

Clustering of Decision Making Units using DEA

Kyeongtaek Kim[†]

Department of Industrial and Management Engineering, Hannam University

DEA를 이용한 의사결정단위의 클러스터링

김 경 택[†]

한남대학교 산업경영공학과

The conventional clustering approaches are mostly based on minimizing total dissimilarity of input and output. However, the clustering approach may not be helpful in some cases of clustering decision making units (DMUs) with production feature converting multiple inputs into multiple outputs because it does not care converting functions. Data envelopment analysis (DEA) has been widely applied for efficiency estimation of such DMUs since it has non-parametric characteristics. We propose a new clustering method to identify groups of DMUs that are similar in terms of their input-output profiles. A real world example is given to explain the use and effectiveness of the proposed method. And we calculate similarity value between its result and the result of a conventional clustering method applied to the example. After the efficiency value was added to input of K-means algorithm, we calculate new similarity value and compare it with the previous one.

Keywords : DEA, Clustering, K-Means Clustering

1. 서 론

Data Envelopment Analysis(DEA)는 유사한 목적을 가지고 운영되는 동일한 다수산출요소/다수투입요소 자료 구조를 갖는 의사결정단위(DMU)들의 상대적 효율성을 평가하는 방법이다. 모든 투입요소를 비례적으로 증가시킬 때 산출도 일정한 비율에 따라 증가하는, 규모에 대한 수익불변(Constant Returns to Scale)을 가정하여, Charnes et al.[6]이 DEA 모델을 제시함으로써 널리 알려지게 되었다.

클러스터링은 동일한 군집(cluster) 내에 있는 객체들의 유사성은 극대화하면서, 서로 다른 군집에 속한 객체들과의 유사성(similarity)은 최소화되도록, 객체들의 군집을 결정하는 절차이다[12]. 일반적으로, 클러스터링 방법으로는

계층적(hierarchical) 클러스터링, 센트로이드 기반(centroid-based) 클러스터링, 밀집도 기반(density-based) 클러스터링, 모델기반(model-based) 클러스터링 등이 있다. 계층적 클러스터링은 덴드로그램(dendrogram)이라 불리는 군집트리(cluster tree)를 생성함으로써 객체들을 군집화한다. 트리를 구성하는 방법에 따라 병합식(agglomerative)과 분할식(divisive)으로 나뉜다. 센트로이드 기반 클러스터링은 K-means 알고리즘이 가장 널리 알려진 예이며, 어떤 알고리즘을 사용하던지 거리(distance)를 정의하는 함수에 따라 클러스터링의 결과가 크게 변한다. 밀집도기반 클러스터링에서, 군집은 밀집도가 나머지 지역보다 높은 지역으로 정의된다. 모델기반 클러스터링은 데이터가 만들어졌을 것으로 추정되는 모델과, 데이터 사이의 적합성(fitness)이 최적화되도록 군집을 구성한다.

DEA와 클러스터링을 결합한 여러 연구들이 있다. Shin and Shon[19]은 모든 고객들을 DMU로 정의하기에는 너무 많아서, 고객들을 클러스터링 한 후, 각 군집을 하나의 DMU로 정의하여 분석하였다. Sharma and Yu[18]는

DMU를 입력요소에 따라 클러스터링하여 추후 벤치마킹 과정에 도움이 되도록 하였다. Schreyogg and Reitzenstein [20]은 DEA를 적용한 후, 각각의 클러스터링 목적에 맞게 일부 요소만으로 클러스터링을 수행하여, 각 목적별로 2개의 전략그룹(strategic group)를 선별하여, 그들의 특징을 비교하였다. Saati et al.[17]은 퍼지 환경하에서 DMU를 클러스터링 하는 Fuzzy DEA 방법을 제안하였다. Marroquin et al.[14]은 DMU의 수가 너무 많아서 가장 널리 사용되는 Excel Solver 같은 소프트웨어로 DEA 최적해를 구하는데 많은 시간이 소요될 때, DMU를 먼저 클러스터링한 다음, 각 군집에서 효율성이 1인 DMU들을 구하고, 이들 DMU들만 모은 다음, 처음부터 다시 절차를 반복하여 시간을 단축하는 방법을 제안하였다.

