Site Location Optimization Based on Genetic Algorithm

Yahui Zuo¹, Shan Huang¹, Zijian Feng¹, Yajie Dou²* and Xiangqian Xu²

¹Department of Operations Research and Mission Planning, National University of Defense Technology, Changsha, Hunan, China
²Department of Management Science and Engineering, National University of Defense Technology, Changsha, Hunan, China
Email: 1832538491@qq.com; *yajiedou_nudt@163.com

Abstract. With the goal of providing food to people in need, the FBST serves nearly 4,000 square miles in six counties in New York State. The MFP program is among the main activities of FBST, using an MFP truck loaded with food to serve 70 sites in New York State. An effective and fair schedule for visiting all 70 sites is necessary to optimize the MFP plan. In order to describe the problem, a triad {G, R, C} of goal, resource and constraint is established according to the decision theory. Based on the established triad, this paper establishes a multi-objective optimization model with the goal of improving validity and fairness, a genetic algorithm has been adopted to solve the problem. The time schedule for visiting sites is finally obtained after the iteration of 2000 times, which took 156.40 seconds. Considering that customers can go to further sites, the number of service sites and the location of sites can be optimized. The K-means clustering algorithm based on the improved demand centroid solution method is used to solve the site location selection. After optimizing the location of the site, the performance of senior-site and regular-site systems is improved by 30.755% and 36.413% respectively.

Keywords. Triad {G, R, C}; multi-objective optimization; genetic algorithm; K-means clustering algorithm.

1. Introduction

1.1. Background

1.1.1. Introduction to Food Distribution System. In the current era of big data with developed Internet, people’s purchase of daily materials is no longer limited to the offline way, most people would choose to buy the daily materials they need online, and then material circulation departments will carry out point-to-point distribution of materials.

In the United States, Feeding America operates a network of organizations dedicated to providing essential material distribution services for people with limited financial means. One of them, the Southern Food Bank (FBST), which is also the focus of this paper, provides distribution services in six counties and nearly 4,000 square miles of New York State.

In the above the operation mode of the network system, mobile food storeroom (MFP) program is one of the main operations of FBST, the goal is to make more accessible to people who are live in relatively more gated communities with nutrition and health food. However, even in places where other organizations to provide assistance, due to the choices of transport conditions or these institutions operating time, can led to serve customers do not always get food they needed.
1.1.2. The Introduction of Specific Distribution Mode. Typically, MFP provides food directly to customers through prepackaged boxes or farmers’ market-style delivery. When an MFP truck arrives at a service area, its staff put the food on tables around the truck. Customers can choose what they need to purchase.

MFP has a total of three trucks, each capable of carrying up to 15,000 pounds, and FBST usually has enough food and volunteers to operate two of them on any given day. A typical mobile pantry visit lasts two hours, and the schedule for all MFP food flow programs is posted months in advance to help customers plan their food purchases accordingly.

Unfortunately, this year FBST was forced to revise its service model due to the impact of the covid-19 outbreak. There are many limitations in the current operating model of MFP, and the operating procedures used by customers are very limited in scope and flexibility. FBST plans to return to pre-pandemic MFP service levels next year, allowing customers to show up when trucks visit without having to register in advance.

1.2. Our Works
After consideration, this paper decides to choose the first, second and fourth tasks as the three problems studied in this paper.

First of all, according to the statistical data of 2019 MFP given in the document of the title, this paper designed an optimized aid schedule for 70 regular stations in view of the actual situation in 2021, so as to achieve the purpose of aid for 70 stations and meet the requirements of the title in terms of effectiveness and fairness.

Then, through analysis, we understand the situation that people are willing to go to conventional sites farther away to receive assistance, so the solution of this paper adopts the method of reducing the number of sites.

At the same time, the location deployment of each site was planned again. After completion of the optimization, a quantitative system was established to evaluate the improvement of the overall MFP system performance by the optimization of the site location mentioned above.

