Facility Location Selection Using Clustering Based Genetic Algorithm

Kümeleme Tabanlı Genetik Algoritmayla Tesis Yeri Seçimi

Abstract

Facility location selection is one of the most important decisions of companies and industries. At the same time, since a business process begins with the selection of a facility location, it is the first step to consider. Everything starts with the facility location selection. If a location far to suppliers, manufacturers or the market is selected, this will lead to increasing costs in the long run for both the company and other items in the supply network. The distant location also affects the mutual contracts in detail. Besides, the facility location has effects on labor costs and other related costs. Almost all of the costs in the company is closely related with the facility location. Based on this mentioned importance, the facility location selection problem is considered in this study, and the clustering based genetic algorithm method is proposed for the solution of facility location selection problem. In the introductory part of the study, facility location selection problem and the related literature is introduced. After, methods used in the solution are presented as K-means clustering algorithm, genetic algorithm and the proposed algorithm respectively. Detailed numerical results of the study is given in the facility location selection section by using Ruspini75 data set from Operations Research Library. All obtained results are interpreted in the results and discussion section and the study is concluded with the suggestions for future works.

Yakup Çelikbilek
Doç. Dr., İstanbul Gelişim
Üniversitesi, İİBF,
Yönetim Bilişim Sistemleri Bölümü,
ycelikbilek@gelisim.edu.tr,
Orcid: 0000-0003-0585-1085

Öz

Tesis yeri seçimi, şirketlerin ve endüstrilerin en önemli kararlarından biridir. Aynı zamanda, bir iş süreci tesis yeri seçimi ile başladığından, üzerinde durulması gereken ilk adımdır. Eğer tedarikçilere, üreticilere veya pazara uzak bir konum seçilirse; bu durum, hem şirket hem de tedarik ağındaki diğer öğeler açısından uzun vadede artan maliyetlere neden olacaktır. Ayrıca uzak konum, karşılıklı yapılan sözleşmeleri de detaylı olarak etkileyecektir. Bununla birlikte tesis yeri, işgücü maliyetleri ve diğer maliyetler üzerinde de etkilidir. Şirketteki maliyetlerin neredeyse tamamı tesisin konumu ile yakından alakalıdır. Bahsedilen bu öneminden yola çıkarak, bu çalışmada tesis yeri seçimi problemi ele alınmakta ve problemin çözümü için, kümeleme tabanlı genetik algoritma yöntemi önerilmektedir. Çalışmanın, giriş bölümünde, tesis yeri seçimi problemi ve ilgili literatür tanıtılmaktadır. Ardından; çözümde kullanılan yöntemler, K-ortalamalar kümeleme algoritması, genetik algoritma ve önerilen algoritma olarak sırasıyla sunulmaktadır. Çalışmanın ayrıntılı sayısal sonuçları, Yöneylem Araştırması Kütüphanesi'nden Ruspini75 veri seti kullanılarak tesis yeri seçimi bölümünde verilmektedir. Tartışma ve sonuçlar bölümünde, elde edilen tüm sonuçlar yorumlanmakta ve gelecekteki çalışmalar için öneriler ile çalışma sonuçlandırılmaktadır.

Introduction

Facility location problem, which is one of the most important vital decisions of companies and industries economically, is firstly introduced by Weber and Friedrich (1962). Facility location selection is the determination of geographical location of a facility to start, relocate or expand the operations of a firm in order to optimize at least one objective i.e. cost, profit, distance, service etc. (Singh, 2016). However, facility location selection is not the only important decision to start operations. Firms implement their manufacturing strategies with the following decisions in their production–distribution system (Verter and Dincer, 1992):

– Facility Location

– Capacity Acquisition
– Technology Selection
– Production Mix
– Time-phasing of Investments
– Financial Planning

When the decisions given above are analyzed, all of them integrated with the facility location selection and starts after it. Therefore, the most important step for companies to start the operations is the selection of the best facility location. A worse selected facility location increases the other costs cumulatively related with the facility location. On the contrary, the best selected facility location also decreases the other costs both partially and cumulatively.