다수의 투입요소를 사용하며 또한 다수의 산출물을 생산하는 DMU의 생산특성을 비교하기 위해서, K-means 알고리즘 같은 전통적인 클러스터링 방법 대신에, Po et al.[16]은 DEA 모델을 이용한 클러스터링 알고리즘을 제시하였다. 이 알고리즘은 다음과 같은 장점을 갖는다 : (1) 투입요소와 산출요소 사이의 함수적 관계의 가정이 필요하지 않다. (2) 군집의 개수가 DEA 적용 결과에 의해 객관적으로 정해진다. Bi et al.[4]은 Po et al.[16]의 DEA 기반 클러스터링 알고리즘을 이용하여, 중국 각 지역을 DMU로 하여 환경에 대한 효율성을 측정하고, 클러스터링을 하였다. 한편, Amin et al.[1]는 최적해가 다수 있을 때 DEA 기반 클러스터링 알고리즘이 어떻게 서로 다른 군집들을 생성하는지를 보여주었다.

그러나, Po의 DEA 기반 클러스터링 알고리즘은 계산이 복잡한 단점이 있다. 본 연구에서는 이러한 단점을 수정하여, 새로운 DEA 기반 클러스터링 알고리즘을 제시하고, 실제 데이터에 적용하였다. 미국 플로리다주에 있는 각 병원을 DMU로 설정하여, 2012년 통계자료에서 투입변수와 산출변수를 결정한 다음 DEA 모형을 이용하여 효율성을 측정하고, 본 연구에서 제안한 알고리즘을 적용하여 클러스터링 한다. 아울러, K-means 클러스터링 방법을 통한 결과와 비교한다.

2. DEA를 이용한 클러스터링

2.1 DEA

효율성(efficiency)은 투입요소의 사용량에 대한 산출물 생산량의 비율을 의미한다. 효율성은 생산조직이 단일 투입요소를 사용하여 단일 산출물을 생산할 경우에는 계산이 매우 간단하다. 그러나 대부분의 생산조직은 다수의 투입요소를 사용하며 또한 다수의 산출물을 생산한다. 이

러한 다수투입 다수산출의 경우 효율성을 계산하려면 다수의 투입요소에 가중치를 적용하여 총괄한 총괄투입과 다수의 산출물에 가중치를 적용하여 총괄한 총괄산출을 계산하는 과정이 필요하다. DEA는 다수산출/다수투입을 하는 DMU에 대하여, DMU가 속한 그룹 내에 있는 베스트 프랙티스(best practice)와 비교한 상대적 효율성을 계산한다.

Charnes et al.[6]은, 모든 투입요소를 비례적으로 증가시킬 때 산출도 일정한 비율에 따라 증가한다고 가정하여, DMU의 효율성은, 나머지 DMU의 총괄산출/총괄투입 비율이 1보다 작거나 같아야 한다는 조건하에서, 가중된 투입에 대한 가중된 산출의 비율의 최대치로 정의하였다. 이 모형은 식 (1)과 같다.

각 DMU k 에 대하여,

$$\begin{aligned} \text{Max} \quad & \frac{\sum_{r=1}^s u_{rk}y_{rk}}{\sum_{i=1}^m v_{ik}x_{ik}} \\ \text{Subject to} \quad & \frac{\sum_{r=1}^s u_{rk}y_{rj}}{\sum_{i=1}^m v_{ik}x_{ij}} \leq 1, \quad j=1, \dots, n \quad (1) \\ & u_{rk} \geq 0, \quad r=1, \dots, s \\ & v_{ik} \geq 0, \quad i=1, \dots, m \end{aligned}$$

여기에서 y_{rj} , x_{ij} (모두 양수)는 j 번째 DMU의 이미 알고 있는 산출변수 및 투입변수이며, u_{rk} , v_{ik} 는 DMU k 의 가변의 가중치이다.

이 모형은 비선형문제인데, Charnes et al.[8]은 이 모형을 아래 식 (2)와 같은 선형계획법(Linear Programming) 문제로 변환하였다.