After completing the analysis and solution of the first two problems, this paper makes executive summaries of the above two tasks, explaining the main advantages and potential disadvantages of the method provided in this paper.

2. The Description of the Problem

2.1. Problem Statement
Based on the data collected by the 2019 FBST in the MFP program, this paper will optimize the access plan for 70 sites in 2021 and come up with a supply schedule suitable for 2021. According to the analysis of historical data, due to the cold winter, the number of people buying food at the site is much less than usual. However, when the weather improves, the number of people be served at each site will reach the maximum.

This information indicates that the serve can go to a remote site to receive the service. Based on this information, this article plans to reduce the total number of sites and optimize the location layout of the sites to improve the performance of the system and to quantify the extent of the improvement.

2.2. Analysis of Specific Issues

2.2.1. Analysis of Problem 1. In the process of solving problem 1 we need to obtain the schedule of visits to 70 sites in 2021, according to the data given in 2019, and the service for each site is required to be effective and fair. This article mainly considers two elements in the schedule:

(1) The number of visits to each site.
(2) Time spent visiting each site.
Due to the lack of information in 2019, the visit time of each site involves too many factors and their influence effects are not clear, so this paper will focus on optimizing the visit times of each site, and the visit time of each site will be determined by the visit times and the time arrangement rules designed in this paper.

2.2.2. Analysis of Problem 2. In the solution of problem two, we know that people may go to further sites to get information about the materials they need. This problem needs to reduce the total number of sites and optimize the location of sites to achieve higher system performance. People may go further afield to get help, but the time and money spent travelling the distance must not be greater than the value of getting the food. People still have to move as little distance as possible.

The movement of the site also imposes a financial burden on the service organization, so the sites cannot be changed on a large scale, and most sites need to be retained for processing. Combined with the analysis idea of the first question, this paper will build a multi-objective optimization model and obtain a satisfactory solution.

3. Basic Assumption
(1) The total demand for each site in 2021 is the same as the total demand for 2019.
(2) The total number of FBST services provided to each site in 2021 is the same as in 2019.

4. Glossary & Symbols

4.1. Glossary
Effective degrees: FBST every time on site service, service effectiveness from average to attend services. The fewer people attending the service at a time, the better the service is, the more effective it is.

Fair degree: each site service to be as fair as possible, the effective degree of each site as consistent as possible, this article adopts the variance of each site effectiveness to represent the sum FBST service for each site to fair degree.

4.2. Symbols

| Symbols | Definition |
|---------|------------|
| $a_{id}$ | The $i_{th}$ point of MFP |
| $x_{i}$ | Number of accesses at point $i_{th}$ of MFP |
| $a_{d_{i}}$ | The average demand for food supplies at point $i$ |
| $D_{i}$ | $i_{th}$ of MFP Total demand for point $i_{th}$ |
| $m$ | Total number of services available by FBST |
| $M$ | Targets of problem |
| $S$ | Schedulable resources |
| $C$ | The problem constraints |

5. Models

5.1. Analysis and Solving of Question One

5.1.1. Model Establishment. In this paper, a triplet $\{G, R, C\}$ is used to describe the problem, where $G$ represents the goal of the problem, $R$ represents the schedulable resource of the problem, and $C$ represents the constraint of the problem. The conceptual diagram drawn is shown in figure 1.
Next, this paper will optimize the visit schedule of each site in 2021, so as to improve the effectiveness and fairness of service for each site, which are the two goals of this problem.

The less people go to the site to receive the service, the more food they will get, and the less time they will spend waiting in line. Therefore, the smaller the average demand for a site, the better the site will serve the people who go to the site to receive the service, that is, the higher the effectiveness will be. Thus, this paper uses the average demand of each site to describe the effectiveness. The sum of the average demand of each site represents the degree of system effectiveness. The smaller the sum of the average demand, the stronger the effectiveness of the system.

