

Genetic Algorithm을 적용한 Fuzzy DEA에 관한 연구 Fuzzy DEA via Genetic Algorithms

최 흥, 손소영

연세대학교 컴퓨터과학 · 산업시스템공학과

Abstract

DEA has been effectively applied to various areas which need the evaluation of relative efficiency. We propose a DEA model based on fuzzy LP combined with Genetic Algorithm in order to consider uncertain synergy effects due to M&A of existing organization. We apply the suggested approach to forecasting the efficiency of merged academic departments in a university in Korea. We expect that our approach can be utilized to effectively realign existing departments.

1. 서론

본 연구에서는 DEA(Data Envelopment Analysis)를 이용해 상대적 효율성을 측정함에 있어, 미래의 성능자료 부재 및 조직통합에 의한 synergy 효과의 불확실성을 고려한 분석을 위해 fuzzy LP를 적용한 DEA 모형을 제시하고, 이 과정에서 생기는 다목적 계획문제에 대한 파레토 최적해를 구하기 위해 유전자 알고리즘(Genetic Algorithms)을 적용하는 모형을 제안하고자 한다. 본 논문의 구성은 다음과 같다. 제 2장에서는 입 · 출력변수가 불확실한 자료로 구성되어 있을 때 효율성을 분석하기 위해, fuzzy LP 모형을 적용한 DEA 모형을 소개하고, 이러한 fuzzy DEA 모형의 최적해를 구하기 위해 유전자 알고리즘을 적용하는 과정에 대해 기술하기로 한다. 제 3장에서는 결론 및 향후연구방향에 관해 언급하도록 한다.

2. fuzzy DEA 모형

DEA에 fuzzy LP를 적용한 모형을 구하기 위해, 우선 CCR 모형에서 각 입 · 출력 변수가 퍼지수로 구성되어 있는 경우를 고려해 보면, 식 (1)과 같음을 알 수 있다. 여기서 첨자“ ' ”는 해당변수들이 퍼지수(fuzzy number)로 구성되어 있음을 의미하며,

$F(R)$ 가 퍼지수들의 집합일 때 $x'_{ij}, y'_{rj} \in F(R)$ 이 성립함을 알 수 있다($i=1, \dots, m; r=1, \dots, s; j=1, \dots, j$).

$$\begin{aligned} \text{Min} \quad & \theta - \epsilon [\sum_{i=1}^m s_i^- - \sum_{r=1}^s s_r^+] \\ \text{s.t.} \quad & \sum_{j=1}^n \lambda_j x'_{ij} + s_i^- = \theta x_{i0}' \\ & \sum_{j=1}^n \lambda_j y'_{rj} - s_r^+ = y_{r0}' \\ & s_i^-, s_r^+, \lambda_j \geq 0 \end{aligned} \quad (1)$$

식 (1)에 일반적인 fuzzy LP의 해법을 적용하기 위해서 각 제약조건을 부등식으로 변환하는 작업이 필요하며, 그 결과 식 (2)과 같은 전형적인 fuzzy LP모형으로 전개할 수 있다.

$$\begin{aligned} \text{Min} \quad & \theta - \epsilon [\sum_{i=1}^m s_i^- - \sum_{r=1}^s s_r^+] \\ \text{s.t.} \quad & \sum_{j=1}^n \lambda_j x'_{ij} + s_i^- - \theta x_{i0}' \geq 0 \\ & - \sum_{j=1}^n \lambda_j x'_{ij} - s_i^- + \theta x_{i0}' \geq 0 \\ & \sum_{j=1}^n \lambda_j y'_{rj} - s_r^+ \geq y_{r0}' \\ & - \sum_{j=1}^n \lambda_j y'_{rj} + s_r^+ \geq -y_{r0}' \\ & s_i^-, s_r^+, \lambda_j \geq 0 \end{aligned} \quad (2)$$

이 때 식 (2)의 각 입력변수 및 출력변수에 관하여 $x_{ij}' = L(x_{ij}, a_{ij})$, $y_{nj}' = L(y_{nj}, b_{nj})$ 과 같은 삼각구성함수를 가지는 퍼지수로 정의할 수 있다. 한편 식 (2)과 같은 퍼지선형계획법에서, 제약조건의 부등식은 식 (3)과 같은 두 개의 퍼지수의 크기를 비교하는데 있어 사용 가능한, 퍼지수들의 믿음의 정도(creditable degree)를 나타내는 퍼지수 D' 에 의해 표현될 수 있다.