각 DMU k 에 대하여,

$$\begin{aligned} \text{Max} \quad & Eff_k = \sum_{r=1}^s u_{rk}y_{rk} \\ \text{Subject to} \quad & \sum_{r=1}^s u_{rk}y_{rj} - \sum_{i=1}^m v_{ik}x_{ij} \leq 0, \quad \forall j \\ & \sum_{i=1}^m v_{ik}x_{ik} = 1 \quad (2) \\ & u_{rk} \geq 0, \quad \forall r \\ & v_{ik} \geq 0, \quad \forall i \end{aligned}$$

위 모형은 CCR 모형이라 부른다. 한편, Banker et al.[2]은 규모수익불변(CRS) 가정 대신에 규모수익가변(Variable Returns to Scale)을 가정하여 아래와 같이 효율성을 측정하는 식 (3)을 제안하였다.

각 DMU k 에 대하여,

$$\begin{aligned}
& \text{Max} && \frac{\sum_{r=1}^s u_{rk} y_{rk} - u_{0k}}{\sum_{i=1}^m v_{ik} x_{ik}} \\
& \text{Subject to} && \frac{\sum_{r=1}^s u_{rk} y_{rj}}{\sum_{i=1}^m v_{ik} x_{ij} - u_{0k}} \leq 1, \quad j=1, \dots, n \quad (3) \\
& && u_{rk} \geq 0, \quad r=1, \dots, s \\
& && v_{ik} \geq 0, \quad i=1, \dots, m \\
& && u_{0k} \quad \text{unrestricted sign}
\end{aligned}$$

위 식에서 u_{0k} 는 효율적 DMU의 규모수익>Returns to Scale)효과를 평가하는 척도로 해석된다. $u_{0k} > 0$ 은 규모수익체증을 나타내고, $u_{0k} < 0$ 은 규모수익체감을 나타낸다.

그 외에 Additive 모형[7], Slack-based measure 모형[21], Russell measure 모형[9] 등이 있다. 본 연구에서는 CCR 모형을 사용하기로 한다.

2.2 DEA 기반 클러스터링 알고리즘

DMU가 투입요소 및 산출요소로 구분되는 특징을 가지고 있다면, DMU를 클러스터링 할 때 이 특성을 반영하여야 한다. 그러나 일반적인 클러스터링 알고리즘은 투입변수와 산출변수 사이의 관계를 전혀 고려하지 않고, 모든 변수들을 동일한 변수로 취급하여 클러스터링을 한다. 이러한 단점을 극복하고자, PO et al.[16]은 생산함수(production function)를 기반으로 클러스터링을 하는 것이 더 타당하다고 주장하고, 이러한 클러스터링을 가능하게 하는 알고리즘을 제시하였다.

Po의 DEA 기반 클러스터링 알고리즘[16]은 3단계로 되어있다. 첫 번째 단계에서 u_{rk}^* 및 v_{ik}^* 가 모두 nonzero인 DMU k 에 대하여 $\sum_{r=1}^s u_{rk}^* y_{rj} - \sum_{i=1}^m v_{ik}^* x_{ij} = 0$ 이면, 새로운 군집 PF(h)로 분류한다. 두 번째 단계에서 u_{rk}^* 및 v_{ik}^* 가 모두 nonzero이지는 않은 DMU k 는 기존의 군집 C에 배정한다. 두 번째 단계를 수행하기 위해서 각 DMU 당 p 번만큼 계산하여 $t(1), \dots, t(p)$ 를 구한다음 그 중 최대치를 구하여 클러스터를 결정한다.

본 연구에서는 Po의 DEA 기반 클러스터링 알고리즘과 동일한 클러스터링 결과를 가져오면서 보다 간편히 계산 할 수 있는 알고리즘을 제시한다.

