

A Hybrid Clustering Method Using Balanced Scorecard and Data Envelopment Analysis

Hêriş Golpîra,

Department of Industrial Engineering, Sanandaj Branch, Islamic Azad University,
Sanandaj, Iran

Abstract: This paper introduces a new hybrid clustering method using data envelopment analysis (DEA) and balanced scorecard (BSC) methods. DEA cannot identify its' input and output by itself and it is a major weakness of the DEA. In the proposed method, this weakness is resolved by integrating DEA and BSC. Number of decision making units (DMUs), needed in DEA method, in compliance with number of inputs and outputs is the major drawback of this integration. To deal with this drawback, the proposed method selects the most important strategic factors, attained from the BSC method. These data considered to be the input data for the DEA method to calculate relative closeness (RC) of each DMU to ideal one. Plotting the scree diagram regarding RC index leads us to the final clustering method. Finally, computational results show the applicability and usefulness of the method.

Key words: Balanced scorecard, data envelopment analysis, ranking method, clustering

1. INTRODUCTION AND LITERATURE REVIEW

Clustering is a statistical method to divide similar objects into the same bunches. There is a vast literature in the field of clustering and there have been attempts to categorize these researches. Fahad et al. (2014) introduce concepts and algorithms related to the area of clustering as well as providing a comparison, not only from a theoretical but also from an empirical perspective. Many algorithms are proposed to cluster data based on minimizing total dissimilarity (Po, Guh, & Yang, 2009), such as hard C-means (HCM) (Ross, 2009), fuzzy C-means (FCM) (Barrios, Villanueva, Cavazos, & Colas, 2016), possibilistic C-means (PCM) (Škrjanc & Dovžan, 2015), interval Type-2 fuzzy possibilistic C-means clustering algorithm (Rubio, Castillo, & Melin, 2015), multiple kernels interval Type-2 possibilistic C-means (Vu & Ngo, 2016), and so on. Thanassoulis (1996), Narasimhan, Talluri, and Mendez (2001), Po et al. (2009), and Hêriş Golpîra and Hajebi (2015) introduce some new clustering methods based on DEA approach. Po et al. (2009) employ Charnes, Cooper & Rhodes (CCR) model and Goudarzi and Ansari (2012) use Banker, Charnes and Cooper (BCC) concept of DEA in their method. This paper proposes a new method based on the (Y.-M. Wang & Luo, 2006) for clustering sub-organizations of a complex organization. The method not only uses this approach, but applies Balanced Scorecard (BSC) to introduce the outputs and rather the inputs to have a complete comprehensive method for organizational clustering.

Resource assessment is taking place in the competitive environment (Bentes, Carneiro, da Silva, & Kimura, 2012), particularly in large, complex organizations. Given the multidimensionality and the complexity of the concept, Venkatraman and Ramanujam (1986), Chakravarthy (1986), and Barney (2002) advise the use of multiple measures for organizational performance measurement.

BSC is one of the well-known multi-dimensional organizational assessment methods (Johnson & Kaplan, 1987). The university based idiom of the method is denoted by Kaplan and Norton (1992). More emphasize on balanced measurement and related factors is documented by Cobbold and Lawrie (2002), and the strategy plan is finally employed to complete the model (Niven, 2011). The model

finds the relation between strategic goals and operational controls, based on four fundamental factors. BSC eliminates information overload and forces the management team to illuminate the organizational strategy. This process makes the model to be more tractable. But the most important drawback of the BSC is its identification of metrics. This paper tries to overcome this problem by using DEA method.