$$\mu_{D'}(A_1', A_2') = \sup_{x \leq y} [\mu_{A_1'}(x) \wedge \mu_{A_2'}(y)] \quad (3)$$

where $A_1', A_2' \in F(R)$

이 때 $\mu_{D'}$ 는 퍼지수 D' 에 의해 정의되는데, $A_2' = L(a_2, b_2)$ 가 $A_1' = L(a_1, b_1)$ 보다 클 믿음의 정도를 의미하며, 식 (4)와 같은 구성함수로 나타낼 수 있다.

$$\mu_{D'}(A_1', A_2') = \begin{cases} 1, & \text{if } a_1 < a_2 - b_1 - b_2 \\ 1 - \frac{a_1 - (a_2 - b_1 - b_2)}{b_1 + b_2}, & \text{if } a_2 - b_1 - b_2 \leq a_1 < a_2 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

여기서 식 (4)의 의미를 구체적으로 살펴보면 다음과 같다.

- ① Max $A_1' < \text{Min } A_2'$ ($a_1 + b_1 < a_2 - b_2$)일 때, $\mu_{D'}(A_1', A_2') = 1$ 이다.
- ② Max $A_1' \geq \text{Min } A_2'$ ($a_1 + b_1 \geq a_2 - b_2$)이고 $a_1 < a_2$ 일 때, $0 < \mu_{D'}(A_1', A_2') < 1$ 이다.
- ③ $a_1 \geq a_2$ 일 때, $\mu_{D'}(A_1', A_2') = 0$ 이다.

이와 같은 원리에 의해 식 (2)의 $2(m+s)$ 개의 제약조건에 관하여, 각각의 투입부등식 및 산출부등식과 관련된 제약조건은 식 (5)~(8)과 같은 구성함수를 가지는, 개별 제약조건에 대한 믿음의 정도를 나타내는 퍼지수 $D_i', D_{-i}', D_r', D_{-r}'$ 로 표현할 수 있다.

$$\mu_{D_i'} \left(0, \sum_{j=1}^m \lambda_j x_{ij}' - \theta x_{io}' + s_i^- \right) = \begin{cases} 1, & \text{if } \sum_{j=1}^m (x_{ij} - a_{ij}) \lambda_j - \theta(x_{io} - a_{io}) > 0 \\ 1 + \frac{\sum_{j=1}^m (x_{ij} - a_{ij}) \lambda_j - \theta(x_{io} - a_{io})}{\sum_{j=1}^m \lambda_j a_{ij} - \theta a_{io} + s_i^-}, & \text{if } \sum_{j=1}^m (x_{ij} - a_{ij}) \lambda_j - \theta(x_{io} - a_{io}) \leq 0 < \sum_{j=1}^m \lambda_j x_{ij} - \theta x_{io} + s_i^- \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$\mu_{-D_i'} \left(\sum_{j=1}^m \lambda_j x_{ij}' - \theta x_{io}' + s_i^-, 0 \right) = \begin{cases} 1, & \text{if } \sum_{j=1}^m (x_{ij} + a_{ij}) \lambda_j - \theta(x_{io} + a_{io}) + 2s_i^- < 0 \\ 1 - \frac{\sum_{j=1}^m (x_{ij} + a_{ij}) \lambda_j - \theta(x_{io} + a_{io}) + 2s_i^-}{\sum_{j=1}^m \lambda_j a_{ij} - \theta a_{io} + s_i^-}, & \text{if } -\sum_{j=1}^m \lambda_j a_{ij} + \theta a_{io} - s_i^- \leq \sum_{j=1}^m \lambda_j x_{ij} - \theta x_{io} + s_i^- < 0 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$\mu_{D_r'} \left(y_{ro}', \sum_{j=1}^s \lambda_j y_{nj}' - s_r^+ \right) = \begin{cases} 1, & \text{if } \sum_{j=1}^s \lambda_j (y_{nj} - b_{nj}) - (y_{ro} + b_{ro}) > 0 \\ 1 + \frac{\sum_{j=1}^s \lambda_j (y_{nj} - b_{nj}) - (y_{ro} + b_{ro})}{\sum_{j=1}^s \lambda_j b_{nj} + b_{ro} - s_r^+}, & \text{if } \sum_{j=1}^s \lambda_j (y_{nj} - b_{nj}) - b_{ro} \leq y_{ro} < \sum_{j=1}^s \lambda_j y_{nj} - s_r^+ \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$\mu_{-D_r'} \left(\sum_{j=1}^s \lambda_j y_{nj}' - s_r^+, y_{ro}' \right) = \begin{cases} 1, & \text{if } \sum_{j=1}^s \lambda_j (y_{nj} + b_{nj}) - (y_{ro} - b_{ro}) - 2s_r^+ < 0 \\ 1 - \frac{\sum_{j=1}^s \lambda_j (y_{nj} + b_{nj}) - (y_{ro} - b_{ro}) - 2s_r^+}{\sum_{j=1}^s \lambda_j b_{nj} + b_{ro} - s_r^+}, & \text{if } -\sum_{j=1}^s \lambda_j b_{nj} + (y_{ro} - b_{ro}) + s_r^+ \leq \sum_{j=1}^s \lambda_j y_{nj} - s_r^+ < b_{ro} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

