IJE TRANSACTIONS C: Aspects Vol. 26, No. 6, (June 2013) 597-604

International Journal of Engineering

Journal Homepage: www.ije.ir

RESEARCH NOTE

A Clustering Based Location-allocation Problem Considering Transportation Costs and Statistical Properties

M. Mohammadkhanloo, M. Bashiri *

Department of Industrial Engineering, University of Shahed, Tehran, Iran

PAPER INFO

Paper history:
Received 10 August 2012
Received in revised form 03 November 2012
Accepted 24 January 2013

Keywords:
Multivariate Clustering Analysis
Location Allocation Problem
Simulated Annealing Algorithm
SA K-means

ABSTRACT

Cluster analysis is a useful technique in multivariate statistical analysis. Different types of hierarchical cluster analysis and K-means have been used for data analysis in previous studies. However, the K-means algorithm can be improved using some metaheuristics algorithms. In this study, we propose simulated annealing based algorithm for K-means in the clustering analysis which we refer it as SA K-means. In this algorithm, an evaluation criterion is used in the clustering stage to have accurate clusters. Then, another cost based criterion has been introduced to have efficient and accurate clusters. The proposed approach has been presented for solving the location allocation problem. To show the effectiveness of the proposed approach, some numerical examples of location allocation problems have been tested by the proposed approach. Comparing the results of the proposed approach with exact solution and another developed GA algorithm for numerical examples of the location allocation problem show that the performance of the proposed SA K-means approach is satisfactory.

doi: *10.5829/idosi.ije.2013.26.06c.05*

1. INTRODUCTION

Given the growing nature of data in the daily business environment, the analysis and implementation of data seems to become a basic task of management. Data mining is a useful and efficient process to analyze such data and clustering is a popular data analysis technique. The term "clustering" is used for different data type can be applied in a wide range of problem especially when the subject of study is about multi facility location with allocation decisions. Although they seem to be as separate problems, both the content and solution styles are the same. Clustering procedures partitionize a set of objects into clusters such that objects in the same cluster are more similar to each other than objects in different clusters according to some predefined criteria. In the location-allocation problem, the clustering method is used in the allocation step of customers to the located facilities in the problem. The problem can be generalized as follows: given a set n data points in d dimensional space, we must determine how to assign n points, called demand nodes to a set C of c points, called cluster centers so as to optimize the problem based on some criteria. In most of cases, it is assumed that n is much greater than c and d is relatively small.

This formulation is an example of unsupervised learning. The system will create groupings based on a new criterion and the information contained in the n data points. K-means is a popular algorithm that was first presented in three decades ago. The mentioned criterion minimizes the total mean-squared distance of each demand node from its cluster center. It is the process of grouping similar data items together based on some measure of similarity (or dissimilarity) between items.

Clustering is a component of exploratory data analysis and is useful for generating hypotheses about data. The standard clustering process consists of the following steps: data preparation and attribute selection, similarity measure selection, algorithm and parameter selection, cluster analysis, and validation.

The performance of K-means algorithm is very sensitive to the initial solution [1, 2] which makes algorithm to fall into the trap of local optimum in most cases. In addition, it has been proven that the metaheuristic algorithms have very good performance in the location problems [3]. Therefore, we used SA algorithm as a solution approach in this paper.

*Corresponding Author Email: *bashiri.m@gmail.com* (M. Bashiri)

TABLE 1. Review of some papers in location allocation problem.

Authors	Problem Titles	Clustering Method (Technique for solving)	Metaheuristic Algorithms
Hsieh and Tien [4]	Location-Allocation	Self organizing maps	-
Bischoff and Dachert [5]	Location -Allocation	K-means	-
Murat et al. [6]	Location–Allocation problem with dense demand	K-means with heuristic algorithm	-
Correa et al. [7]	Maximal covering location-allocation problem	K-means	-
Venkata Reddy Muppani (Muppant) et al.[8]	Warehouse storage location	-	Simulated annealing Algorithm(SA)
Marvin A.Arostegui Jr et al.[3]	Multi facilities problems	-	Simulated Annealing, Tabu search & Genetic Algorithms
Vincent F. Yu et al. [9]	Capacitated location routing problem	-	Simulated Annealing
Current research	Location-Allocation problem	K-means	Simulated annealing algorithm (SA)

The review of some related studies about the location allocation problem has been reported in Table 1. It is shown that the cluster analysis approach has not been considered with metahuristics for location allocation problem before. Thus, in this study we propose the hybrid simulated annealing and k-means clustering for the location allocation problem.

