For office use only	Team Control Number	For office use only
T1 _____	**41125**	F1 _____
T2 _____		F2 _____
T3 _____		F3 _____
T4 _____	Problem Chosen **A**	F4 _____

2015

Mathematical Contest in Modeling (MCM/ICM) Summary Sheet

(Attach a copy of this page to your solution paper.)

Abstract

Ebola has killed thousands of lives since it was found at 1976. Fortunately, scientists have discovered new medication to stop Ebola and cure patients. Hence, it's critical to optimize the eradication of Ebola. Our group generalized several models to solve this optimization problem which considered some critical factors in reality.

To optimize the eradication of Ebola, we consider about how to site the drug delivery center and how to make a good schedule scheme to deliver the drugs to epidemic areas. Obviously, a good location of the drug delivery center and a proper schedule scheme to deliver the drugs are important factors to improve the efficiency of the delivery system and meet the need of the epidemic areas.

The first model we established is SIR model which is a common model to consider the transmission of infectious disease. Applying SIR model is convenient for us to predict the seriousness of the epidemic in the future in order to choose an optimal delivery route in a modified MDVRP (multiple depot vehicle routing problem) model which is a model based on VRP but modified to adjust the problem of optimizing the eradication of Ebola. Later we will talk about this modified MDVRP model in detail.

The second model is to explore a proper site for the drug delivery center considering different factors which obviously affect the eradication of Ebola such as the seriousness of the epidemic area gained from SIR model, the distance to the epidemic area and so on. We generalized a modified K-Means algorithm to select several centers considering not only the distance from the center to the epidemic area but also some important factors mentioned above.

The third model is to achieve an optimal scheduling scheme for vehicles in drug delivery centers. In order to optimize the scheduling problem, we established a modified MDVRP model mentioned above which is a natural extension of VRP (vehicle routing problem). VRP and MDVRP have been proven to be NP-hard problems. Hence, they're hard to be solved using traditional mathematical methods when the scale of the problem grows large. In this paper, we proposed a hybrid genetic algorithm to optimize the MDVRP while considering not only the distance to the epidemic area (though the disaster is serious, it's still important to consider the cost of distance because more time is needed when the medicine is delivered to a remote area and it will also cause the increase of some other areas because of the late arriving of the medicine) but also the seriousness of the epidemic area predicted by SIR(that's to say, in spite of the long distance, if the seriousness of a certain epidemic area is at a high level, it may be considered at a high priority).

Fighting against Ebola

--- A Heuristic Approach

1 Introduction

1.1 Restatement of the problem

Ebola virus disease outbreaks in West Africa, so we need to establish a model to help optimize the eradication of Ebola using the new vaccines and drugs. This is an absolute open-ended problem, but in the view of the differences of the epidemic areas, the speed of manufacturing of the vaccines or drugs and so on, we must take into account the following questions when we establish our models:

- In present circumstances, according to the laws of the transmission of epidemic, how can we retard the speed of the spread.
- Because the condition of medicine manufacturing is limited, the present speed of manufacturing can't meets the needs of every area. Then here comes the problem how to distribute the vaccines and drugs.
- In order to alleviate the epidemic in the shortest time, the shorter the route, the better. And we also need consider the different severity of the epidemic of each area. Hence, we have to take all the prerequisites into account to choose the delivery centers and design the delivery system.

1.2 Background

Ebola virus disease (EVD; also Ebola hemorrhagic fever), or simply Ebola, is a severe, often fatal disease in humans and other primates caused by Ebola viruses. The average EVD case fatality rate is around 50%. Case fatality rates have varied from 25% to 90% in past outbreaks. [1]The virus is transmitted to people from wild animals and spreads in the humans by direct contact with the blood or body fluids of an infected person.

Ebola first appeared in 1976[2] in 2 simultaneous outbreaks, one in Nzara, Sudan, and the other in Yambuku, Democratic Republic of Congo. The latter occurred in a village near the Ebola River, from which the disease takes its name.

The current outbreak in west Africa, (first case notified in March 2014), is the largest and most complex Ebola outbreak since the Ebola virus was first discovered in 1976. There have been more cases and deaths in this outbreak than all others combined. It has also spread among countries starting in Guinea then spreading across land borders to Sierra Leone and Liberia, by air (1 traveller only) to Nigeria, and by land (1 traveller) to Senegal. The most severely affected countries, Guinea, Sierra Leone and Liberia have very weak health systems, lacking human and infrastructural resources, having only recently emerged from long periods of conflict and instability. On August 8, the WHO Director-General declared this outbreak a Public Health Emergency of International Concern.

According to latest statistics from the WHO, there have been almost 22 500 reported confirmed, probable, and suspected cases of EVD in Guinea, Liberia and Sierra Leone, with almost 9000 reported deaths (outcomes for many cases are unknown). [3]

Fortunately, the world medical association has announced that their new medication could stop Ebola and cure patients whose disease is not advanced. In order to take good use of the new vaccines and drugs, WHO must take scientific rescue measures.

Thus, based on the analysis of the trend of Ebola's transmission by SIR epidemic model, this paper designs and simulates a possible delivery system (considering locations of delivery, speed of manufacturing of the vaccine or drug, routes and so on), aiming at the eradication of Ebola, or at least its current strain.

[1][2] http://www.who.int/mediacentre/factsheets/fs103/en

[3] http://apps.who.int/ebola/en/ebola-situation-report/situation-reports/ebola-situation-report-4-february-2015

2 Assumptions

We make the following assumptions about 3 models in this paper.

2.1 SIR Model

- We assume the population size is large and constant, $S(t) + I(t) + R(t) = N$. We do not take into account the birth, death (caused by other reasons except Ebola), immigration or emigration.
- The influence of latent period can be neglected.
- We define the number of susceptible people S(t) proportional to the number of the product of I(t) and S(t),and μ is positive and constant.

2.2 Weighted K-means Algorithm

- Two significant factors are taken into account to decide the location in this module, magnitude of epidemics and distance between the medicine delivery center and epidemic areas. Ignore the other factors.
- The number of medicine delivery center is set as a constant with proper value.
- Cities with different numbers of patients have a corresponding ‘weight’ which affects the position of medicine delivery center.

2.3 Modified MDVRP Model

- In order to simplify the model, we ignore the effect of the amount of vehicles and assume the amount in every delivery center is unlimited.

3 SIR Epidemic Model (spread of disease)

Ebola virus disease is a typical epidemic. We take advantage of a classical epidemic model, SIR model[①], to analyze the trend of Ebola and the changes after receiving the medicines.

