Facility Recommendation System Using Domination Set Theory in Graph

B A Sabarish, Vidhya. S

Abstract: ***Facility location problem has gained importance with the increased applications involving infrastructure development which involves placing facilities in right positions. With the help of GPS based services the analysis of locations and traffic which is vital input for the problem of facility location. The problem of facility location recommender is a multi-objective problem of reducing the transportation cost and increasing the coverage in the geographical region. Conditions to place a facility for a better coverage and reduced cost will differ from facility to facility. For the purpose algorithms such as route finder, Fastest clustering algorithm are used to cluster geographical region for improved infrastructure and better Quality of Service. In this paper analyzes facility of locating schools, hospital and police station in a bounded geographical region. The algorithm uses domination set and k-means clustering algorithm to choose the facility and its corresponding cluster in the region. Clustered are validated using index measures including DBI and Dunn Index values. An experimental analysis is conducted for Coimbatore city and results are evaluated against real facilities.***

Index Terms: ***Facility location recommendation system, Domination set theory, Clustering, Multi-objective Problem***

I. INTRODUCTION

. Facility location problem (FLP) is a process of selecting potential facility location set "S" from complete geographical region "C" where the facility can be provided for a set of demand "D". The goal is to pick a subset of points to minimize the sum of distances between points selected for the facilities "S" and Location point set "C", where "S" is a subset of "C". Facility set "D" will be placed in selected point set "S". Facility location is a critical part of the strategic plan for any kind of firms to set up warehouses or showrooms. Various factors including resources, Transportation, accessibility influence locations where the facility influences in choosing the right location. Many models are proposed for the facility location problems which involves Set Covering Problem (SCP) and Maximal Covering Location Problem (MCLP) [1].

Facility location decisions plays a pivotal role in strategic planning for a wide range of firms. Facility location problem takes into account multiple conflicting criteria and a set of alternatives therefore, various Multiple Criteria Decision Making (MCDM) techniques are extensively used. MCDM can be classified into two types, they are Multi Attribute Decision Making (MADM) and Multi Objective Decision Making (MODM). Multi-Criteria Decision-Making (MCDM) is used for assessing trade-offs between multiple criteria and based on the assessment we decide upon what to be chosen, prioritised among alternatives. Identified alternatives are evaluated both qualitatively and qualitatively and decisions are made based on these measures. Facility location models are build on greedy algorithms such as median based or center based. Multi-objective problem can be solved by scalarization problem (combine to make a single objective) or Pareto optimal approach to reduce the optimal set involved in the process [1].

The facilities can be of different types, undesired facilities such as chemical plants, garbage dumps and desired facilities such schools and hospitals. Facilities like garbage dumps have a negative impact on the environment, hence that facilities have to be placed relatively in the higher distance when comparing to the desired facilities including schools and hospitals which should be placed in a location which is approachable from various regions. The algorithm to choose the various facilities will vary based on it's a desired or undesired facility which is being placed [1].

Facility location models can be classified as: Shape and topography of the facility and demand sites, objectives, Restriction/Constraints, Solution methods. Our problem mainly focuses on discrete network models, where the facilities are going to be get placed in discrete locations. One of the most popular models among facility location models is covering problem. Location Set Covering Problem involves finding the smallest number of facilities and their locations so that each demand is covered by at least one facility. The main objective of Set Covering Problem (SCP) is to minimize location costs satisfying a specified level of coverage. Location Set Covering Problem (LSCP) model has can be either implicit or explicit. Implicit model is built assuming that each demand area can be covered not only by one; it can be covered by two or more to cover a threshold percentage of demand, whereas explicit model considers a specific combination of facilities. Our proposed work uses Domination Set Theory to choose a facility and k-means algorithm to choose a cluster of candidates around the chosen facility. Facility location problem is modelled as a network/graph of crucial elements involved in planning. If all locations are candidate facility location and client location, we have the Dominating Set problem. Domination set theory is an approach which can be applied to the geographical region to choose the right location for various facilities based on the algorithm. Domination set of a graph G= (V, E), is a subset of V where each vertex is adjacent to at least one vertex in the subset (Dg). In Independent sets are close to domination sets. Where independent set can also be dominating set if only if it's a maximal independent set. Placing of generic candidates/facilities such as hospitals, schools, police stations can be visualised as Generic Access Facility

Revised Manuscript Received on July 05, 2019

Sabarish B.A , Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

Vidhya.S, Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India.

(GAF) location problem and can be implemented by clustering algorithm. Clustering is grouping similar objects, hence we can identify centroid point of a regional locations to choosing right location for various facilities. Facilities will be have various constraints in selecting location in turn choosing centroid for various clusters. GAF in turn also can be viewed as partitioning framework based problem where facilities are located across various partitions of geographical region. Partitioning is done by clustering algorithms using Euclidean distance measure [1].

