4th International Conference on Industrial Engineering and Industrial Management
XIV Congreso de Ingeniería de Organización
Donostia- San Sebastián , September 8th -10th 2010

Selecting the best option in the stochastic hub location problem: Using Simulation and Data Envelopment Analysis

Rashed Sahraeian[1], Somayeh Mehrabi Gooravan[2]

[1] Assistant Professor of Industrial Engineering Department, Shahed University, Tehran, Iran. sahraeian@shahed.ac.ir

[2] Post graduated of Industrial Engineering, Shahed University, Tehran, Iran. s_mehr81@yahoo.com

abstract

The hub location problem is a subgroup of network optimization, in which, among existing nodes, p nodes are selected as hubs, and the remaining hubs are allocated to hub nodes. In the real world, demand, i.e., flow, through the nodes is probable and the simulation methods are usually used to solve problems in probability space. Thus, we review the hub location problem in probability space and some measuring criteria for existing options. To choose a suitable option, the Data Envelopment Analysis method is used as the ranking tool.

Keywords: *Hub Location, Simulation, Data Envelopment Analysis*

1. Introduction

Hub Location problems have been very important in recent two decades because of their application in modern transportation and Communications systems (Campbell et al., 2002).

Network design is a very important and basic issue in many transportation and communication systems because of its great effect on efficiency and final cost. In the case of direct communication among nodes is very expensive; it is better for the goods to move through nodes that act as hubs using indirect connections (connecting through the hub) instead of using direct connections from the origin to the destination. Therefore, the economies of scale can be used (Rodriguez and Salazar, 2008). In an aerial network, the economies of scale mean that bigger and more efficient airplanes can be used to connect hubs together. In communication networks, using light fibers with higher capacity are considered to connect hubs. Hub location problems may be found in three basic areas (Alumar and Kara, 2007):

- Switching, which includes computer systems, power distribution, phone networks
- Trans-shipment, such as aerial services, special and fast cargo
- Sorting points, e.g., postal centres.

These problems are subgroups of network optimization and location-allocation problems. The hubs act by concentrating or compounding: i.e., each hub divides a big entrance flow into small components and sends them to their individual destinations (or vice versa) (Campbell et al., 2002). Thus, there are two notable issues: determining location of hub facilities among existing nodes and allocating remaining nodes to hubs. In the studies in this field, three assumptions are usually considered (Alumar and Kara, 2007):

- The network between the hubs is a complete network
- The economies of scale has been considered by using the discount factor (α) for the connections between hub nodes
- There is no direct connection among non-hub nodes

To allocate non-hub nodes to hub nodes, there are two posible cases: single allocation (SA) or multiple allocations (MA). In SA, each non-hub node connects to exactly one hub node but in MA, each non-hub node can connects to more than one hub.

Usually in studies, the target function is to minimize flow and set up costs. Although decreasing traveling time (Kara and Tansel, 2000), increasing amount of coverage (Kara and Tansel, 1999) and decreasing congestion (Rodriguez et al, 2007) are also considered. Another problem is specifying the number of hubs. If the number of hubs is defined, the problem is known as *p-hub*. In the following, the simplest mathematical formulation has been proposed for hub location problems. (O'Kelly 1987)

$$Min \sum_i \sum_j W_{ij} \left(\sum_k X_{ij} C_{ik} + \sum_m X_{jm} C_{jm} + \alpha \sum_k \sum_m X_{ik} X_{jm} C_{km} \right) \quad (1)$$

$$\text{s.t.} \quad (\text{n-p+1}) X_{kk} - \sum_i X_{ik} \geq 0 \quad ,\text{forall k} \quad (2)$$

$$\sum_k X_{ik} = 1 \quad ,\text{forall i} \quad (3)$$

$$\sum_k X_{kk} = p, \quad (4)$$

$$X_{ik} \in \{0,1\} \quad (5)$$

In this model, X_{ij} will be 1, if node i is allocated to hub j. C_{ik} shows the cost per unit of flow from node i to k. Relation (2) prevents non-hub nodes from being connected to each other. Relation (3) shows that each node should only be allocated to one hub. Relation (4) provides the number of hubs. For more information on studies in this field, Alumar and Kara (2007) and Campbell et al. (2002) have written useful articles.