3.1 Terms, Definitions and Symbols

- N --- the population of the city
- t --- time
- $S(t)$ --- the time-based parameters, denoting the number of susceptible people
- $I(t)$ --- the time-based parameters, the number of infected people
- $R(t)$ --- the time-based parameters, the number of recovered people and death
- λ --- the propagation coefficient of disease
- μ --- the removal rate constants from groups I to group R

3.2 The Foundation of Model

According to the spread principle of the infectious disease, establish a classical SIR model.

i. According to the assumption, each infected person infects $\lambda S(t)$ susceptible people daily. The total number of patients is $N*I(t)$, so $\lambda S(t)I(t)$ susceptible people are infected every day, then the rate of the decrease of susceptible people is

$$\frac{dS(t)}{dt} = -\lambda S(t)I(t)$$

ii. The change of the number of the infected people include the increase of new infected people and the decrease of the removed people, thus the rate of the increase of infected people is

$$\frac{dI(t)}{dt} = \lambda S(t)I(t) - \mu I(t)$$

iii. The change of the removed people (including recovered and death) equals to the decrease of the infected people every day, so

$$\frac{dR(t)}{dt} = \mu I(t)$$

In summary, the variables above satisfy the following model relationship:

$$\begin{cases} \frac{dS(t)}{dt} = -\lambda S(t)I(t) \\ \frac{dI(t)}{dt} = \lambda S(t)I(t) - \mu I(t) \\ \frac{dR(t)}{dt} = \mu I(t) \end{cases}$$

$S(t) + I(t) + R(t) = N, I(t_0) = I_0, S(t_0) = S_0.$

3.3 The Analysis of Model

By $\begin{cases} \frac{dS(t)}{dt} = -\lambda S(t)I(t) \\ \frac{dI(t)}{dt} = \lambda S(t)I(t) - \mu I(t) \end{cases}$ (*)

we have $\frac{dI(\mathrm{t})}{dS(\mathrm{t})} = \frac{\lambda S(\mathrm{t})\,\mathrm{I}(\mathrm{t}) - \mu\,\mathrm{I}(\mathrm{t})}{-\lambda S(\mathrm{t})\,\mathrm{I}(\mathrm{t})} = -1 + \frac{\mu}{\lambda S(\mathrm{t})}$.

Then $I(\mathrm{S}) = -\mathrm{S} + \frac{\mu}{\lambda}\ln S + c$.

According to the initial conditions, $I(t_0) = I_0, S(t_0) = S_0$, and define $\rho = \frac{\mu}{\lambda}$ as the threshold value.

So, $c = I_0 + S_0 - \rho \ln S_0$.

Thus, $I(\mathrm{S}) = \mathrm{I}_0 + S_0 - S + \rho \ln \frac{S}{S_0}$ (**).

Now let's discuss the nature of the curve(**),

from $\frac{dI(\mathrm{t})}{dS(\mathrm{t})} = -1 + \frac{\rho}{S(\mathrm{t})} \begin{cases} < 0 & S(\mathrm{t}) > \rho \\ = 0 & S(\mathrm{t}) = \rho \\ > 0 & S(\mathrm{t}) < \rho \end{cases}$,

we know, when $S(\mathrm{t}) < \rho$, $I(\mathrm{S})$ is an increasing function of $S(\mathrm{t})$, and while $S(\mathrm{t}) > \rho$, $I(\mathrm{S})$ is a decreasing function of $S(\mathrm{t})$.

And $I(0) = -\infty, \mathrm{I}(\mathrm{S}_0) = \mathrm{I}_0 > 0$, by the intermediate value theorem and monotonicity of continuous function ,there must be a sole point S_∞ ,

$0 < S_\infty < S_0$,so that $I(S_0) = 0$, but while $S_\infty < S \le S_0$, $I(S) > 0$.

By equations (*), we learn that when $I = 0$, $\frac{dS}{dt} = 0, \frac{dI}{dt} = 0$,so $(S_\infty, 0)$ is the equilibrium point of the equations (*).

While $t \ge t_0$, the graph of the equation (**) is shown in Figure 3-1.

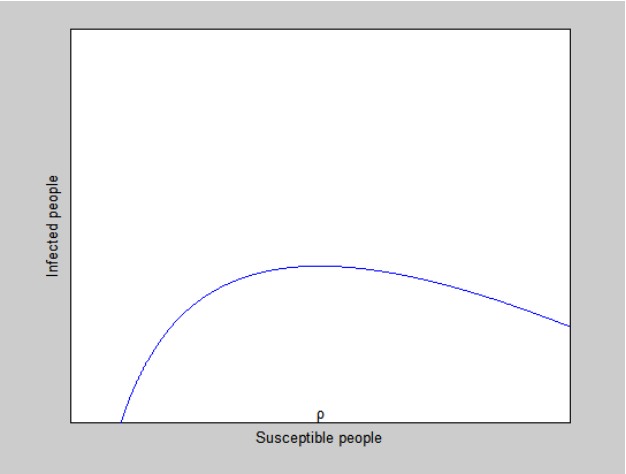


Figure 3-1

When t increases from t_0 to ∞ , the point $(S(t), I(t))$ moves along the curve (**) to the S decreasing direction, cause $S(t)$ monotonically decreases with the increase of time. So if $S_0 < \rho$, $I(t)$ reduces to 0 monotonically, and $S(t)$ to S_∞ . So if few infectives I_0 scatter into the susceptibles S_0 and $S_0 < \rho$, the condition of the disease (Ebola) will get alleviated.

On the contrary, if $S_0 > \rho$, $I(t)$ increases while $S(t)$ decreasing to ρ , and when $S = \rho$, $I(t)$ achieves the maximum value. Only if $S(t) < \rho$, $I(t)$ starts to decrease. Later we will apply this SIR model into a modified MDVRP model to solve the problem of optimizing the eradication of Ebola.

4 Weighted K-Means Algorithm (siting of delivery center)

As we mentioned above, a plenty of people have died because of Ebola, and recently with the improvement of medical level, the control of outbreaks can be realized by a kind of new medicine created by the World Medical Association. Because whether the patients can be cured depends on how fast or how much the medicine they can get, it is vital to select a proper location as the medicine delivery center.

In order to satisfy the word ‘proper’ many factors should be taken into consideration, including the distance between medicine delivery center and disaster areas, magnitude of the disaster, health facility and so on. In the next part we attempt to obtain some proper locations which suits the requirement best with less cost and higher efficiency. We achieve it by K-means algorithm and its improved model, the weighted K-means algorithm.

4.1 K-means algorithm

K-means algorithm is a classical cluster method based on partition. Its basic idea is that for all the given points, there are several central points and each of them represents a cluster. The given points would find out which cluster they belong to by the distance between them and central points. After finishing the classification, each cluster should produce a new central point according to the points inside it. Then repeat these steps until the central points do not change anymore. Finally, we can get several clusters. They are all obtained from partitioning the given points and each point inside them have the less distance to its central point than those in other clusters. Next we are going to clarify this algorithm through an example.

Assume that the number of central points is k, and the number of given points is n with their coordinates as follows,

$$C_1(x_1, y_1), C_2(x_2, y_2), ..., C_k(x_k, y_k), \quad A_1(x_1, y_1), A_2(x_2, y_2), ..., A_n(x_n, y_n)$$

Firstly we initialize the central points by assigning values to them. There are many choices about the values. You can obtain from the given points or set as special coordinates. But in most cases we get the values by applying a random function to make sure of its equality.

$$C_i(x, y) = (\min(x) + (\max(x) - \min(x)) * \text{rand}(), \min(y) + (\max(y) - \min(y)) * \text{rand}());$$

Next, we calculate the distance between the given points and central points. Each time we select one given point and calculate the distance between it and all the central points to find least distance. Through the least distance we can find the corresponding central point the given point belongs to. So we successfully classify the given point into a cluster. By repeating these steps we can classify all the given points into clusters.

