

A heuristic approach to handle capacitated facility location problem evaluated using clustering internal evaluation

G R Sutanto^{1,2*}, S Kim¹, D Kim¹, H Sutanto³

¹Department of Industrial and Management Engineering, Myongji University, Republic of Korea

²Department of Industrial Engineering, University of Buddhi Dharma, Indonesia

³Department of Manufacturing Engineering, Atma Jaya Catholic University, Indonesia

Corresponding author: gerardrandi@yahoo.com

Abstract. One of the problems in dealing with capacitated facility location problem (CFLP) is occurred because of the difference between the capacity numbers of facilities and the number of customers that needs to be served. A facility with small capacity may result in uncovered customers. These customers need to be re-allocated to another facility that still has available capacity. Therefore, an approach is proposed to handle CFLP by using *k*-means clustering algorithm to handle customers' allocation. And then, if customers' re-allocation is needed, is decided by the overall average distance between customers and the facilities. This new approach is benchmarked to the existing approach by Liao and Guo which also use *k*-means clustering algorithm as a base idea to decide the facilities location and customers' allocation. Both of these approaches are benchmarked by using three clustering evaluation methods with connectedness, compactness, and separations factors.

1. Introduction

The facility location problem (FLP) is a research topic as a branch of Operations Research and Computational Geometry, which has been a research topic for a long time. One of the main issues in this topic is about space equity and capacity limitation of suppliers (Nauss, 1978; Chudak and Williamson, 2005; Liao and Guo, 2008). In FLP, facilities need to be placed to cover a certain number of customers. Finding the most optimal placements for facilities is the main issues in this topic. Various approaches from various fields of study had been invented to find the most optimal FLP solution. Most common approaches are from Cluster Analysis (Clustering). One example of clustering application to solve FLP is the solution to solve Logistics Center Location Problem (Qingfang Ruan et al., 2010). However, many approaches are invented to solve FLP with stochastic environment or constraint. One example is FLP under small sample or no-sample cases (Wena et al., 2015). Or another FLP with capacity limitation as constraint, which is defined as Capacitated Facility Location Problem (CFLP). CFLP is a branch of Facility Location Problem (FLP), which is concerned with the optimal placement location of each facility, to cover a certain number of customers. In addition, each facility also has limited number of capacity as a constraint.

To solve CFLP, results must cover both the location and the coverage of each facility. Allocating and re-allocating customers is one way to solve the CFLP, because the existing facilities are used into their full extent, without necessities to make a large amount of investment. One of the existing approaches using customers' allocation and re-allocation to solve CFLP is based on the *k*-



means clustering algorithm (Liao and Guo, 2008). The k -means clustering algorithm is an algorithm which aims to partition n objects into k clusters. In Liao's Approach clusters are the coverage area of the facilities, with each cluster centers act as the facility location.

In this research, another approach is proposed to solve CFLP. This approach also use k -mean clustering algorithm to allocate customers into their designated facilities. However, if customers' re-allocation is needed, it'll be done based on the average distance between customers and available facilities. This approach is benchmarked to Liao's Approach, using clustering evaluation method. Without any benchmark dataset to be compared with, the only suitable evaluation method is the Internal Evaluation.

An evaluation package for R language is done by taking Internal Evaluation as one of its evaluation method (Brock et al., 2008). Other factors which selected for the clustering Internal Evaluation are compactness, connectedness, and separations. Therefore, in this research, Liao's approach is evaluated with evaluation methods based on those factors.

2. Liao's Approach

This existing approach is based on paper "A Clustering-Based Approach to the Capacitated Facility Location Problem" published in 2008 by Liao and Guo. This particular approach is adapted from k -means clustering algorithm (figure 1). Similar with k -means clustering algorithm, centroids (facility locations), are randomly initialized. Capacity constraints are considered when allocating customers into facility. Customers are allocated into facility based on their distance. These steps will be repeated until the new facility location is same as the previous facility location.

