology to overcome the limitation of DEA. We present our research framework, which is a hybrid approach utilizing C5.0 to supplement the limitation of DEA. In this methodology, DEA is repetitively used to evaluate the efficiency of DMUs and cluster them together according to their efficiency level (Tier Analysis). We generate the rules for classifying new DMUs into each tier and discriminate among the input and output variables by the degree of affecting the efficiencies of the DMUs (C5.0).

II. Theoretical Background

2.1 DEA

DEA was developed by Charnes et al. as a generalization of the framework of Farrell [14] on the measurement of productive efficiency. DEA, as a non-parametric approach, evaluates relative efficiency of inputs and outputs and determines a set of Pareto-efficient DMUs with an objective of calculating a discrete piecewise frontier. Details of the methodology as well as description of DEA can be found in Charnes et al.[8].

Several characteristics that make DEA powerful are as follows: First, it can handle simultaneously multiple inputs and multiple outputs of a DMU. Second, it does not require an assumption of a functional form relating inputs to outputs. Third, DMUs are directly compared against a peer or combination of peers and it provides managers with a procedure to differentiate between efficient and inefficient DMUs. Fourth, it pinpoints the sources and the amount of deficiency for each of the inefficient DMUs. Fifth, it can be used to detect specific inefficiencies that may not be detectable through other techniques such as linear regression or ratio analyses. Finally, inputs and outputs can have different units of measurement.

Despite of these powerful advantages of DEA, DEA also has many problems mentioned in section I. These problems have been remedied with extension of basic DEA model by many scholars. One of them is to apply different input and output variables to each DMUs tier using decision tree algorithms for classification. and clustering. The most recent remedy study of DEA is evaluation of efficiency in the big data context[9][10].

2.2 Decision Tree Algorithms for Classification

Decision tree classifier provides a hierarchical decomposition of the training data space to divide the data using a condition on the attribute value[11]. For the classification, researchers have developed various decision tree algorithms such as ID3, C4.5 and C5.0 over a period of time with enhancement in performance and ability to handle various types of data. Some algorithms are summarized below [12].

Table 1. Summarization of Decision Tree Algorithms [13]

	C5.0	C4.5	ID3
Type of Data	Continuous & Categorical, dates, times, timestamps	Continuous & categorical	Categorical
Speed	Highest	Faster than ID3	Low
Pruning	Pre-pruning	Pre-pruning	No
Boosting	Supported	Not supported	Not supported
Missing Values	Can deal with	Can't deal with	Can't deal with
Formula	Use split information and gain ratio	Use split information and gain ratio	use information entropy and information gain

2.3 Review of the Efficiency evaluation factors of container terminals

Data Envelopment Analysis (DEA) has been applied to a widely diverse set of business fields and in particular to container terminals for the measurement of financial and operational efficiency. Al-Eraqi, et al.[14] analyzed 22 seaports in the Middle East and East African region for 6 years (2000-2005) with Berth Length(m), Storage Area(m2), and Handling Equipment as inputs, and Ship Calls(Units) and Throughput(Tons) as outputs. Carvalho[15] analyzed forty one ports from eleven European countries using DEA models, and using Operational Expenses(OPEX) and Capital Expenses(CAPEX) as inputs; and conventional general cargo, containerized cargo, roll on-roll off cargo, dry bulk cargo, liquid bulk cargo and passengers as outputs. The study concluded that all the Portuguese ports had very low efficiency scores except Lisbon which was deemed as efficient due to a very high volume of passenger traffic. Wang and Cullinane[16] focused on measuring the efficiency of container terminals in Europe. They proposed using DEA with CCR and BCC models to evaluate efficiency. Using those they came to the conclusion that management skills are crucial and emerge as a core in terms of a port's business competence. Cullinane, et al.[17] used DEA to highlight the major objective of port privatization to improve the efficiency of this sector, with data of the container throughput as output and the area and length of the terminal, quay crane, yard crane, straddle as inputs. All the above authors concluded that public and private/public ports perform better than public/private and private ports. Barros [18] evaluated the performance of 24 Italian seaports for the 2002 -2003 period using multiple efficiency models, such as DEA, CCR, BCC, Cross efficiency DEA, and DEA Super efficiency, whereas previously published articles were limited to only one or two analysis models. The end result was a general conclusion emerging purporting that the Italian companies display relatively high management skills, with most of them being Variable Return to Scale (VRS) efficient. This study also provides benchmarks that will help to improve the functioning of the port especially in terms of efficiency. Lee, et al.[19] analyzed and compared efficiency by RDEA and DEA method for 16 ports in Asia Pacific region, using the No. of Cranes, No. of Container berths, No. of tugs, Terminal area(m2), Delay time(h) and Labor(units) as inputs, and the TEUs handled and Ship rate as outputs. Cullinane, et al.[20] applied window analysis in order to evaluate the efficiency score of the world's major container ports over time by using panel data and cross-section data for 2003. They concluded that the cross-section method is poor because it does not provide details of port performance, whereas the panel data with window analysis reflect a variation of the absolute performance of a port over time, and the relative performance of that port in comparison to the others at the same time. Barros & Manolis[21] compared the efficiency of ports of two European countries, Greece and Portugal. They took data from several ports of each of these countries during the 1998-2000 periods. Their paper is intended to evaluate the efficiency of major seaports in two small European countries using the CCR and BCC models.