After we get *k* clusters, we need to find the new central point for each of them. The new central point of a cluster is exactly the center of all the given points inside it which can be calculated by midpoint formula

$$C_i(x, y) = ((x_1 + x_2 + ... + x_j)/j, (y_1 + y_2 + ... + y_j)/j)$$

With the new central points, we can once again obtain the distance between central points and given points and once again classify the given points into clusters. And of course we can one again create the new central points. With constant iterations, in the end, the central points will not change anymore. And until then, we can get several clusters as well as their central points which satisfy that the distance between a given point and its central points is always no longer than the distance between it and other central points.

4.2 Improved K-means algorithm

According to our thought, the location of medicine delivery center plays a very important part in this module. If we ignore the magnitude of epidemics and

only consider the distance factor, the medicine delivery center should be built on the place near the epidemic area, which helps make the transportation cost as low as possible. We regard the medicine delivery center as central points and epidemic areas as the given points. By applying K-means algorithm we can easily find out the nearest medicine delivery center for an epidemic area. So if we decide the location of medicine delivery center through that method, we can make sure that the total cost of transportation between epidemic areas and delivery center can be reduced to minimum. Assume that there are 10 points to partition into 3 clusters. Figure 4-1 shows the result of location by applying K-means algorithm in the following.

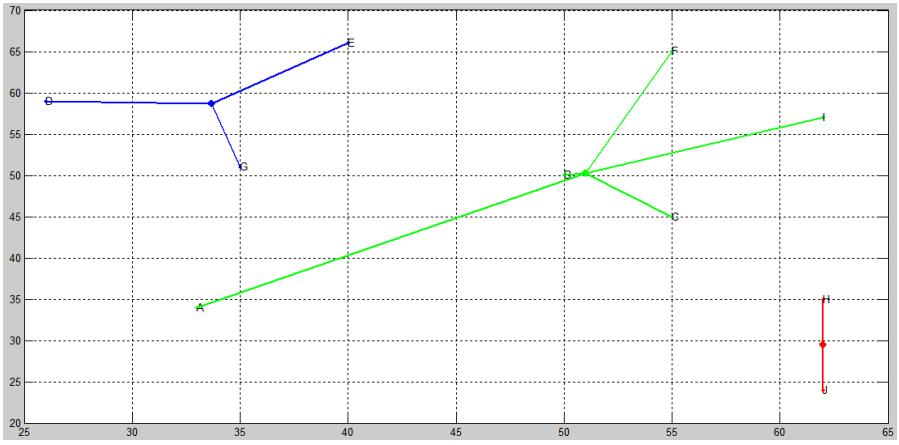


Figure 4-1: Traditional K-means algorithm

However, the solution mentioned above has not taken the magnitude of epidemics into account. So it is kind of unscientific and unsuitable in practical use. After all, our purpose to build the medicine delivery center is to save as many people as we could in the shortest time. You cannot build the center near a city with 100 patients but far away from the city with 1,000 patients in it. But when we start to consider about the magnitude of epidemics, things become different. You have to handle these two factors well, using both of them reasonably to judge the best location of medicine delivery center.

Here is our solution, which we call improved K-means algorithm. At first, we add an extra attribute, ‘weight’, to each of the epidemic area. The value of weight is decided by the infectious patients in the epidemic area. Generally the more the

infectious patients in the epidemic area, the more weight the area can get. Table 1 shows the stricken cities and the number of patients there.

As the number of patients and the distance between epidemic areas and delivery center are in different units, we thought it would be better for us to process the data we get by data normalization in the beginning. However, after trying out some experiments we found that whether or not we operate data normalization on the distance between delivery center and epidemic areas does not change the length relationship between them. Therefore it is kind of meaningless. According to the situation we change our thought and create an attribute, the weight, for each of the epidemic areas based on the number of patients there. The relationship between weight and number of patients is shown as follows,

$$w(\mathrm{i}) = k * \mathrm{NumberOfPatients}$$

In the formula, k is a constant ranging from 0 to 1. It can be modified according to actual demand. Generally speaking, the bigger the value of k is, the greater it would influence the final result. To see the result more clearly, in our module we set the value of k as 1. We have download the number of patients in different epidemic areas, which is shown table 1. So we can get the weight of all the epidemic areas through the formula above.

In the next step, we are going to calculate the distance between epidemic areas and delivery center. With weighted epidemic areas as well as the distance between epidemic areas and medicine delivery center, we can finally decide the central point. This is the key part of Improved K-means algorithm, distinguishing it from traditional one. The Improved K-means algorithm works like traditional K-means algorithm in many parts. However, in the last step when calculating the central points for each cluster, the value of central points are not the center of the inside given points, but the weighted center of the inside given points. It works as follows,

$$C_i(x,y) = (\sum A_k(x_k,1) * w(k) / \sum w(k), \sum A_k(x_k,2) * w(k) / \sum w(k),)$$

As is shown above, each time calculating the location of delivery center, the weight of the epidemic areas would be taken into consideration. Therefore, with the increase in the number of iteration, the location of delivery center will gradually move or have the trend to move towards the epidemic areas with larger weight. This ensures that on the basis of distance between epidemic areas and medicine delivery center, the location of delivery center will be modified according to the magnitude of epidemics, which finally returns the best location of medicine delivery center. See figure 2 as follows. We add a weight with the value of 2 to point A while the value of other points remain 1. Compare between figure 1 and figure 2, and it is obvious that after adding weight to the algorithm, the central point which contains point A moves from origin place towards point A.

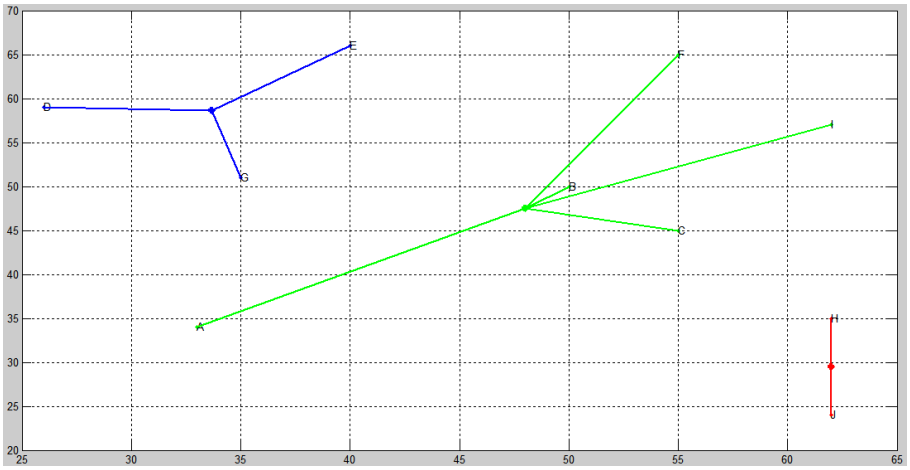


Figure 4-2: Improved K-means algorithm

4.3 Conclusion

During this part, our group attempt to work out a proper location of drugs delivery center which can bring us less transportation cost but high cured rate. We start from the traditional cluster method, K-means and by apply this algorithm we obtain a best location which could reduce the distance between

delivery center and epidemic areas to the minimum when regardless of the magnitude of epidemics. Then we are inspired by this algorithm and create the Improve K-means algorithm which has taken both of the two factors into consideration to decide the location. It is implemented by adding ‘weight’ to each of epidemic areas according to their magnitude of epidemics when calculating the locations. In this way, we can make sure that on the basis of distance, the location of delivery center will gradually move towards the epidemic areas with larger weight, letting more people be cured in time.