1. $FR = \{\}; BFR = \{\}$
 각 DMU k 에 대하여
 - 1-1. DEA CCR 모델 식 (2)를 실행하여 효율성(Eff_k)을 계산한다.
 - 1-2. $\sum_{r=1}^s u_{rk} y_{rj} - \sum_{i=1}^m v_{ik} x_{ij} = 0$ 이 되는 모든 j 의 집합을 F_k 라 한다.
 - 1-3. 만일 $F_k \in FR$ 이 아니면, $FR = FR \cup \{F_k\}$

2. $FR = \{fr_1, fr_2, \dots, fr_{p+q}\}$ 이라 하자.
 각 원소 fr_t 에 대하여
 $|fr_t| < m$ 이면, $FR = FR - \{fr_t\}$, $BFR = BFR \cup \{fr_t\}$
3. 각 원소 fr_t 에 대하여 $C(fr_t) = t$ 로 놓는다.
4. $BFR = \{bfr_1, bfr_2, \dots, bfr_q\}$ 이라 하자.
 각 원소 bfr_h 에 대하여
 $bfr_h \subset fr_t$ 이면, $C(bfr_h) = t$ 로 놓는다.
5. 각 DMU k 에 대하여 클러스터 $C(F_k)$ 를 배정한다.

Po 알고리즘의 1단계는 선형계획 문제를 풀어서 효율성을 계산하고 클러스터를 결정하는 단계로 계산횟수는 제시된 알고리즘의 1단계보다 계산 횟수가 많다. 제시된 알고리즘의 2단계는 $p+3q$, 3단계는 p , 4단계는 최대 pq 이다. 4단계에서 각 DMU k 에 대하여 $C(F_k)$ 를 계산하여야 하므로, 최대 p 만큼의 계산(비교)이 필요하며 따라서 4단계에서 최대 pm 만큼의 계산을 하게 된다. 제시된 알고리즘의 2~5단계의 계산 횟수를 합하면, $2p+3q+pq+pm$ 이고, 일반적으로 $n \gg p$ 이고, $n \gg q$ 이므로, $2p+3q+pq+pm < 2pn$ 이다. Po 알고리즘의 2단계에서는, 먼저 각 DMU 당 $t(1), \dots, t(p)$ 를 구해야 하는데, $t(h)$ 마다 $2(s+m)$ 만큼의 계산이 필요하므로, 각 DMU마다 $2(s+m)p$ 만큼의 계산이 필요하다. 따라서 전체적으로, $2(s+m)p(n-p)$ 만큼의 계산이 필요하다. $s+m \geq 2$ 이고 일반적으로 $n \gg p$ 이므로, Po의 알고리즘 2단계는 $2(s+m)p(n-p) \geq 4p(n-p) \cong 4pn$ 의 계산이 필요하게 되어, 본 논문에서 제시한 알고리즘이 더 효율적이라 할 수 있다.

3. 적용 사례

3.1 데이터

병원은 DEA의 관점에서 쉽게 DMU로 정의될 수 있기 때문에, 보건 의료 분야는 DEA 분석에 아주 적합한 분야로 여겨져 왔다[3]. 따라서, 보건 의료 분야에서 DEA를 이용한 많은 연구들이 수행되었다[13].

본 연구에서는, Florida Agency for Health Care Administration에서 공개한 2012 Florida Hospital Financial Data [11] 및 AHCA Inpatient Data[10]에 있는 데이터에 대하여, 제시된 알고리즘을 적용하여 미국 플로리다주에 있는 각 병원을 DMU로 하여 효율성을 측정하고, 클러스터링을 한다. 2012 Florida Hospital Financial Data[11]에서는 병원을 16개 그룹으로 나누었는데, 본 연구에서는

Group 5(Medium Urban Hospital Group)로 분류된 33개 병원 전체를 적용 대상으로 하였다. 본 사례는 제시된 알고리즘을 쉽게 설명하고, 결과를 이차원 좌표에 표시하기 위하여, 투입변수를 1개, 산출변수를 2개로 제한한다.

