A Florida County Locates Disaster Recovery Centers

Xinran Hao s2040564@ed.ac.uk
Peng Jin s2087060@ed.ac.uk
Zhaowei Zhang s2043293@ed.ac.uk

1 Background

Discrete Location Problem is a very classic problem. With the publication of the first site selection theory by German scholar Weber in 1909 as the symbol, discrete site selection problem officially entered the scientific research era [1]. According to the different needs of the problem to be solved by the model, Discrete Location Problem mainly includes several different subproblems: Median Problem, Covering Problem, Central Problem, etc.[2]. In this paper, the Federal Emergency Management Agency (FEMA) need to provide effective recovery assistance to disaster-stricken areas after natural disasters occur. It required each county in Florida to establish at least three potential disaster recovery centres (DRCs). The problem belongs to the discrete facility location problem of the emergency system. For the problem of the emergency system location, it is suitable to use maximizing the number of customers supported within the specified time (distance), minimizing the costs of building facilities and minimizing the maximum distance of any customer to the centres as objective functions.

2 Analysis of the Problem

Alachua County is a country located in the north-central portion of the U.S. state of Florida. According to the 2005 data, the population in this country was approximately 96,000. The east-west direction and north-south direction of Alachua County is about 32 miles (51.52 kilometres) and 30 miles (48.3 kilometres). The land area is about 874 square miles (2266 square kilometres).

In general, there are four optimization objectives [3]:

1 Minimize the average travel distance to a closest DRC.

2 Minimize the maximum travel distance to a closest DRC.

3 Minimize the total number of DRCs needed, subject to each county resident being within a radius r of a nearest DRC.

4 Maximize the probability that at least one DRC will be usable following a disaster.

Our article tends to identify the objective that the number of DRCs required is minimized as much as possible (the third objective), that is, a set covering location problem with all building cost equal to one. In this case, the distance between each resident of the county and the nearest DRC does not exceed a fixed distance r.

After converted the special distance measurement of original data (See Section 3 for details) into miles unit distance measurement, we then used 1-norm distance (for any two points (x_1, y_1), (x_2, y_2), the Manhattan distance between them is $|x_1 - x_2| + |y_1 - y_2|$) which is more realistic, instead of the Euclidean distance. The first goal of our article is to replicate the model of the original paper as much as possible and obtain an optimal solution. Therefore, we chose the same set of values of r as in the original paper, i.e., r : 10, 15 and 20 miles.

The original paper was roughly divided into two steps. In the first step, they used a clustering algorithm to reduce the data size from 6000 to 200. Then, by building an integer programming model, they got the ideal optimal solution. In the second step, they classify and score the buildings around these selected optimal locations according to a list of mainly subjective constraints with unknown weights. Hence, a more realistic location can be achieved by modifying step 1 location slightly according to a list of the score for around buildings. However, Our article does not intend to design this building grade system mainly for the following reasons. The scores are too subjective, and little information is available about the building grade, so it is quite difficult to implement this process one to one. Therefore, our article only focuses on step 1, that is, solving idealized covering location problems to obtain the best optimal solution mathematically.

3 Data Collection and Processing

The data we selected is from a GIS website [1], which contains the parcels information data of Alachua County in 2005. The observation of data is every parcel of land. Every observation contains important information such as the number of buildings, the total square footage of the building, the x and y coordinates of the parcel centre (In GIS special measurement), etc. However, the number of observation is too large than 7000, which is the number of parcels in the original article. We need to preprocess this data and reduce the number of observation.

3.1 Data Preprocessing

At first, we reduce the data to 6988 parcels (i.e., demand points in the mathematical model) from 90,000 parcels by using the following ways:

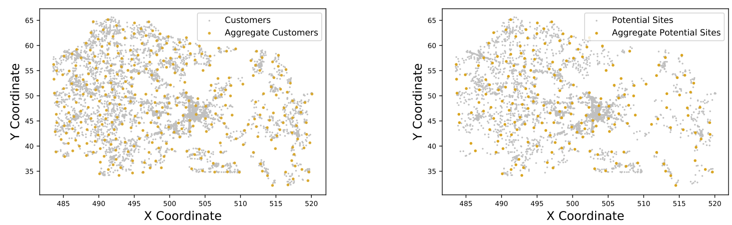
[1] https://www.arcgis.com/home/item.html?id=e140df77a81a49a9a1b0ce69868a5097

- Remove observations with outlier and missing data on the total square footage (**SQFT**) of the building and the number of buildings.
- Remove those data with the number of buildings (**NUMBLDG**) equal to 0.
- Drop the data involving **Residential** in **CATEGORY 2** [2].