The contribution of this paper can be devided in two items. The first one is developing a new criterion for location allocation problem solution approach based on the clustering method. The other one is developing a hybrid Simulated Annealing and K-means clustering method for efficient solving the location allocation problem.

The remainder of this paper has been organized as following: In the next section, the proposed hybrid algorithm based on the simulated annealing for multivariate clustering analysis (SAK-means) will be discussed. The proposed algorithm has been tested by several hypothetical numerical examples which have been reported in section 3. Finally, the conclusion has been mentioned in the last section.

2. PROPOSED ALGORITHM (SAK-means)

Location-Allocation (LA) problem is to locate a set of new facilities so the transportation cost from facilities to customers is minimized and an optimal number of facilities have to be placed in an area of interest in order to satisfy the customer demand. This problem occurs in many practical settings where facilities provide homogeneous services such as the determination and location of warehouses, distribution centers, communication centers and production facilities. The basic decisions of location allocation problems consist of determine the number of facilities, their locations and finally allocation of customers to facilities.

In this paper, the following model for location-allocation problem has been studied.

variables:

y_i is the location binary variable and is one if the facility is located at node *i*.

Z_{ij} is the allocation binary variable and is 1 if customer of node *j* is assigned to facility located in node *i*.

parameters:

$d(x_i, a_j)$: is the distance between the nodes of *i* and *j*

r_{ij} : is the flow between nodes of *i* and *j*

fc_i : is the fixed establishment cost of the facility in node *i*

The objective and constraints are as follows:

$$\min TC = \sum_{i=1}^{c}\sum_{j=1}^{n} Z_{ij} r_{ij} d(x_i, a_j) + \sum_{i=1}^{c} fc_i y_i \tag{1}$$

Subject to:

$$\sum_{i=1}^{c} Z_{ij} = 1 \qquad j = 1, 2, \ldots, n \tag{2}$$

$$Z_{ij} \in \{0,1\} \qquad \begin{cases} i = 1, 2, \ldots, c \\ j = 1, 2, \ldots, n \end{cases} \tag{3}$$

In this model, Equation (1) minimizes the total fixed and transportation cost between customers and facilities while fc_i is the fixed cost such as opening and establishment cost of facility *i*. Equation (2) ensures that all customer demands to be satisfied. Variables and parameters used in this model are mentioned in Table 2. Equation (3) defines the variables type.

As mentioned before, we use the clustering analysis as a tool for allocation decision determination. This concept can reduce number of decisions variables in our problem and consequently can decrease the computational time. To better illustration of clustering analysis, Figure 1 shows an example for clustering problem with c=3 (number of clusters or facilities) and some of customers have been served by facilities. In Figure 1, the red triangle symbol shows the located facility and the blue diamond symbol shows the customers. It is obvious that there are 3 clusters, so each customer will be allocated to the facility in its related cluster.

TABLE 2. Variables and parameters that used for the location allocation model.

TC	Total cost for transportation the services between facilities and customer
c	Number of facilities (clusters)
n	number of customer
Z_{ij}	1 if a customer j is assigned to a new facility i ;0 otherwise
r_{ij}	The cost of servicing between a customer j to facility i
x_i	The coordinates of the new facilities, $i=1, 2,\ldots, c$
a_j	the coordinates of the customers, $j=1, 2,\ldots, n$
$d(x_i,a_j)$	The distance between customer j and new facility i
fc_i	Fix cost to build any facilities
y_i	Whether facility for cluster i is established ($yi=1$) or not $yi=0$