<Table 1> Data from 33 Florida Hospitals

DMUs	CMI Adjusted Discharge	Staffed Bed	Employees FTE
DMU1	5831.1	112	583.4
DMU2	45708.8	515	2957.5
DMU3	45691.1	466	2165.4
DMU4	24878.5	1375	1355.7
DMU5	11119.2	210	1928.7
DMU6	20040.1	309	2159.9
DMU7	14843.5	139	778.6
DMU8	6841.1	168	471.9
DMU9	16498.8	155	997.2
DMU10	39560.9	407	2269.5
DMU11	1112.3	396	1572.8
DMU12	8159.5	120	493.0
DMU13	17982.6	312	994.3
DMU14	31477.6	559	2505.9
DMU15	16782.4	254	731.0
DMU16	27431.6	316	1505.8
DMU17	6965.5	307	472.4
DMU18	14497.7	208	694.6
DMU19	12020.7	155	466.2
DMU20	35222.6	475	2219.1
DMU21	16918.5	199	723.9
DMU22	17087.2	270	1041.1
DMU23	20926.0	323	1561.2
DMU24	17381.2	224	700.6
DMU25	17600.3	238	724.7
DMU26	14198.6	288	652.2
DMU27	4530.4	162	265.9
DMU28	16783.5	163	1254.5
DMU29	23895.5	290	900.1
DMU30	20077.5	234	846.7
DMU31	9938.4	199	962.3
DMU32	8137.2	119	424.2
DMU33	20269.1	289	881.7

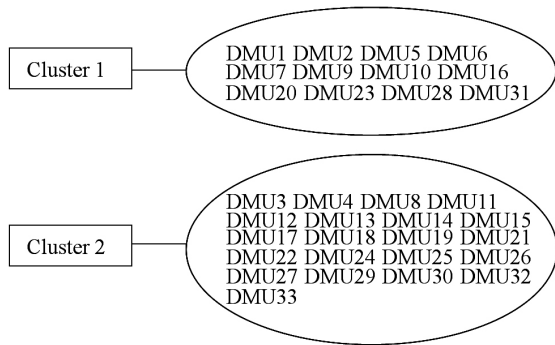
<Table 2> Summary Statistics of Inputs and Outputs

Variable	Mean	S.D.	Min	Max
CMI Adjusted Discharge	18497.2	10885.6	1112.3	45708.8
Staffed Bed	301.7	222.0	112.0	1375.0
Employees FTE	1159.5	691.6	265.9	2957.5

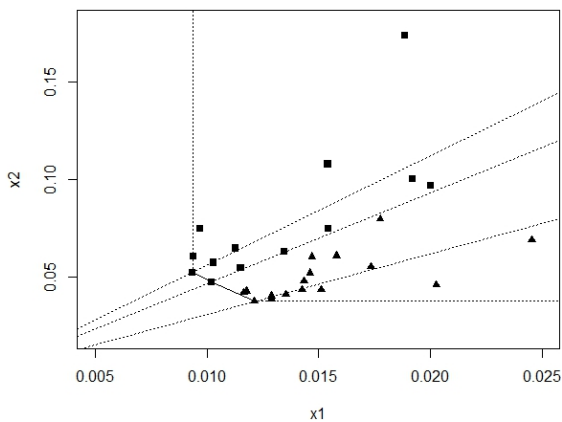
<Table 3> DEA Optimal Solution

k	Eff_k	u_{1k}^*	v_{1k}^*	v_{2k}^*	F_k
1	0.5045	0.000087	0.004803	0.000792	3, 7
2	0.8311	0.000018	0.001942	0	7
3	1	0.000022	0.001114	0.000222	3, 29
4	0.6912	0.000028	0	0.000738	29
5	0.4958	0.000045	0.004762	0	7
6	0.6073	0.000030	0.003236	0	7
7	1	0.000067	0.00374	0.000617	7
8	0.5461	0.000080	0	0.002119	29
9	0.9968	0.000060	0.006452	0	7
10	0.9122	0.000023	0.00128	0.000211	3, 7
11	0.0308	0.000028	0.00141	0.000281	3, 29
12	0.7344	0.000090	0.004583	0.000913	3, 29
13	0.6924	0.000039	0.00196	0.000391	3, 29
14	0.5842	0.000019	0.000945	0.000188	3, 29
15	0.8648	0.000052	0	0.001368	29
16	0.8757	0.000032	0.001772	0.000292	3, 7
17	0.5554	0.000080	0	0.002117	29
18	0.8221	0.000057	0.002887	0.000575	3, 29
19	0.9712	0.000081	0	0.002145	29
20	0.7545	0.000021	0.001189	0.000196	3, 7
21	0.9682	0.000057	0.002914	0.00058	3, 29
22	0.703	0.000041	0.002095	0.000417	3, 29
23	0.6494	0.000031	0.001723	0.000284	3, 7
24	0.9389	0.000054	0.00275	0.000548	3, 29
25	0.9148	0.000052	0	0.00138	29
26	0.82	0.000058	0	0.001533	29
27	0.6418	0.000142	0	0.003761	29
28	0.9642	0.000058	0.006135	0	7
29	1	0.000042	0	0.001111	29
30	0.9793	0.000049	0.002483	0.000495	3, 29
31	0.5005	0.000050	0.002796	0.000461	3, 7
32	0.7853	0.000097	0.004914	0.000979	3, 29
33	0.8659	0.000043	0	0.001134	29