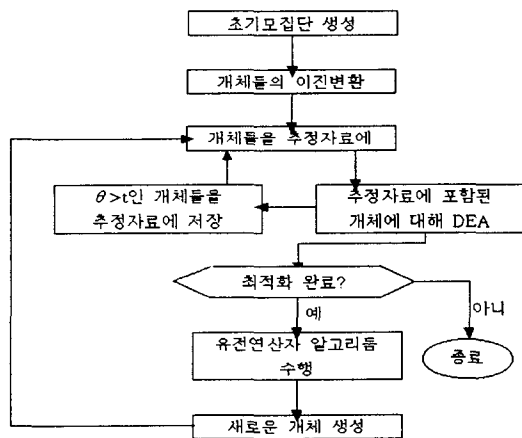
이 때 식 (2)의 최적해는 목적함수는 최소화하고, 각 제약조건에 대한 믿음의 정도는 최대화하는 과정에서 구해진다고 볼 수 있다. 그런데 각 제약조건에 대한 믿음의 정도를 최대화하는 과정에서, 믿음의 정도의 값이 1 혹은 0의 값을 가질 때는 전형적인 단일목적함수를 가지는 선형계획법으로 풀 수 있는 식으로 단순화되는 반면, 1과 0사이의 값을 가지는 경우에는 식 (9)와 같이 $1+m+s$ 개의 목적함수를 가지고 $2(m+s)$ 개의 제약조건을 가지는 다목적 비선형계획문제가 됨을 알 수 있다.

$$\begin{aligned}
 & \text{Min } \theta - \epsilon [\sum_{i=1}^n s_i^- - \sum_{i=1}^n s_i^+] \\
 & \text{Max } \mu_{D_i'} = 1 + \frac{\sum_{j=1}^n (x_{ij} - a_{ij})\lambda_j - \theta(x_{i0} - a_{i0})}{\sum_{j=1}^n \lambda_j a_{ij} - \theta a_{i0} + s_i^-} \\
 & \text{Max } \mu_{-D_i'} = 1 - \frac{\sum_{j=1}^n (x_{ij} + a_{ij})\lambda_j - \theta(x_{i0} + a_{i0}) + 2s_i^-}{\sum_{j=1}^n \lambda_j a_{ij} - \theta a_{i0} + s_i^-} \quad (9) \\
 & \text{Max } \mu_{D_r'} = 1 + \frac{\sum_{j=1}^n \lambda_j (y_{rj} - b_{rj}) - (y_{r0} + b_{r0})}{\sum_{j=1}^n \lambda_j b_{rj} + b_{r0} - s_r^+} \\
 & \text{Max } \mu_{-D_r'} = 1 - \frac{\sum_{j=1}^n \lambda_j (y_{rj} + b_{rj}) - (y_{r0} - b_{r0}) - 2s_r^+}{\sum_{j=1}^n \lambda_j b_{rj} + b_{r0} - s_r^+}
 \end{aligned}$$

Subject to

$$\begin{aligned}
 & \sum_{j=1}^n (x_{ij} - a_{ij})\lambda_j - \theta(x_{i0} - a_{i0}) \leq 0 < \sum_{j=1}^n \lambda_j x_{ij} - \theta x_{i0} + s_i^- \\
 & - \sum_{j=1}^n \lambda_j a_{ij} + \theta a_{i0} - s_i^- \leq \sum_{j=1}^n \lambda_j x_{ij} - \theta x_{i0} + s_i^- < 0 \\
 & \sum_{j=1}^n \lambda_j (y_{rj} - b_{rj}) - b_{r0} \leq y_{r0} < \sum_{j=1}^n \lambda_j y_{rj} - s_r^+ \\
 & - \sum_{j=1}^n \lambda_j b_{rj} + (y_{r0} - b_{r0}) + s_r^+ \leq \sum_{j=1}^n \lambda_j y_{rj} - s_r^+ < b_{r0}
 \end{aligned}$$