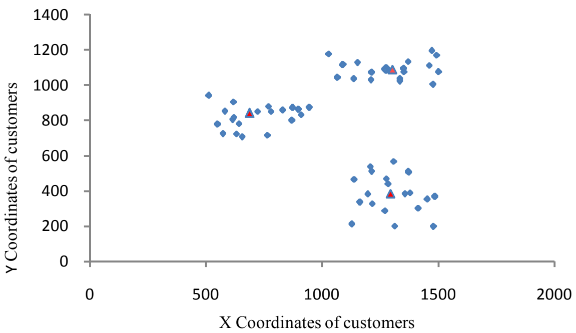

Figure 1. An example for location allocation problem

As it was seen in previous sections, one of the most important problems in center-based clustering is convergence to the local optimum. For combinatorial optimization problems there are some metaheuristic methods to find the global optimum point instead of local one. Tabu search [10], simulated annealing [11], evolutionary methods, swarm optimization methods are examples of these mataheuristic methods. In order to overcome the local optima problem lots of studies have been done in clustering and vector quantization literature. Steinley declares that the popular K-means clustering methods generally provide solutions that are only locally optimal for a given set of data [12]. Hansen and Mladenovic [13] proposed a method, J-means, based on local search heuristic.

ZulalGungor and AlperUnler [14] proposed K-harmonic means data clustering with simulated annealing metaheuristic. Frnti et al. [15, 16] proposed randomized local search, Cano et al. [17] proposed greedy randomized adaptive search procedure (GRASP) and Kanade and Hall [18] proposed an ant colony optimization method to solve local optima problem in clustering. Some authors [15, 19-22] used Tabu search method, while some others [23, 24] using simulated annealing, and some other [25, 26] using hybrids like genetic simulated annealing or Tabu search with simulated annealing, to overcome this problem. The main idea of their algorithms is to use some heuristics and metaheuristics (Tabu search, simulated annealing, random search etc.) to generate non-local moves for the cluster centers and to select the suitable best solution. All of the algorithms mentioned above use either K-means for the clustering problems.

In the clustering analysis usually the validation criterion is the ratio of between clusters sum of square and each cluster sum of square. It shows that, if we find a solution with similar observations in the cluster and different cluster from each other, the criterion should be as large as possible.

Mentioned criterion can be computed according to the following equation:

$$\lambda = SS_b / SS_w \tag{4}$$

where, SS_b and SS_w are between and within the clusters sum of squares, respectively. In this paper, we have used a modified the following criterion for our proposed the clustering algorithm:

$$V = (\lambda/\lambda_{max})/(OFV/OFV_{max}) \tag{5}$$

We know that:

$$0 \le \lambda \le 1 \tag{6}$$

So:

$$\lambda_{max} = 1 \tag{7}$$

where, OFV is the objective function value of the current solution for the location allocation problem. In Equation (5), we used λ criterion because the accuracy of clustering is important as well. Moreover, total transportation cost should be considered. However, in the classic clustering algorithm the statistical aspect is only considered. Nevertheless, in our proposed algorithm we need to cluster customers with higher similarity and lower total transportation cost. Therefore, the objective function is also involved in the clustering criterion. In the proposed searching algorithm, the clustering solution can be accepted as a best solution when it returns the maximum value of λ (high statistical accuracy) and minimum OFV (lower total location and allocation cost). In mentioned integrated criterion, the statistical aspect (λ value) and operational aspect (OFV) can be integrated after the normalization as described in Equation (5). The classic clustering algorithm consumes high computational time, so it's composition to simulated annealing can increase the efficiency of proposed solution approach. The resean is that the simulated annealing has a diverse search characteristic and the less quality solutions are accepted with a probability. In another word, inspite of classic *k*-means a metaheuristic based *k*-means is presented in this paper. Moreover, according to the proposed solution algorithm, the solutions are moving toward best clustering and lower location and allocation costs simultaneously during the clustering.

Simulated annealing (SA) algorithm needs an initial solution to start. In this study, we can give initial solution in two forms to the SA algorithm. The first one is the final K-means algorithm solution and the other is the random generated solution. If random initial solution is used for the clustering in SA, the initial λ value will be assumed to be zero. After the initial clustering, the proposed algorithm calculates the objective function value (OFV) of the current cluster. It is also considered as the maximum objective function value (OFV_{max}) as an upper bound for all iterations of the algorithm, the algorithm begins to change the customers allocation to another cluster. If new V value was better than previous V values, this value is new best V value. But if new V value is smaller than the previous one, so it is accepted by a computed probability. Therefore, SAK-means is not convergence to local optima. Finally, λ and objective function values and optimum clusters can be reported according to the final V value. Since in this study, we described V value for the clustering validation, therefore, if V value is increased, we can conclude that the clustering has appropriate accuracy with lower total cost. The proposed SAK-means algorithm structure has been illustrated in Figure 2. It should be mentioned that the OFV_{max} is an upper bound of the objective function and the algorithm is based on the new presented criterion (V).