The fairness requirement requires that on the basis of solving the validity, the validity of services provided by different sites should be as same as possible. Therefore, this paper uses the sum of the validity variance of each site to describe the fairness of the system. Here, this paper believes that the smaller the sum of the validity variance, the better the fairness of the whole system.

G has two goals.

The first goal is the effectiveness of the system:

\[ \min z_1 = \sum_{i=1}^{n} ad_i \]

\(ad_i\) represents the number of evaluated services at the \(i\)th site, \(ad\) represents the average demand. The second goal is system fairness:

\[ \min z_2 = \sum_{i=1}^{n} (ad_i - ad)^2 \]

\(R\) is the dispatchable resource in this problem. According to the information in the question, the MFP plan has three trucks, sufficient food, 722 services, and personnel capable of operating two trucks at the same time.

Then, constraint C of the problem can be obtained: three or more trucks cannot be operated at the same time, the total number of services cannot exceed 722, and the number of services at each station is at least one time.

\[ \min z_1 = \sum_{i=1}^{n} ad_i \]  
\[ \min z_2 = \sum_{i=1}^{n} (ad_i - ad)^2 \]  
\[ \sum x_i \leq m \]  
\[ ad_i = \frac{D_i}{x_i} \]
ad_i represents the number of evaluated services at the i-th site, ad represents the average demand. The validity goal of the service is represented in equation (1).

Equation (2) represents the fairness goal of the service.

In equation (3), m is the total number of visits to the site, and represents the constraint of the number of visits.

5.1.2. Model Solving. According to the data in 2019, some sites specifically serve the senior, and the data are quite different from that of the regular sites.

In this paper, for the purpose of practical and make the model more accurate, the problem is divided into service for the senior of the site with conventional two sites, according to multi-objective optimization model of this article, use the elite preserving genetic algorithm for the site distribution compared to the original plan better access (visits) frequency, and then through the algorithm of random time distribution of 2021 schedules.

Frequency allocation based on genetic algorithm, the algorithm flow chart is shown in figure 2:

![Figure 2. Flowchart of genetic algorithm.](image)

Genetic algorithm (GA) is a heuristic algorithm that searches for the global optimal solution by randomly generated possible solutions. Each possible solution is represented by a coded chromosome, and based on the principle of “survival of the fittest and survival of the fittest”, individuals with low fitness are eliminated, and individuals with high fitness are selected for genetic operator operation to generate a new generation of population, until the local optimal solution or optimal solution is obtained.

Genetic algorithm consists of several important elements, such as coding, population size, fitness function, genetic operator, stop criterion, etc. In order to prevent the loss of the optimal individuals in the current population in the next generation, the algorithm also adopts the elite reservation strategy. The following introduces some specific implementation details of the algorithm. The crossover process of genetic algorithm is shown in figure 3.

- The encoding
Encoding target problem space can be mapped to the algorithms of coding space, this paper USES the symbol encoding will be a site for the assigned frequency as a gene, its index represents the site number, for example, there is a “5 12 7 6 20” gene fragments, this part of the gene fragment represents its index corresponds to the site (site) respectively arranged 5,12,7,6,20 event in 2021.

- Population formation
In order to compare with the data of 2019 on the same level as possible, each gene generation is required to meet the length of 70, and the total is 2019. For all numbers in the gene, the sum is the sum of the total number of visits by the corresponding sites in 2019, and cannot be zero. The population size is set to 100.

![Figure 3. Diagram of the algorithm crossover.](image)

- Fitness function
Fitness reflects the degree of adaptation of a species to its living environment. The lower the fitness of an individual, the easier it is to be eliminated. The fitness can guide the search of the algorithm to improve the direction. In this problem, according to the objective function:

$$\min z = \frac{\sum_{i=1}^{n} ad_i}{n} + \frac{\sum_{i=1}^{n} (ad_i - \bar{a}d)^2}{n}$$

The optimization direction of objective function is “min”, so the fitness function is constructed as follows:

$$\delta(x) = 1000 \times \left( \sum_{i=1}^{n} \frac{D_i}{x_i} + \sum_{i=1}^{n} \left( \frac{D_i - \bar{D}}{x_i} \right)^2 \right)^{-1}$$

- Genetic operators
Selection operator, crossover operator and mutation operator constitute genetic operator. These three operators are the core of the whole genetic algorithm, and make the genetic algorithm has a very powerful search ability. According to these three operators, genetic algorithm can simulate the genetic process of biological in nature.