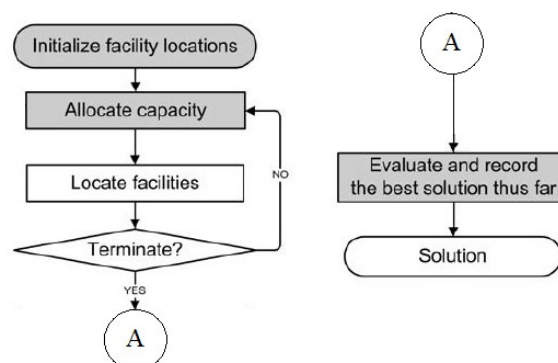


Figure 1. Flowchart for Liao's Approach

In the Initialize Facility Location step, facility locations are initialized by generating random value of x and y coordinates. Total number of facility locations is also given in this step. Next, in the Allocate Capacity step, a capacity constraint is given to each facility. In this research, all facilities have an equal number of capacities. Then, locations for each facility are determined in the Locate Facilities step. Locations for each facility are located by calculating the means value of all customers which each facility covered. Facilities only covering a certain number of customers based on their capacity, and priority is given to the closest customer from their locations. Those processes will always be repeated until the locations for each facility are same with their locations from previous iteration. The solution is the facility locations after the whole steps are terminated.

3. Proposed Approach

This proposed approach also use k -means clustering algorithm to allocate customers into facility. In addition, k -means also used to determine the facility locations. However, if there is one or more facility serving more customers than its capacity, re-allocation will happen. Re-allocation will be done only for the un-served customers, based on average distance method between the customers and the available facilities.

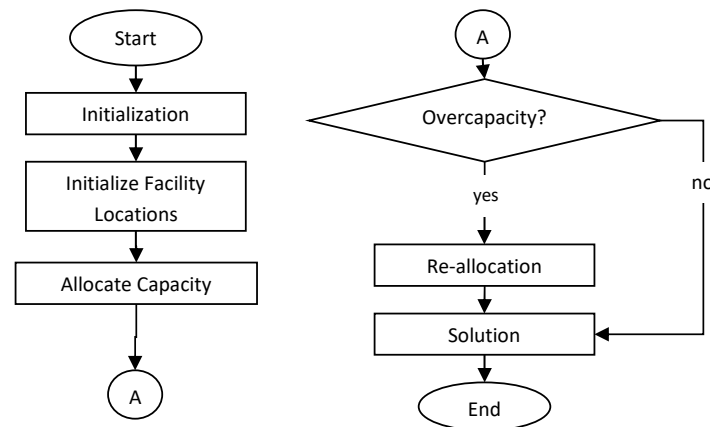


Figure 2. Flowchart for the proposed approach

Initialization is the first step to assign the desired number of facility location, and its capacity (figure 2). Capacity should be uniform for each facility. This step also read customer locations data, and put it into a matrix. Customer locations should be consists of x and y coordinates. In Initialize Facility Location step, k -means Clustering Algorithm is used to determine facility location. The k -means Clustering Algorithm will run to process the customer locations data, with desired number of facility as a constraint. The centroids from k -means Clustering Algorithm's result are the initial facility locations, as can be seen on figure 3.

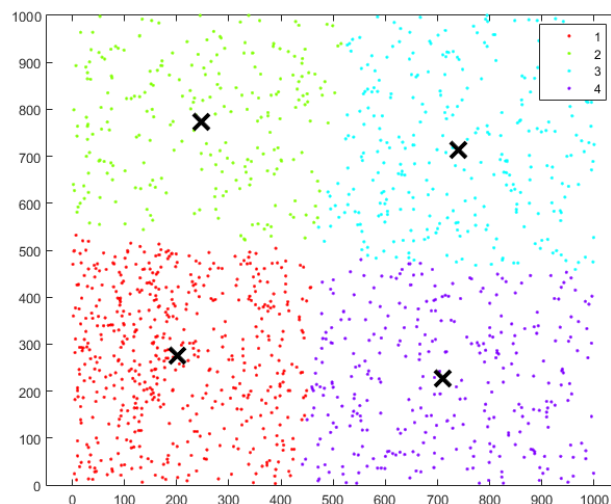


Figure 3. An example if a company decided to build 4 facilities ($k=4$)

Then, for each facility location, capacity constraints are applied. After applying the capacity constraint, each facility needs to be checked for overcapacity. If there's any facility containing more than its capacity, some customers from that particular capacity will be omitted (in figure 4). These customers need to be designated into another facility which still has available capacity. Omitted customers will be re-allocated based on average distance of its neighboring customers within a certain range.