침상수 및 종업원 FTE(Full Time Equivalent)를 투입변수로 사용하고, CMI 보정 퇴원환자수(CMI Adjusted Discharge)를 산출변수로 사용하였다. CMI(Case Mix Index, 환자구성지표)는 병원의 환자중증도를 보정하기 위한 지표로, 일반적으로 환자중증도가 높을수록 더 많은 의료서비스를 제공받게 된다. 따라서, 병원의 효율을 측정할 때, CMI로 보정을 해 주는 것이 보다 정확한 측정의 전제조건이다[5, 15]. 본 사례에서는 퇴원 환자수에 CMI를 곱하여 CMI 보정 퇴원환자수를 구하였다. 종업원 FTE는 파트타임 종업원을 포함하여 임금을 받는 종업원들의 연간 총 근무시간을 2080으로 나눈 값으로, 정규직종업원 몇 명에 해당하는지를 나타내는 값이다. <Table 1>은 본 연구에서 사용된 데이터를 보여주며, <Table 2>는 이들의 통계치를 보여준다.



<Figure 1> Clustering Results of Suggested Algorithm

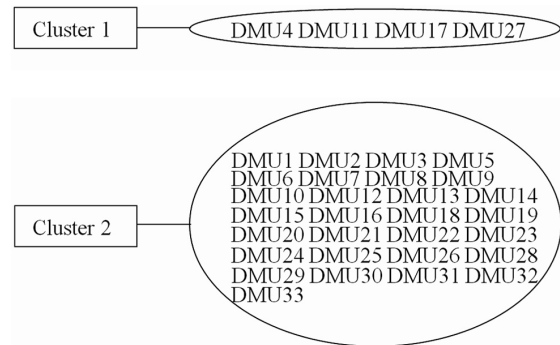


<Figure 2> An Illustration of Suggested DEA-based Clustering Algorithm

3.2 결과

<Table 3>은 제시된 알고리즘의 1단계를 실행한 결과이다. 2단계에서, $FR = \{\{7\}, \{3, 7\}, \{3, 29\}, \{29\}\}$ 임을 알 수 있다. 그러면, 4단계에서, $C(\{7\}) = 1$, $C(\{3, 7\}) = 1$, $C(\{3, 29\}) = 2$, $C(\{29\}) = 2$ 가 된다. 알고리즘을 모두 수행하면 <Figure 1>과 같은 군집을 얻는다. 이는 Po의 알고리즘을 적용한 결과와 일치한다. <Figure 2>는 제시된 알고리즘을 적용한 결과를 이차원 좌표에 나타내었다 (단, DMU4, DMU11, DMU17, DMU27은 나타내는 점들의 해상도를 위하여 <Figure 2>에 나타내지 않았다). 여기에서 x_1 , x_2 는 산출요소 1단위를 산출하기 위하여 투입된 입력요소의 크기를 나타낸다.

