Clustering Based Model For Facility Location In Logistic Network Using K-Means

Shaju Varghese [1], Gladston Raj S [2],
[1]Research Scholar, Bharathiar University, Coimbatore, Tamilnadu, India, shajubpc@gmail.com
[2]Head, Department of CS, Govt. College, Nedumangad, Thiruvananthapuram, Kerala, India.

Abstract- The optimum location for facilities is one of the major challenges in logistics network. These locations are scattered in various places in urban area, places in highways, populated areas in order to speedup package deliveries and minimize the overall transport cost and time. Identifying and deciding the optimum locations and the minimum number of distribution centers (DCs) are the major goals in the design of any type of logistics network. The number of DCs will basically depend upon various types of attributes related to the network. The optimum locations of DCs will reduce the overall transport cost. As operations globalize, location decisions become more complex. This is a soft computing based model for facility location in logistics network design. In this work we propose a k-means clustering model for facility location in logistics.

1. INTRODUCTION

The facility location problem or location analysis is a k center problem, which deals with the optimal placement of facilities to minimize transportation costs by satisfying some constraints. The techniques also apply to cluster analysis so that the clustering algorithms used in data mining can also be applied for location analysis. In this work, we design a soft computing based model for facility location for logistics analysis

1.1. Facility Location Problem (FLP)

The facility location-allocation problem will identify and locate the best as well as optimal location to implement one or more facilities that the users can effectively utilizes with the minimum service distance and period. In general the "facility" consists of the bases, units, equipment, weapon systems, logistics, civil objects, etc. The location-allocation problem corresponds to the solutions of finding the best configuration to install one or more facilities in order to attend the largest subset of users within a service distance.

The facility location problem is a well-known problem in the areas of production and operations management and combinatorial optimization. The problem finds an optimal location of facilities considering facility construction costs, transportation costs, etc. This problem is very popular because it is faced by many companies

Location models are often difficult to solve, especially for large problem instances [3]. Besides location models are application dependent. Their objectives, constraints and variables are determined by particular problem under study. Therefore, there does not exist an all-purpose location model that is appropriate for all potential or existing applications [3].

There are two types of techniques in general [3]

1. Exact Solution Techniques [3]
2. Heuristic Solution Techniques [3]

The following are the some of the techniques used for as evaluation of heuristics [3]

- Bound on Optimal Solutions
- Worst Case Analysis

- Statistical Evaluation

1.2. Applications of FLP

The research in this area has not only concentrated on the mathematical modeling and algorithmic aspects but also in the area of innovative applications. The applications include various optimization areas such as parameter calibration, land use optimization, transport route searching, retail shop outlet location selection, spatial objects optimization, performance optimization of urban modeling. When applied to the design of various types of telecommunication networks plays an important role in this process. It is well known that optimization techniques can be used to improve telecommunication network design. Specifically for example many decisions have to be made concerning trade-off between concentrator location and cable expansion. Many analysts have claimed that more than 50% of capital and operating costs are a consequence of the provision of Local Access Telecommunication Networks.

II. FACILITY LOCATION PROBLEM

This problem is very popular because it is faced by many companies. There are many studies going on to find the optimal solution for this problem and proposed solution approaches.

The purpose of this model is to generate a decision support system. Also, this model helps the authority to locate and fix an appropriate location for the facility for logistic management.

Here, we give a mathematical formulate of the un-capacitated facility location problem

Let I and J denote a set of facilities and customers, respectively ($i = 1,\ldots,m$ and $j = 1,\ldots,n$), and let f_i denote the fixed cost of locating a facility at location i. Let c_{ij} be the cost of supplying customer j from facility i. The decision variables are as follows:

x_{ij}: denotes the fraction of customer j's demand satisfied by facility i

$$y_i = \begin{cases} 1 & if\ we\ locate\ a\ facility\ at\ location\ i \\ 0 & otherwise \end{cases}$$

The following is a mixed-integer linear programming formulation of the un-capacitated facility location problem.