5 Modified MDVRP Model (delivery system)

To decide a feasible and efficient delivery system while considering the quantity of the medicine needed and the speed of manufacturing of the vaccine or drug, we established a modified MDVRP model. Our mission is to design a schedule scheme for all of our drugs delivery centers each day. The goal is not only to minimize the distance as short as possible but also to consider about the seriousness of the epidemic area and deliver drugs to them in order to optimize the eradication of Ebola.

5.1 Assumptions and Conditions

- A certain schedule scheme for each day and all the delivery missions of the schedule scheme will be finished in that day.
- Unlimited amount of vehicles in each delivery center
- Limited amount of drugs in each delivery center (We add this constraint in order to fit with the limited speed of manufacturing of drugs in reality. So n ot all the epidemic areas will be served.)
- *m* drugs delivery centers established by our modified K-Means
- *n* epidemic areas
- V_i maximum load of vehicle (each vehicle can only transport V_i drugs at most)
- Q_i certain demand of epidemic area
- D_i certain amount of drugs in a certain drugs delivery centers
- Modify the λ and μ of each epidemic area according to the drugs they receiv e each day. Because of the lack of actual data which can illustrate the affecti on of drugs on the λ and μ, we assume that λ and μ are updated as below eac h day after the delivery finish. $\lambda=\lambda+1/Q_i/10000000$, $\mu=\mu+1/Q_i$.
- Each epidemic area can be only served by one vehicle.

- To simplify the problem, if the rest load of a vehicle is smaller than *Di* of a certain epidemic area, this epidemic area will not be served.
- P_1, P_2, P_3 represent for three penalties added on the distance
- W_1 , W_2 , W_3 represent for three multiplier on P_1, P_2, P_3

5.2 Introduction of Basic MDVRP

Multi-Depot Vehicle Routing Problem (MDVRP), which is the natural extension of the Vehicle Routing Problem (VRP). VRP was first proposed by Dantzing and Ramser in 1959. VRP provides a situation, where there are some clients with requirements of the goods and a delivery center to provide the clients goods. A proper route should be arranged in order to meet the clients' needs under a certain constraint, at the same time achieve that the cost is minimized. VRP has been proved to be a NP-hard problem.

Traditional VRP contains only one depot. Differently, MDVRP contains multiple depots as it is named. It aims at finding a route with the least cost under a certain constraint. Generally, VRP or MDVRP is solved with heuristic algorithm because of their property of NP-hard. For instance, paper③implements a saving algorithm to optimize the VRP and obtain a better solution. Other papers ④~⑦choose Tabu Search Algorithm to optimize it. Besides, paper④and ⑧ deal with it using Simulated Annealing Algorithm. In these kinds of heuristic algorithm, some use Cluster First-Route Second method while others apply Route First-Cluster Second method.

In reality, a medicine delivery system generally contains more than one delivery center. In a consequence, MDVRP model is more suitable for our drug delivery system in optimizing the eradication of Ebola. Although MDVRP model is close to the situation in our problem of delivering drugs to epidemic areas, there’re still some difference between the traditional model and the real situation. There’s no limitation of the amount of goods in the delivery center in traditional MDVRP and it’s obviously not proper enough to model our problem because the speed of manufacturing of the drugs is not high enough to provide all the patients

infected by Ebola. Moreover, because of no limitation on the amount of goods in traditional MDVRP, all of the clients will be served. Also, we can not only consider the cost of distance now because we may prefer to deliver the drugs to a remote area far from the drugs delivery center due to the high value out of the bound in SIR model in that area (review that in SIR model). Obviously, we need to make some improvements on the traditional model to adjust our special drugs delivery system well.

Figure 5-1 shows that the basic MDVRP model which we apply on 41 epidemic areas and 3 drugs delivery centers generated by K-Means without constraint of the amount of drugs in each center. Although it's not appropriate enough to fit with our problem, it's important for us to do further attempts on this basic model. And obviously we can see that the routes of vehicles are relatively reasonable.

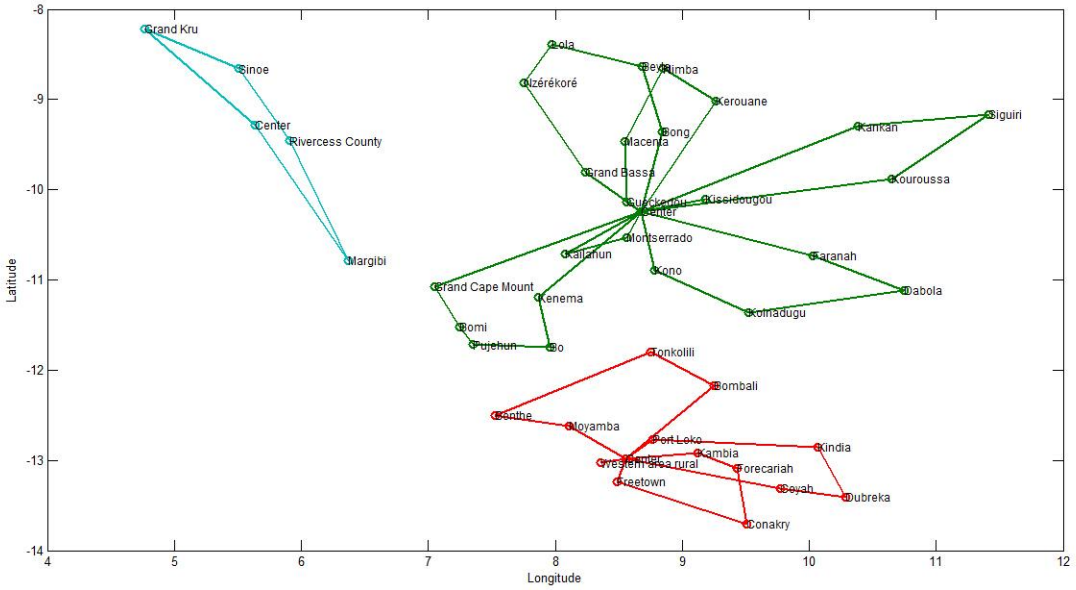


Figure 5-1 basic MDVRP model on 41 epidemic areas

5.3 Modified MDVRP

In order to adjust our problem, we modified the basic MDVRP with adding constraint on the amount of goods (drugs) at our delivery center. By adding constraint on the amount of drugs at our delivery center, the situation becomes that not all the clients (epidemic areas) will be served and it's closer to real life.

Moreover, we adjust the value of λ and μ of each center in our SIR model according to the drugs each center receive each day. As mentioned in the assumption and to simplify the problem, assume that each day we finish all the delivery missions of the schedule scheme and each day we generate a new schedule scheme because the threshold of each epidemic area will be updated each day due to the increase of μ and the decrease of λ determined by the drugs each epidemic area receive.

5.4 Hybrid Genetic Algorithm Solving Modified MDVRP

Because of its property of NP-hard, we use genetic algorithm combining with local search to deal with it instead of traditional mathematical methods.