In this study, facility location selection using clustering based genetic algorithm is studied and applied to the data set named Ruspini75 from OR–library (Operations Research Library). The outline of the study is as follows. The related literature is introduced in the first section. Then, k-means clustering algorithm, genetic algorithm and the proposed algorithm are presented in the following sections respectively. Detailed numerical results of the study are given in the facility location selection section. In the results and discussion section, all obtained results are interpreted and the study is concluded with the suggestions for future works in the conclusion section.

1. Literature Review

Before introducing the theoretical background and the proposed algorithm, some of the important facility location studies will be presented in this part to clarify the importance of the problem. Even though facility location selection problem is one of the most important problem, there are not enough effective studies in the literature. Almost all of the studies in the literature were done by using multi criteria decision making (MCDM) methods, which are not really effective and enough to meet the needs of today's companies. Because the companies have too many criteria and data that are also more complex for MCDM solutions in today's world.

In 2003, Kahraman et al. (2003) presented four different fuzzy multi-attribute group decision-making approaches to select facility locations. The authors also applied all approaches with a numeric example for the comparative analysis. Arogundade et al. (2005) introduced two different methods including branch and bound techniques. They applied the proposed algorithms for the fire and emergency service facility location selection in Nigeria. A fuzzy outranking method for facility location selection was proposed by Kaya and Çinar (2006) in 2006. To deal with uncertainty in decision making problems, the fuzzy set theory was included in this proposed model. The model was also supported by a numerical example. Ertuğrul and Karakaşoğlu (2008) applied fuzzy analytical hierarchy process (AHP) and fuzzy technique for order preference by similarity to ideal solution (TOPSIS) methods for the facility location selection. They preferred fuzzy version of the MCDM methods to deal with uncertainty of linguistic assessment. Same year, a fuzzy simple additive weighting system with group decision making was introduced by Chou et al. (2008). The authors used different fuzzy and additive models to select the facility location. Another fuzzy approach with group decision making process was also presented by Shen and Yu (2009). The proposed approach included a risk judgement process to select the best location.

Another MCDM method, preference ranking organization method for enrichment evaluation II (PROMETHEE II), was used to select the real time facility location by Athawale and Chakraborty (2010). Safari et al. (2012) preferred to use fuzzy TOPSIS for the facility location selection. The authors selected three alternatives among five criteria. Wang and Watada (2012) presented a hybrid modified particle swarm optimization for the location selection of facilities with capacities. The authors supported their solution with the numerical experiments. An integrated Delphi and fuzzy AHP method was proposed by Kabir and Sumi (2013) and introduced with a numerical example for facility location selection problem. Fuzzy C-Means and Gustafson-Kessel algorithms were applied to clustering analysis of facility location selection by Büyüksaatçı and Esnaf (2014). Then, selection of the best facility location was done by minimizing CO_2 emission levels. Temur et al. (2014) presented a type-2 fuzzy TOPSIS to select the best facility location in reverse logistics. Another fuzzy

TOPSIS method using interval type-2 fuzzy sets was introduced by Çebi and Otay (2015). The authors applied the method to a site selection problem of a cement factory as a real life problem. Ray et al. (2015) applied ELECTRE I (ELimination Et Choix Traduisant la REalite) with various MCDM methods for the facility location selection problems. Basti and Sevkli (2015) used an artificial bee colony algorithm to select the best facility location in the p-median facility location problem. Galvao benchmark problems from OR library was applied in the study and the results were compared with the similar studies. Combinative Distance-based Assessment (CODAS) method integrated with interval-valued intuitionistic fuzzy sets was introduced for facility location selection problems by Bolturk and Kahraman (2018). A wave energy facility location selection problem was solved with this introduced method. Rahman et al. (2018) used AHP method to select the best facility location for a company in Bangladesh. In 2019, Kheybari et al. (2019) applied the best and the worst method for the facility location selection problem of energy production in Iran. Hakli and Ortacay (2019) studied on uncapacited facility location problems and improved the scatter search algorithm. The proposed method was also applied with twenty other different methods in the literature to compare the results. Obtained comparative results show that the method improved the performance of the basic algorithm. A new hierarchical group compromise ranking methodology using hesitant fuzzy sets was introduced by Mousavi et al. (2019). The proposed method was applied to a facility location selection and the best alternative was selected by using the new ranking index, which was also introduced by the authors. Seker and Aydın (2020) proposed an interval valued Pythagorean fuzzy TOPSIS method to handle with the uncertain data. Hydrogen production facility location selection problem as a real life problem was solved by using this proposed method. An integrated fuzzy AHP and fuzzy TOPSIS method was presented by Kaul et al. (2020) for facility location evaluation. The application of the presented method was conducted on an Indian company.