Most studies about these problems were done in definite space and their probability were not been studied . For stochastic situation in hub location problem,see the papers by Sim et al. (2008), Yang (2009). Sim et al. (2008) presented the stochastic p-hub center problem (Sp-HCP) with service-level constraints. They assumed that delivery times on the links are normally distributed and formulated this problem to minimize the maximum delivery times on the path with determined service-level. Yang (2009) introduced a stochastic programming model to solve the air freight hub location and flight routes planning under seasonal demand variations. They formulated this problem to minimize the set up and transportation costs.

To approach reality, considering the probability of flow is necessary. Accordingly,using simulation to find the best option is suggested. Because we can present long time results by considering the real condition of the problem in a short time, the decision-maker is able to decide correctly. Therefore the aim of this present research is to realize the problem and make the solution applicable. According to applications of hub location problems, several factors are important when determining the best answer for the decision-maker (the previous studies did not address this issue). Although some studies may have goals in addition to costs, it is rare to consider several goals or measuring criteria simultaneously. Prior to 2009, only one case considered the two goals of expense and servicing time simultaneously (Costa et al, 2008). Of course, this study is more theoretical than applicable. In some research, the number of limitations has been added instead of considering multiple factors (Marianov and Serra, 2000), (Bryan, 1998),(Klincewicz, 1998). We note that using the criteria instead of limitations, causes increasing options and increasing fields of possible answers, and therefore, we can obtain the best answers. In this study, several criteria have been considered simultaneously. To choose the best option, Data Envelopment Analysis is used as a ranking tool. It seems that, according to positive specificities, Data Envelopment Analysis (DEA) can be helpful in choosing the best option. This paper includes the following sections. The second section introduces the proposed method and explains the problem. The third section presents the specificities of a real system in a postal centre, while the fourth section analyses the results. Finally, the fifth section contains conclusion.

2. Proposed method

In this section, the proposed method is presented according to the states in reality. The hub location problem is applied to the postal system in this paper. Therefore, the hypothesis is close to this application. In a postal system, the letters are sent to centers, called hubs, after they are received in the non-hub post offices. In the hubs, the letters are arranged according to their destinations and then they are sent to their destinations. Several different states may exist:

- The destination is the same as its origin (non-hub centre);
- The origin and destination of the letter are allocated to one hub;
- The origin and the destination of the letter are allocated to two separate hubs.

The different conditions can be seen in Figure 1. The origin and destination nodes and the path are displayed in black highlight color. In state 1 ("A" in Figure 1), node 1 is considered to be the origin and destination of the letter. In this case, the letter goes from node 1 to the hub node and then returns back to node 1; in this condition, there are two routes. In state 2 ("B" in Figure 1), the origin node is node 1, and the destination node is node 2; both of them are allocated to one hub. In this case, the letter goes from node 1 to the hub node and then to node 2; therefore there are two routes. In state 3 ("C" in Figure 1), node 1 is the origin node and node 7 is the destination node. The letter goes from the node 1 to its connected hub node and then to the hub allocated to node 7; finally, it is transferred to node 7. In this condition, there are three routes. We have considered them all in this paper.

