Using Spatial Data Analysis in Distribution System Design

Mahmood Pariazar (Msc)[1]
m_pariazar@aut.ac.ir
Mohammad Saeed zaeri (Msc)[2]
Zaeri@aut.ac.ir
Jamal Shahrabi (Phd)[3]
Shahrabi@dal.ca
[1,2,3] Industrial Engineering Department, Amirkabir University of Technology (AUT)
Postal Code: 424 Hafez Ave., Tehran, Iran. 15875-4413
Phone: +98-2122497336

1. Introduction

Supply chain management is the term used to describe the management of the flow of materials, information, and funds across the entire supply chain, from suppliers to component producers to final assemblers to distribution (warehouses and retailers), and ultimately to the consumer. The efficient and effective movement of goods from raw material sites to processing facilities, component fabrication plants, finished goods assembly plants, distribution centers, retailers and customers is critical in today's competitive environment. Supply chain management entails not only the movement of goods but also decisions about where to locate plants and distribution centers (Simchi-Levi 2000). Location decisions may be the most critical and most difficult of the decisions needed to realize an efficient supply chain.
Consider a set of Distribution centers and retailers are geographically dispersed in a region. Each retailer experiences demands for a variety of products which are manufactured at the plants. A set of distribution centers must be located in the distribution system from a list of potential sites.

Research on facility location models has mainly dealt with assigning facilities to serve their nearest demand areas, in order to minimize aggregate operational costs (Bramel and Simchi-Levi, 1997). To achieve this goal, appropriate decisions in terms of the number and location of facilities as well as the demand area potentially served by each facility, should be made using the facility allocation logic rules.

As it progresses, in section 2 we have introduced cluster analysis. In section 3, Multicommodity Capacitated Facility Location model has been scrutinized. In section 4, with the help of stated concepts in section 2 and 3, a case study has been done and finally in section 5, results have been obtained.

2. Cluster analysis

In this section cluster analysis have been introduced to be one of the most useful techniques of multivariate analysis. Cluster analysis involves categorization: dividing a large group of observation into smaller groups so that the observation within each group relatively similar and the observations in different groups are relatively dissimilar. Clustering algorithms have been the subject of a great deal of research (Hand David et al 2001).

there are many measures of similarity, many ways of measuring how well the subdivision reflects the character of the data, and even various concepts of subdivision. Many clustering algorithms assume that the points exist in some metric space, often real space under the Euclidean (or another Lp) norm; the similarity is then defined simply by the distance metric. For some clustering problems, the distance function need not obey the triangle inequality, be symmetric, be real-valued, or even be defined on all pairs of points (Grabmeier Johannes 2003).

3. Multicommodity Capacitated Facility Location Problems (MCFLP)

We introduce the following model, which considers Distribution system and multi-product families. The model we consider has the following features:

- A set of customers are geographically located in a region.
- Each customer demands variety of products that are manufactured at the plants.
- Products are shipped from distribution centers to customers. For each shipping link and product there is a per unit shipping cost.
- A given number of distribution centers must be located among a list of potential sites.
- A fixed opening and operating cost must be paid for each distribution centers that is opened.
- Each distribution center has a capacity not to be exceeded.

In other word, we have a set of customers with known demands of a certain commodity as well as a set of potential sites where facilities that satisfy the demands of the customers can be located. The capacity of the facility at each potential site is known. The problem is to determine the set of facilities to open as well as the manner in which the demand of each customer is satisfied without violating the capacity restriction of any facility, so that the total cost is minimized. The indices (sets) are as follows:

Let we denote the set of product families $i \in I$ and (in addition to the fixed cost f_j) g_{ij} the fixed product cost. C_j is capacity of distribution center located at site *j* and S_k is volume of one unit of product *k*. Then an capacitated multi-activity model, also called multi-commodity or multi-activity capacitated facility location problem (MCFLP) can be given as follows:

$$\min \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} q_{ijk}\, w_{ijk} + \sum_{i\in I}\sum_{j\in J} g_{ij} z_{ij} + \sum_{j\in J} f_j y_j$$

$$\text{s.t.} \quad \sum_{j\in J} w_{ijk} = 1 \qquad \forall\, i \in I, k \in K \qquad (1)$$

$$z_{ij} - y_j \le 0 \qquad \forall\, i \in I, j \in J \qquad (2)$$

$$w_{ijk} - z_{ij} \le 0 \qquad \forall\, i \in I, j \in J, k \in K \qquad (3)$$

$$\sum_{i\in I}\sum_{k\in K} S_k d_{ik} \le c_j y_j \qquad (4)$$

$$z_{ij}, y_j \in \{0,1\} \qquad \forall\, i \in I, j \in J \qquad (5)$$

$$w_{ijk} \ge 0 \qquad \forall\, i \in I, j \in J, k \in K \qquad (6)$$

Here, z_{ij} is a binary variable which equals 1 if product/service type *i* is provided at depot *j*. The variable w_{ijk} denotes the fraction of demand d_{ik} of demand node *k* for product *i* which is covered by depot *j*. Likewise q_{ijk} denotes the cost of providing d_{ik} units of product *i* from depot *j* to demand node $k \in K$. The objective function measures the transportation cost from distribution centers to customers and the fixed cost of locating and operating distribution centers. Constraints (1) require that the demand of each customer is covered. The coupling constraints (2) and (3) forbid to assign products to closed depots and to deliver product *i* to node *k* from depot *j* if product *i* is unavailable at the depot. Constraints (4) ensures that the costumers demand serviced by a distribution center doesn't exceed its capacity (Klose 2005).