As a brief review of the literature, it can be easily seen that almost all of the studies in the literature are related with MCDM methods. In addition, most of the presented MCDM methods are also different fuzzy extensions of existing MCDM methods. There are few examples using or presenting different methods. Therefore, in this study, clustering based genetic algorithm method is proposed for the solution of facility location selection problem, apart from multi-criteria decision making methods.

2. Theoretical Background

This section is allocated to explain the basis of the proposed algorithm, which is introduced in the following section. K-means clustering algorithm is explained step by step in detail in the first sub-section. Then, genetic algorithm is presented with its basic techniques in the second sub-section.

2.1. K-Means Clustering Algorithm

K–means clustering provides simple procedures to develop lexigraphic classification systems for a large sample of data (MacQueen, 1967). It starts with random centroids of a group of clusters. Then, it computes the centroids and the members of each cluster iteratively by using Euclidean distance measure. It repeats the procedures until the stopping criterion is satisfied.

The basic K–means clustering algorithm used in this study is as follows:

Step 1: Determine the number of clusters, *K*.

Step 2: Let $C_1, C_2, \ldots\ldots, C_K$ be the representation of the cluster centroids of $\{X_1, X_2, \ldots\ldots, X_N\}$ and determine the K–means cluster centroids randomly.

Step 3: Assign each element to the closest cluster by using Euclidean distance given in the Eq. (1) for M-dimension space.

$$d_{ij} = \sqrt{\sum_{k=1}^{M}(X_{ik} - X_{jk})^2} \tag{1}$$

Step 4: Re-determine the cluster centroids by using the Eq. (2).

$$C_j' = \frac{1}{n_j}\sum_{X_i \in C_j} X_i \tag{2}$$

where n_j represents the number of elements in the cluster with the centroid C_j.

Step 5: Repeat from Step 3 until the stopping criterion is satisfied.

2.2. Genetic Algorithm

Solutions based on Genetic Algorithm (GA) were firstly introduced by Holland (1975). The method models the problems inspired by genetics. Variables in the problems are defined as genes and chromosomes, and solutions are obtained by using techniques such as crossover, mutation etc. like in the genetics.

In the following paragraphs, the basic genetic operators used in this study and their definitions are given briefly.

Fitness Function (FF): FF is calculated by using the objective function, which is derived from Euclidean distance function in this study. The Objective function is also known as penalty function. Generally, in the literature, FF is equal to the objective function in maximization problems. In contrast in this study FF is used for minimizing the objective function given in Eq. (3), to obtain the minimized distance from the cluster centroids.

$$FF = \frac{1}{\sum d_j} \tag{3}$$

Mutation: Altering the random genes by using a predefined percentage value. Mutation is mostly preferred in GA–based solutions, because it eliminates trapping to the local optimums. In problems defined by binary system, mutation operator alters 0 to 1 or 1 to 0. But, if the problem defined by different systems, the mutation operator is also defined according to the problem or the solution. For example, in a shortest path problem defined by the combination of the routes, if the routes are defined as genes, the mutation operator can alter a route randomly with another one. Examples of the mutation operator are shown in the Figure 1.