For more illustration, a pseudo code of the proposed algorithm has been depicted in Figure 3. More analysis on the λ and objective function values (OFV) have been considered in the numerical analysis section. All the above formulas and model have been coded in MATLAB 7.8.0(R 2009a). System specifications that we have used for this paper are as follows: RAM: 2GB DDR3, CPU: Intel Pentium inside (dual core).

3. PARAMETER SETTING

To tune the proposed algorithm (SAK-means), extensive experiments are conducted with different sets of parameters. In other words, these parameters are set by conducting a comprehensive sensitivity analysis. In this paper, we have used the Taguchi method in MINITAB14 software to find the sensitive parameters in the proposed algorithm. In the taguchi designed experiments 4 factors with three levels were analyzed by a *L9* design. Results showed that the algorithm is sensitive to the initial temperature and Markov chains values. We set the following parameters for each group of testing data (n=500, 200, 100, 30) whit c=3 (number of clusters or facilities). The tuned and selected parameters for the solution algorithm have been reported in Table 3.

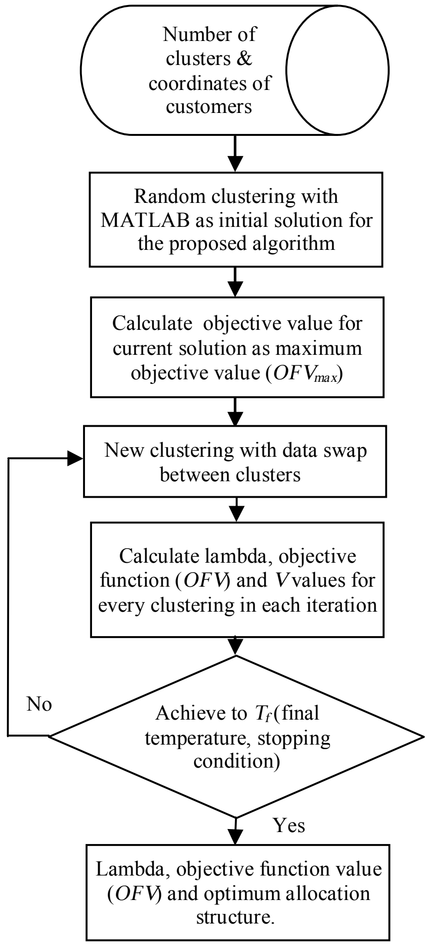

Figure 2. Main structure of the proposed SAK-means algorithm.

TABLE 3. Parameters in SAK-means.

Parameter	Description	n=200, 100, 30	n=500
T_0	Initial temperature	15000	1000
Max_nover	Markov chain value (The number of algorithm's iteration in each temperature)	400	270
Max_nlimit	The number of accepting the establishment vector	400	265
R	Reducing the temperature coefficient	0.99	0.99

Begin

- Input c (number of clusters or facilities) and matrix x (customers coordinates)
- Choose c as initial centers { Final solution of K-means / Randomly }
- Set $T=T_0$
- Set T_f
- Set cooling schema (r) and max_nover and max_nlimit
- Calculate OFV_{max}

Repeat:

Repeat:

Change the allocation of observation in to the clusters (Neighborhood solution)

Calculate the clustering index (V-value)

If $\Delta V>0$ *then swap old and new solution*

Else

Generate a random number (s)

If $e^{-\Delta V/T}>s$ then swap old and new solution

End if

End if

Until maximum iteration

- $T=r*T$

Until ($T=T0$) stopping condition

End

Figure 3. Pseudo code for the proposed algorithm.