II. LITERATURE SURVEY

A. Review on Multiple Criteria Undesirable Facility Location Problems

This paper deals with facilities of selecting right location for landfilling and placing treatment centre in a region. Facility may be desirable or undesirable, algorithms will be different for both the type of facilities. Objectives and attributes will differ for each of the facilities has to be extracted before decision making process. MODM normally works with a set of SMART objectives, specific constraints. FLP can be median based or centre based for fixed number of facilities (p) or unlimited or exponential problem where facilities are expected to be varying. Multi-objective can be solved by either scalarization or pareto optimizing model. In scalarization, multiple objectives are fused to make a single new objective, in contrast to pareto, which tries to reduce number of alternates [1].

Warehouse problems which is known as determining and minimizing number of warehouses which includes a constraints of capacity, cost matrix for transportation. Warehouse location identification problem is solved using a cost matrix by decomposing problem into various linear sub-problems [2].

A variation of facility location problem in which multiple facilities in the same site are allowed is discussed. The model identified features is that a variant set-up cost is considered based on the size of the facility and the other important feature is that there are multiple types of facilities that can be located at different sites. The problem was modelled as a non-linear integer programming [3].

Network design application also can be viewed GAF problem, by means of deciding access points for networking facility based on crowd and available constraint on bandwidth resources. Objective of making effective and economic connection for increased efficient performance and reduced cost by proper selection of critical points (crowded points)[5].

Hazardous materials disposal and transportation is a problem considering fact of radiation and issues caused. Hence problem to identify locations to reduce cost and risk involved in transportation of explosives. Location has to be selected should be selected considering transportation risk and cost [6].

Proposed a two level Supply Chain Management for satisfying complete set of customers by means dividing problem of single warehouse into multiple warehouses which can serve whole customer set. It involves creating a distributed environment for facilities to reduce cost and increase accessibility for customers [7].

Capacitated facility location problem (CFLP) is a multi-objective decision making problem. It considers objective of minimization of cost without exceeding capacity of plants in automotive industry. Problem gains significance in terms inappropriate positioning will leads to excessive cost in long term (life times) , hence transportation, inventory cost should be optimized. CFLP includes a customer market and profit based on rate of return, hence attracted many practitioners and gained importance in research [8].

III. PROBLEM STATEMENT

A. Problem Definition

Facility location problem is multi-objective decision making problem for selection of facilities to satisfy specified constraints. For the basic problem of facility location takes a finite number of known locations and decision variables. Problem is defined with Set of Facilities (F) for which each facility involves a cost which is normally distance (accessibility) for customers (C). FLP can be formulated as a optimization problem with two cases either minisum or minimax. Minisum tries to place facility (Subset of F) in a place by which total distance between existing and new facilities distance is reduced.

$$\min f(x) = \sum_{i=1}^{n} dist(Fnew, Fi) \quad (1)$$

F_i is the list of existing facilities. F_new is location of placing new facilities. Minimax tries to maximize the distance in contrast to minisum problem.

$$\max f(x) = \sum_{i=1}^{n} dist(Fnew, Fi) \quad (2)$$

B. Problem Statement

Geographical region of a city represented as graph in which each vertex represents the junction points and edges represents the distance between the various points. City will be represented as graph C (J, D). J represents the points as vertex (Longitude, Latitude) and D represents the distance as edge. Graph is represented as adjacency matrix (N X N). In which each element in the matrix represent where the vertices are directly connected or not. If the vertices are connected then represented as '1' and if not directly connected then represented as '0'.

IV. EXPERIMENTAL SETUP

A. Dataset

Dataset is collected from various route maps. For the implementation, dataset is collected using google maps for Coimbatore City. From which various junction points and distance are calculated from the Google map. Routes for various locations were collected from various students and faculty of our institution and other institutions in and around Coimbatore.

B. Preprocessing

Dataset collected from google data are mapped and converted to form a graph in which each city represents node and edges represents the path between the two city. Weights are assigned to edges which is the distance between points represented. Node selection is a process of identifying critical points or junction points is done based on simple decision making locations and crowd movement in specified geographical regions. Junction points identified are given as input for facility location recommendation system.

C. Framework

Facility recommendation system is a multi-objective problem which considers various factors including cost of transportation and accessibility. Implementation of facility recommender system start with representation of complete geographical location in form of graph. Junction points identified using preprocessing stage is represented in form of graph in which each node represents critical point of city. Application domain is implemented for a geographical region of Coimbatore City. By preprocessing analysis, 18 critical points identified for Coimbatore region which covers entire region of city.