Ⅲ. METHODOLOGY

In this chapter, we present our research framework as shown in [Fig. 1]. We generate the rules for classifying new DMUs into each tier and determine the input and output variables that will discriminate the best choice between the tiers by the degree of affecting the efficiencies of the DMUs (discriminant descriptor).

3.1 Definition of input and output data set for DMUs

We propose a port evaluation and improvement model with four inputs and one output as shown in [Table 2]. We select throughput (TEU) used in most papers of literature reviews as output and no. of berths, port depth, yard area, and no. of container cranes, commonly used in seaport reviews as the inputs.

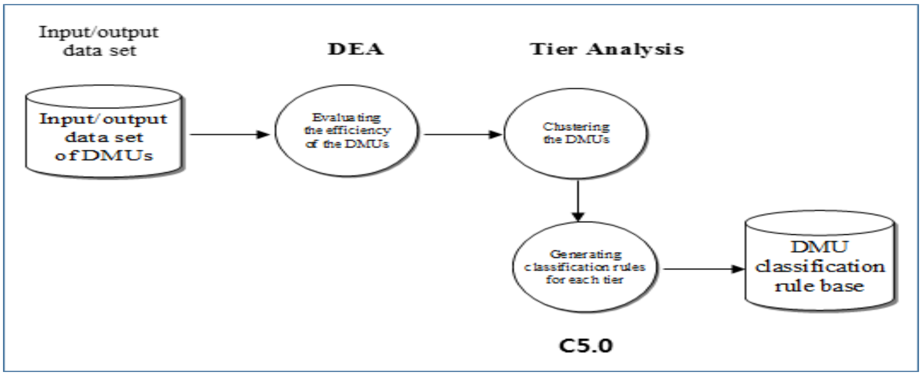

Fig. 1. Research Framework

Table 2. Variable Description

Variable		Description
Input	No. of Berth (#)	Number of berth in container terminal
	Port Depth(m)	Depth of sea in port.
	Yard Area (m2)	Total area of Container Terminal
	No. of Container Crane(#)	Number of Container Crane
Output	Through (TEU)	Total throughput of container (year)

3.2 Evaluating the Efficiencies of DMUs using DEA

A DEA involves an alternative principle for extracting information about a population of observations. In contrast to parametric approaches whose object is to optimize a single regression plane through the data, DEA optimizes on each individual observation with an objective of calculating a discrete piecewise frontier determined by the set of Pareto-efficient DMUs. Both the parametric and non-parametric (mathematical-programming) approaches use all the information contained in the data.

In parametric analysis, the single optimized regression equation is assumed to apply to each DMU. DEA, in contrast, optimizes the performance measure of each DMU. This results in a revealed understanding about each DMU instead of the depiction of a mythical “average” DMU. In other words, the focus of DEA is on the individual observations as represented by the noptimizations (one for each observations) required in DEA analysis, in contrast to the focus on the averages and estimation of parameters that are associated with single-optimization statistical approaches. DEA calculates a maximal performance measure for each a DMU relative to all the DMUs in the observed population with the sole requirement that each DMU lie on or below the extreme frontier. Each DMU not on the frontier is scaled against a convex combination of the DMUs on the frontier facet closest to it.

The solid line represents a frontier derived by applying DEA to data on a population of DMUs, each utilizing different amounts of a single input to produce various amounts of a single output. It is important to note that DEA calculations, because they are generated from actual observed data for each DMU, produce only relative efficiency measures. The relative efficiency of each DMU is calculated in relation to all the other DMUs, using the actual observed values for the outputs and inputs of each DMU.