By enlarging the vector figure and comparing some customers distribution areas that can be recognized by the naked eye, we found that the distribution of some discrete areas in the graph we obtained is significantly different from the original paper. After many times of manual comparison, we found that if we try to get the same points, it is possible to lose many of the valid data in the paper graph, which means it will reduce the accuracy of the data. Therefore, based on our observation, we selected the data to around 4400 parcels as the potential DRC sites in the following ways.

- Select parcels satisfying the total square footage (**SQFT**) greater than 2000 square feet and (**ACRES**) exceeding 0.0459.
- Choose parcels with stories and activity year all exceeding 0.

The 6988 demand points and 4400 potential DRC sites are respectively distributed as below:



(a) Customer and Aggregate Customers in Alachua

(b) Potential Sites and Aggregate Potential Sites in Alachua

Figure 1: Demand Points and Potential DRC Sites

3.2 PTF Algorithm

The PTF algorithm in the original paper is to aggregate demand points with a given aggregation error b so that the distance from any demand point to its nearest DRC would not exceed $r + b$. The aggregated points are shown in the form of yellow points in Fig.1.

One of the advantages of the PTF algorithm is that the exceeded distance is controllable. Another important reason is that the aggregated points distribute uniformly on the map using this algorithm to aggregate demand points. That means any two aggregated potential sites would not too close to each other. The PTF algorithm is as follows:

[2] Attributes in **CATEGORY 2**: Agricultural; Campgrounds; Commercial; Commercial, Golf Courses and Driving Ranges; Government; Government, Parks and Recreation; Residential; Industrial; Institutional; Miscellaneous; Mobile Home or Park Misc. Residential.

STEP1 Find an arbitrary demand point q. Add q into the aggregated demand point set Q.

STEP2 Calculate the distance between every unaggregated point to set Q following minimum distance rule. i.e., $D(c, Q) = minD(c, q)\ \forall q \in Q$.

STEP3 Let Ω be the set of all points. Then define $d = max(distance(c, Q)), \forall c \in \Omega \backslash Q$.e., the maximum distance d and then add the point which is the farthest from set Q into Q.

STEP4 Repeat step 2 and 3 until $d < b$.

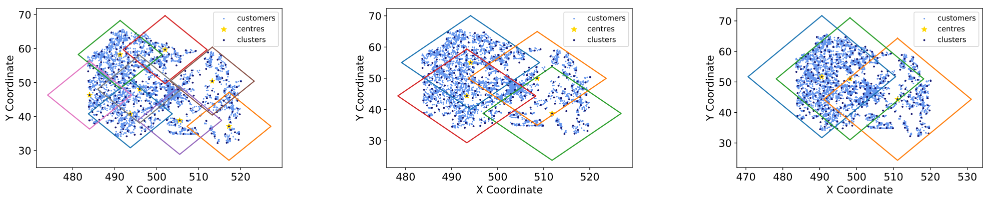
4 Model

To solve the covering facility location problem, we can model it as an integer programming problem: $min \sum_{j=0}^{N} x_j \quad s.t. AX \geq e$. A is a matrix with rows for customers and columns for potential DRC sites. If the 1-norm distance between customer i and potential DRC site j is less than r, $a_{ij} = 1$, otherwise, $a_{ij} = 0$. Decision variable X is a vector of 0–1 values. If the potential DRC site j is picked as a DRC, $x_j = 1$; otherwise, $x_j = 0$. e is a vector with all elements equal to 1. In this way, by solving the problem, it can obtain the minimum number of DRCs.

5 Result

In section 3.2, we have aggregated 6988 customers into 200 clusters, and the distance of each customer to the nearest cluster is no more than b. From Fig 2, we find all the clusters are covered by the DRCs in all of the three cases of r: 10, 15, 20 miles and the minimal number of DRCs is 3,4 and 8, respectively. Even in the case of $r = 10$ or 15 miles, some customers have not been covered. The distance from these customers to the nearest DRC will not exceed $r + b$ miles.