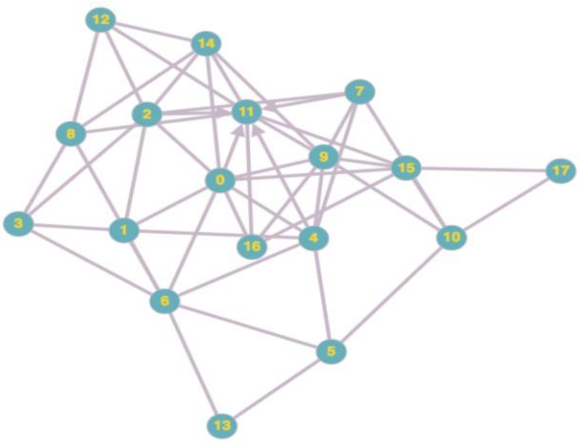

FIG .1 GRAPH REPRESENTATION

Node Numbers	Critical point
0	Townhall
1	Gandhipuram
2	Rspuram
3	Sbc
4	Ganapathy
5	Saravanampatti
6	Kavundampalayam
7	Peelamedu
8	Vadavalli
9	Singanallur
10	Chinniyampalayam
11	Kovaipudur
12	Thondamuthur
13	Thudiyalur
14	Selvapuram
15	Ettimadai
16	Podanur
17	Neelambur

TAB1. NODE LABELS

D. Distance calculation

Distance matrix is generated by using Euclidiean distance between various junction points generated. Recommendation system is chosen for three different facilities including hospital, police station and schools. Accessibility and distance measure will be varying for facilities. Comparing to school and police station, hospital should be in a more shortest and accessible. In case of police station and schools which can be in approachable distance, need not be a shorter distance.

E. Clustering and Validation

Clustering of locations are performed by means of k-means clustering using Euclidean distance measure. Validation of clustering is done using inter cluster and intra cluster distance measures along with correlation index for merging larger clusters to form a better clustering.

i. K-means clustering, K-means algorithm is a two staged algorithm on which in first phase of assignment deals with selection of initial centroid points and form a cluster based on distance between centroid and points. In second phase of updating new centroid and update details of points which may move from one cluster to other is done. This iteration of update process is repeated till update phase not making any change in clusters.

ii.Davies Bouldin Index, Dunn Index with a small variance between members of the cluster, and well separated, where the means of different clusters are sufficiently far apart, as compared to the within cluster variance. For a given assignment of clusters, a higher Dunn index indicates better clustering. Higher Dunn index indicates better clustering [9].

$$DB_k = \frac{1}{k}\sum_{i=1}^{k} R_i \quad (3)$$

iii. Dunn Index, Dunn Index with a small variance between members of the cluster, and well separated, where the means of different clusters are sufficiently far apart, as compared to the within cluster variance. For a given assignment of clusters, a higher Dunn index indicates better clustering. Higher Dunn index indicates better clustering [9].

$$D_{nc} = \min_{i=1,2,\ldots nc} \left\{ \min_{j=i+1,\ldots nc} \left[\frac{d(c_i, c_j)}{\max_{k=1,2,\ldots nc} diam(c_k)} \right] \right\} \quad (4)$$

V. RESULTS

Intra-cluster distance is measure of how well clusters are well organized. Cluster are formed in such a way that intra cluster distance should be minimal to indicate clusters are closely packed and inter cluster distance should be maximized to indicate clusters are well separated. Following graph represents variation of inter and intra cluster distance between clusters generated by k-means clustering.Cluster generated are using two different distance values measures based on facility chosen, in case of schools and hospital clusters are generated using average distance measures since accessibility factor can be enough. For facility of hospital travel distance should be minimal so cluster for hospital is generated using minimal distance algorithm.

Clusters	Hospital	School	Police Station
Set 1	[7,1,2,4,5,6,9,14]	[11,6,8,10,16,18]	[12,1,3,5,8,9,10,13,16,17]
Set2	[3,1,2,4,8,12,13,15]	[7,1,2,4,5,6,9,14]	[16,1,10,11,12,17,18]
Set 3	[7,1,2,4,5,6,9,14]	[12,1,3,5,8,9,10,13,15,16,17]	[16,1,10,11,12,17,18]
Set 4	[7,1,2,4,5,6,9,14]	[12,1,3,5,8,9,10,13,15,16,17]	[18,11,16]