동일한 투입변수 및 산출 변수에 k-means 알고리즘을 적용하여 DMU를 클러스터링하여, 제시된 알고리즘의 클러스터링 결과가 유사한지 살펴보자. 이를 위하여 군집수를, 제시된 알고리즘의 클러스터링 결과와 동일하게 2로 정하고, 각 변수를 정규화한 후, k-means 알고리즘을 적용하여, 클러스터링을 한 결과, <Figure 3>과 같은 결과를 얻었다.



<Figure 3> Clustering Results of K-means Algorithm

두 클러스터링의 결과가 얼마나 유사한지를 나타내는 척도중의 하나로 유사도(similarity measure)가 있다. Torres et al.[22]은 두 클러스터링 결과, $C = \{C_1, C_2, \dots, C_m\}$, $D = \{D_1, D_2, \dots, D_n\}$ 에 대하여, 유사도 $Sim(C, D)$ 를 다음과 같이 정의하였다.

$$Sim(C, D) = \frac{\sum_{i=1}^m \sum_{j=1}^n S_{ij}}{\max(m, n)}$$

여기에서 S_{ij} 는 C_i 와 D_j 에 대한 자카드 유사도 계수(Jaccard's Similarity Coefficient)이다. 이 정의를 이용하여 본 예제에 대한 DEA 기반 클러스터링의 결과와 k-means 클러스터링의 결과와의 유사도를 측정하면 0.5597이다. 이 값으로부터 두 클러스터링의 결과가 유사하지 않음을 알 수 있다.

k-means 클러스터링 알고리즘은 투입변수와 산출변수 사이의 관계를 전혀 고려하지 않고, 모든 변수들을 동일한 요소로 한다. 따라서, 거리가 가까운 DMU들은 동일한 군집으로 분류된다. 이 결과는 동일한 군집에 속한 DMU들이 비슷한 생산특성도 가지고 있다는 잘못된 메시지를 전달할 수 있다.

4. 결론

본 연구에서는 DEA 모델을 이용한 새로운 클러스터링 알고리즘을 제시하였다. 제시된 모델은 Po의 알고리즘과 동일한 결과를 가져오나, 계산이 더 간편함을 보였다. 아울러, 플로리다 병원데이터에 적용한 DEA 기반 클러스터링의 결과와 k-means 알고리즘을 적용한 결과는 유사도가 작아, DEA 기반 클러스터링과 k-means 클러스터링은 전혀 다른 결과를 가져옴을 보였다.

본 논문의 한계로는 최근에는 computing power가 높아짐에 따라 대용량의 자료가 아닌 한 알고리즘의 효율성의 의미가 점차 중요하지 않게 되어가고 있다는 점이다. 아울러, 사례연구에서 사용된 33개 DMU의 클러스터링은 제시된 알고리즘의 간편성을 설명하기에는 부족한 수준

이며, 보다 대용량의 데이터에 대한 연구가 필요하다.

본 논문은 동일한 투입변수, 산출변수들을 사용하고도 K-means 알고리즘의 결과와 완전히 다른 클러스터링 결과를 산출함을 보여주고 있다. 이에 따라, 추후 연구 과제로 DMU의 효율을 비교할 때, 전체를 하나의 클러스터로 보고 비교하는 것이 타당한지, 클러스터링을 한 후 클러스터 내에서 각 DMU의 효율성을 비교 하는 것이 나은지에 관한 연구가 있을 수 있다. 또한 각 클러스터내에서 각 DMU의 효율성을 비교할 때, 어떤 클러스터링 결과가 더 타당한지에 관한 연구 등으로 확장 될 수 있다. 이러한 추후연구의 결과는 학계 및 산업계에서 DMU의 효율성 비교에 큰 영향을 끼칠 수 있다.