The DEA calculations are designed to maximize the relative efficiency score of each DMU, subject to the condition that the set of weight. Obtained in this manner, DMU must also be feasible for all the other DMUs included in the calculation. For each inefficient DMU (one that lies below the frontier), DEA identifies the sources and level of inefficiency for each of the inputs and outputs. The level of inefficiency is determined by comparison to a single referent DMU or a convex combination of other referent DMUs located on the efficient frontier that utilize the same level of inputs and produce the same or a high level of outputs. Details of the methodology as well as descriptions of data envelopment analysis can be found in Charnes, et al.[8] We evaluate the efficiencies of the Asia pacific seaports with a DEA.

3.3 Clustering the DMUs through the Tier Analysis

In previous section, we used DEA to evaluate the efficiencies of Ports. DEA determines the most productive group of the DMUs and the group of less-productive DMUs. That is, the DMUs are clustered into an efficient group or an inefficient one by DEA. A similar approach to clustering DMUs by DEA was presented by Thanassoulis[22]. However, the clusters on that study were not made by their efficiency levels but by the characteristics of the input resource mix. Tier analysis that we propose is a kind of technique that can be used to cluster DMUs together according to their efficiency levels.

In the first step of tier analysis, we obtain the efficiency scores of the set of entire DMUs. The result of the first step should reveal the most efficient group of DMUs by indicating their scores are equal to 1.0. I call this group "Tier 1". In the second step, we proceed DEA again only with the inefficient DMUs which are not part of Tier 1. DMUs whose efficiency scores in the second step are equal to 1.0 are Tier 2. The same procedure can be repeated during the number of remaining inefficient DMUs is at least three times multiple of that of inputs along with outputs (4 + 1 = 5), as Banker et al. [23] have proposed, which makes it possible to appropriately discriminate efficient DMUs from inefficient ones. we call this procedure "the tier analysis" because DMUs that belong to each tier form the efficient production frontier in each step.

3.4 Generating classification Rules for each tier using C5.0

A typical decision tree learning system, C5.0, which is going to be used to generate the rule set for classifying 35 ports, adopts a supervised learning scheme that constructs decision trees from a set of examples. A decision tree is a directed graph showing the various possible sequences of questions (tests), answers, and classifications. The method first chooses a subset of the training examples (window) to form a decision tree. If the tree does not give the correct answer for all the objects, a selection of the exceptions (incorrectly classified examples) is added to the window and the process continues until the correct decision set is found. The eventual outcome is a tree in which each leaf carries a class name, and each interior node specifies an attribute with a branch corresponding to each possible value of that attribute.

C5.0 uses an information theoretic approach aiming at minimizing the expected number of tests to classify an object. The attribute selection part of C5.0 is based on the assumption that the complexity of the decision tree is strongly related to the amount of information. An information based heuristic selects the attribute providing the highest information gain ratio, i.e., the ratio of the total information gain due to a proposed split to the information gain attributable solely to the number of subsets created as the criterion for evaluating proposed splits. The C5.0 system uses an information gain ratio as the evaluation function for classification, with the following equation[24]. We generate the rules for classifying new DMUs into each tier and determine the input and output variables that will discriminate best between the tiers by the degree of affecting the efficiencies of the DMUs (discriminant descriptor).

Ⅳ. RESULTS OF ANALYSIS

4.1 Evaluating the Efficiency of 35 ports using DEA and Clustering 35 ports through the Tier Analysis

We choose 35 Asia-pacific ports of top 70 ports based on throughput in the Containerization International Yearbook 2016 [25] in order to evaluate efficiency and provide an improvement model for inefficient ports. According to Banker, et al. [18], the minimum number of DMUs to analyze is greater than max {m*n, 3(m+1)}, where m is the number of input factors and n is the number of output factors. Hence, 35 ports as DMUs is a sufficient number for the DEA analysis. We group 35 ports together into four tiers by the tier analysis. The efficiency score itself is not important in this time. Only what matters is to which tier each port belongs. (Refer to [Table 3]).

The table shows that four ports, including such No. 24, 25, 26 and 33 are best-practiced companies with DEA productivity rating of 100 percent. Port No. 1 (Los Angeles) is less productive with DEA productivity rating of 31 percent, suggesting that it could provide its current mix and volume of outputs with only about 31 percent of the resources it actually consumes and belongs tier 4. In fact, 31 of the 35 ports are using excess resources. These findings indicate that there is room that the 31 ports could make substantial productivity improvements and cost reductions.