$$\min : \sum_{i\in I} f_i y_i + \sum_{i\in I}\sum_{j\in J} c_{ij} x_{ij}$$

Subject to:

$$\sum_{i\in I} x_{ij} = 1 \qquad for\ all\ j \in J, \qquad (1)$$

$$y_i - x_{ij} \geq 0 \qquad for\ all\ i \in I, j \in J, \qquad (2)$$

$$x_{ij} \geq 0 \qquad for\ all\ i \in I, j \in J, \qquad (3)$$

$$y_i \in \{0,1\} \qquad for\ all\ i \in I. \qquad (4)$$

The objective is to minimize the total costs. The first set of constraints shows that the demand of customer j (for $j = 1,\ldots,J$) will be satisfied. The second set of constraints shows that if a facility is not located at a potential location i, there will be no shipments from that location. The third set of constraints is the non-negativity constraints, and the last set of constraints is the integrality constraints.

The following is the dual formulation of the linear programming relaxation of the un-capacitated facility location problem.

$$\max : \sum_{j \in J} v_j$$

Subject to :

$$\sum_{j \in J} w_{ij} \le f_i \qquad \text{for all } i \in I, \qquad (5)$$

$$v_j - w_{ij} \le c_{ij} \qquad \text{for all } i \in I, j \in J, \qquad (6)$$

$$w_{ij} \ge 0 \qquad \text{for all } i \in I, j \in J. \qquad (7)$$

v_j are the dual variables for constraint set (1), and w_{ij} are the dual variables for constraint set (2).

Time complexity of problems of the above mentioned problem is a obvious one, so that linear programming based methods will take much time with respect to the size of the problem and the constraint. In this work, we will design a soft computing model for solving a simple facility location problem with minimum constraints.

2.1. Facility Location in Logistics Network

In this work, the Public Logistic Network (PLN) designed using the standard tools. We use a simplified version of that design. So that instead of using the "average package delivery time" as a metric for optimization, we used simple distance as the metric in the fitness function of the soft computing model. This approach was used to minimize the optimization time. Since the distance is directly proportional to package delivery time, we believe that this approach also will lead to equal results, logically with in lesser time.

- Generation of the Underlying Road Network (URN)
- Developing the network of Distribution centers (DCs),
- Finding public DC locations that minimize the distance between the DCs and the User locations.

2.2. Generation of the Road Network of USA

The following map shows the road network that we created from the US census data set.

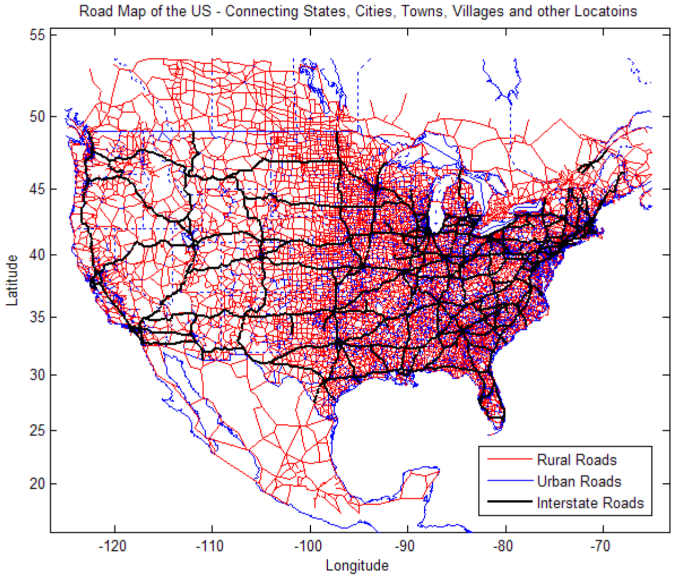

Figure 1: Road Network of USA

2.3. Generation of the Underlying Road Network of regional Distribution Centers.

The population in RDC is represented by total 925 U.S. census blocks that are plotted on the map of RDC. A sub-graph of the road network was generated that is then followed by the removal of two-degree nodes from the network. Each point in this network is a potential location for a DC

The following graph/map shows the road network of Alabama (AL), USA that will be the example of a sub graph we created and used to create the regional distribution centers (RDC) that we are interested.