1. Select operator
   Selection operator is the process of simulating biological evolution “survival of the fittest”.
   In this paper, roulette selection method is used to select individuals. When the population size is n, the fitness function of the $i_{th}$ individual in the population is $f(x_i)$, then the probability $P(k)$ of the individual being selected is as follows:
\[ P(k) = \frac{f(x_i)}{\sum_{j=1}^{n} f(x_j)} \]

(2) Crossover operator

The crossover probability is the likelihood that the chromosomes of two individuals to mate will exchange genes.

After selecting an individual \( \delta_1 \) in the population according to the selection result, the random number is used to determine whether to carry out crossover operation and the segment to carry out crossover operation. Then, a gene \( \delta_2 \) is selected in the population and a random position is determined to exchange this segment of the two genes, and the difference value of the exchanged segment is recorded.

If the interpolation is not 0, it continues to find another location to swap and update the interpolation until the interpolation is 0. If no interpolation is found, the crossover is abandoned. The crossover probability is set to 70\% based on experience [1].

(3) Mutation operator

Mutation operation is to simulate the phenomenon of genetic mutation of a gene on individual chromosome in the natural biological evolution. The mutation operation is carried out in the selected individual, and the two gene values to be operated need to be determined by random number. The value of one gene fragment is +1, and the value of the other is -1. If 0 occurs, the mutation will be carried out again. In some cases, mutation operation and crossover operation are beneficial to prevent genetic algorithm from falling into local optimum. The probability of mutation is set to 10\% [26]. The mutation process of genetic algorithm is shown in figure 4.

Figure 4. Diagram of the genetic algorithm mutation method.

- Elite retention strategies

In order to prevent the loss of the optimal individuals of the current population in the next generation, which leads to the failure of the genetic algorithm to converge to the global optimal solution, the management strategy is adopted to directly copy the best individuals (i.e., the individuals with the highest fitness) that have emerged so far in the evolutionary process of the population to the next generation without matching and crossing [2-4].

- Stop criteria

When the number of iterations reaches the preset value, the iteration will be stopped and the results will be output. After several runs of the program and observation of the results during the period, we find that the results are generally stable around 1200 times, so the number of inheritances is set to 2000 times. The iterative process of genetic algorithm is shown in figure 5.

Figure 5. Diagram of the best gene’s grade of every generation.
Random time allocation algorithm:
Based on the frequency of generation, a random time allocation algorithm is used to generate a one-year schedule.
When assigning tasks, there are a few rules:
• Try to evenly distribute the number of visits to each site over the course of a year.
• No more than three sites are active each day.
• Some statutory holidays and Sundays are taken into account.

5.1.3. Results. The paper takes the first six days of January to show the results. The results are shown in table 1.

| Date     | Mission                                    | Date       | Mission                                   |
|----------|--------------------------------------------|------------|-------------------------------------------|
| 1-Jan    | Rest                                       | 4-Jan      | MFP Senior - Conifer Village              |
|          | 1: MFP College Corning Community College   | 2: MFP Senior - Dayspring                      |
| 2-Jan    | MFP Redeemer Lutheran Church               | 5-Jan      | MFP Senior - Village Square/Manor         |
|          | 3: MFP Senior - Northern Broome             |            | MFP Waverly                               |
| 3-Jan    | Rest                                       | 6-Jan      |                                          |
|          | 1: MFP Pratt-burgh                          |            |                                          |
|          | 2: MFP Senior - Wells Apartments            |            |                                          |

After 2000 iterations of the genetic algorithm (Iteration), the best gene’s grade is 0.891228, and the corresponding Part of the frequency table about Senior sites is obtained.
After 2000 iterations of the genetic algorithm (Iteration), the best gene’s grade is 0.095789, and the corresponding Part of the frequency table about other sites is obtained.