Table 2. Clusters generated

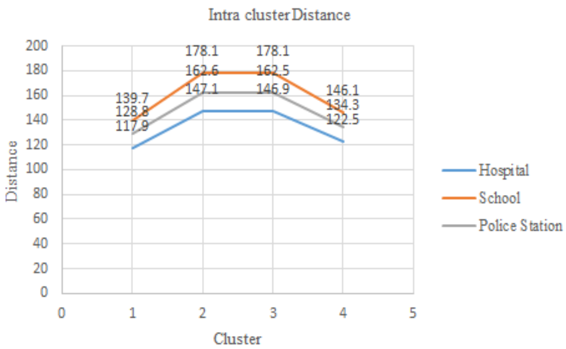

Fig .2 Intra-cluster Distance

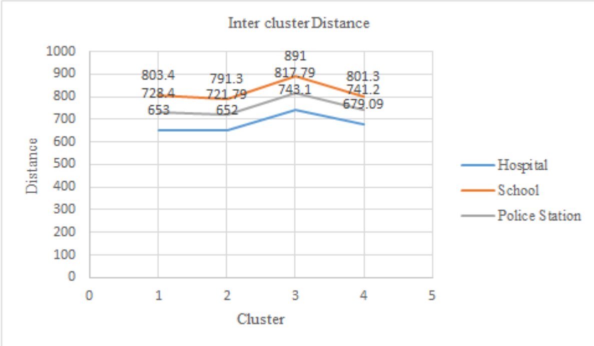

Fig .3 Inter-cluster Distance

Fig. 1. & 2 specifies the inter cluster and intra cluster distance values of clusters generated by using k-means clustering. Intra-clustering distance is measured by sum of distance between all points involved in the cluster, Inter cluster distance is measured by sum of distance between the cluster points. From analysis it shows set 3 provides high inter cluster distance showing well separated clusters and in contrast set 1 provides a minimal intra-cluster distance showing highly compacted clusters.

Davies –Bouldin Index	Hospital	School	Police Station
Set 1	**0.989**	**1.165**	**1.077**
Set2	5.995	4.999	4.536
Set 3	1.247	1.509	1.378
Set 4	0.9890	1.165	1.0772

Table 3. Davies Bouldin Index

Dunn Index	Hospital	School	Police Station
Set 1	1.9759	1.8198	1.88
Set2	**4.0739**	**6.2405**	**6.1259**
Set 3	5.675	5.9146	5.8027
Set 4	4.475	4.5510	4.5186

Table 4. Dunn Index

From analysis of complete clusters analysis, high Dunn value is generated for set 2 and Low DBI is identified for set1. From analysis of complete region of various clustering index values its identified that set 2 is optimum for making right location for various facilities including hospital, school and police station

VI. CONCLUSION

Approach proposed provides an effective and efficient method to identify the dominating sets in a graph which can play a major role in developing smart cities and provide better infrastructure for better quality of living. The results obtained are accurate and the same procedure can be applied to any geographical location and it can be applied as preprocessing steps for designing smart city applications to place various facilities to accommodate minimum cost with maximum accessibility.

REFERENCES

1. Tang SH, Boyer O, Pedram A, Yusuff RB, Zulkifli N. A review on multiple criteria undesirable facility location problems. Journal of Basic and Applied Scientific Research. 2013;3(8):708-13.
2. Bagherpoor Z, Parhizi S, Hoseininia M, Heidari N, Yaghin RG. Warehouse Location Problem. InFacility Location 2009 (pp. 295-314). Physica, Heidelberg.
3. Wu LY, Zhang XS, Zhang JL. Capacitated facility location problem with general setup cost. Computers & Operations Research. 2006 May 1;33(5):1226-41.
4. Melkote S, Daskin MS. Capacitated facility location/network design problems. European journal of operational research. 2001 Mar 16;129(3):481-95.
5. McGregor P, Shen D. Network design: An algorithm for the access facility location problem. IEEE Transactions on Communications. 1977 Jan;25(1):61-73.
6. Berglund PG, Kwon C. Robust facility location problem for hazardous waste transportation. Networks and spatial Economics. 2014 Mar 1;14(1):91-116.
7. Chen CH, Ting CJ. Combining lagrangian heuristic and ant colony system to solve the single source capacitated facility location problem. Transportation research part E: logistics and transportation review. 2008 Nov 1;44(6):1099-122.
8. Yu H, Gao L, Lei Y. Model and solution for capacitated facility location problem. InControl and Decision Conference (CCDC), 2012 24th Chinese 2012 May 23 (pp. 1773-1776). IEEE.
9. Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. Journal of intelligent information systems. 2001 Dec 1;17(2-3):107-45.

AUTHORS PROFILE

Sabarish. B.A currently serves as Assistant Professor at Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India. He received her B. Tech. degree in Information Technology from Cape Institute of Technology Levengipuram, and M. Tech. degree in Information Technology from Anna University, Coimbatore. His areas of research include Wireless Sensors and Data Mining. https://www.amrita.edu/faculty/ba-sabarish

Vidhya S. joined School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, as a teaching faculty in July 2014. She received her B. Tech. degree in Computer Science and Engineering from Government Engineering College, Sreekrishnapuram, Palakkad, and M. E. degree in Computer Science and Engineering from Anna University. She currently serves as Assistant Professor in the Department of Computer Science and Engineering, School of Engineering, Coimbatore Campus. Her areas of interest include Data Structures and Algorithms, Operating System. Her areas of research include Social Computing, Graphs. https://www.amrita.edu/faculty/s-vidhya