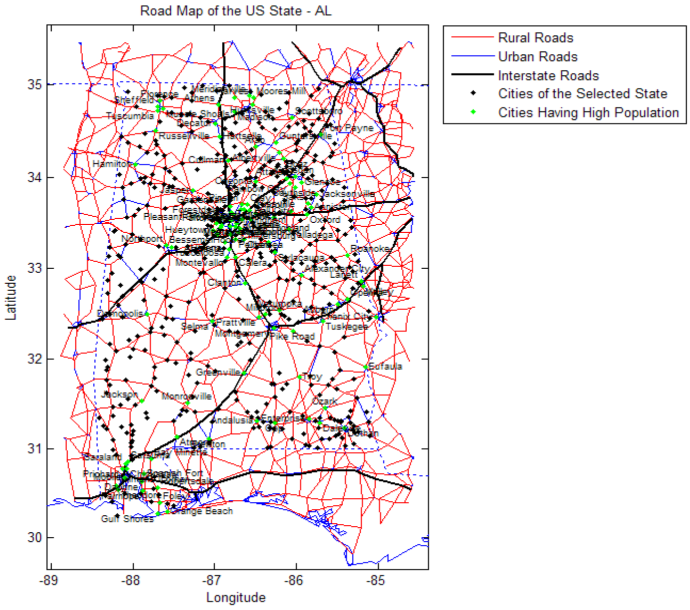

Figure 2. Underlying Road Network of RDCs

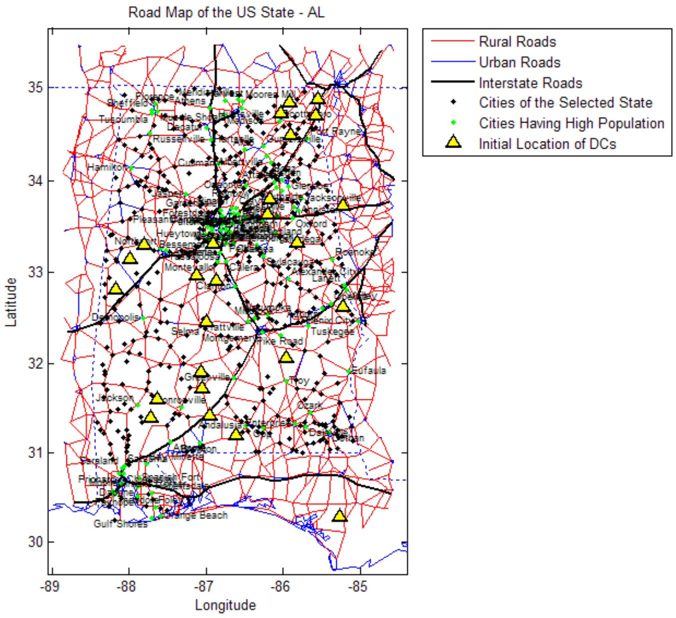

Figure 3: Initial Location of Distribution Centers

2.3. K-Means Clustering

K-Means Clustering is an algorithm among several that attempt to find groups in the data. In pseudo code, it follow this procedure:

The vector **m** contains a reference to the sample mean of each cluster. **x** refers to each of our examples, and **b** contains our "estimated class labels.

```
K-Means Algorithm

Initialize m_i,  i = 1,…,k, for example, to k random x^t
Repeat
        For all x^t  in X
        b_i^t ← 1 if || x^t - m_i || = min_j || x^t - m_j ||
                b_i^t ← 0 otherwise
        For all m_i,  i = 1,…,k
                m_i ← sum over t (b_i^t x^t) / sum over t (b_i^t )
Until m_i converge
```

Explained perhaps more simply in words, the algorithm roughly follows this approach:

1) Choose some manner in which to initialize the m_i to be the mean of each group (or cluster), and do it.
2) For each example in your set, assign it to the closest group (represented by m_i).
3) For each m_i, recalculate it based on the examples that are currently assigned to it.
4) Repeat steps 2-3 until m_i converge.