Then, Table 1 shows the DRCs performances with three cases. All customers are covered when $r = 20$ miles.



(a) b-Termination: DRC Locations with a 10 mile Travel Distance

(b) b-Termination: DRC Locations with a 15 mile Travel Distance

(c) b-Termination: DRC Locations with a 20 mile Travel Distance

Figure 2: DRC Locations with b-Termination

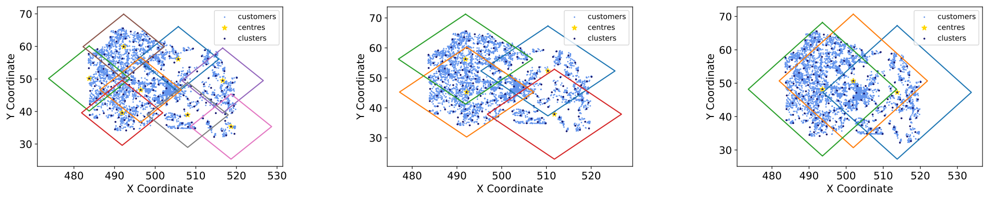
Table 1: Comparison of Different Radii with b Termination

Travel limit radius :	10 miles	15miles	20miles
Maximum travel distance (miles)	10.311	15.272	19.934
Average travel distance (miles)	6.222	7.921	8.907
%parcels within travel limit radius of a centre :	0.999	0.999	1.000
Average distance in excess of travel-limit for parcels farther from any centre than the travel limit (miles)	0.164	0.212	0.000

6 Analysis and Improvement

6.1 PTF Algorithm with p-centre Termination

During the experiment, we found the computational time is unstable, when termination parameter b is small. Thus, we tried to use PTF algorithm but stop with given size Q = 200. These DRC sites can not cover all customers with r = 20 miles (Fig 3(c)). This is due to the responding value for b is around 2.0 miles which provide a larger aggregate error. In addition, the covering of DRCs with a radius of 10 miles and 15 miles is shown in the Fig 3(a) and Fig 3(b) respectively. Comparing with Table 1 and Table 2,



(a) p-centre Termination: DRC Locations with a 10 mile Travel Distance

(b) p-centre Termination: DRC Locations with a 15 mile Travel Distance

(c) p-centre Termination: DRC Locations with a 20 mile Travel Distance

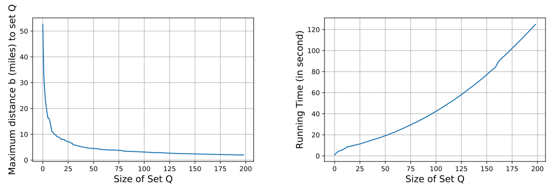
Figure 3: DRC Locations with p- centre Termination

Table 2: Comparison of Different Radii with p-centre Termination

Travel limit radius :	10 miles	15miles	20miles
Maximum travel distance (miles) :	10.859	15.741	20.932
Average travel distance (miles) :	6.281	8.698	8.606
%parcels within travel limit radius of a centre :	0.999	0.999	0.999
Average distance in excess of travel-limit for parcels farther from any centre than the travel limit (miles) :	0.416	0.286	0.624

6.2 $Q - b$ Analysis

Our article analyses the relationship between the size of the Q set and aggregation error b in Fig 4(a). It can be found that if the size of set Q increases, the parameter b tends to decrease exponentially, but the iteration time to obtain set Q will increase as shown in Fig 4(b).