DEA, introduced by Charnes, Cooper, and Rhodes (1978), often evaluates the decision making units (DMUs) from the best possible relative efficiency. Entani, Maeda, and Tanaka (2002) and Y. Wang, Greatbanks, and Yang (2007) acquire the model to look at both the optimistic and pessimistic points, until Y.-M. Wang and Luo (2006) propose their model based on the relative closeness (RC) index to the ideal DMU (IDMU). J.-X. Chen (2012) proposes a corrective notes on the method introduced by Y.-M. Wang and Luo (2006) in the area of ranking. Hêriş Golpîra (2012) Employs this version of DEA for formulating the problem of project monitoring and achieve correct comprehensive project success measurement. H Golpîra and Mohajeri (2012) employ the same concept in compliance with the BSC model in order to assess the organizations. This paper uses the same approach not only for evaluating sub-organizations, but also for clustering. Coelli, Rao, O'Donnell, and Battese (2005) promote 11 major drawbacks encountered in conducting the DEA. They illustrate that the exclusion of an important input or output can result in biases. In other words, the main drawback of the DEA is its weakness on identifying Input and output factors. It is noteworthy that Banker, Chang, Janakiraman, and Konstans (2004) use the combination of the DEA and the BSC to evaluate the trade-offs among different performance indexes. T.-Y. Chen and Chen (2007) use this combination to measure the performance of a semiconductor industry. Chiang and Lin (2010) apply it to measure the performance of two distinct industries. Min, Min, and Joo (2008) use the same method to measure the performance of Korean hotels and Macedo, Barbosa, and Cavalcante (2009) apply it in banking. Amado, Santos, and Marques (2012) apply DEA to measure the performance of DMUs in only one company. It is clear that the focus of these scholars is on the performance assessment; however, this paper employs this combination to introduce a new powerful method for clustering. Besides, Cooper (2000) proposes a generally accepted principle to ensure satisfactory discrimination of the DEA method as follows:

$$n \geq \max\{m \times s, 3(m + s)\} \quad (1)$$

Where n is the number of DMUs, m and s are the number of inputs and outputs. However, such conditions may not be satisfied in many applications.

BSC literature shows that the strategy map of organizations should contains at least two or three indexes for each level of the factors which is consequently introduces at least 8 factors as the outputs. Optimistically, with considering only two inputs, the methods may contain more than 30 DMUs to make a clear satisfying result. This makes the traditional methods to be complex and impractical. It is the superiority of the proposed method that is not be limited by this principle and may be used for ranking and clustering the DMUs with any number of outputs or inputs. So, in this paper, a hybrid method is proposed that handle the advantages of BSC and DEA all together and encounter the disadvantage of the DEA by the relative advantages of the BSC and TOPSIS. In other words, The BSC method is used to determine two or three most important factors in any field of its four basic fundamental factors. The factors are then used as the input data for the DEA method to make ranking and clustering in any organizations with and number of sub-organizations.

2. PROPOSED METHOD

Suppose n DMUs, each consumes m inputs, denoted by $x_{ij} (i = 1, \dots, m, j = 1, \dots, n)$, to produce outputs denoted by $y_{rj} (i = 1, \dots, s, j = 1, \dots, n)$. A virtual DMU which uses the least inputs, $x_i^{\min} (i = 1, \dots, m)$, to produce the most outputs, $y_r^{\max} (r = 1, \dots, s)$, and a DMU, which consume the most inputs, $x_i^{\max} (i = 1, \dots, m)$, to generate the least outputs, $y_r^{\min} (r = 1, \dots, s)$ can be defined as ideal decision making unit (IDMU) and anti-ideal decision making unit (ADMU) respectively. To completing the model, the LP model shown in Equations (2) and (3) must be solved for all DMU_s such as DMU₀ to

calculate the $\theta_{j_0}^*, \phi_{j_0}^*$, where j_0 is the DMU under evaluation (denoted by DMU₀), u_r, v_i are decision variables, ε is the non-Archimedean infinitesimal, θ_{IDMU}^* is the optimum efficiency of IDMU that may be calculated by Equation (4) and ϕ_{ADMU}^* is the worst efficiency of the ADMU that may be calculated by Equation (5). It is obvious that the relative closeness index of DMU₀ to IDMU is defined by Equation (6). It is clear that the bigger the RC_{j_0} value is the better the performance of DMU₀.

$$\begin{aligned}
 & \text{Max } \theta_{j_0} = \sum_{r=1}^s u_r y_{rj_0} \\
 & \text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1 \\
 & \sum_{r=1}^s u_r y_j^{\max} - \sum_{i=1}^m v_i (\theta_{IDMU}^* x_i^{\min}) = 0 \\
 & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j = 1, \dots, n \\
 & u_r, v_i \geq \varepsilon \quad \forall r, i
 \end{aligned} \tag{2}$$

$$\begin{aligned}
 & \text{Min } \phi_{j_0} = \sum_{r=1}^s u_r y_{rj_0} \\
 & \text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1 \\
 & \sum_{r=1}^s u_r y_j^{\min} - \sum_{i=1}^m v_i (\phi_{IDMU}^* x_i^{\max}) = 0 \\
 & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j = 1, \dots, n \\
 & u_r, v_i \geq \varepsilon \quad \forall r, i
 \end{aligned} \tag{3}$$