5.4.1 Genetic Algorithm

Creatures live and propagate in nature, revealing its excellent adaptability to the environment. Inspired by it, people are dedicated to research the mechanism of the creatures' living and simulate their behavior, providing a broad prospect for the design and development of artificial adaptive systems. Genetic Algorithm (in short, GA) is one of the most remarkable achievements of the computational simulation of this biological behavior. Based on computational simulation of these process, the GA makes all kinds of artificial genetic algorithm system more adaptive and optimized.

The basic operations of the Genetic Algorithm are as follows:

(1)Population initialization. Set the counter of the revolution *generation* = 0 and the max value of *generation* is *T*. We generate *M* individuals as the primitive population $P(0)$ randomly or with certain method.

(2)Evaluation of the individuals. Evaluate the fitness for every single individual of the population *P(generation)*.

(3)Selection. Select among the current generation in order to last the

individual with higher fitness to the next generation or make mate with each other and produce a better generation.

(4)Crossover. Select the individuals from two or more populations and let them mate, which is core of the Genetic Algorithm.

(5)Mutation. Make every single individual within the population mutate with a certain probability in order to escape from the local optimization.

Genetic algorithm used in this paper and the above steps are similar, except for our promotion for the selection and mutation operations in order to achieve better results. As for selection operations, we choose to select the better individuals to do the cross operation and the children they produce will take place of the relatively poor individuals, instead of the traditional methods such as Roulette. In the mutation operation, we add a variety of local search methods, traverse-like locally searching the optimized individuals and retaining the better solutions. At the same time, we keep the poor solution with a certain probability to avoid being stuck in the local optimized solution. The procedures of our algorithm are described in details below.

5.4.2 Genetic Encoding

We adopt an encoding method which is similar to the encoding method in TSP, in which the order of code represents the sequence of visiting car and the code represents the drugs delivery centers belonging to them, and this method will be introduced in following paragraph in detail.

To simplify our statement, "epidemic area" will be replaced by "ea" for short. Assuming that there are *m* drugs delivery centers and *n* epidemic areas, we encode them from 0, which means 0~*m-1* represents m drugs delivery centers, and *m*~*m*+*n-1* represents *n* epidemic areas, then the genetic code of each elements is a sequence from 0 to *m*+*n-1*, in which the start is always 0. This order is such a sequence :(0 $centerY_0$ eaX_0 eaX_1 … eaY_1 eaX_k … $centerY_2$… $centerY_{m-1}$ …), in which $centerY_n$ represents the delivery center whose number is Y_n , eaY_n represents the epidemic area whose number is X_n , and the meaning

is that $centerY_{k-1}$ serves the epidemic areas between $centerY_{k-1}$ and $centerY_k$, and $centerY_{m-1}$ serves the remaining epidemic areas in the sequence. If $centerY_{k-1}$ and $centerY_k$ is adjacent, it means $centerY_{k-1}$ does not serve any epidemic area.

For example, if *m*=3 and *n*=6, then the number of delivery centers is 0,1,2, and the number of epidemic is 3,4,5,6,7,8. Supposing the solution of such a code sequence (0 4 5 8 1 7 3 2 6). It means that epidemic area 4, 5, 8 are served by center 0, epidemic area 7,3 are served by center 1, and epidemic area 6 is served by center 2. For each center the order of vehicle visiting is fixed. Consider the possible situations of center 0 in the example above. Figure5-2 shows its four possible situations.

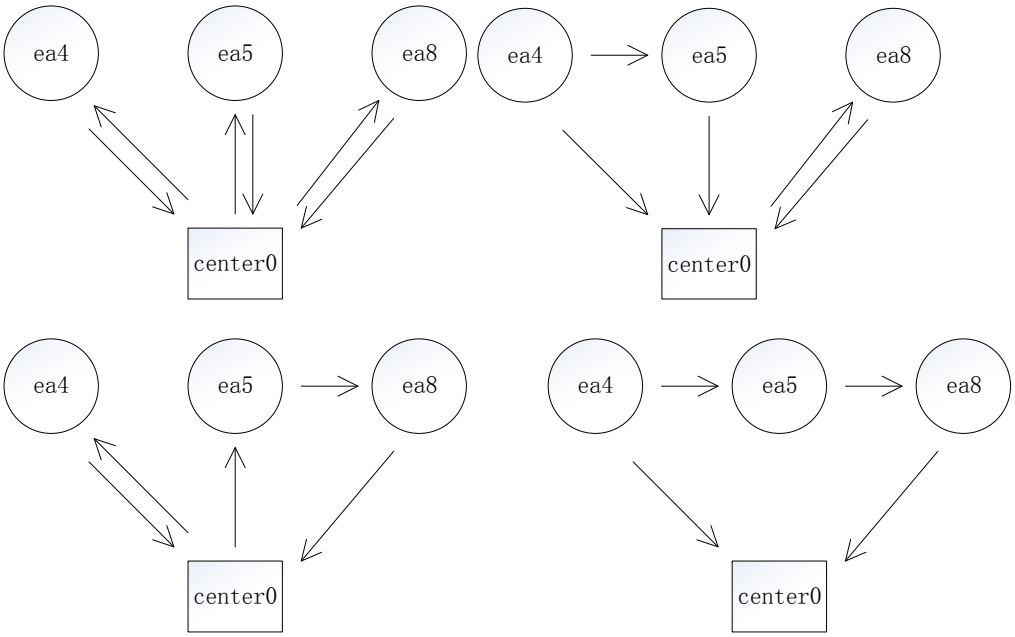


Figure5-2 four possible situations of center0

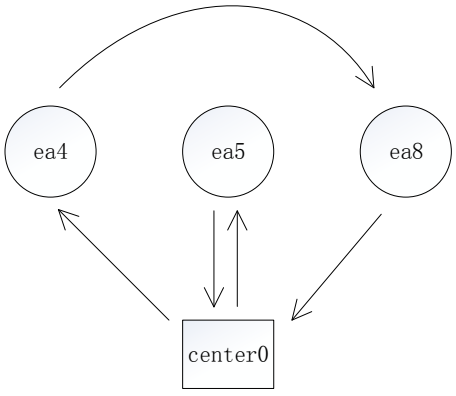


Figure5-2 an illegal visiting order of center 0

As shown in figure5-2, for each drugs delivery center, though the visiting order is fixed, there exists different strategies to arrange the vehicles. In addition, we have to consider the capacity limit of vehicles and demand of each epidemic area when applying it in reality.

Thus, we will introduce a split method in next section to find an optimal strategy to arrange the visiting order of vehicles which will satisfy the constraint we include.

5.4.3 Calculation of Fitness

The genetic encoding in the last section determines the visiting order of the vehicles, then we can use a split method to find an optimal strategy to arrange the visiting order of vehicles (vehicles in banned list will not be considered). The time complexity of which is $O(n^2)$[1]. After splitting, we are able to calculate the fitness of this solution. Solutions with higher fitness will be preferred. The algorithm below arranges all the epidemic areas which are not added into the banned list to a certain vehicle of its drugs delivery center.