We group 35 ports together into four tiers by the tier analysis. The efficiency score itself is not important in this time. Only what matters is to which tier each company belongs.

Table 3. DEA and Tier Analysis Results

No.	DMU	Score	TIER	Ref. Set
1	Los Angeles	0.31	4	Shanghai, Shenzhen
2	Long Beach	0.24	5	Shanghai, Shenzhen
3	New York/New Jersey	0.22	6	Shanghai, Shenzhen
4	Savannah	0.15	5	Shanghai, Shenzhen
5	Oakland	0.11	6	Shanghai, Shenzhen
6	Virginia	0.12	6	Shanghai, Shenzhen
7	Seattle	0.01	6	Shanghai, Shenzhen
8	Tacoma	0.01	5	Shanghai, Shenzhen
9	Houston	0.14	6	Shenzhen
10	Charleston	0.01	6	Shanghai, Shenzhen, Hong Kong
11	Port Everglades	0.17	5	Shenzhen
12	Miami	0.01	6	Shenzhen
13	Kaohsiung	0.56	3	Shanghai, Shenzhen
14	Keelung	0.18	4	Shenzhen, Lianyungang
15	Taichung	0.15	5	Shenzhen
16	Busan	0.56	2	Shanghai, Shenzhen, Hong Kong
17	Gwangyang	0.15	6	Shenzhen
18	Incheon	0.21	4	Shenzhen
19	Tokyo	0.23	4	Shanghai, Shenzhen, Hong Kong
20	Yokohama	0.16	5	Shanghai, Shenzhen, Hong Kong
21	Nagoya	0.13	5	Shanghai, Shenzhen, Hong Kong
22	Kobe	0.14	6	Shanghai, Shenzhen, Hong Kong
23	Osaka	0.12	6	Shanghai, Shenzhen, Hong Kong

24	Shanghai	1	1	
25	Hong Kong	1	1	
26	Shenzhen	1	1	
27	Qingdao	0.60	3	Shanghai, Shenzhen, Hong Kong
28	Ningbo	0.86	2	Shenzhen
29	Guangzhou	0.65	2	Shanghai, Shenzhen
30	Tianjin	0.51	3	Shanghai, Shenzhen, Hong Kong
31	Xiamen	0.68	2	Shenzhen
32	Dalian	0.26	4	Shanghai, Shenzhen, Hong Kong
33	Lianyungang	1	1	
34	Yantai	0.11	5	Shenzhen
35	Fuzhou	0.13	6	Shenzhen

4.2 Generating classification Rules for each tier using C5.0

C5.0 is a program that generates a decision tree. The decision tree is a directed graph showing where new data belongs in predetermined classes by the rule set for dividing classes. This requires data that defines the class in order to generate the rule set, algorithm to make rules, and how to present the generated rules.

In order to generate the decision tree, this study uses C5.0 well known and developed by Quinlan. [Table 4] shows sample cases for training C5.0, and [Fig. 2] shows a decision tree generated with [Table 4] for 35 ports using C5.0.

Table 4. Training cases for C5.0.

Input No. / Port	No. Berth	Depth	Yard Area	No. C/C	Tier
1	29	13.84	6,477,336	71	4
2	34	14.62	4,889,227	60	5
3	36	13.39	5,566,100	70	6
:	:	:	:	:	:

The output of the decision tree generator in our instance appears in [Fig 2]. Note that in the numbers at the leaves, of the form (N) or (N/E), N is the sum of the fractional cases that reach the leaf; E is the number of cases that belong to classes other than the nominated class.C5.0 generates classification rules and determines which input and output factors affect most the efficiency of ports. In the decision tree, the sequence of leaf indicates the order of influences on efficiency score. In [Fig 2], we find that the most important factor on efficiency are No. of container cranes, Yardarea, Depth, and No. of berths. Decision trees are usually simplified by discarding one or more sub-trees and replacing them with leaves; as when building trees, the class associated with a leaf is found by examining the training cases covered by the leaf and choosing the most frequent class. C5.0 also allows replacement of a sub-tree by one of its branches. From the decision tree, we can extract classification rules (refer to [Fig. 3]).

```
C/C > 75 : 1 (3.0/2.1)
C/C <= 75 :
|   Total_Area <= 1.1403e+006 :
|   |   Total_Area > 757000 : 5 (5.0/2.3)
|   |   Total_Area <= 757000 :
|   |   |   Depth <= 12.85 : 4 (3.0/2.1)
|   |   |   Depth > 12.85 : 2 (2.0/1.0)
|   Total_Area > 1.1403e+006 :
|   |   Total_Area <= 1.42137e+006 :
|   |   |   Depth <= 13 : 6 (2.0/1.0)
|   |   |   Depth > 13 : 3 (4.0/2.2)
|   |   Total_Area > 1.42137e+006 :
|   |   |   Berths <= 14 : 6 (6.0/2.3)
|   |   |   Berths > 14 :
|   |   |   |   Berths <= 16 : 4 (3.0/2.1)
|   |   |   |   Berths > 16 :
|   |   |   |   |   Berths <= 22 : 5 (2.0/1.0)
|   |   |   |   |   Berths > 22 : 6 (5.0/3.2)
```