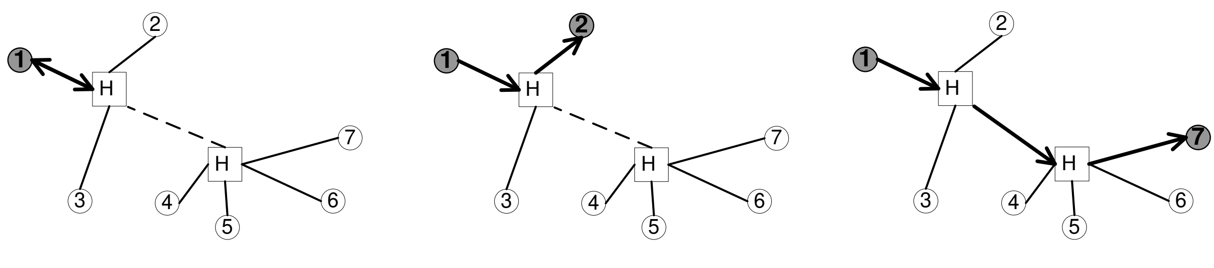

Figure 1.The deferent conditions for a letter's origin and destination

The factors that may influence the hub location problem are as follows: flow through the nodes, servicing time in nodes and the time to pass through nodes. The proposed method can respond to stochastic states of all three factors. The probability of these factors will be considered using probability distribution functions in the model and method. In this method, the entrance of letters to postal centres has a probable distribution. The destination of each letter is also considered probabilistically. According to these two probabilities, the flow through the nodes also has a probabilistic distribution. In each node, according to the kind of postal centre (non-hub or hub), the servicing time for each letter has a probabilistic distribution. Therefore, the system is made to be as close as possible to the real postal system. To consider possible options, the options that are impossible because of limitations are removed. In all states, we hypothesise that after choosing the hubs from among the nodes, the allocation phase is based on the nearest hub. Although previous research has proven that this allocation method does not necessary lead to the best solution, such a strategy is used for allocation in most studies (Aykin, 1990). According to the kind of postal system and previous studies, four criteria have higher importance with respect to time:

- The percentage of on-time letters;
- The average time in the system;
- The average expense of transporting each letter;
- The average distance of each hub from non-hub postal centres allocated to that hub.

It is clear that among these four criteria, the percent of on time letters is a positive criteria and its development is desirable. However, the three remaining criteria are negative and should be reduced as much as possible. Accordingly, after considering the postal system and their options, the problem changes to a multiple criteria problem related to decision making. Therefore, using MCDM (Multiple Criteria Decision Making) is necessary to rank options. As stated previously, the technique used in this article is Data Envelopment Analysis. This technique is based on linear programming, which has the ability to measure the partial efficiency of units with several similar inputs and outputs. This technique can consider criteria that are completely different from the view point of scale and size. This positive feature increases the efficiency of the current research. In this technique, the efficiency is considered to be the ratio of the total of the output weights to the total of the input weights in a unit or system (DMU). In the next step, based on the mathematical models, the efficiency of a unit is maximized relative to other units. Initially, DEA was used to review the efficiency of units, However, it was also exploited as a technique to rank the units (Sarkis, 2000). The first suggested model for Data Envelopment Analysis is the CCR. This model was presented by Charnes, Cooper and Rohdes and is known as the CCR model (Charnes et al, 1978). This model is as follows:

$$\text{Max } Z = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \tag{6}$$

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \qquad j=1,\ldots,n \tag{7}$$

$$u_r, v_i \geq 0$$

In the above model, y_{rj} is the *rth* output of unit j and x_{ij} is *ith* the input of unit j. u_r and v_i are, respectively, the *rth* output weight and *ith* input weight. It is observed that the target function, which is maximised, is the efficiency of the reviewed unit. In this condition, we consider the efficiency limitation to be smaller than 1 in all units. This limitation is given in Equation (7). This model must be formed and solved with equal units. It should be mentioned that, after presenting the above model, other models are also presented. For more information it is possible to refer Cook and Seiford (2009). The returns to scale structure of units are important in DEA model. In fact, the returns to scale reacts growth ratio in outputs toward growth in inputs. Two kinds of returns to scale are considered:

- Constant returns to scale: growth in outputs is as growth in inputs

- Variable returns to scale: growth in outputs is less or more than in inputs.

In this article, according to the introduced criteria and many differences among them, an efficiency model to variable scale is used to rank options. Another point that is effective in determining the model type is hub input or output. In other words, we can analyse efficiency by concentrating on inputs or outputs. According to these two views, researchers have introduced efficiency as follows (Charnes et al, 1978):

- Input-oriented: a unit is inefficient when there is a possibility of decreasing each input by making other inputs and keeping outputs fixed.

- Output-oriented: a unit is inefficient when there is a possibility of increasing each output by making other outputs and keeping inputs fixed.