1	0	0	1	1	1	0

↓

1	1	0	1	1	1	0

1	2	3	4	5	6	7	8

↓ ↓

1	2	3	7	5	6	4	8
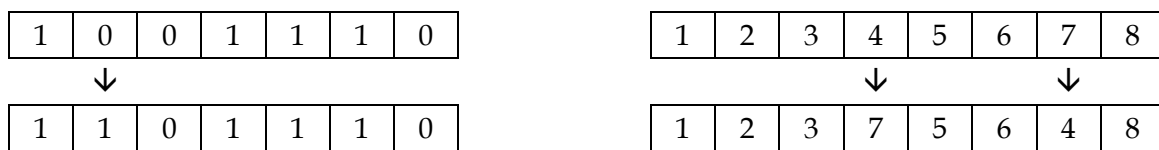
Figure 1. How the Mutation Operator Works?

Roulette Wheel Selection (RWS): RWS is used for the selection of parents to form the next generations. Two of the population is selected with the technique, and then a child solution is generated by using crossover the parent solutions. RWS technique increases the selection chance of better solutions, and worse solutions vanished faster from the population. An example of RWS possibility is given in Eq. (4) with the FF, which means that minimum value is better.

$$p_j = \frac{\frac{1}{\sum d_j}}{\sum \left(\frac{1}{\sum d_j} \right)} \tag{4}$$

where p_j is selection chance of the parent solution *j* and $\sum d_j$ is the penalty cost of parent *j*.

Crossover: Crossover techniques are used to generate child solutions from parent solutions. There are three main crossover techniques, which are one-point crossover, two-point crossover and uniform crossover, have been using in the literature. Two-point crossover is used in this study. An example for all three main crossover techniques is given in the Figure 2 respectively.
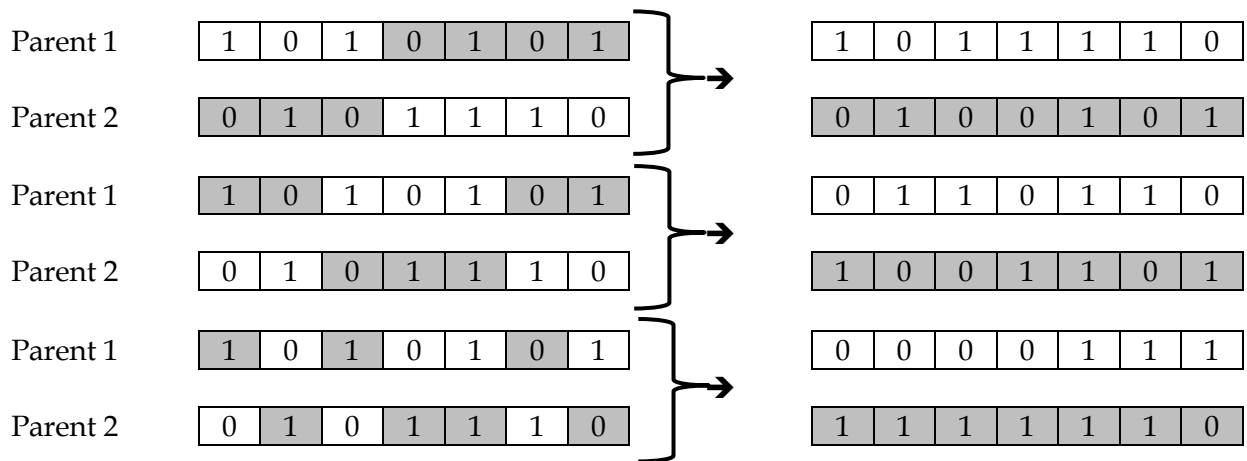
Figure 2. How the Crossover Operator Works?

3. The Proposed Algorithm

The proposed algorithm for the solution of facility location selection problem is given in this section. The proposed algorithm consists of two parts, which are K–means clustering algorithm and genetic algorithm.

The beginning of the algorithm is the determination of the number of clusters, which is also number of facilities in the handled problem. After the determination, **k**–means clustering algorithm is run and the memberships of each element for the clusters are obtained. Then, GA is executed to improve the best facility location for each cluster. Flowchart of the proposed algorithm is shown in Figure 3 and the pseudocode of the proposed algorithm is given in detail in Figure 4.