TABLE 4. λ and objective function values for numerical examples

					Cost Objective Function Values (OFV)		
Problem	***n***	***c***	***n/c***	***λ-value(obtained from SAK-means)***	**SAK-means**	**LINGO8**	**GA**
#1	500	10	50	0.0621	124310	173424.6	235400
#2	500	5	100	0.0194	221180	261909.3	272960
#3	500	3	166.67	0.0022	334630	353863.4	355190
#4	200	10	20	0.4244	61840	75020.49	86270
#5	200	5	40	0.2663	80727	103261.5	105750
#6	200	3	66.67	0.0248	138440	177742.1	138480
#7	100	10	10	0.9558	38410	40714.77	39297
#8	100	5	20	0.1624	53529	87518.62	56227
#9	100	3	33.33	0.0460	70534	87318.62	76353
#10	30	10	3	0.2558	10453	26733.03	11592
#11	30	5	6	0.2351	13966	35723.76	14419
#12	30	3	10	0.2023	18453	35523.76	19443

4. NUMERICAL ANALYSIS

In this study, we used V value to consider the statistical and operational aspects simultaneously, given the Table 4. It can be seen that the proposed algorithm has been tried on several numerical examples. In the numerical examples, we have considered 4 groups of customers with numbers 500, 200, 100 and 30 customers that we want to cluster them into 3 cluster numbers of c (10, 5, 3). For considering the efficiency of the proposed algorithm, in location-allocation problem, the problem was solved by an optimization software (LINGO 8.0) for each numerical example. Because of the numerical examples size, most of reported results by LINGO 8.0 were as local optimum solution and the reported results is the best found results by the Lingo software in a predefined computational time. Morever, the results of the proposed algorithm for each numerical examples has been reported in Table 4, the comparison of 2 approaches considering the objective functiion value (*OFV*) show that in all cases, the proposed algorithm can find better solutions in both aspects of statistical and operational.

Figure 5 illustrates mentioned comparison more clearly, so that objective function values obtained from SAK-means are better than *OFV* obtained from LINGO 8.0. so, we can conclude that SAK-means algorithm has better performance than an exact solution approach.

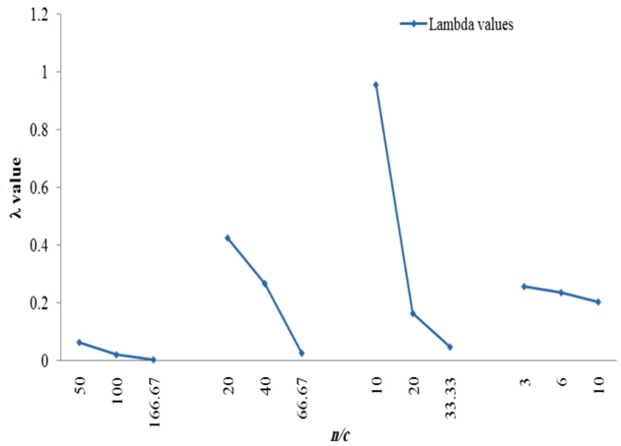

Figure 4. Graph of relation between *n/c*, *λ* in proposed algorithm.

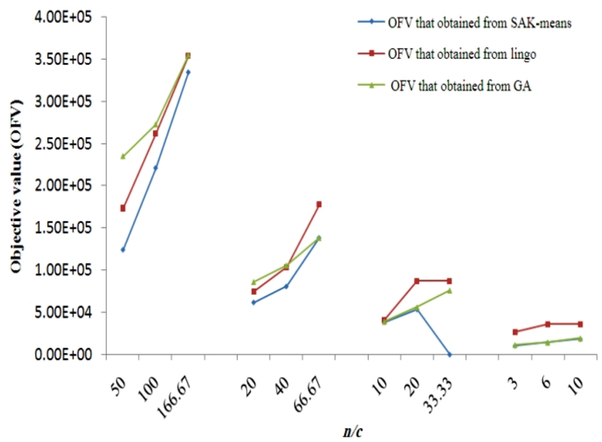

Figure 5. graph of relation between *n/c*, objective function value (*OFV*) in proposed algorithm.