5.1.4. Analysis of the Result
• Part of Frequency Table and Schedule have been listed above, and all the results can be found in the attachment.
• We picked using the algorithm on the basis of the fitness function of the frequency of the original data for evaluation, pro 2019 sites frequency (grade) is 0.5938549, the fitness of other sites frequency (grade) of the fitness of 0.083943, comparing the optimized frequency arrangement of 2021, the Senior sites frequency (grade) increased by 50.075%, the fitness of at the same time, Senior sites frequency fitness (grade) increased by 14.111%. It can be seen that after optimization, the activity frequency of each site is more reasonable, which can make the fairness and effectiveness of the activity better.

5.2 Analysis and Solving of Question Two

5.2.1. Model Establishment. We hope to improve the performance of the system by reducing the total number of sites and optimizing the location of sites, based on the additional information provided that “people can go further to obtain services”. People travel further to obtain services that increase the cost of time, and the cost of travel adds to the burden. FBST is an organization that reduces the burden of living for people in need, so reducing the cost of travelling people will be a goal of this inquiry.

The adjustment of the site location will involve the organization and the new site negotiation time and other cost issues, the original site users are not satisfied with the problem, then the adjustment of the site should be a small range of adjustment, the range of adjustment cannot be too large. The problem concept diagram is shown in figure 6.
Here we set the constraint that 60% of the original number of sites should be saved. The multi-objective optimization model of the first question is continued, but the above analysis shows that the triplet \{G, R, C\} describing the problem will be changed. The adjustment of site location will increase the cost of time and money for the users of the original site, so the goal of minimizing people’s travel cost is added in G.

According to the actual situation, the adjustment range should not be too large, so the constraint of at least 60% should be retained when the original site is added in C.

On the basis of Problem 1 model, a new multi-objective optimization model is established by adding targets and constraint conditions as follows:

\[
\begin{align*}
\text{min } z_3 &= \sum_{i=1}^{n} |L_i - L'_i| \\
\frac{|L_i|}{|L|} &\leq 0.4
\end{align*}
\]  

In equation (5), \(L_i\) represents the position vector of the original site to be replaced, and \(L'_i\) represents the position vector of the new site to be replaced by \(L_i\). This formula indicates the minimum distance between the original site replaced by the new site and the new site.

In equation (6), \(L_i\) represents the original site set replaced, and \(L\) represents the original site set. This formula indicates that no more than 40% of the original site should be replaced.

5.2.2. Model Solving. For this problem, due to too much unknown information, the detailed parameters of the multi-objective optimization model, such as the priority relationship between targets, cannot be accurately determined, resulting in a relatively brief model. In this paper, the demand traction method is used to solve the second problem to simplify the model calculation and get a satisfactory solution.

In this paper, an improved K-means [5] clustering algorithm based on site requirements and location is adopted to optimize site location, so that a satisfactory solution with low cost of distance consumption can be obtained under the condition of satisfying the constraints. The range of activities people can move can be seen as a circle, and the closer the location of the new site is to the center of the circle, the lower the cost of people going to the new site.

In order to reduce the distance people had to move more than the original site location and reduce the total number of sites at the same time to improve system performance, this paper uses the improved
K-means [6] clustering method to gather some points together to reduce the total number of sites, and according to the average demand and location of the original site, the clustering center is the location of the new site.

\[
\min E = \sum_{i=1}^{k} \sum_{x \in C_i} |x - \mu_i|^2
\]  \hspace{1cm} (7)

\[
\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x
\]  \hspace{1cm} (8)

\[
\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} ad_i x
\]  \hspace{1cm} (9)

In equation (7), \(C_i\) is the \(i\)th cluster of the cluster, and \(U_i\) is the center of mass of the cluster \(C_i\).