Algorithm 1：

```
V[path[S]] := 0 //path[S] is the number of drugs delivery centers
for i := S+1 to E do V[path[i]] := +∞ endfor //initialize the distance for other epidemic area points as infinity
for i := S+1 to E do
    load := 0; cost := 0; j := i
    while (j<=E) and (load<=L) //enter loop while sum of current car demand less than limit and there exists epidemic area points not being considered
        load := load + Qj //load represents for the sum of demand of current epidemic area
            if i == j then //calculate the distance if there is only one epidemic area
        cost := dis[path[S]][path[i]] + dis[path[i]][path[S]] //cost represents
```

```
for the distance
            else //otherwise add epidemic area path[j] into path, calculate its
distance
              cost := cost - dis[path[j-1]][path[S]] + dis[path[j-1]][path[j]] +
dis[path[j]][path[S]]
            endif

            if (load<=L) then //if number of current car is no more than limit
          if (V[path[i-1]] + cost < V[path[j]]) then
            V[path[j]] := V[path[i-1]] + cost//update the optimal solution
                  pre[path[j]] := path[i-1]//record the path
           endif
           j := j + 1
            endif
     endwhile
  endfor
```

In this algorithm, when calling this function, the *S-th* point is supposed to be delivery center and the *E+1-th* point is also supposed to be delivery center or the *E-th* is the end of the sequence. That is to say in the sequence from *S* to *E*, delivery center *S* must serve all epidemic areas among node *S+1* and *E*. Then we divide the epidemic areas in the sequence, and arrange a certain vehicle to serve them, then we can calculate the fitness of each delivery center so that we can obtain the fitness of this solution by summing. In which the expression of fitness is *fitness = worst_dis - dis*, and *worst_dis* is a big constant which makes fitness nonnegative, dis is the minimal distance we attained by summing all the fitness of the delivery centers in the sequence.

In this algorithm, the *S-th* point is the drugs delivery center, so the total of epidemic areas is *E-S*. *V[path[i]]* in this program means the minimal distance when only considering from the *S-th* point to the *i-th* point. The basic idea of this split algorithm is that: each loop of variable *i* can determine the optimal solution

(minimal distance) of front *i-S* epidemic area points served by drugs delivery center path[S]. When only the first epidemic area point *S+1* is considered, *V[path[S+1]]* has only one condition which is $S \rightarrow S+1 \rightarrow S$, thus the optimal solution of first epidemic area point can be calculated in the first loop. When taking the second client point into consideration, *V[path[S+1]]* has already got the optimal solution. And then we can get the value of *V[path[S+2]]* by the following equation:

$$V[path[S+2]]=Max\{ V[path[S+1]] + Cost(S \rightarrow S+2 \rightarrow S) , V[path[S+2]] \}$$

In a similar fashion, when looping from *S+1* to *E*, we can finally find the optimal partitioning solution which is required.

It performs well when n is small, Because the time complexity of the solution mentioned above is $O(n^2)$.

5.4.4 Penalty added on the distance:

1. The amount of drugs in drugs delivery center is under a certain limit. So not all the epidemic areas will be served. But how to choose which epidemic areas to be served? We take both the distance from the center to the epidemic areas and the seriousness of the epidemic areas into account. Define the threshold we get from SIR model as *Th* and the susceptible people of the epidemic areas as *S*. We add a penalty factor $P_1 = Th - S$ onto the distance. If $S > Th$, the seriousness of the epidemic area will increase. By adding the penalty factor, the distance will be shortened and it will be preferred to serve. On the contrary, if $S < Th$, the distance will be lengthened because its situation is relatively better. In the genetic encoding mentioned above, the center which a certain epidemic area is determined. Because of the limitation amount of the drugs in delivery center, not all the epidemic area will be served. So we generate a banned list for each delivery center. If a delivery center is not able to serve all its areas, those areas with longer “distance” (after adding the penalty) will be added into banned list first until delivery center is able to serve all those areas who are not in the banned

list. By applying this "Banned List" method, our model is able to balance the cost of distance and the seriousness of the epidemic area to choose appropriate areas to deliver drugs.

2. "Banned List" method seems to solve the problem of the balance between the cost of distance and the seriousness of the epidemic area. But a serious problem exists. Since our genetic algorithm prefer the solution with shorter distance which has been added penalty on it, those solutions which have a great number in its banned list would have a high fitness because the total epidemic areas they serve is less. But actually those solutions serve less epidemic areas are bad because many drugs in the delivery center would not be delivered to the epidemic areas although the disaster is serious. Hence, we add another penalty $P_2 = drugsLeft$. *drugLeft* represents for the amount of drugs which are not delivered at that day in the delivery center. By adding this penalty, those solutions serving few areas will have a low fitness and will be less possible to be preserved in our genetic algorithm.

3. Moreover, the amount of infective people is also an important factor and we also add a penalty $P_3 = MaxConstant - I$ onto the distance. I represents for the amount of the infective people in the epidemic area. *MaxConstant* is a large constant to adjust that epidemic area with a larger I would be preferred to serve.

In reality, the penalty need to be adjusted by multiplying a parameter W because of the different datasets. Experiments in the later section show that our method is not only effective but also efficient.

5.4.5 Population Initialization

We initialize the population randomly by generating a random sequence from 1 to $m+n-1$ and adding 0 to the head of the sequence. We draw a conclusion that setting the size of the population to 30 performs better.

5.4.6 Crossover

We apply the typical OX crossover operator, and Figure5-1 shows the procedure of the crossover.

Rank	:	1	2	3	4	5	6	7	8	9
					i=4		j=6			
					↓		↓			
P1	:	1	3	2	6	4	5	9	7	8
P2	:	3	7	8	1	4	9	2	5	6
C1	:	8	1	9	6	4	5	2	3	7
C2	:	2	6	5	1	4	9	7	8	3

Figure5-1 OX crossover

Pay attention that in this procedure, we do not take the start position 0 into crossover. Because 0 always represents for the first drugs delivery center and it should be placed at the start of the sequence. Two positions are selected randomly in parent genes sequences and the sequences between them are kept in the same place while the left and the right part interchange. It's possible to generate illegal sequences might occur, so we need to check its validity and start exchanging from $j+1$. Assume that P_1, P_2 are parent sequences and i, j are the positions produced randomly, the offspring C_1, C_2 will keep the sequences between i and j of P_1, P_2. Then we think about C_1 firstly. Start looping from the $j+1$ element of P_2 to fill the lost ones in C_1, while those existing elements will be ignored. Similarly, we do the same thing on C_2 --- filling the lost elements in C_2 starting from the $j+1$ element of P_1.

5.4.7 Population Selection

During the experiment, we found that the result from the method that selecting individual to form a new population by Roulette in traditional genetic algorithm performs not well because it will create many identical individuals and operating crossover on them won't create a new individual. As a result, it shrinks the search scope of feasible solution. In order to improve the quality of solutions, we abandon Roulette but apply a new method which works as follow: firstly we sort the individuals according to their fitness and then we choose the individuals in the first half to crossover with each other. Their offspring take the place of

second half. Then a new population with relatively high diversity is generated.

5.4.8 Local Search and Variation

Traditional genetic algorithm only apply simple mutation which change the gene of an individual and accept it without considering its quality. To find a better solution, we need to strengthen the search of neighborhood of every solution. Local search traverses part of neighborhood of a solution and decides whether there is a better individual. If not, just keeping the old individual. We uses three neighborhood search operators and our local search is set to traverse all situations:

(1)2-opt

2-opt is short for 2-optimization (also called 2-exchanged), which represents for optimization of two elements. 2-opt algorithm was put forward by Croes[2] at the earliest to solve TSP problems.