References

- [1] Amin, G., Emrouznejad, A., and Razaeei, S., Some Clarifications on the DEA Clustering. *European Journal of Operations Research*, 2011, Vol. 215, p 498-501.
- [2] Banker, R.D., Charnes, A., and Cooper, W.W., Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 1984, Vol. 40, p 1078-1092.
- [3] Ben-Arieh, D. and Gullipalli, D., "Data Envelopment Analysis of Clinics with Sparse Data : Fuzzy Clustering Approach. *Computers and Industrial Engineering*, 2012, Vol. 63, p 13-21.
- [4] Bi, G.-B., Song, W., and Wu, J., A Clustering Method for Evaluating the Environmental Performance based on slack-based Measure. *Computers and Industrial Engineering*, 2014, Vol. 72, p 169-177.
- [5] Bjorkgren, M.A., Fries, B.E., Hakkinen, U., and Brommels, M., Case-Mix Adjustment and Efficiency Measurement. *Scandinavian Journal of Public Health*, 2004, Vol. 32, p 464-471.
- [6] Charnes, A., Cooper, W.W., and Rhodes, E.L., Measuring the Efficiency of Decision Making Units. *European Journal of Operations Research*, 1978, Vol. 2, p 429-444.
- [7] Charnes, A., Cooper, W.W., Golany, B., Seiford, L.M., and Stutz, J., Foundation of Decision Envelopment Analysis and Pareto-Koopmans Empirical Production Functions. *Journal of Econometrics*, 1985, Vol. 30, p 91-107.
- [8] Charnes, A., Cooper, W.W., and Rhodes, E.L., Evaluating Program and Managerial Efficiency : An Application of Decision Envelopment Analysis to Program Follow Through. *Management Science*, 1981, Vol. 27, p 668-697.
- [9] Fare, R.S. and Lovell, C.A.K., Measuring the Technical Efficiency of Production. *Journal of Economic Theory*, 1978, Vol. 19, p 150-162.
- [10] Florida Agency for Health Care Administration, AHCA Inpatient Data, 2012.
- [11] Florida Agency for Health Care Administration, Florida Hospital Financial Data, 2012.
- [12] Jain, A.K., Murty, M.N., and Flynn, P.J., Data Clustering : A Review. *ACM Computing Surveys*, 1999, Vol. 31, No. 3, p 264-323.
- [13] Liu, J.S., Lu, L.Y.Y., Lu, W.-M., and Lin, B.J.Y., A Survey of DEA Applications. *Omega*, 2013, Vol. 41, p 893-902.
- [14] Marroquin, M.G.V., Pena, M.L., Castro, C.E., Castro, J.M., and Cabrera-Rios, M., Use of Data Envelopment Analysis and Clustering in Multiple Criteria Optimization. *Intelligent Data Analysis*, 2008, Vol. 12, p 89-101.
- [15] O'Neill, L., Rauner, M., Heidenberger, K., and Kraus, M., A Cross-National Comparison and Taxonomy of DEA-based Hospital Efficiency Studies. *Scio-Economic Planning Sciences*, 2008, Vol. 42, p 158-189.
- [16] Po, R., Guh, Y., and Yang, M., A new Clustering Approach using Data Envelopment. *European Journal of Operations Research*, 2009, Vol. 199, p 276-284.
- [17] Saati, S., Hatami-Marbini, A, Tavana, M., and Agrell, P.J., A Fuzzy Data Envelopment Analysis for Clustering Operating Units with Imprecise Data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based System*, 2013, Vol. 21, No. 1, p 29-54.
- [18] Sharma, M.J. and Yu, S.J., Performance based Stratification and Clustering for Benchmarking of Container Terminals. *Expert Systems with Applications*, 2009, Vol. 36, p 5016-5022.
- [19] Shin, H.W. and Sohn, S.Y., Multi-attribute scoring method for Mobile Telecommunication Subscribers. *Expert Systems with Applications*, 2004, Vol. 26, p 363-368.
- [20] Schreyogg, J. and Reitzenstein, C., Strategic Group and Performance Differences among Academic and Medical Centers. *Health Care Management Review*, 2008, Vol. 33, No. 3, p 225-233.
- [21] Tone, K., A Slack-based Measure of Efficiency in Data Envelopment Analysis. *European Journal of Operations Research*, 2001, Vol. 130, p 498-509.
- [22] Torres, G.J., Basnet, R.B., Sung, A.H., Mukkamala, S., and Ribeiro, B.M., A Similarity Measure for Clustering and Its Applications. *International Journal of Electrical, Computer and System Engineer*, 2009, Vol. 3, No. 3, p 164-170.