따라서 본 연구에서는 이러한 다목적비선형계획 문제에 대한 비지배해(nondominated solution) 집합을 구하기 위해, 다지점 탐색기법으로 최적해를 구하고, 벌금함수 등을 통해 비선형인 제약식을 다룸으로써 전체최적(global optimum)을 탐색할 수 있는 방법인 유전자 알고리즘을 적용하고자 하였다. 이때 개체들의 효율적인 적용도 평가를 위하여, 개체들의 적용도를 목적함수들의 최적화에 기여하는 효율성의 개념으로 간주하여, DEA를 통해 각 개체들의 효율성 지수를 적용도로 삼고자 하였다(Arakawa et al., 1998). 먼저 본 연구에서 유전자 알고리즘을 적용해 식 (9)의 최적해를 탐색해 나가는 과정을 도식화하면 <그림 1>과 같다.



<그림 1> 유전자 알고리즘을 이용한 다목적비선형계획문제의 최적해 탐색과정

이러한 과정에 대하여 구체적으로 설명하면 다음과 같다. 먼저 유전알고리즘은 개체(염색체)들로 구성된 모집단을 운영하므로 초기에 모집단이 생성되어야 한다. 초기의 모집단을 만드는 방법으로는 문제의 특성을 이용한 기존의 발견적 방법이나 임의생성방법이 있을 수 있다. 그러나 본 문제의 경우에 있어, 발견적 기법에 의해 모집단을 생성할 경우, 해들이 조기 수렴하여 비지배해를 찾기 위한 해공간의 다양한 탐색을 방해할 수 있는 경향이 있을 것으로 판단되어, 임의생성방법을 사용하기로 하였다. 다음으로, 임의 생성하는 경우 모집단을 구성하는 개체들을 유전적 표현으로 나타내야 하므로, 개체원소(유전자)의 수만큼 0과 1을 임의로 발생하여야 하나의 개체가 형성된다. 따라서 개체를 이진수 벡터로 표시하기 위해 먼저 벡터의 길이를 요구되는 정밀도에 의해 결정하였다. 본 연구의 경우, 결정변수 λ_j 및 h_0 가 소수점 3자리까지의 정밀도를 요구한다고 가정하고, λ_j 및 h_0 가 모두 $[0, 1]$ 의 구간에 존재하므로 1×1000 개의 등간격 구간이 필요함을 알 수 있다. 그러면 $514 = 2^9 \leq 1000 \leq 2^{10} = 1024$ 이므로, 10 bit로 이루어진 이진수 벡터 $(b_9 \ b_8 \ \dots \ b_0)$ 를 십진수 값으로 변환하면 $(b_9 \ b_8 \ \dots \ b_0)_2 = \left(\sum_{i=0}^9 b_i \cdot 2^i \right)_{10} = x'$ 이다. 따라서 $[0, 1]$ 구간으로 변환하면 결정변수 λ_j 및 h_0 의 값은 x (x 은 λ_j 혹은 h_0) $= 0 + 1 \cdot \frac{x'}{2^{10} - 1}$ 와 같이 구할 수 있다.

이제 이렇게 이진변환을 통해 유전적으로 표현한 자연 개체들에 대한 생존능력을 알기 위해 적용도 평가를 실시한다. 일반적으로 적용도를 평가하는 함수로는 최적화문제의 목적함수 자체를 흔히 사용하나, 본 문제의 경우에는 비선형의 제약조건을 내포하고 있으므로 일반적인 적용도 평가 방법을 적용하기 어려울 뿐만 아니라, 최적화하고자 하는 목적함수의 형태가 단일목적함수가 아닌 다목적함수로 구성되어 있다는 특징이 있다. 따라서 본 연구에서는 이러한 복합적인 문제를 해결하기 위하여, 먼저 각 목적함수에 제약조건을 어기는 정도를 벌금으로 부과하는 벌금함수를 적용함으로써 제약조건을 포함하게끔 하였다. 또한 평가 개체에 대한 각 목적함수의 값을 평가 개체가 모든 목적함수의 최적화에 기여하는 정도, 즉 그 개체의 효율성의 개념으로 간주하고, DEA 평가를 통해 평가 개체의 상대적 효율성 지수를 구함으로써 이를 해당 개체의 적용도로 삼았다. 이 경우, DEA 분석의 비교