In most problems that use DEA as a ranking tool, one of input or output models is utilized according to the condition of the system. According to the criteria in this problem, both viewpoints are used because of two reasons: first, a unit is logically efficient (superior rank) when proved according to both viewpoints presented above (Campbell et al, 2002); second, the decision-maker does not want to concentrate on just one of the inputs (negative criteria) or outputs (positive criteria). Accordingly, both models of input-oriented with variable returns to scale and output-oriented with variable returns to scale have been used. In the next section, to model the problem, Enterprise Dynamic software is used and to rank the options, deap software is used.

3. A Postal centre Problem

The problem is a type of single allocation with several definite hubs, and its symbol is " p-hub/D/SA/•/ criteria". The presented symbol is based on classification issues (Hamacher &

Nickel, 1998). First, the type of problem is specified. In this section, "hub" is written, and if the number of hubs is specified, "p" is used. Second, the problem is described in terms of continuous or discrete points. Usually, "D" refers to discrete. In the third part, restrictions and added variables are written: "SA" means single allocation problem. Relationships between new and existing facilities are given in the fourth subsection. " $\bullet$ " symbolises the discount coefficients for the distance between nodes. The last section shows the type of the objective function, and "criteria" displays the use of the criteria to assess the options. The postal system includes ten nodes as postal centres. The goal is to choose two centres among these ten nodes as hubs and allocate the remaining eight nodes as non-hub nodes. Therefore, there are 45 possible options. According to the positions of the postal centres in the four regions, choosing two nodes as hubs among the nodes in one region is not acceptable. (There are two nodes in areas 1 and 2, and three nodes in areas 3 and 4). Therefore, 8 options are eliminated from the total options. Moreover, according to the limited capital available to develop and change non-hub postal centres to hub centres, it is hypothesised that choosing hub nodes from special areas (Like 1&4, 3&4 and 2&3) is not possible simultaneously. Therefore we cannot accept 21 options, and they are eliminated. Thus, there are 16 possible options, which are shown in Table 1.

Table 1. Possible Options

Options	Hub	Allocated nodes to hub	Options	Hub	Allocated nodes to hubs
1st option	1	2, 5, 6, 7	9th option	2	1, 3, 5, 6
	3	4, 8, 9, 10		9	4, 7, 8, 10
2nd option	1	2, 5, 6, 7, 8	10th option	2	1, 3, 5, 6
	4	3, 9, 10		10	4, 7, 8, 9
3rd option	2	1, 5, 6, 7, 8	11th option	3	1, 4, 8, 9, 10
	3	4, 9, 10		5	2, 6, 7
4th option	2	1, 5, 6, 7, 8	12th option	3	1, 2, 4, 8, 9
	4	3, 9, 10		6	5, 7, 10
5th option	1	2, 6	13th option	3	1, 2, 4
	8	3, 4, 5, 7, 9, 10		7	5, 6, 8, 9, 10
6th option	1	2, 3, 5, 6	14th option	4	3, 8, 9, 10
	9	4, 7, 8, 10		5	1, 2, 6, 7
7th option	1	2, 3, 5	15th option	4	1, 2, 3, 8, 9, 10
	10	4, 6, 7, 8, 9		6	5, 7
8th option	2	1, 3, 5, 6	16 th option	4	1, 2, 3, 9
	8	4, 7, 9, 10		7	5, 6, 8, 10

Transporting letters among nodes is considered to be continuous in the model. To get closer to realize, the transportation rate among hub nodes is four times rate among non-hub nodes. On the other hand, the transportation cost among hub nodes is considered to have a coefficient of 0.25. To calculate the transportation cost of letters, the cost is divided as multiple packages. Table 2 includes possible functions of letter entrances to each node. Table 3 shows the destination distribution of each letter that enters the nodes.

Table 2. Possible functions of letter entrance to original nodes (second)

Original node number	1	2	3	4	5
Percentage of letters	Normal (220.40)	Normal (180.35)	Normal (280.20)	Normal (300.50)	Normal (120.40)
Original node umber	6	7	8	9	10
Percentage of letters	Normal (150.29)	Normal (250.50)	Normal (320.10)	Normal (360.45)	Normal (280.50)

Servicing time in nodes according to hub or non-hub nodes is shown in Table 4. In Table 5, the distance between nodes is considered. Each model (option) is performed for 24 hours, and all data are based on the calculations. The results of the system simulation are in Table 6 for different options. Considering that increasing the criteria of "percentage of on-time letters"is favourable, this criterion is considered as output, and the other criteria are considered as input. Using Deap software, the input and output model of the Data Envelopment Analysis with variable returns to scale for each option is performed, and the results are shown in Table 7.