This equation is the objective function of k-means clustering and represents the minimum squared error \(E\). In equation (8) is the original centroid equation of K-means [7-9] algorithm.

In equation (9) adds the centroid formula of site demand information to the improved K-means algorithm.

5.2.3. Results. The locations of all sites are not shown here. See attachment for all site locations. The paper takes the first ten days of January to show the results. The results are shown in tables 2 and 3.

| Table 2. Location table. |
|---------------------------|
| MFP-SNEW1 42.1029 -75.91 | MFP-NEW1 42.0058 -76.538 |
| MFP-SNEW2 42.1661 -77.091 | MFP-NEW2 42.4578 -76.537 |
| MFP-SNEW3 42.4299 -76.499 | MFIP-NEW3 42.0882 -76.801 |
| MFP-SNEW4 42.4971 -76.316 | MFP-NEW4 42.0961 -75.91 |

| Table 3. The timetable. |
|-------------------------|
| Date | Mission | Date | Mission |
| 1-Jan | Rest | 6-Jan | 1: MFP Senior-Ellis Hollow |
| 2: Freeville |
| 3: tmp1 |
| 2-Jan | 1: Canisteo | 7-Jan | 1: Richford |
| 2: Woodhull | 2: Windsor |
| 3: Woodhull | 3: tmp1 |
| 3-Jan | Rest | 8-Jan | Rest |
| 4-Jan | 1: Owego | 9-Jan | 1: Corning |
| 2: Windsor | 2: Freeville |
| 3: Woodhull | 3: Van Etten |
| 5-Jan | 1: Canisteo | 10-Jan | Rest |
| 2: Woodhull |
| 3: Woodhull |

5.3. Analysis and Solving of Question Two
The following is the executive summary of this paper, to analyze the advantages and disadvantages of model building and problem solving.
6. Conclusions
The problem described in this paper is based on the triple \( \{g, R, C\} \) canonical analysis of this problem. Based on the objective function of triple \( \{g, R, C\} \), the cost of quadratic distance addition and related constraints are analyzed, and the multi-objective optimization model of this problem is established. Finally, the improved k-means clustering algorithm and genetic algorithm are used to solve the model, and the access plan and site location optimization scheme of each site in 2021 are obtained. This paper provides a clear idea and feasible scheme to solve such problems.

References
[1] Guo J, Zhao P and Ni J 2020 Fire station location planning model based on genetic algorithm Journal of Computer Applications 40 (S1) 41-44.
[2] Zhang Y, Zhang B, Zeng M, et al. 2020 Location selection model of logistics park based on cost optimization Journal of Nanjing University of Science and Technology 44 (3) 372-378.
[3] Hamza A, Nassim M, Abderrahmane B, et al. 2021 CSFCM: An improved fuzzy C-Means image segmentation algorithm using a cooperative approach Expert Systems with Applications 166.
[4] Sharma M and Garg R 2020 HIGA: Harmony-inspired genetic algorithm for rack-aware energy-efficient task scheduling in cloud data centers Engineering Science and Technology: An International Journal 23 (1).
[5] Abualigah L and Alkhrabsheh M 2021 Amended hybrid multi-verse optimizer with genetic algorithm for solving task scheduling problem in cloud computing The Journal of Supercomputing.
[6] Kirti K and Prakash S J 2020 Identification of cyberbullying on multi-modal social media posts using genetic algorithm Transactions on Emerging Telecommunications Technologies 32 (2).
[7] Hartigan J A and Wong M A 1979 Algorithm AS 136 Journal of the Royal Statistical Society Series C (Applied Statistics) 28 (1) 100-108.
[8] Wang C and Wu D H 2021 A K-means clustering-based multiple importance sampling algorithm for integral global optimization Journal of the Operations Research Society of China 1-19.
[9] Hartigan J A and Wong M A 1979 A K-means clustering algorithm J R Stat Soc Ser C Appl Stat 28 100-108.