Now that we have some rudimentary understanding of what k-means

```
Function d= EuclideanDist(XY_U, XY_F)
Begin
// Compute the Euclidean distance with each coordinate
[R,C]=size(XY_U);
//sum squared data - save re-calculating repeatedly later
XYsq=repmat(sum(XY_U.^2,2),1,NoDCs);
// The distance Function d^2 = (x-c)^2 = x^2 + c^2 -2xc
Dist = XYsq + repmat (sum ( (XY_F.^2)' ,1), R, 1) - 2. * (
XY_U *( XY_U '));
//label points
[d,Classes]=min(Dist,[],2);
d =sqrt(sum(d));
return (d)
End
```

2.4. The Fitness Function

The following function is used to find the fitness at the set of facility location $XY_F= (X_i, Y_j)$, where i,j = 1 to n. The set of points which has the lowest fitness vaue will be the optimum location for placing facilities. XY_U is the locations of all the customers(cities, towns, villages)

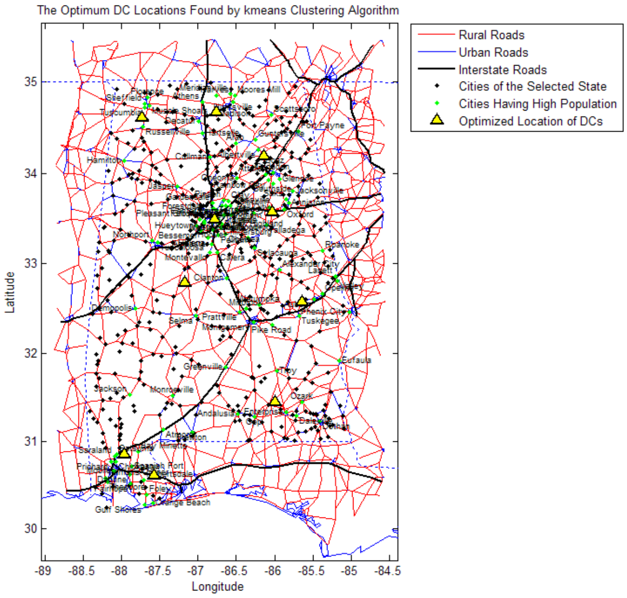

Figure 4: Distribution Centers found by K-means

III. RESULTS AND DISCUSSION

We have implemented the proposed soft computing based models for facility location in logistics analysis using Matlab software version R2012s. We tried to use almost equal input parameters for each and every evaluated method. We used the USA census data and map data which is much suitable for this kind of research. We decided to use USA data because, it is the only data refereed in some of the previous works and there seems no such detailed data available for any other country for validating the methods of facility location and logistics analysis.

3.1. The Parameters of the Soft Computing Models

K-means Clustering

MaxRepetitions : 10

MaxIter : 20

The following table shows the overall results of this work. Since the performance of a soft computing model will depend up on several factors and some random conditions, we run the algorithm several times and only selected the values which are minimum.

Table 1 : The Overall Performance for Locating 25 Facilities

Facility Location Method	Avg. Distance	Time Consumed
Random	6.16	-
K-means	3.95	1.87

IV. CONCLUSION

Facility location for logistics is a wide area for research. In this work we addressed the possibilities of using soft computing based model for facility location in for logistics. We used soft computing based clustering approach for Facility Location Problem & Logistic Analysis.

In this work, we used a simple Euclidean distance function as a fitness function in the design of soft computing based location optimization model. But, there are much more constraints and parameters in a practical logistics problem that can be included in the design of the fitness function such as (1) travel time with respect to road type, (2) loading unloading time at DCs, (3) different modes of travel times such as air travel time. This kind of more constraints and parameters can be included in future design of soft computing based optimization models. Our future works will address these issues.