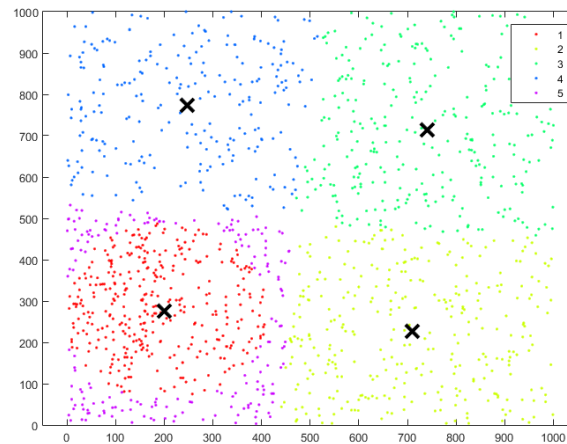


Figure 4. An example of omitted customers (purple points), from the initial data in Figure 3

3.1. Re-allocation step. Priorities will be given for each un-served customer. Priorities will be based on distance between every un-served customer with another customer in a certain range. How wide the range for each un-served customer is decided using Elbow Method. Ranges are the k , and average distances of all customers in that particular range are the cost functions. The value of range in “elbow” position of the graph would be used for deciding priorities. In a circle where the un-served customer act as the center and the value of range act as the borderline, every customer inside the circle will act as the deciding factor for re-allocation.

For example, for re-allocating customer X, a certain k range is selected by using Elbow Method, and containing another customer A, B, and C. Customer A and B are served by Facility 1 with average distance to customer X is 10. C is served by Facility 2 with average distance to customer X is 20. So customer X would be served by Facility 1, because the average distance value is the smallest one compared to another average distance value in that particular range. This example is shown in figure 5, with A and B are the triangles which means from the same facility (Facility 1), and C is square which means C is from another facility (Facility 2). In addition, re-allocation only will happen if the designated facility still have empty slot of capacity.

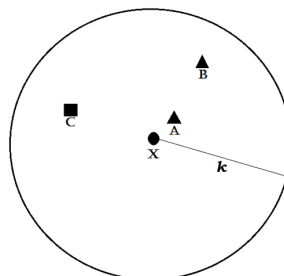


Figure 5. An example of re-allocation based on average distance

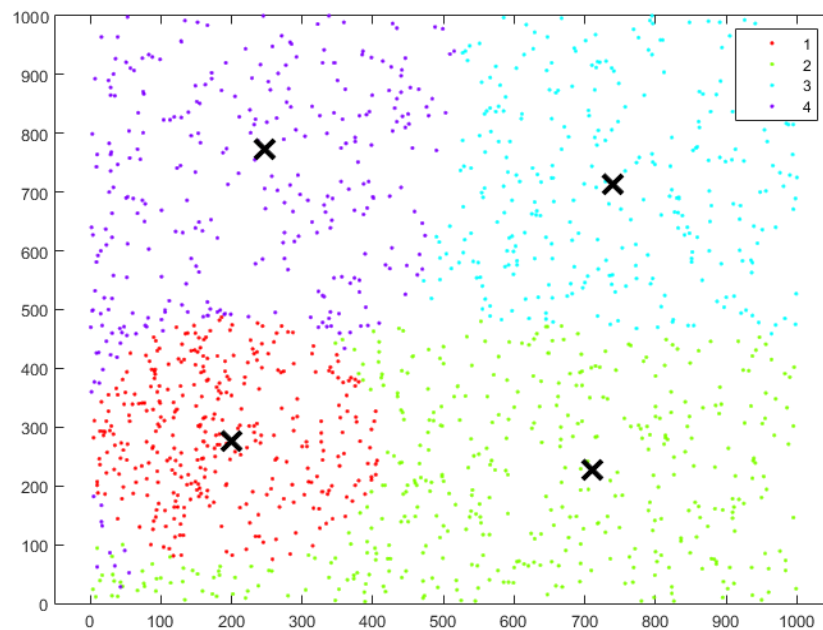


Figure 6. Solution after re-allocation for data in Figure 4

4. Evaluation Method

CFLP problems are replicated by randomly generating a number of datasets. These datasets are consisting of potential locations for facilities and a set of customer locations. Customer locations are consisting of x and y coordinates. Two types of synthetic datasets are used. First dataset is evenly distributed, and the other is clustered. And then, For evaluating the approach's solutions, a variety of evaluation methods are selected. These evaluation methods are aimed to determining results from a cluster analysis. All selected evaluation methods are categorized into the Internal Evaluation, which means evaluation for evaluating the results based on the results condition only. There is no comparison with another or existing result.