4. Case study

In a case study in Tehran, one of good distribution corporations has been investigated. This corporation is responsible for distribution of three product family of dried nuts, cereals and meat products. Regarding existing records, this corporation is faced with 640 demand points for these products. The amount of demand for each product family is determined in all points. Now, This Corporation is investigating its distribution centers' locations. There are 4 places that have been determined as alternatives. Management is intended to determine which of these alternative distribution centers that should be selected and how different product families should be allocated to these demand points. Regarding that examining all these 640 points in the model mentioned in section 3 is kind of NP-Hard problem, Cluster analysis has been used for categorizing demand points. With the help of geographical maps and ARCGIS software, we have tried to identify coordinates of each demand points. Criterion for similarity that has been used is average distance. This approach can be considered a sort of compromise between single and complete linkage. Some researchers prefer this method because it comes closest to fitting a tree that satisfies a least square minimization criterion. Instead of using the minimum (single linkage) or the maximum (complete linkage).

Based on characteristics of location data of demand points, 640 demand points have been divided into 7 clusters. This is considered to be the best advantage of using spatial data mining in this study. Centroid of each cluster is deemed to be the representative of that cluster so that the demand of all categories of products of that cluster has been paid attention to in it(figure1).

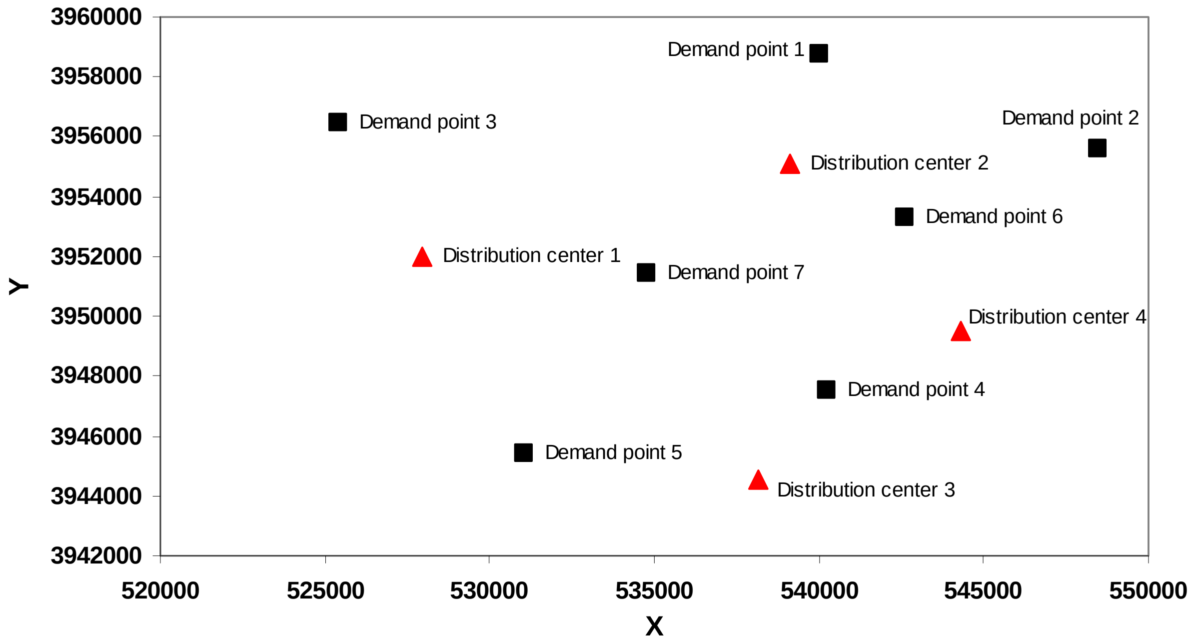

Figure 1: coordinates of demand points and distribution centers

In this step, based on MCFLP model and by using LINGO software, problem is solved. It should be mentioned that unit shipping cost from DC j to customer i of product k has been considered like a function of geographical distance between 2 points of origin and destination. Calculation of transportation cost between 2 points based on location data input due to compatibility with real geographical conditions of the city illustrates the importance of using spatial data. After designating all input parameters of problem, model has been solved. From amongst 4 DC candidates, centers 1, 2 and 4 have been chosen. Obtained results from model solving have been brought in table 1. For example, in cluster 3, 63% of demand for product 1 from distribution center 1 and 37% from distribution center 2 will be supplied.

Number of cluster	DC 1			DC 2			DC 4		
	Product 1 %	Product 2 %	Product 3 %	Product 1 %	Product 2 %	Product 3 %	Product 1 %	Product 2 %	Product 3 %
1	57	11	0	43	78	59	0	11	41
2	0	5	0	65	43	44	35	52	56
3	63	74	58	37	24	42	0	2	0
4	13	35	17	20	12	29	67	53	54
5	38	26	41	41	32	35	21	42	24
6	10	0	5	49	61	53	41	39	42
7	43	61	33	37	39	50	20	0	17

Table 1 : results of solving MCFLP

5. Conclusions

In this article with provision of combination methodology including spatial analysis and MCFLP model, we tried to design distribution network for a case study in Tehran. In this article, characteristics of location data input for multivariate analysis and calculation of transportation costs for transferring goods have been well utilized.

6. References

Bramel J and Simchi-Levi D, Springer Series in Operations Research, 1997, *The Logic of Logistics: theory, algorithms, and applications for logistics management*.

Grabmeier J and Rudolph A, 2003, Techniques of cluster algorithms in data mining. *Data Mining and Knowledge Discovery*, 6(4):303. 360.

Hand D, Mannila H and Padhraic S, The MIT Press, 2001, *Principles of data mining*.

Klose A and Drexl A ,2005, Facility location models for distribution system design. *European Journal of Operational Research*, 162 4–29

Simchi-Levi, D and Kaminsky P and Simchi-Levi E, Irwin McGraw-Hill, 2000, *Designing and Managing The Supply Chain*.