Table 3. Distribution of letters

Number of destination node	1	2	3	4	5	6	7	8	9	10
Percentage of letters	9%	5%	8%	12%	14%	12%	10%	10%	7%	13%

Table 4. Servicing time in nodes

Kind of node	**Possibility function of servicing time**
Non-hub node	Normal (60.15)
Hub node	Normal (30.10)

Table 5. The distance among nodes

	1	2	3	4	5	6	7	8	9	10
1	--	5	10	13	11	14	19	10	16	21
2		--	9	11	9	12	16	8.5	12.5	15
3			--	4	22	27	23	9.5	12.5	14
4				--	18	23.5	21	9	8	11
5					--	6	5.5	10	13	15
6						--	4.5	15	18	12
7							--	9	11.5	9.5
8								--	7.5	7
9									--	3.5

4. Result Analysis

In both models, the 12th, 13th and 14th options are suitable options with efficiency 1. In the Data Envelopment analysis technique, many methods have been suggested to differentiate among efficient units. One of the simple and useful methods is calculating the number of times a unit is determined as a reference for inefficient units. The reference unit is one that an inefficient unit is expected to reach. According to this approach, a superior unit is one that is reached more times than the reference. Known reference units for each option are determined through Deap software and are shown in Table 7. According to the results, in the hub input model, the 12th option in 2 cases and the 13th option in 11cases and 14th option in 13 cases are references. In the hub output model, the 12th option in 13 cases, the 13th option in no cases, and the 14th option in 7 cases are known as references. To differentiate among 12th, 13th and 14th options and to use both hub input and output models, the results of both models are analysed. For this purpose, we use total times in which each option of two models is determined to be efficient. Considering the models (hub input and output), the 12^{th} option in 15 cases, the 13^{th} option in 11 cases and the 14^{th} option in 20 cases are known as reference units. Therefore, the 14^{th} option is the best one. Thus, the best option in the above postal system is: 4^{th} and 5^{th} nodes as hubs and to allocate 3^{rd}, 8^{th}, 9^{th} and 10^{th} nodes to 4^{th} node and allocate 1^{st}, 2^{nd}, 6^{th} and 7^{th} nodes to 5^{th} node.

Table 6. Results of system simulation for different options

Option	1st criteria	2nd criteria	3rd criteria	4th criteria
1^{st} option	0.717466	3758	16.805	9.125
2^{nd} option	0/290323	12765	20.1158	11
3^{rd} option	0.441159	6518	17.568	9.625
4^{th} option	0.389747	107.20	16.82	9.188
5^{th} option	0.430982	7087	1605	8.875
6^{th} option	0.844842	2682	17.01	**9.313**
7^{th} option	0.603821	4160	17.27	8.688
8^{th} option	0.731248	3402	15.18	8.438
9^{th} option	0.717829	3271	15.6067	8.188
10^{th} option	0.683817	3512	16.01	8.25
11^{th} option	0.82965	2949	15.805	8.813
12^{th} option	1.00	1486	16.1736	8.438
13^{th} option	0.465033	6525	16.1	7.875
14^{th} option	0.753391	3248	13.83	7.938
15^{th} option	0.608239	4536	15.45	8.313
16^{th} option	0.544538	4484	14.9807	8.063

5. Conclusion

The goal of this study is to get the best option in probability space to choose hubs and to allocate non-hub nodes. Reviewing the studies in hub location problem shows that probable states are less noticed. The necessity of reviewing such cases can be considered in real problems. In the current article, a simulation method has been used to find the best option because two reasons: hub location problems have many complexities in definite states, and simulation is suggested as a common applicable method in probable states. After simulating a system, results analysis is necessary and important step in solving problem. In conditions in which we cannot determine the superiority of a specific option on the basis of results and criteria, MCDM techniques are used. Thus, DEA is used to determine the best option. There is a possibility that each multiple-criteria technique results in the same ranks for some options. In the Data Envelopment Analysis technique, different methods have been presented to differentiate the same ranks. Finally, it has been attempted to obtain a special ranking and differentiate options by combining two suitable models. The results of the suggested method show that, to reach a suitable location relevant to the mentioned criteria in the article, 4th and 5th nodes should be the hubs. Therefore, using simulation tools and considering probability space, and considering the existence of several criteria for suitable options in the hub location problem, are innovative aspects of this paper.