5. CONCLUSION

Clustering is an efficient way to reach the information from raw data and K-means is a basic method for it. Although it is easy to implement and understand, K-means has serious drawbacks. Over the past 30 years, many papers have presented ways to improve K-means. Some focus on creating good initialization methods, while others look for finding optimal values of *K*, and some others to find globally optimal solutions. In this study, we have proposed a hybrid K-means data clustering algorithm with simulated annealing metaheuristic refered as SAK-means. The algorithm has been implemented and tested on several numerical examples in location allocation problem. We have presented with the experiments that our algorithm (SAK-means) outperforms to K-means considering the statistical and operational aspects. On the other hand, SAK-means has appropriate accuracy and lower cost than the exact solution obtained by LINGO 8.0. As a future research considering of K-means data clustering with other metaheuristic based search methods like swarm optimization, Imperialist Competitive Algorithm and etc. is recommended.

6. REFERENCES

1. Bradley, P. S. and Fayyad, U. M., "Refining initial points for k-means clustering", in Proceedings of the fifteenth international conference on machine learning, San Francisco, CA, USA. Vol. 66, (1998).
2. Fahim, A., Salem, A., Torkey, F., Saake, G. and Ramadan, M., "An efficient k-means with good initial starting points", ***Georgian Electronic Scientific Journal: Computer Science and Telecommunications***, Vol. 2, No. 19, (2009), 47-57.
3. Arostegui Jr, M. A., Kadipasaoglu, S. N. and Khumawala, B. M., "An empirical comparison of tabu search, simulated annealing, and genetic algorithms for facilities location problems", ***International Journal of Production Economics***, Vol. 103, No. 2, (2006), 742-754.
4. Hsieh, K.-H. and Tien, F.-C., "Self-organizing feature maps for solving location–allocation problems with rectilinear distances", ***Computers & Operations Research***, Vol. 31, No. 7, (2004), 1017-1031.
5. Bischoff, M. and Dächert, K., "Allocation search methods for a generalized class of location–allocation problems", ***European Journal of Operational Research***, Vol. 192, No. 3, (2009), 793-807.
6. Murat, A., Verter, V. and Laporte, G., "A multi-dimensional shooting algorithm for the two-facility location–allocation problem with dense demand", ***Computers & Operations Research***, Vol. 38, No. 2, (2011), 450-463.
7. Correa, F. d. A., Lorena, L. A. N. and Ribeiro, G. M., "A decomposition approach for the probabilistic maximal covering location-allocation problem", ***Computers & Operations Research***, Vol. 36, No. 10, (2009), 2729-2739.
8. Adil, G. K., "Efficient formation of storage classes for warehouse storage location assignment: A simulated annealing approach", ***Omega***, Vol. 36, No. 4, (2008), 609-618.
9. Yu, V. F., Lin, S.-W., Lee, W. and Ting, C.-J., "A simulated annealing heuristic for the capacitated location routing problem", ***Computers & Industrial Engineering***, Vol. 58, No. 2, (2010), 288-299.
10. Glover, F. and Laguna, M., "Tabu search", Springer, Vol. 22, (1997).
11. Kirkpatrick, S., Jr., D. G. and Vecchi, M. P., "Optimization by simulated annealing", ***Science***, Vol. 220, No. 4598, (1983), 671-680.
12. Steinley, D., "Local optima in k-means clustering: What you don't know may hurt you", ***Psychological Methods***, Vol. 8, No. 3, (2003), 294.
13. Hansen, P. and Mladenovic, N., "J-means: A new local search heuristic for minimum sum of squares clustering", ***Pattern Recognition***, Vol. 34, No. 2, (2001), 405-413.
14. Gungor, Z. and Unler, A., " K-harmonic means data clustering with simulated annealing heuristic", ***Applied Mathematics and Computation***, Vol. 184, No. 2, (2007), 199-209.
15. Franti, P., Kivijarvi, J. and Nevalainen, O., "Tabu search algorithm for codebook generation in vector quantization", ***Pattern Recognition***, Vol. 31, No. 8, (1998), 1139-1148.

16. Franti, P. and Kivijarvi, J., "Randomised local search algorithm for the clustering problem", ***Pattern Analysis & Applications***, Vol. 3, No. 4, (2000), 358-369.

17. Cano, J., Cordon, O., Herrera, F. and Sánchez, L., "A greedy randomized adaptive search procedure applied to the clustering problem as an initialization process using k-means as a local search procedure", ***Journal of Intelligent and Fuzzy Systems***, Vol. 12, No. 3, (2002), 235-242.