Fig. 2. A Decision Tree generated by C5.0

"IF the value of C/C is less than or equal to 75
AND that of Total_Areais less than or equal to 1.14
AND that of Depthis less than 12.85,
THEN the resulting group may be tier 1 with the confidence level of (90%)."

Fig. 3. Rule induced from the decision tree

V. CONCLUSION

In this paper, we aim to show that DEA can be used to evaluate the efficiency of Ports and to suggest the methodology to overcome the limitation of DEA. We present our research framework as shown in [Fig 1], which is a hybrid approach utilizing C5.0 to supplement the limitation of DEA. We generate the rules for classifying new DMUs into each tier and determine the input and output variables that will discriminate the best choice between the tiers by the degree of affecting the efficiencies of the DMUs(discriminant descriptor).

The methodology we proposed can be summarized like below. We apply a DEA to evaluate the efficiency of the DMUs with their multidimensional inputs and outputs. After that, we clustered the DMUs together through the tier analysis, which recursively apply the DEA to the remaining inefficient DMUs, and then generated the DMU classification rules using the C5.0, the decision tree classifier, with the DMU tiers that had identified by the tier analysis. In conventional DEA, it only (1) identifies inefficiencies, (2) identifies comparable efficient units, and (3) locates slack resources. But, we provide more information about discriminant descriptors among input and output variables, which affects the efficiency of DMUs and rules for classifying new DMUs than other decision tree algorithms. We can generate classification rules C5.0 in order to classify any new Port without perturbing previously existing evaluation structures by proposed methodology.

The study has some limitations. Data for input and output variables is focused on particular year and does not reflect ports' efficiency trend in this paper. Secondly, DEA has fundamental problem that does not guarantee ports measured as efficient are actually efficient.

참 고 문 헌

[1] W. W. Cooper, R. G. Thomson, and R. M. Trall, "Introduction: Extensions and new developments in DEA," Annals of Operations Research, Vol.66, pp.3-45, 1996,

[2] M. J. Farrell, "The measurement of productivity efficiency," Journal of the Royal Statistical Society Series A, Vol.120, No.3, pp.253-281, 1957.

[3] B Golany and Y Roll, "An Application Procedure for DEA," OMEGA Int. J. of Mgmt Sci., Vol.17, No.3, pp.237-250, 1989.

[4] D. L. Day, A. Y. Lewin, and H. Li, "Strategic leaders or strategic groups: A longitudinal DEA of the U.S.brewingindustry," European Journal of Operational Research, Vol.80, pp.619-38, 1995.

[5] Sabrina Sestito and Tharam S. Dillon, *Automated Knowledge Acquisition,* Prentice Hall, 1994.

[6] J. E. Beasley, "Comparing university departments," OMEGA Int. J. of ManagementScience, Vol.18, No.2, pp.171-183, 1990.

[7] R. G. Thompson, E. J. Brinkmann, P. S. Dharmapala, and M. D. Gonzalez-Lima, "DEA/AR profit ratio and sensitivity of 100 large U.S.banks," European Journal of Operational Research, Vol.98, pp.213-22, 1997.

[8] A. Charnes, C. T. Clark, W. W. Cooper, and B. Golany, "A development study of Data Envelopment Analysis in measuring the efficiency of maintenance units in the U.S." Air Force, Annals of Operations Research, Vol.2, pp.95-112, 1985.