In Internal Evaluation, three factors were taken into consideration: (1) Compactness, (2) Separation, and (3) Connectedness. Connectedness is a factor that shows how well each observation connected to another observation in the same. Compactness is a factor that relates into homogeneity between observations in the same cluster. Separation is a factor that relates with how far are the clusters from each other. Compactness and separations are combined into single evaluation method, Silhouette (Rousseuw, 1987). Another evaluation method which also combines compactness and separation in one evaluation method is Davies-Bouldin Index (Davies and Bouldin, 1979). On the other hand, evaluation method for Connectedness is Connectivity cluster (Handl et al., 2005). All of these methods are common to determining how good a cluster analysis result. Silhouette is a method of evaluation based on how well each observation clustered into its cluster. Silhouette (ASW) can be calculated by:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (1)$$

$$ASW = \frac{1}{n} \sum_{i=1}^n s(i) \quad (2)$$

where:

n = the number of observations

$a(i)$ = average distance of observation i with all other observations in the same cluster (which cover the compactness factor)

$b(i)$ = average distance of i with all other observations in the nearest another cluster (which cover the separations factor)

ASW = average silhouette width

The result of Silhouette evaluation method (ASW) is between -1 and 1, with the result closer to 1 is meaning the better result. Davies-Bouldin Index is a method of evaluation based on the average distance of all observations in a cluster to its center, and also based on distance between clusters centers. This index can be calculated by:

$$DB = \frac{1}{m} \sum_{i=1}^m \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \quad (3)$$

where:

m = the number of clusters

c_i = centroid of cluster i

c_j = centroid of cluster j

σ_i = average distance of all observations in cluster i to centroid c_i (which cover the compactness factor)

σ_j = average distance of all observations in cluster j to centroid c_j (which cover the compactness factor)

$d(c_i, c_j)$ = distance between centroid c_i and c_j (which cover the separations factor)

Smaller number in the result of Davies-Bouldin Index evaluation method (DB) is meaning the better result. Connectivity is a criterion of evaluation based on the value of certain number of the closest neighbors from a certain observation. This index can be calculated by:

$$Conn(C) = \sum_{i=1}^n \sum_{j=1}^l f(x_i, z_{i(j)}) \quad (4)$$

where:

n = the number of observations

$z_{i(j)}$ = the j -th nearest neighbor of observation i

$f(x_i, z_{i(j)}) = 0$, if i and $z_{i(j)}$ are in the same cluster, or

$f(x_i, z_{i(j)}) = 1/j$, if i and $z_{i(j)}$ are in different clusters

l = number of neighbors that contribute to the connectivity calculation

Smaller number in the result of Connectivity evaluation method (C) is meaning the better result. Each approach is iterated 5 times for each dataset. Then, for each dataset result, the average of 5 time iteration results is taken to compare performances as shown in tables below. For Connectivity, extra iterations are done to accommodate its variations. Extra iterations are done for $l = 1$, $l = 5$, and $l = 10$. $l = 1$ means only 1 neighbor is contributing to Connectivity calculation, so $l = 5$ means 5 neighbors are contributing, and $l = 10$ means 10 neighbors are contributing.

5. Analysis Results

The results are evaluated using Silhouette, BI Index, and Davies-Bouldin Index evaluation methods. Each approach is tested into 2 randomly generated datasets. First dataset is normally distributed with 1400 customers to serve, and sufficient capacity is given in this test. Number of facility available is 4, and each has 350 capacities, so it's enough to serve all 1400 customers. Table 1 contains comparison in evenly distributed dataset.

Table 1. Results comparison in evenly distributed dataset

	Liao's Approach			Proposed Approach		
Silhouette	0.2337			0.4559		
BI Index	L = 1	L = 5	L = 10	L = 1	L = 5	L = 10
	120	537.2667	894.0032	164	607.2	911.1063
Davies-Bouldin	2.0818			1.1045		

From results in table 1, in terms of Silhouette evaluation, the result from Liao's approach is worse than the proposed approach, because in Silhouette evaluation, bigger value result is the better result. Same thing happened in Davies-Bouldin Index evaluation, the result from Liao's approach is also worse, because smaller value result is the better result in this evaluation. But in BI Index evaluation, the result from Liao's approach is better than the proposed approach, because in BI Index evaluation smaller value result is the better result.

The second dataset is a clustered dataset with 800 customers to serve, and sufficient capacity is given in this test. The number of facility available is 4, and each has 200 capacities. Table 2 contains comparison in clustered dataset.