Symbols used in the proposed algorithm and the pseudocode of the proposed algorithm are given in the Table 1 with descriptions.

Table 1. List of Symbols

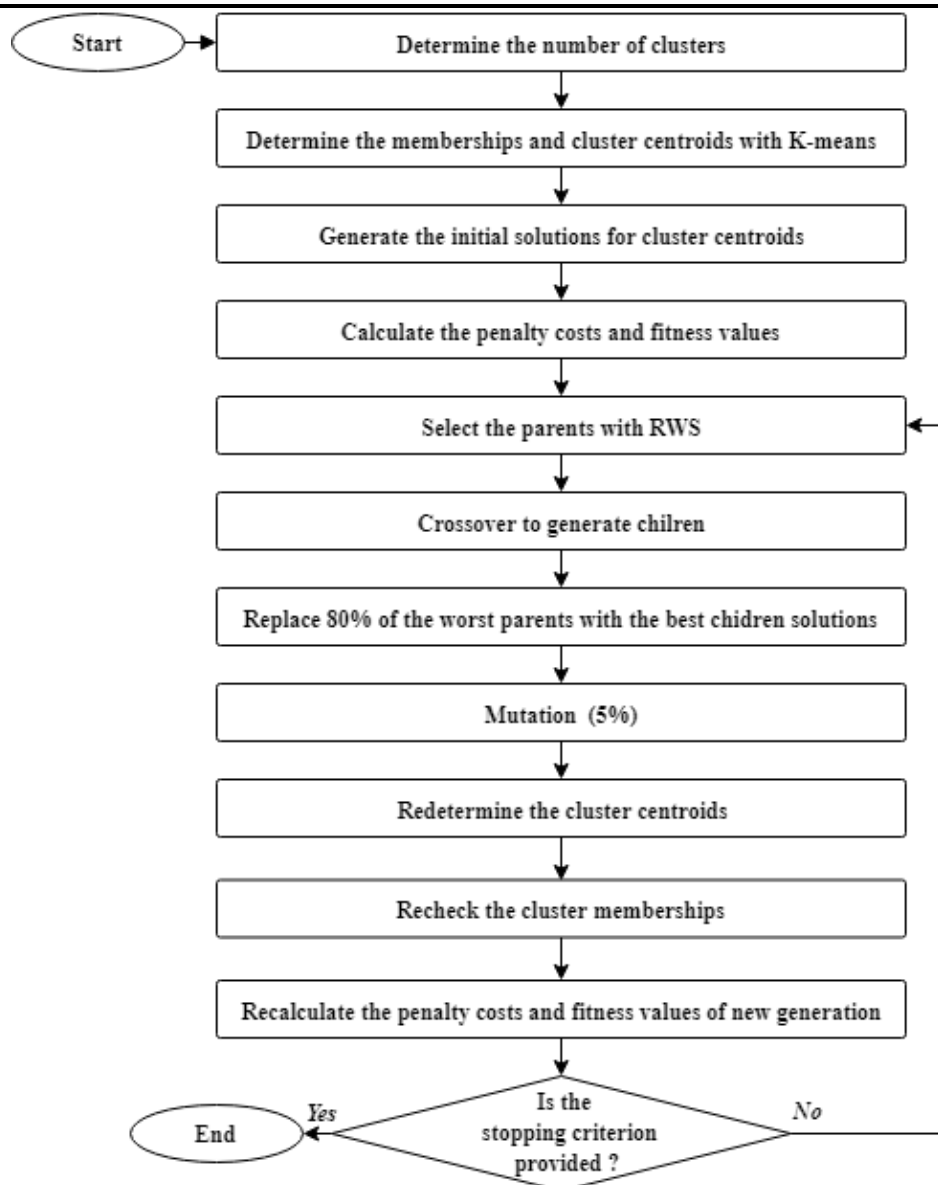|			
C_j	The centroid of cluster j	n	Number of clusters
C_j'	New centroid of cluster j	p_i	Selection possibility of parent I in *RWS*
d_{ij}	Distance between element i and element j	X_i	Coordinate vector of element i
d_j	Sum of the distances of elements in cluster j to the centroid	X_{ik}	Coordinate of the k. dimension of element i
FF	Fitness function	$\sum d_j$	Penalty cost of parent j
N	Number of elements in the problem	$\sum$	Summation