We have designed the proposed models as a single objective problem. But there are facility location and logistics situations where there may be more than one objective during optimization. Future works may address the design of soft computing based optimization models for multi objective optimization scenarios.

REFERENCES

[1] Erlenkotter, D., "A Dual-Based Procedure for the Uncapacitated Facility Location." Operations Research, Vol 26(6), pg. 992-1009, 1978.

[2] Francis, R.L., McGinnis, F.L., Jr., White, J.A., "Facility Layout and Location: An Analytical Approach." Prentice Hall, 2nd Ed., 1974.

[3] Juliana Karakaneva, "A LOCATION PROBLEM MODELING AND SOLVING", Trakia Journal of Sciences, Vol 1, No 4, pp 1-7, 2003,ISSN 1312-1723, Copyright © 2003 Trakia University

[4] HK Smith, G Laporte and PR Harper, "Locational analysis: highlights of growth to maturity", http://eprints.soton.ac.uk/68930/1/Locational_Analysis_-_Smith,_Laporte_and_Harper.doc

[5] Michael J. Bucci, Michael G. Kay, Donald P. Warsing†, Jeffrey A. Joines, "Metaheuristics for Facility Location Problems with Economies of Scale", IIE Transactions

[6] Michael J. Bucci, Ryan Woolard, Jeffrey Joines, Kristin Thoney, Russell E. King, "An Application of Heuristics Incorporating Economies of Scale to Facility Location Problems in Carpet Recycling"

[7] Biehl, M., Prater, E., Realff, M.J., 2007, Assessing performance and uncertainty in developing carpet reverse logistics systems, Computers and Operations Research, 34, 443-463.

[8] Brimberg, J.,Hansen, P., Mladenovic, N.,Taillard, E.D., 2000, Improvements and Comparison of Heuristics for Solving the Uncapacitated Multisource Weber Problem, Operations Research, 48, 444-460.

[9] Bucci, M.J., Kay, M.G., Warsing, D.P., Joines, J.A., 2009, Metaheuristics for Facility Location with Economies of Scale, Fitts Department of Industrial and Systems Engineering working paper, North Carolina State University, Raleigh NC.

[10] Cooper, L., 1963, Location-allocation problems, Operations Research, 11, 331-343.

[11] CARE (Carpet America Recovery Effort) annual report, 2007, http://www.carpetrecovery.org/pdf/annual_report/07_CARE-annual-rpt.pdf

[12] CARE (Carpet America Recovery Effort) network website,2009, http://www.carpetrecovery.org/pdf/reclamation_centers/Carpet_Reclamation_Center s.pdf

[13] Daskin, M.S., 1995, Network and discrete location: models, algorithms, and applications, John Wiley and Sons, New York.

[14] De Brito, M.P., Dekker, R., Flapper, S.D.P., 2003, Reverse Logistics - a review of case studies, ERIM Report Series.

[15] Fleishchmann, M., Krikke, H.R., Dekker, R., Flapper, S.D.P., 2000, A characterization of logistics networks for product recovery, Omega, 28, 653-666.

[16] Louwers, D., Kip, B.J., Peters, E., Souren, F., Flapper, S.D.P., 1999, A facility location allocation model for reusing carpet, Computers and Industrial Engineering,

[17] Mirchandani, P.B., Francis, R.L., 1990, Discrete Location Theory, Wiley, NewYork.

[18] Realff, M.J., Ammons, J.C., Newton, D., 1999, Carpet Recycling: Determining the 38:3, 547-567.

[19] Realff, M., Systems Planning for Carpet Recycling. 2006, Recycling in Textiles, Editor Youjiang Wang, CRC Press

[20] Shaju Varghese, Gladston Raj S, “A Genetic Algorithm Based Optimization Model for Facility Location in Logistic Network Design”, International Journal of Applied Engineering Research” Vol 10, No. 69, pp-338-344, 2005, ISSN 0973-4562