$$\begin{aligned}
 & \text{Max } \theta_{IDMU} = \sum_{r=1}^s u_r y_r^{\max} \\
 & \text{s.t. } \sum_{i=1}^m v_i x_i^{\min} = 1 \\
 & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j = 1, \dots, n \\
 & u_r, v_i \geq \varepsilon \quad \forall r, i
 \end{aligned} \tag{4}$$

$$\begin{aligned}
 \text{Min } \phi_{ADMU} &= \sum_{r=1}^s u_r y_r^{\min} \\
 \text{s.t. } \sum_{i=1}^m v_i x_i^{\max} &= 1 \\
 \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} &\leq 0 \quad j = 1, \dots, n \\
 u_r, v_i &\geq \varepsilon \quad \forall r, i
 \end{aligned} \tag{5}$$

$$RC_{j0} = \frac{\varphi_{j0}^* - \varphi_{ADMU}^*}{(\varphi_{j0}^* - \varphi_{ADMU}^*) + (\theta_{IDMU}^* - \theta_{j0}^*)} \tag{6}$$

The *RC* indexes are sorted in descending order and plotted in the way which is similar to the scree plot in the hierarchical clustering method. In this diagram, the sharp increase in *RC* illustrates a new cluster in DMUs. As per validating the method, it is successfully installed in 10 sub organization of Kermanshah Regional Water Organization Company which is illustrated in the last section.

3. EMPIRICAL STUDY

The data for this study are taken in from the research done by H Golpîra and Mohajeri (2012). The data included 53 creditable performance indexes that factor analysing in SPSS software classifies them into four levels of factors. Data are classified as: (1) 10 financial indexes, (2) 7 internal business process indexes, (3) 7 customer Indexes, (4) 24 innovation and learning indexes. Indexes are given to experts to give a privilege to them according to organizational predefined strategies. Consequently, “five point Likert” and “factor analysis” methods are used to prove the classification. Then the most important indexes in each four levels are chosen. After linking the factors in BSC procedure, the strategic map is given as shown in Fig. 1. These indexes are used as the outputs for the DEA method. Seven inputs which are strongly related to these outputs are also selected and the real data are collected from the 10 sub-organization of the Kermanshah Water Regional Organization which are illustrated in Table 1. Finally, DEA is used to rank these sub-organizations using factors which are indicated on the strategy map. The results are illustrated in Table 2.

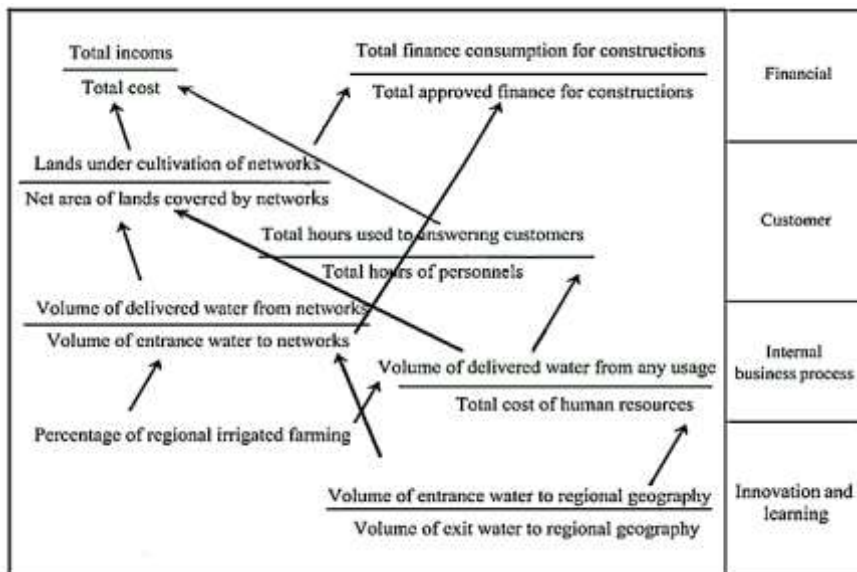


Figure 1: Strategy map of Kermanshah Regional Water Organization (H Golpîra & Mohajeri, 2012)

What is indicated in column five (RC) of Table 2 shows the difference of the sub-organizations. So managers not only can clearly recognize the differences between their organizations to others, but also the related distances can show the intensity of these differences. This information helps the manager to have a better view to perceive the position of his/her organization and enhance an ability to compare it with other similar ones in terms of the organizational strategic goals that may be changed and updated over its life cycle. This ranking is based on the other similar organizations that make it possible and acceptable for any others.