(a) Relation between Size Q and b

(b) Iteration Time for Different Size Q

Figure 4: Analysis for Different Termination

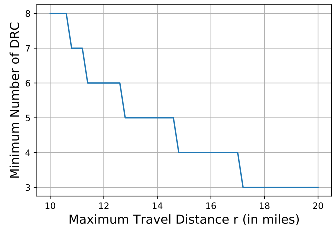
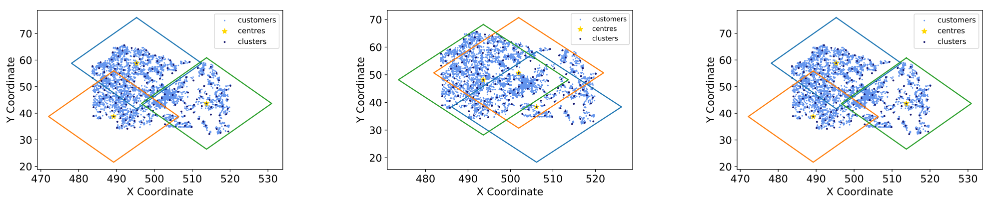


Figure 5: Relation Between r and Minimum Number of Building

6.3 The Relation Between r and Minimum Needed DRCs

From Fig 2(c), it is obvious that the optimal solution is not unique. For a fixed number of optimal value(i.e. the minimum number of DRC buildings), we estimate the less value of r is given, the less quantity of optimal solution can be obtained.It is expected to provide the roughly lower bounds of r for every different optimal value. [3]As shown in Fig 5, the lowest value of r to keep the minimum number of DRCs remaining three is around 17.2 miles. Actually, this optimal solutions are equivalent to the optimal solution for The Centre Problem with the number of building $p = 3$. More details are in Section 6.5.

We obtained one optimal solution by given r = 17.2 as shown in Fig 6(a). But this is not at a good location since the DRC we selected in the lower-left corner deviated seriously from the central area and it cover a lot of space where is not contain any customer. Therefore, we try to find more optimal solutions by given r = 17.2 by using this heuristic algorithm which might has a better location than this one. After removing these three optimal locations one-by-one and resolve them, we found that this problem has an only optimal solution that provides three DRC locations. The reason for this is we selected DRCs from the aggregated usable points. The number of potential DRCs is relatively small, and these points distribute randomly. Thus, it difficult to find another optimal point by shifting along with a few available directions limited by covered boundary clusters.



(a) Original Model with $r = 17.2$

(b) The Variation of Maximum Covering Model with $r = 20$

(c) Centre Problem Model

Figure 6: DRC Locations in SCM, VMCM and CPM

[3]For reduce the computational cost, we selected PTF algorithm with p-centre termination ($|Q| = 200$) in the following context.

We also tried to find all optimal solutions to the original problem by implement this algorithm. Given an initial optimal solution, in each iteration, one of DRC location will be randomly removed, until the minimum number of DRCs become 4. Then, we reset anything and do this operation many times to obtain as many optimal solutions as possible. The more loop times we did, the more accurate number of optimal solution we will obtain. In our last attempt with four loops, there are 74 new optimal solutions in the original problem. Since in practical, the problem often require the approximate DRC locations. Find all optimal solution will increase the computational cost and also useless. In this case, the 75 optimal locating strategies can be plotted and evaluate manual to adjust the practical condition, and it is better than only one optimal solution obtained by the solver.

6.4 Variation of Maximum Covering Problem

When a disaster occurs, the more DCRs within the travel distance for each customer, the more assistance will provide for each customer in time. Therefore, we apply a variation of The Maximum Covering Problem (VMCP) for this optimization objective:

$$
\begin{aligned}
& \max \sum_i z_i \\
s.t. \quad & \sum_j a_{ij} x_j \geq z_i \quad i \in I, \\
& \sum_j x_j = 3, \\
z_i \geq 1 \quad & z_i \in \mathbb{Z} \quad x_j \in \{0,1\} \quad i \in I, \quad j \in J.
\end{aligned}
$$

The objective function maximizes the sum of all potential DRCs that can provide assistance to each customer within a specified radius r when a disaster occurs. The decision variables z_i are integers larger or equal to 1, representing the selected number of DRCs that cover customer i. I and J are the sets of all the aggregated customer and is aggregated potential DRCs. The first constraint indicates the selected number of DRCs that cover each customer is no more than the maximum number of DRCs within radius r. The second constraint fixed the total number of DRCs. In this case, we set it to be 3, 4 and 8 to compare with the original problem. The third constraint forces all the customers are covered by at least one DRC. For r equal to 20 miles, we find that there is only one optimal solution by using the heuristic method (terminate when optimal value change) in 6.2 (Fig 6(b)). Compared to the original performance, VMCP provides a better repeat cover rate, as shown in Table 3.