Figure 3. Flowchart of the Proposed Algorithm.

```
Determine n
Determine the Cj values
Determine the memberships of each parent for the clusters
Calculate Σ dj and FF
do {
   for (%80 of the population)
      Select the parents of the new generation with RWS
      Crossover the parents for the new generation
      Replace the worst parent solutions with the best children
solutions
      }
   Mutation with 5% for the new generation solutions
   Redetermine the Cj values
   Recheck the memberships and update
   Recalculate Σ dj and FF of the new generation
```
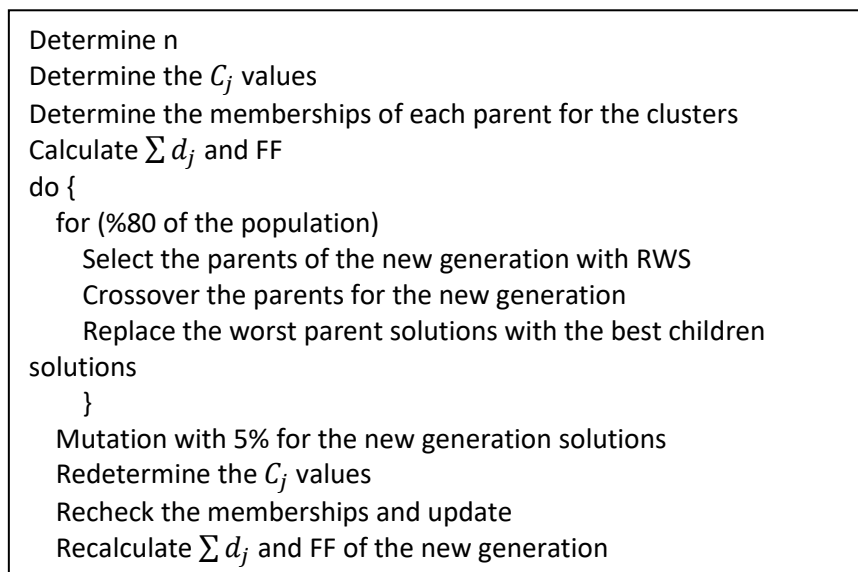
Figure 4. Pseudocode of the Proposed Algorithm.

4. Facility Location Selection

In this section, computational results of the study are given. Calculations of the proposed algorithm were implemented in Java and executed on a computer with Intel(R) Core (TM) i7–4720HQ 2.60 GHz, 16 GB RAM and windows 10 professional 64 bit operating system.

Computational results of the Ruspini75 (Ruspini, 1970; OR Library, 2019) dataset by using the proposed algorithm are given in Table 2 with the comparison of the best results in the literature. The Ruspini75 data set is a benchmark data set, and the best results are known and exact results for the comparison of the new developed algorithms. The best results known in the literature was taken from the study of Mladenovic et al. (1996).

The first column of the table 2 is the number of the facilities in the solution. The second column shows the solution time of the proposed algorithm. The third column indicates the best results known in the literature. The fourth column presents the results obtained by the proposed clustering based genetic algorithm and the last column of the table (Difference (%)) points the differences between the obtained results and the best results.

Table 2. Computational Results of Ruspini75 Data Set with the Proposed Algorithm.