Table 1: Inputs and outputs data for 10 sub-organization of Kermanshah Water Regional Organization

DMU	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	
1	46963.00	54.00	990.00	169934.10	75357.00	273600.00	16	
2	37570.40	9.00	99.00	118953.87	52749.90	164160.00	16	
3	16437.05	12.60	264.00	101960.46	45214.20	68400.00	14	
4	16437.05	32.40	330.00	33986.82	33910.65	54720.00	14	
5	18785.20	7.92	231.00	254901.15	22607.10	76608.00	14	
6	28177.80	6.48	198.00	169934.10	33910.65	109440.00	16	
7	37570.40	14.04	264.00	339868.20	15071.40	191520.00	16	
8	11740.75	16.74	330.00	254901.15	36171.36	54720.00	14	
9	7044.45	13.68	429.00	169934.10	37678.50	41040.00	12	
10	14088.90	13.14	165.00	84967.05	16578.54	27360.00	14	
Max	46963.00	54.00	990.00	339868.20	75357.00	273600.00	16	
Min	7044.45	6.48	99.00	33986.820	15071.40	27360.00	12	
DMU	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈
1	3.00	0.14	0.007	9.18	0.34	0.51	0.86	16
2	0.75	0.33	0.005	6.45	0.20	0.54	0.56	16
3	0.40	0.04	0.003	9.50	0.23	0.41	0.57	14
4	0.33	0.03	0.006	2.38	0.19	0.56	0.39	14
5	0.51	0.12	0.004	7.31	0.26	0.34	0.56	14
6	0.38	0.11	0.002	4.55	0.39	0.43	0.67	16
7	0.13	0.20	0.006	10.05	0.23	1.45	0.56	16
8	0.60	0.06	0.006	9.82	0.19	0.64	0.48	14
9	0.93	0.10	0.008	3.83	0.18	0.34	0.56	12
10	0.30	0.05	0.005	4.61	0.25	0.44	0.45	14
Max	3.00	0.33	0.008	10.05	0.39	1.45	0.86	16
Min	0.13	0.03	0.002	2.38	0.18	0.34	0.39	12

Table 2: DEA Results

DMU	φ*(ADMU)	Θ*(IDMU)	RC	rank
1	1.16903	1	0.095	6
2	3.573291	0.95	0.257	1
3	1	0.6497418	0.079	8
4	1	0.5973795	0.078	9
5	1.308757	0.7616327	0.104	4
6	1.018273	0.5963	0.080	7
7	1	0.4182025	0.077	10
8	1.257319	0.8997455	0.102	5
9	1.346556	1	0.110	3
10	1.647861	1	0.133	2
11	-	10.92647	-	-
12	0.1231	-	-	-

$$\varepsilon = 1 \times 10^{-6}$$

As per completing the proposed clustering procedure the scree diagram is plotted in Fig. 2. One can see that the diagram has sharp increasing shape in some points which produces 4 partitions. The clustering is graphically obvious but the hierarchical clustering method is used to define the number of clusters and this number of clustering is used as the input of the hard C-means method to have a clear predefined valid clustering. This process is done by using SPSS software which its results are shown in Table 3. The results show that the optimal number of clusters is 4 clusters which are used to have final clustering by using hard C-means method. The results are shown in Table 4, Table 5 and Table 6. The results are clearly emphasizing on what is achieved in the proposed method.