Table 3: Performance of SCP and VMCP

Model	10 miles		15 miles		20 miles	
	SCP	VMCP	SCP	VMCP	SCP	VMCP
%parcels covered by at least one centre	0.9991	0.9991	0.9991	0.9991	0.9993	0.9959
%parcels covered by at least two centre	0.2155	0.2274	0.5102	0.5102	0.7873	0.7785
%parcels covered by at least three centre	0.0127	0.0167	0.0311	0.0311	0.4389	0.4662

6.5 Centre Problem

To reduce the loss caused by long-distance assistance. We consider a classic problem: The Centre Problem (CP).

$$
\begin{aligned}
& \min \sum_i z_i \\
s.t. \quad & y_{ij} \leq x_j \quad i \in I, \quad j \in J, \\
& \sum_j y_{ij} = 1 \quad i \in I, \\
& z_i \geq \ d_{ij} y_{ij} \quad i \in I, \quad j \in J, \\
& \sum_j x_j = 3, \\
x_j \in \{0,1\} \quad & y_{ij} \in \{0,1\} \quad z_i \in \mathbb{R}, \quad i \in I, \quad j \in J.
\end{aligned}
$$

The objective function is to minimize the maximum distance z between any resident and its selected DRC. The maximum distance can be satisfied through the third constraint. The first and second constraint represent each resident can only be allocated by one decided DRC. The fourth constraint means we establish exactly three DRCs as the same. For the decision variable, $y_i j$ = 1 if and only if the resident i is allocated to DRC j. If the DRC site j is picked to build, x_j = 1; otherwise, x_j = 0. We define I as the set of all aggregated customers and J as the set of all the aggregated potential DRCs. We solve it by using FICO Xpress and the optimal solution is 17.12 miles, which is very close to 17.2 miles. The result we obtained in 6.2 and the graph for this solution is shown in Fig 6(c). Furthermore, comparing with Fig 6(a) and Fig 6(c). they provide the same locations, which verify our hypothesis in section 6.3.

7 Conclusion

We preprocess the data and use it to replicate the first step, which contains PTF algorithm and Set Covering Problem (SCP) Model. Furthermore, through analyzed the original problem with discussing the relation between the size of cluster set Q and aggregate error b, a new method is introduced to control the iteration time by use p-centre termination. Then we provide The variation of Maximum of Covering Problem (VMCP) and The Centre Problem (CP) as model improvement. VMCP improve the quality of the optimal solution from cover rate respect. CP improve it from optimizing edge customer. By analysing the staged lower bound value of travel distance r, we found CP optimal solution can be achieved by using lower bound value of r with solving multiple times of SCP.

References

[1] M. L. Brandeau and S.S. Chiu. An overview of representative problems in location research. *Management Science*, 35(6):645–674, 1989.

[2] Fei Wang, Yu Xu, and Yixue Li. Review of natural sciences-mathemates-operations research - discrete facility location problem. *Chinese Journal of Academic Journal Abstracts*, 13(7):1–1, 2007.

[3] J. Dekle, M. S. Lavieri, E. Martin, H. Emir-Farinas, and R. L. Francis. A Florida County Locates Disaster Recovery Centers. *Interfaces*, 35(2):133–139, April 2005.

Summary

The article we chosen to replicate is *A Florida County Locates Disaster Recovery Centers*. In this article, we tend to find potential disaster recovery centres (DRCs) for Alachua County, Florida, which makes sure at least one DRC supports each customer in the event of a disaster. Our paper has completed the first two parts of the original paper: used the Pick-the-Farthest (PTF) Algorithm to cluster customers and usable sites and modeled The Set Covering Problem (SCP) to find the optimal solution. Besides, we also analysed the clustering algorithm of the original paper. We tried a new termination with a given size of aggregation set named PTF algorithm with p-centre termination. We compared the results and discussed the relationship between two different terminations. Then analyzed the relationship between the size of the demand point set Q, and the aggregation error b. We found the critical value of the covering radius and the corresponding optimal solution by given fixed number of DRCs. Different models are used to solve this covering problem and evaluated their results, such as the variation of The Maximum Covering Problem and The Central Covering Problem.