[9] Qingyuan Zhu, Jie Wu, and Malin Song,

"Efficiency evaluation based on data envelopment analysis in the big data context," Computers & Operations Research, Vol.98, pp.291-300, 2018.
[10] Dariush Khezrimotlagh, Joe Zhu, Wade D. Cook, and Mehdi Toloo, "Data envelopment analysis and big data," European Journal of Operational Research, Vol.274, No.3, pp.1047-1054, 2019.
[11] J. R. Quinlan, "Induction of decision trees," Machine Learn, Vol.1, pp.81-106, 1986.
[12] Y. Li and A. Jain, "Classification of text documents," Comput J., Vol.41, pp.537-46, 1998.
[13] Brijain R Patel and Kushik K Rana, "A Survey on Decision Tree Algorithm For Classification," International Journal of Engineering Development and Research, Vol.2, No.1, pp.1-5, 2014.
[14] A. S. Al-Eraqi, A. Mustafa, A. T. Khader, and C. P. Barros, "Efficiency of Middle Eastern and East African Seaports: Application of DEA Using Window Analysis."European Journal of Scientific Research, Vol.23, No.4, pp.597-612, 2008.
[15] Carvalho, *Performance Evaluation of the Portuguese Seaports-Evaluation in the European Context,* Dissertation, Universidade Tecnica de Lisboa, 2007.
[16] T. Wang and K. Cullinane, "The Efficiency of European Container Terminals and Implications for Supply Chain Management," Maritime Economics & Logistics, Vol.8, pp.82-99, 2006.
[17] K. Cullinane, D. Song, and T. Wang, "The Application of Mathematical Programming Approaches to Estimating Container Port Production Efficiency," Journal of Productivity Analysis, Vol.24, pp.73-92, 2005.
[18] C. P. Barros, "A Benchmark Analysis of Italian Seaports using Data Envelopment Analysis," Maritime Economics & Logistics, Vol.8, pp.347-365, 2006.
[19] H. S. Lee, M. T. Chou, and S. G. Kuo, "Evaluating Port Efficiency In Asia Pacific Region with Recursive Data Envelopment analysis," Journal of the Eastern Asia Society for Transportation Studies, Vol.6, pp.544-559, 2005.
[20] K. Cullinane, P. Ji, and T. Wang, "An Application of DEA Windows Analysis to Container Port Production," Review of Network Economics, Vol.3, No.2, pp.184-206, 2004.
[21] C. P. Barros and A. Manolis, "Efficiency in European Seaports with DEA: Evidence from Greece and Portugal," Maritime Economics & Logistics, Vol.6, pp.122-140, 2004.
[22] E. Thanassoulis, "A Data Envelopment Analysis Approach to Clustering Operating Units for Resource Allocation Purposes," Omega, Int. J. Mgmt Sci, Vol.24, No.4, pp.463-476, 1996.
[23] R. D. Banker, A. Charnes, and W. W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," anagement Science, Vol.30, pp.29-40, 1984.
[24] J. Ross Quinlan, "C4.5: Programs for Machine Learning," Morgan Kaufmann Publishers, 1993.
[25] J. Fossey, J. Degerlund, and L. Jones, "Containerization International Yearbook 2016," Informal UK. 2017.

저 자 소 개

홍 한 국(Han-Kuk Hong) 정회원

- 1988년 : 고려대 통계학과(경제학사)
- 1990년 : KAIST 산업공학과(공학석사)
- 2000년 : KAIST 경영공학과(공학박사)
- 1990년 ~ 1996년 : 삼성경제연구소 선임연구원
- 1997년 ~ 1998년 : 삼성화재 과장
- 2000년 ~ 현재 : 동의대학교 경영정보학과 교수

〈관심분야〉 : DEA, 데이타마이닝, CRM, SCM

임 병 학(Byung-hak Leem) 정회원

- 1987년 : 고려대 산업공학(공학사)
- 1996년 : KAIST 경영공학과(공학석사)
- 2002년 : University of Texas at Arlington(공학박사)
- 1989년 3월 ~ 1997년 : 삼성전자 과장
- 2002년 ~ 현재 : 부산외국어대학교 경영학부 교수

〈관심분야〉 : 생산관리, 공급망관리, 성과관리

김 삼 문(Sam-Moon Kim) 정회원

- 1990년 : 동명대 경영정보학과(학사)
- 2011년 : 동의대 경영정보학과(석사)
- 2014년 : 동의대 경영정보학과(박사)
- 1996년 ~ 2016년 : (주) SK 브로드밴드
- 2016년 ~ 현재 : 동의대학교 응용소프트웨어공학과 교수

〈관심분야〉 : DEA, 데이타마이닝, CRM, SCM