Number of Facilities	Time (Sec.)	Best Known Results	Proposed Algorithm	Diff. (%)
1	1	4141.21	4141.21	0.000
5	13	779.68	783.72	0.005
10	29	512.21	513.82	0.003
20	33	314.10	316.26	0.006
30	49	199.41	199.68	0.001

Obtained results shows that the proposed algorithm works fast and effectively. All results are close to the best known results with difference less than 0,5% and solution for only one facility was obtained exactly same result in one second. At the same time, the solutions were obtained less than a minute, which is really effective as a heuristic and dynamic method. In heuristic methods, reaching the solution faster is at least as important as reaching the best solution. In this way, obtaining the results, which are close to the best known results less than 0,5% difference and solved under a minute, with the proposed algorithm is applicable and realistic.

5. Results and Discussion

As a result of the findings, less number of facility locations can be solved in a less time with the proposed algorithm. However, more time is required for larger number of facilities. If there are more facility locations, less units are connected to the selected facility locations separately. This seems like problem can be solved in less time because of less number of units. But, it also means that more facility locations and more cluster memberships have to be determined. With all of these, all problems are solved less than a minute with the proposed facility location selection algorithm.

Furthermore, Table 2 shows that obtained results for almost all of the number of facilities are so close to the best known results in the literature. All problems are solved with less than 0.01% difference, which is not a significant difference. This means that the best known results can be achieved with minor improvements to the proposed algorithm.

Conclusion

Facility location selection is one of the most important decisions of companies and industries. It is also first step to start a business process. The aim of the problem is determination of facility location in order to optimize the total distance to the facility location. In this study, facility location selection problem has two main parts; clustering of the elements for facilities and determination of the locations of the facilities.

This study focused on the implementation of clustering analysis to genetic algorithm. First part of the proposed algorithm is the determination of initial clusters and their centroids by k-means clustering algorithm. Second part is the improvement of the initial solution by genetic algorithm. It

improves both elements of the clusters and their centroids. The proposed algorithm is applied to Ruspini75 dataset from OR–library. Computational results are compared with the best results in the literature. The comparison shows that the proposed algorithm has a satisfactory performance and the results are competitive with the literature.

For further research, the proposed algorithm can be also implemented to other heuristic techniques or adapted with different operators to improve. The proposed algorithm can be implemented to various location selection problems from different fields with small modifications depending on problem types and constraints.

References

Arogundade, O. T., Akinwale, A. T., Adekoya, A. F. and Awe Oludare, G. (2005) A 0-1 model for fire and emergency service facility location selection: a case study in Nigeria, *J. Theor. Appl. Inf. Tech*, 9, 50-59.

Athawale, V. M. and Chakraborty, S. (2010, January) Facility location selection using PROMETHEE II method, *In Proceedings of the 2010 international conference on industrial engineering and operations management*, 9-10, Bangladesh Dhaka.

Basti, M. and Sevkli, M. (2015) An artificial bee colony algorithm for the p-median facility location problem, *International Journal of Metaheuristics*, 4(1), 91-113.

Bolturk, E. and Kahraman, C. (2018) Interval-valued intuitionistic fuzzy CODAS method and its application to wave energy facility location selection problem, *Journal of Intelligent & Fuzzy Systems*, 35(4), 4865-4877.

Büyüksaatçi, S. and Esnaf, Ş. (2014) Carbon Emission Based Optimisaton Approach for the Facility Location Problem, *Turkish Online Journal of Science & Technology*, 4(1), 9-20.

Chou, S. Y., Chang, Y. H. and Shen, C. Y. (2008) A fuzzy simple additive weighting system under group decision-making for facility location selection with objective/subjective attributes, *European Journal of Operational Research*, 189(1), 132-145.

Çebi, F. and Otay, İ. (2015) Multi-criteria and multi-stage facility location selection under interval type-2 fuzzy environment: a case study for a cement factory, *international Journal of computational intelligence systems*, 8(2), 330-344.

Ertuğrul, İ. and Karakaşoğlu, N. (2008) Comparison of fuzzy AHP and fuzzy TOPSIS methods for facility location selection, *The International Journal of Advanced Manufacturing Technology*, 39(7-8), 783-795.

Hakli, H. and Ortacay, Z. (2019) An improved scatter search algorithm for the uncapacitated facility location problem, *Computers & Industrial Engineering*, 135, 855-867.