2-opt works as follows: select two position *i, j* and add the path in front of *i* and behind *j* to the new path, as well as the turned-over path between *i* and *j*. For example, assume that the original path is ABCDEFG. If we select *i*=3, *j*=6 (CDEF), applying 2-opt algorithm we will obtain a new path ABFEDCG.

(2)Remove part of the path and insert it to the front or to the back

For example, assume that the original path is ABCDEFG, if we select the interval (CDEF), and select another position A outside this interval and insert into the back of it. Then we can obtain a new path ACDEFBG.

(3)Exchange 2 nodes

For example, assume that the original path is ABCDEFG. We select B and F and interchange their position, we can obtain a new path AFCDEBG.

Experiment shows that the three neighborhood operators mentioned above is relatively effective. In addition, we have tried to find many other neighborhood

search operators, but it performs not so well and only increases the execution time.

Meanwhile, in order to reduce the time of execution, the traversing local search can only apply to part of individuals whose fitness are relatively better. If we apply the traversing local search, the result would not be improved but the time of execution increase. Experiments show that applying the traversing local search on the top 50% individuals with high fitness can reduce the time in execution but keep the quality of optimal solution. Of course, our algorithm also preserve the mutation part of traditional genetic algorithm which accept the solution without considering its quality. Experiments show that our program performs better if setting the probability of mutation to 20%.

5.4.9 Program Framework

The general program framework is shown in Algorithm 2. *update_fitness()* uses the splitting method mentioned above to update the fitness of all individuals among population and *sort()* sorts the individuals from the largest to smallest in order of their fitness. And then *preserve_best()* preserves the best individuals, *crossover()* execute the procedure of crossover and update the population. *mutation_localSearch()* execute the procedure of mutation and local search and accept the bad solution at a certain probability.

During the experiment, we set the value of *maxGeneration* as 50. But in fact, algorithm converges after 30 iterations.

Algorithm 2: Program Framework

```
while (generation <= maxGeneration)
 update_fitness();
 sort(population);
 preserve_best_solution();
 crossover();
 mutation_localSearch ();
 generation++;
endwhile
```

5.5 Result

To examine the effectiveness of our modified MDVRP model in drugs delivery system, we generalize two groups of data. Now we do not care its future development because we focus on observing the effectiveness of the modified MDVRP model. So we observe the schedule scheme of this day whose SIR data is as below. There're two drugs delivery centers numbered by 0 and 1. And there're 18 epidemic areas with SIR data. The SIR data in two groups is totally different. In the first group, the situation is more serious in epidemic area 2~9. In the second group, the situation is more serious in epidemic area 12~18. See as chart5-1.

Chart5-1

Epidemic	Susceptible	Infective	Recovered
2	0	100	0
3	0	100	0
4	0	100	0
5	0	100	0
6	0	100	0
7	0	100	0
8	0	100	0
9	0	100	0
10	99	1	0
11	99	1	0
12	99	1	0
13	99	1	0
14	99	1	0
15	99	1	0
16	99	1	0
17	99	1	0
18	99	1	0

Epidemic	Susceptible	Infective	Recovered
2	99	1	0
3	99	1	0
4	99	1	0
5	99	1	0
6	99	1	0
7	99	1	0
8	99	1	0
9	99	1	0
10	99	1	0
11	99	1	0
12	0	100	0
13	0	100	0
14	0	100	0
15	0	100	0
16	0	100	0
17	0	100	0
18	0	100	0

The schedule scheme is shown as below:

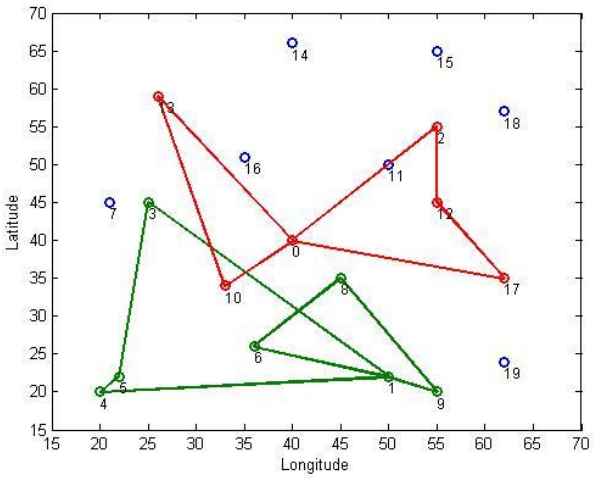


Figure x1 schedule scheme of the first group of data

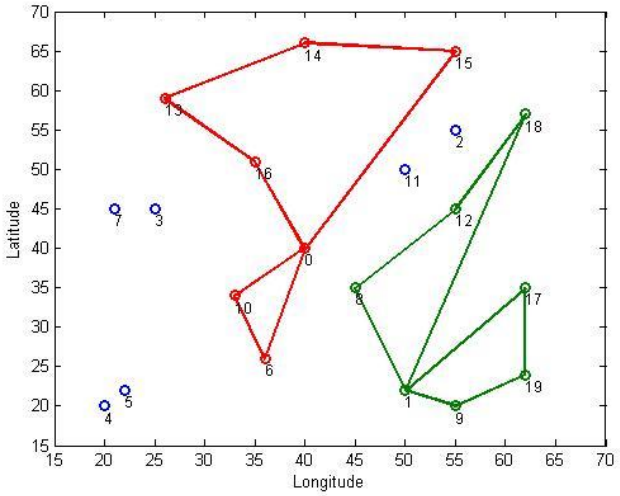


Figure x2 schedule scheme of the second group of data

We can see that in the schedule scheme of the first group of data, drugs delivery centers prefer to select the epidemic areas with serious situation. In a conclusion, our delivery system based on MDVRP could select appropriate epidemic areas to deliver drugs while considering both the cost of distance and the seriousness of epidemic areas.

6 Experiment

In order to find out the best location of medicine delivery center, we need to handle the data properly, including the number of patient in epidemic areas and the distance of transportation the medicine. We firstly download the number of patients who have been infected by Ebola virus from the Internet[4], shown in the following table 1.