18. Kanade, P. M. and Hall, L. O., "Fuzzy ants as a clustering concept", in Fuzzy Information Processing Society, NAFIPS 2003. 22nd International Conference of the North American, IEEE. Vol., No., (2003), 227-232.

19. Al-Sultan, K. S., "A tabu search approach to the clustering problem", ***Pattern Recognition***, Vol. 28, No. 9, (1995), 1443-1451.

20. Al-Sultan, K. S. and Fedjki, C. A., "A tabu search-based algorithm for the fuzzy clustering problem", ***Pattern Recognition***, Vol. 30, No. 12, (1997), 2023-2030.

21. Delgado, M., Skarmeta, A. G. and Barbera, H. M., "A tabu search approach to the fuzzy clustering problem", in Fuzzy Systems, Proceedings of the Sixth IEEE International Conference on, IEEE. Vol. 1, No., (1997), 125-130.

22. Sung, C. and Jin, H., "A tabu-search-based heuristic for clustering", ***Pattern Recognition***, Vol. 33, No. 5, (2000), 849-858.

23. Kovesi, B., Boucher, J.-M. and Saoudi, S., "Stochastic k-means algorithm for vector quantization", ***Pattern Recognition Letters***, Vol. 22, No. 6, (2001), 603-610.

24. Wang, X.-Y. and Garibaldi, J. M., "Simulated annealing fuzzy clustering in cancer diagnosis", ***Informatica***, Vol. 29, No. 1, (2005), 61-70.

25. Chu, S.-C., Roddick, J. F. and Australia, A., "A clustering algorithm using the tabu search approach with simulated annealing", in In Data Mining II-Proceedings of Second International Conference on Data Mining Methods and Databases, Citeseer, (2000).

26. Huang, H.-C., Pan, J.-S., Lu, Z.-M., Sun, S.-H. and Hang, H.-M., "Vector quantization based on genetic simulated annealing", ***Signal Processing***, Vol. 81, No. 7, (2001), 1513-1523.

A Clustering Based Location-allocation Problem Considering Transportation Costs and Statistical Properties

RESEARCH NOTE

M. Mohammadkhanloo, M. Bashiri

Department of Industrial Engineering, University of Shahed, Tehran, Iran

PAPER INFO

Paper history:
Received 10 August 2012
Received in revised form 03 November 2012
Accepted 24 January 2013

Keywords:
Multivariate Clustering Analysis
Location Allocation Problem
Simulated Annealing Algorithm
SA K-means

چکیده

تجزیه و تحلیل خوشه‌ای، از تکنیک‌های مفید در تجزیه و تحلیل آماری چند متغیره می‌باشد. در مطالعات گذشته از انواع مختلف خوشه بندی سلسله مراتبی و الگوریتم خوشه بندی K-means استفاده شده است. بنابراین، عملکرد الگوریتم خوشه بندی K-means را می‌توان با استفاده از الگوریتم های فرا ابتکاری بهبود داد. در این مطالعه، ما الگوریتم شبیه سازی تبرید مبتنی بر خوشه بندی K-means را پیشنهاد دادیم و الگوریتم پیشنهادی را SAK-means نامیدیم. در این الگوریتم، یک معیار ارزیابی در مرحله خوشه بندی استفاده شده است که این معیار برای ارزیابی دقت خوشه ها می‌باشد. سپس، هزینه خوشه بندی مبتنی بر معیار تعریف شده برای خوشه بندی بدست می آید. الگوریتم پیشنهادی برای حل مسأله مکان یابی–تخصیص بکار برده می‌شود. برای نشان دادن عملکرد و تأثیر الگوریتم پیشنهادی، تعدادی مثال عددی که در حوزه مسأله مکان یابی–تخصیص هستند به وسیله الگوریتم پیشنهادی حل شدند. مقایسه نتایج حاصل از الگوریتم پیشنهادی با جواب های دقیق و جوابهای حاصل از الگوریتم ژنتیک برای همان مثال ها نشان داد که الگوریتم پیشنهادی از عملکرد قابل قبولی برخوردار است.

doi: *10.5829/idosi.ije.2013.26.06c.05*