Table 7. Option efficiencies and reference units through deap software for hub input and output models with efficiency to variable scale

Option	Efficiency of input-oriented model	Efficiency of output-oriented model	Reference unit in input-oriented model	Reference unit in output-oriented model
1st option	0.870	0.717	13&14	12
2nd option	0.720	0.290	13&14	12
3rd option	0.823	0.441	13&14	12
4th option	0.862	0.390	13&14	12
5th option	0.893	0.437	13&14	12&14
6th option	0.882	0.845	13&14	12
7th option	0.912	0.604	13&14	12
8th option	0.942	0.817	12&14	12&14
9th option	0.972	0.819	13&14	12&14
10th option	0.962	0.754	13&14	12&14
11th option	0.921	0.863	12&14	12&14
12th option	1.000	1.000	12	12
13th option	1.000	1.000	13	13
14th option	1.000	1.000	14	14
15th option	0.952	0.658	13&14	14&12
16th option	0.982	0.668	13&14	12&14

6. References

Alumar, S. Kara, B.Y., (2007), Network hub location problems: The state of the art, European Journal of Operational Research

Aykin, T.,(1990), On a quadratic integer program for the location of interacting hub facilities. European Journal of Operational Research 46 (3), 409–411

Bryan, D.L., (1998). Extensions to the hub location problem: Formulations and numerical examples. Geographical Analysis 30, 315–330.

Campbell, J.F., Ernst, A., Krishnamoorthy, M., (2002). Hub location problems. In: Drezner Z, Hamacher H, editors. Facility location: applications and theory. Heidelberg, Germany: Springer; 373–408.

Charnes A. Cooper W. W. and Rhodes E. (1978) Measuring the Efficiency of Decision Making Units. EJOR 2, 429-444.

Cook, W. D. Seiford, L. M. (2009) Data envelopment analysis (DEA) -Thirty years on. European Journal of Operational Research 192, 1–17.

Costa, M., Captive, M., And Climaco, J., (2008), Capacitated single allocation hub location problem a bi-criteria approach, Computers & Operations Research, 35, 3671-3695

Hamacher, H.W., Nickel ,S., (1998), Classification of location models, Location Science, 6, 229-242

Kara, B.Y., Tansel, B.C., (1999). On the single-assignment p-hub covering problem, Technical report, Department of Industrial Engineering, Bilkent University, Bilkent 06533, Ankara, Turkey.

Kara, B.Y., Tansel, B.C., (2000). On the single-assignment p-hub center problem, European Journal of Operational Research 125, 648–655.

Klincewicz, J.G., (1998). Hub location in backbone/tributary network design: A review. Location Science 6, 307–335.

Marianov, V., Serra, D.,(2000), Location models for airline hubs behaving as m/d/c queues, Technical report 453, Department of Economics IET and GREs, Universitat Pompeu Fabra, Barcelona, Spain.

O'Kelly, M.E., (1987). A quadratic integer program for the location of interacting hub facilities. European Journal of Operational Research , 32, 393 - 404.

Rodriguez, V., Alvarez, M. J., And Barcos, L.(2007), Hub location under capacity constraints, Transportation Research, Part E, 43, 495-505

Rodriguez-Martin, I., Salazar-Gonzales, J. J. (2008), solving a capacitated hub location problem, European Journal of Operational Research, 184, 468-479

Sarkis, J. A (2000) comparative analysis of DEA as a discrete alternative multiple criteria decision tool, European Journal of Operational Research 123, 543±557

Sim, T., Lowe, T. J., Thomas, B. W., (2008). The stochastic p-hub center problem with service-level constraints, Computers and Operations Research 12, 3166-3177

Yang, T. H.(2009). Stochastic air fright hub location and flight routs planning. Applied Mathematical Modeling 12, 4424-4430