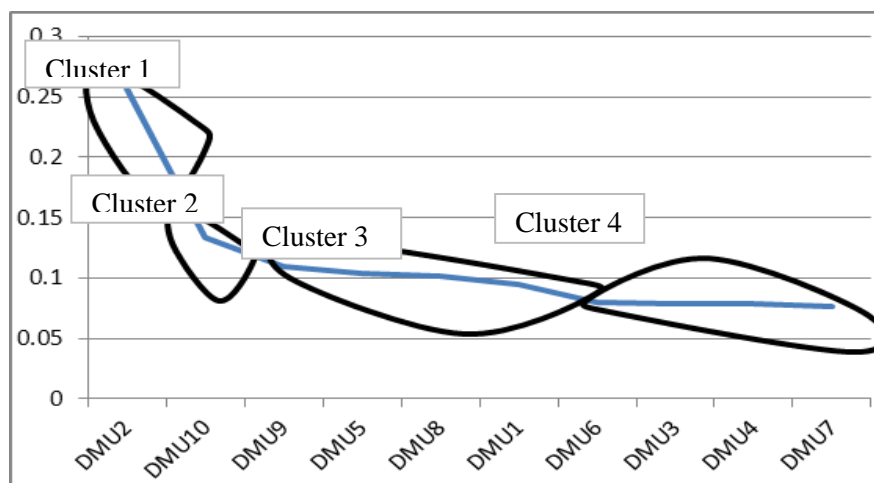


Figure 2: Final clustering results

Table 3: Hierarchical clustering results

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	8	9	0.000	0	0	2
2	7	8	0.000	0	1	3
3	7	10	0.000	2	0	7
4	4	5	0.000	0	0	5
5	3	4	0.000	0	4	6
6	3	6	0.000	5	0	7
*7	3	7	0.001	6	3	8
8	2	3	0.002	0	7	9
9	1	2	0.026	0	8	0

Table 4: Hard C-means cluster membership

Case Number	VAR00001	Cluster	Distance
1	DMU2	1	0.000
2	DMU10	2	0.000
3	DMU9	3	0.007
4	DMU5	3	0.002
5	DMU8	3	0.001
6	DMU1	3	0.007
7	DMU6	4	0.001
8	DMU3	4	0.000
9	DMU4	4	0.000
10	DMU7	4	0.001

Table 5: Hard C-means final clustering centers results

	Cluster			
	1	2	3	4
VAR00002	0.256966	0.133153	0.102784	0.078410

Table 6: ANOVA results

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
VAR00002	0.009	3	0.000	6	470.700	0.000

4. CONCLUSION

This paper introduces a new hybrid clustering method using data envelopment analysis (DEA) and balanced scorecard (BSC) methods. The basic BSC is employed to define the important factors in organizational performance which leads the system having valid and strategic factors. These factors are used as the outputs of the DEA method and trying the relative inputs with no limitations. The model introduced by Y.-M. Wang and Luo (2006) is employed afterward to determine RC indexes for all of the DMUs. Finally the DMUs are classified by using the scree plot and focusing upon sharpness of the diagram. The results are validated by using two well-known traditional clustering methods. The noticeable superiority of the model is its ability to encounter with clustering problems without any limitation from number of inputs/outputs or the number of DMUs. The other superiority of the model is its comprehensiveness and practical characteristics. The simple graphical process is the other advantage of the method that makes it understandable and acceptable in addition to its capability to be used as the ranking, benchmarking and clustering method synchronously. The numerical results are clearly validating the method and make it practical.

REFERENCES

- Amado, C. A., Santos, S. P., & Marques, P. M. (2012). Integrating the Data Envelopment Analysis and the Balanced Scorecard approaches for enhanced performance assessment. *Omega*, 40(3), 390-403, [CrossRef](#)