Holland, J. H. (1975) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, Ann Arbor, University of Michigan Press.

Kabir, G. and Sumi, R. S. (2013) Integrating modified Delphi with fuzzy AHP for concrete production facility location selection, *International Journal of Fuzzy System Applications (IJFSA)*, 3(3), 68-81.

Kahraman, C., Ruan, D. and Doğan, I. (2003) Fuzzy group decision-making for facility location selection, *Information sciences*, 157, 135-153.

Kaul, A., Darbari, J. D. and Jha, P. C. (2020) A Fuzzy MCDM Model for Facility Location Evaluation Based on Quality of Life, *In Soft Computing for Problem Solving*, 687-697, Springer, Singapore.

Kaya, I. and Çinar, D. (2006) Facility location selection using a fuzzy outranking method, *In Applied Artificial Intelligence*, 359-366.

Kheybari, S., Kazemi, M. and Rezaei, J. (2019) Bioethanol facility location selection using best-worst method, *Applied energy*, 242, 612-623.

MacQueen, J. (1967, June) Some methods for classification and analysis of multivariate observations, *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 281-297.

Mladenovic, N., Moreno, J. A. and Moreno-Vega, J. M. (1996) Chain-interchange heuristic method, *Yugosl J Oper Res*, 6(1), 41-54.

Mousavi, S. M., Gitinavard, H., Vahdani, B. and Foroozesh, N. (2019) Hierarchical group compromise ranking methodology based on Euclidean–Hausdorff distance measure under uncertainty: An application to facility location selection problem, *Journal of Optimization in Industrial Engineering*, 12(2), 93-105.

OR Library (2019) Location Problems, Multi Source Weber Problems. http://mistic.heig-vd.ch/taillard/problemes.dir/location.html. Access Date: 15.02.2020.

Rahman, M. S., Ali, M. I., Hossain, U. and Mondal, T. K. (2018) Facility location selection for plastic manufacturing industry in Bangladesh by using AHP method, *International Journal of Research in Industrial Engineering*, 7(3), 307-319.

Ray, A., De, A. and Dan, P. K. (2015) Facility location selection using complete and partial ranking MCDM methods, *International Journal of Industrial and Systems Engineering*, 19(2), 262-276.

Ruspini, E. H. (1970) Numerical methods for fuzzy clustering, *Information Sciences*, 2(3), 319-350.

Safari, H., Faghih, A. and Fathi, M. R. (2012). Fuzzy multi-criteria decision making method for facility location selection. *African Journal of Business Management*, 6(1), 206.

Seker, S. and Aydin, N. (2020) Hydrogen production facility location selection for Black Sea using entropy based TOPSIS under IVPF environment, *International Journal of Hydrogen Energy*, doi: 10.1016/j.ijhydene.2019.12.183.

Shen, C. Y. and Yu, K. T. (2009) A generalized fuzzy approach for strategic problems: The empirical study on facility location selection of authors' management consultation client as an example, *Expert Systems with Applications*, 36(3), 4709-4716.

Singh, R. K. (2016) Facility location selection using extent fuzzy AHP, *International Advanced Research Journal in Science, Engineering and Technology*, 3(2), 47–51.

Temur, G. T., Kaya, T. and Kahraman, C. (2014) Facility location selection in reverse logistics using a type-2 fuzzy decision aid method, *In Supply chain management under fuzziness*, 591-606, Springer, Berlin, Heidelberg.

Verter, V. and Dincer, M. C. (1992) An integrated evaluation of facility location, capacity acquisition, and technology selection for designing global manufacturing strategies, *European Journal of Operational Research*, 60(1), 1-18.

Wang, S. and Watada, J. (2012) A hybrid modified PSO approach to VaR-based facility location problems with variable capacity in fuzzy random uncertainty. *Information Sciences*, 192, 3-18.

Weber, A. and Friedrich C. J. (1962) *Theory of the location of industries*, Chicago, Illinois: The University of Chicago Press.