대상인 DMU은 자연 개체 및 유전연산자를 작용한 후 새로이 생성된 개체들로 구성된 모집단이고, 입력변수는 최소화하고자 하는 목적함수의 벌금함수 값이며, 출력변수는 최대화하고자 하는 목적함수의 벌금함수 값이 된다. 이러한 방식으로 해당 잠재해들에 대한 적응도 평가가 끝나면, 효율성 지수가 1 혹은 의사결정자의 주관에 의해 결정된 임의의 효율성 지수를 만족하는 개체들은 선별하여 저장해둠으로써 다음 세대에도 생존하게끔 한다.

그 후에는 일반적인 유전알고리즘의 과정인 선별 및 유전 연산자로서의 교차, 돌연변이 등을 통해 새로운 자손 개체를 생성한 다음, 이들 생성 개체들에 대한 적응도 평가를 통해 문제의 종료조건인 최적화에 도달했는가를 판단하고 그렇지 않으면 선별과정과 유전 연산 및 평가를 반복하게 된다.

3. 결론

본 연구에서는 조직의 상대적 효율성을 분석하는 DEA모형에서 synergy 효과에 의한 불확실성을 고려하기 위해 CCR 모형을 변형한 fuzzy DEA를 제시하였다. 향후 더욱 심도 있게 연구되어야 할 것으로는, 먼저 입·출력변수 이외에 효율성 지수의 확률적인 측면(stochastic aspect)을 고려하여 효율성 지수의 분포를 분석함으로써 유용한 정보를 획득할 수 있는 분석이 이루어져야 할 것이다. 이러한 연구는 Sengupta(1996)이 효율성 지수의 분포를 분석하기 위하여 index number 접근법과 정보이론의 엔트로피(entropy) 접근법을 제안하고, 이들과 DEA 접근법과의 연계를 통해 효율성 지수의 분포에 대한 분석의 실질적 유용성을 예증한 바 있다. 또한 환경적 원인에 의해 변수의 수준이 고정되어야 하는 외생고정변수(exogenously fixed variable) (Banker & Morey, 1985)을 고려한 DEA 분석이 이루어져야 할 것이다.

Reference

- 김여근, 윤복식, 이상복, 메타휴리스틱, 영지문화사, 3-56.
- Arakawa, M., Nakayama, H., Hagiwara, I., Yamakawa, H.(1998), Multiobjective Optimization Using Adaptive Range Genetic Algorithms with Data Envelopment Analysis, Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, P.3, 2074-2082.
- Banker, R. D., Morey, R. C.(1985), Efficiency Analysis for Exogenously Fixed Inputs and Outputs, Operational Research, 34, 513-521.
- Charnes, A., Cooper, W. W., Rhodes, E.(1978), Measuring the Efficiency of Decision Making Units, European Journal of Operational Research, 2(6), 429-444.
- Charnes, A., Clark, C. T., Cooper, W. W., Golany, B.(1985), A Development Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U. S. Air Forces, Annals of Operational Research, 2, 95-112.
- Kahraman, C., Tolga, E.(1998), Data Envelopment Analysis Using fuzzy Concept, Proceedings of the 28th IEEE International Symposium on Multiple-Valued Logic (ISMVL'98), 338-343.
- Kao, C. K., Liu, B.(1998), Data Envelopment Analysis with Missing Data: An Application to University Libraries in Taiwan, National Cheng Kung University.
- Kim, S. H., Park, C. G., Park, K. S.(1999), An Application of Data Envelopment Analysis in Telephone Offices Evaluation with Partial Data, Computers and Operations Research, 26, 59-72.
- Maeda, Y., Entani, T., Tanaka, H.(1998), fuzzy DEA with interval Efficiency, Proceedings of the 6th European Congress on Intelligent Techniques & Soft Computing, 2, 1067-1071.
- Sengupta, J. K.(1996), Entropy, Efficiency and the Productivity Index Numbers, International Journal of Systems Science, 27(12), 1195-1204.
- Tong, S.(1994), interval number and fuzzy number linear programming, fuzzy Sets and Systems, 66, 301-306.