- Banker, R. D., Chang, H., Janakiraman, S. N., & Konstans, C. (2004). A balanced scorecard analysis of performance metrics. *European Journal of Operational Research*, 154(2), 423-436, [CrossRef](#)
- Barney, J. B. (2002). Gaining and sustaining competitive advantage.
- Barrios, J. A., Villanueva, C., Cavazos, A., & Colas, R. (2016). Fuzzy C-means Rule Generation for Fuzzy Entry Temperature Prediction in a Hot Strip Mill. *Journal of Iron and Steel Research, International*, 23(2), 116-123, [CrossRef](#)
- Bentes, A. V., Carneiro, J., da Silva, J. F., & Kimura, H. (2012). Multidimensional assessment of organizational performance: Integrating BSC and AHP. *Journal of business research*, 65(12), 1790-1799, [CrossRef](#)
- Chakravarthy, B. S. (1986). Measuring strategic performance. *Strategic management journal*, 7(5), 437-458, [CrossRef](#)
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444, [CrossRef](#)
- Chen, J.-X. (2012). A comment on DEA efficiency assessment using ideal and anti-ideal decision making units. *Applied Mathematics and Computation*, 219(2), 583-591, [CrossRef](#)
- Chen, T.-Y., & Chen, L.-h. (2007). DEA performance evaluation based on BSC indicators incorporated: The case of semiconductor industry. *International journal of Productivity and Performance management*, 56(4), 335-357, [CrossRef](#)
- Chiang, C.-Y., & Lin, B. (2010). An Integration of Balanced Scorecards and Data Envelopment Analysis for Firm's Benchmarking Management. *Quality control and applied statistics*, 55(1), 61-62.
- Cobbold, I., & Lawrie, G. (2002). The development of the balanced scorecard as a strategic management tool. *Performance measurement association*.
- Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*: Springer Science & Business Media.
- Cooper, W. (2000). Seiford. LM and Tone, K.(2000). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*: Boston: Kluwer Academic Publishers.
- Entani, T., Maeda, Y., & Tanaka, H. (2002). Dual models of interval DEA and its extension to interval data. *European Journal of Operational Research*, 136(1), 32-45, [CrossRef](#)
- Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *Emerging Topics in Computing, IEEE Transactions on*, 2(3), 267-279, [CrossRef](#)
- Golpîra, H. (2012). Real project success measurement by using data envelopment analysis. *Scientific Committee–Editorial Board*, 43.
- Golpîra, H., & Hajebi, S. (2015). Clustering approach for organizational evaluation project: integrating BSC and DEA Practice and Perspectives, 117.
- Golpîra, H., & Mohajeri, A. (2012). A New Method to Organizational Ranking: Integrating BSC and DEA. *International Journal of Research in Industrial Engineering*, 1(3), 39-47.
- Goudarzi, M., & Ansari, J. (2012). Clustering Decision Making Units (DMUs) Using Full Dimensional Efficient Facets (FDEFs) of PPS with BCC Technology. *Applied Mathematical Sciences*, 6(29), 1431-1452.
- Johnson, H. T., & Kaplan, R. S. (1987). *Relevance lost*: Boston: Harvard Business School Press.
- Kaplan, R., & Norton, D. (1992). The Balanced Scorecard-Measures That Drive Performance. *Harvard Business Review*, 1(70).
- Macedo, M., Barbosa, A., & Cavalcante, G. (2009). Performance of bank branches in Brazil: applying data envelopment analysis (DEA) to indicators related to the BSC perspectives. *E&G—Revista Economia e Gestão*, 19(19), 65-84.

- Min, H., Min, H., & Joo, S.-J. (2008). A data envelopment analysis-based balanced scorecard for measuring the comparative efficiency of Korean luxury hotels. *International Journal of Quality & Reliability Management*, 25(4), 349-365, [CrossRef](#)
- Narasimhan, R., Talluri, S., & Mendez, D. (2001). Supplier evaluation and rationalization via data envelopment analysis: an empirical examination. *Journal of supply chain management*, 37(2), 28-37, [CrossRef](#)
- Niven, P. R. (2011). *Balanced scorecard: Step-by-step for government and nonprofit agencies*: John Wiley & Sons
- Po, R.-W., Guh, Y.-Y., & Yang, M.-S. (2009). A new clustering approach using data envelopment analysis. *European Journal of Operational Research*, 199(1), 276-284, [CrossRef](#)
- Ross, T. J. (2009). *Fuzzy logic with engineering applications*: John Wiley & Sons.
- Rubio, E., Castillo, O., & Melin, P. (2015). A new Interval Type-2 Fuzzy Possibilistic C-Means clustering algorithm. Paper presented at the Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC), 2015 Annual Conference of the North American, [CrossRef](#)
- Škrjanc, I., & Dovžan, D. (2015). Evolving Gustafson-kessel Possibilistic c-Means Clustering. *Procedia Computer Science*, 53, 191-198, [CrossRef](#)
- Thanassoulis, E. (1996). A data envelopment analysis approach to clustering operating units for resource allocation purposes. *Omega*, 24(4), 463-476, [CrossRef](#)
- Venkatraman, N., & Ramanujam, V. (1986). Measurement of business performance in strategy research: A comparison of approaches. *Academy of management review*, 11(4), 801-814, [CrossRef](#), [CrossRef](#)
- Vu, M. N., & Ngo, L. T. (2016). A Multiple Kernels Interval Type-2 Possibilistic C-Means Recent Developments in Intelligent Information and Database Systems (pp. 63-73): Springer, [CrossRef](#)
- Wang, Y.-M., & Luo, Y. (2006). DEA efficiency assessment using ideal and anti-ideal decision making units. *Applied Mathematics and Computation*, 173(2), 902-915, [CrossRef](#)
- Wang, Y., Greatbanks, R., & Yang, J. (2007). Measuring the efficiency of decision making units using interval efficiencies. *European Journal of Operation Research*.