Table 2. Results comparison in clustered dataset

	Liao's Approach			Proposed Approach		
Silhouette	0.4328			0.4219		
BI Index	L = 1	L = 5	L = 10	L = 1	L = 5	L = 10
	52	185.4	287.0397	158	372.8	509.1063
Davies-Bouldin	1.5878			1.435		

Table 2 shows based on Silhouette evaluation, the result from Liao's approach is better than the proposed approach, because the result's value is biggest one, but only by slight difference. In evaluation based on BI Index also shows that Liao's Approach have better result than the proposed approach, because the result's value is smaller than the proposed approach. On the contrary, based on Davies-Bouldin Index evaluation, the proposed approach is better than Liao's Approach, because the result's value is smaller compared to others results.

6. Conclusions

A clustering approach to solve CFLP also needs to be evaluated as a clustering method. In Internal Evaluation, three factors were taken into consideration: (1) Compactness, (2) Separation, and (3) Connectedness. Compactness and Separation are able to be evaluated using Silhouette and Davies-Bouldin Index Evaluation Method. And Connectedness is able to be evaluated using Connectivity Evaluation Method. Two heuristic approaches are evaluated and compared. One is an existing approach based on k -means clustering algorithm (Liao's Approach), and another is a new approach with different method in customers re-allocation step. Both approaches are evaluated in CFLP with evenly distributed customer locations and clustered customer locations.

In terms of CFLP with evenly distributed customer locations, the new proposed approach had good results in general. The proposed approach was better in Silhouette and Davies-Bouldin Index

evaluation methods. In other hand, in terms of CFLP with clustered customer locations, all results from the Liao's approach are better than the first proposed approach.

References

- [1] Brock, G., Pihur, V., Datta, S. and Datta, S. (2008), clValid, an R package for cluster validation, *Journal of Statistical Software*, vol. 25, pp. 1-22.
- [2] Chaves, A. A. and Lorena, L. A. N. (2010), Clustering search algorithm for the capacitated centered clustering problem, *Computers & Operations Research*, vol. 37(3), pp. 552-558.
- [3] Chudak, Fabian A., Williamson, David P. (2005). Improved Approximation Algorithms for Capacitated Facility Location Problems, *Journal Mathematical Programming: Series A and B*, vol. 102, Issue 2, pp. 207 – 222.
- [4] Davies, David L., Bouldin, Donald W. (1979). A Cluster Separation Measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PAMI-1 vol. (2), pp. 224–227.
- [5] Hongtao Yu, Liqun Gao, Yanhua Lei. (2012), Model and solution for capacitated facility location problem, In: *Control and Decision Conference (CCDC)*, 2012 24th Chinese, pp. 1773-1776, 23-25 May 2012.
- [6] Julia Handl, Joshua Knowles, Douglas B. Kell. (2005), Computational cluster validation in post-genomic data analysis, *Bioinformatics*, vol. 21 (15), pp. 3201-3212.
- [7] Kaufman, L. and P. J. Rousseeuw (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, USA.
- [8] Liao, K. and Guo, D. (2008), A Clustering-Based Approach to the Capacitated Facility Location Problem, *Transactions in GIS*, vol. 12, pp. 323–339
- [9] Mahmoodi Darani, N., Ahmadi, V., Saadati Eskandari, Z., & Yousefikhoshbakht, M. (2013). Solving Capacitated Clustering Problem by a Combined Meta-Heuristic Algorithm, *Journal of in Computer Research*, vol. 4(1), pp. 89-100.
- [10] Qingfang Ruan, Lixin Miao, Zhijun Zheng. (2010). A novel clustering-based approach for the location of multi-logistics centers, In: *8th International Conference on Supply Chain Management and Information Systems (SCMIS)*, 2010, 1-5, 6-9 Oct. 2010.
- [11] Robert M. Nauss. (1978). The 0–1 knapsack problem with multiple choice constraints, *European Journal of Operational Research*, vol. 2, Issue 2, pp.125-131.
- [12] Seyed Mohsen Mousavi and Seyed Taghi Akhavan Niaki. (2013). Capacitated location allocation problem with stochastic location and fuzzy demand: A hybrid algorithm, *Applied Mathematical*, vol. 37, Issue 7, pp. 5109-5119.
- [13] Silva, F. and Serra, D. (2007). A Capacitated Facility Location Problem with Constrained Backlogging Probabilities, *International Journal of Production Research*, vol 45, Issue 21.
- [14] Verter Verdat. (2011). *Foundation of Location Analysis*, Springer, USA.
- [15] Wena, M., Qinc, Z., Kangb, R. (2015). The Capacitated Facility Location-Allocation Problem under Uncertain Environment, *Journal of Intelligent & Fuzzy Systems*, vol. 29, no. 5, pp. 2217-2226.