Table 1

City(Lon, Lat)	Patients	City(Lon, Lat)	Patients
Kouroussa, Kankan, Guine (10.65, -9.88)	4	**Grand Bassa, Liberia (8.24, -9.81)**	202
Dabola,Faranah,Guinea (10.75 ,-11.12)	5	**Rivercess County, Liberia(5.91, -9.46)**	211
Kindia, Guinea (10.07, -12.85)	9	**Sinoe, Liberia (5.51, -8.66)**	234
Dubreka, Kindia, Guinea (10.29, -13.41)	20	**Grand Kru, Liberia (4.77, -8.22)**	235
Conakry, Guinea (9.51, -13.71)	26	**Faranah,Guinea (10.03, -10.73)**	251
Coyah, Kindia, Guinea (9.77, -13.31)	31	**Koinadugu, Northern Province, Sierra Leone (9.52, -11.36)**	256
Siguiri, Kankan, Guinea (11.42, -9.17)	32	**Kono, Eastern Province, Sierra Leone (8.78, -10.89)**	271
Forecariah, Kindia, Guinea (9.43, -13.08)	32	**Kerouane, Kankan, Guinea (9.27, -9.02)**	294
Kambia, Northern Province, Sierra Leone (9.13, -12.92)	37	**Moyamba, Southern Province, Sierra Leone (8.11, -12.62)**	314
Kankan, Guinea (10.38, -9.30)	46	**Bo, Southern Province, Sierra Leone (7.95, -11.74)**	381
Kissidougou, Faranah, Guinea (9.18,-10.10)	59	**Macenta, Nzérékoré, Guinea (8.55, -9.47)**	434
Gueckedou, Nzérékoré, Guinea (8.57, -10.13)	70	**Nzérékoré, Guinea (7.75, -8.82)**	442
Beyla, Nzérékoré, Guinea (8.68, -8.63)	86	**Grand Cape Mount, Liberia (7.05, -11.07)**	498
Kailahun, Eastern Province, Sierra Leone (8.08, -10.71)	91	**Bong, Liberia (8.84, -9.36)**	523
Kenema, Eastern Province, Sierra Leone (7.86, -11.20)	103	**Bombali, Northern Province, Sierra Leone (9.25, -12.17)**	693
Lola, Nzérékoré, Guinea (7.97, -8.39)	104	**Port Loko, Northern Province, Sierra Leone (8.77, -12.77)**	744
Bonthe, Southern Province, Sierra Leone (7.53, -12.51)	104	**Tonkolili, Northern Province, Sierra Leone (8.75, -11.80)**	979
Pujehun, Southern Province, Sierra Leone (7.35, -11.72)	113	**Western area rural,Sierra leone (8.36, -13.03)**	1292
Bomi, Southern Province,	140	**Montserrado, Liberia**	1974

Sierra Leone (7.24, -11.53)		**(8.56, -10.53)**	
Nimba, Liberia (8.85, -8.66)	161	**Freetown, Western Area, Sierra Leone (8.49, -13.23)**	3097
Margibi, Liberia (6.37, -10.79)	178		

To show the effectiveness of the weighted K-Means. We apply basic K-Means algorithm without considering the magnitude of disaster at first to obtain the location of medicine delivery center. See as Figure 6-1.

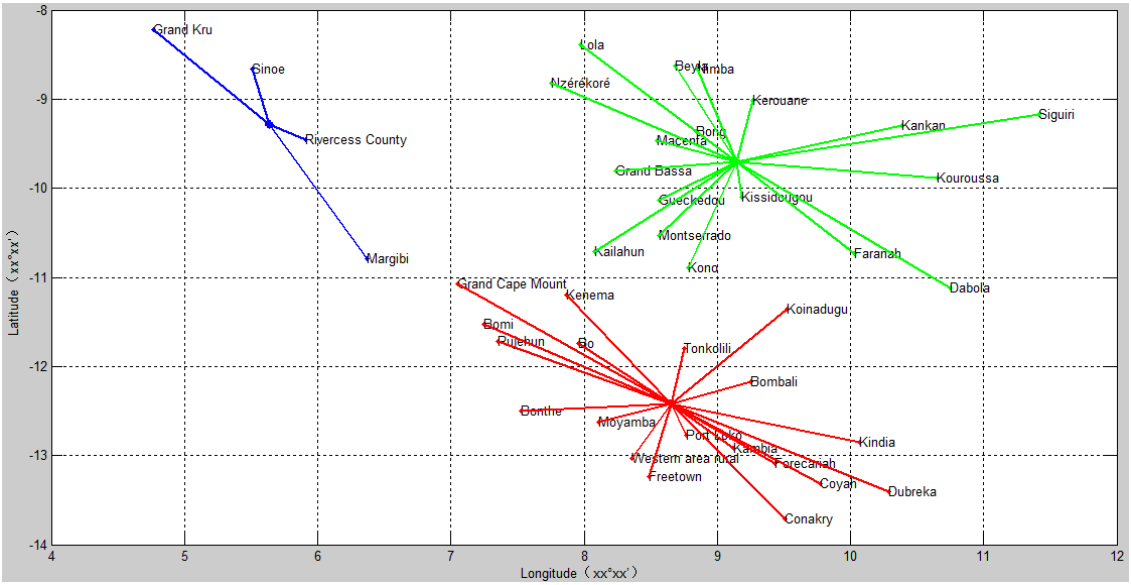


Figure 6-1 basic K-Means generating three delivery centers

Then in order to use weighted K-Means, we obtain the weight of each epidemic area using the data above by the formula

$$w(\mathrm{i}) = k * \mathrm{NumberOfPatients}$$

In our model, we set the value of k to be 1. As a result, we can simply use the number of patients as the weight. We give each of the epidemic areas a weight according to their numbers of patients, which can be used to modify the location of drug delivery center by the formula

$$C_i(\mathrm{x},\mathrm{y}) = (\sum A_k(\mathrm{x}_k,1) * w(\mathrm{k}) / \sum w(\mathrm{k}), \sum A_k(\mathrm{x}_k,2) * w(\mathrm{k}) / \sum w(\mathrm{k}),)$$

Then three drugs delivery centers are generated again. See as figure 6-2.

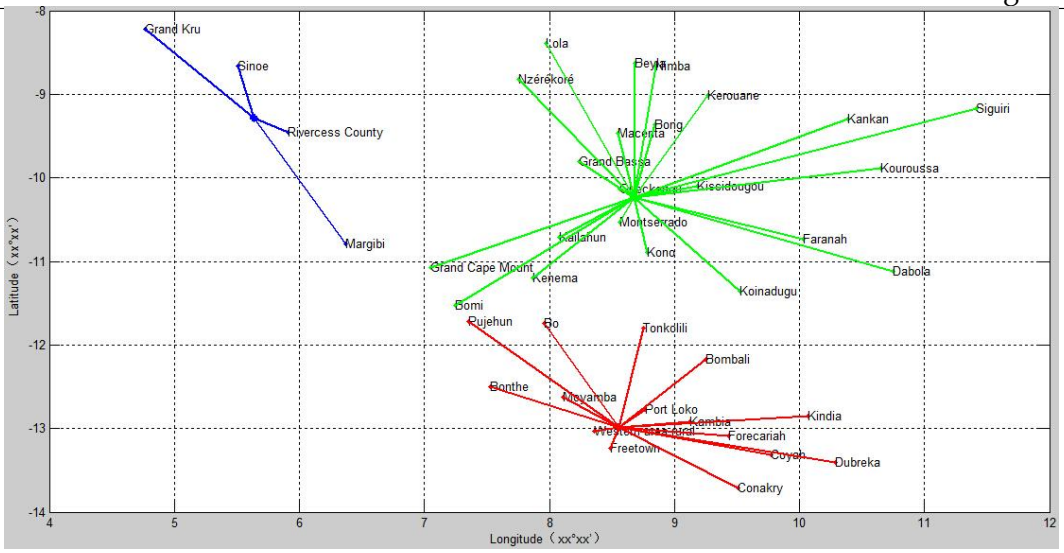


Figure 6-2 weighted K-Means generating three delivery centers

We can see that the locations are a little bit different from figure 6-1. Observe the delivery center at the bottom. By checking the chart we know that the disaster in the city Freetown is serious. And by adding the weight on basic K-Means, the center skews a little to the bottom. It illustrates our model works.

- Selecting some good delivery centers is important for the drugs delivery system which we use a modified MDVRP model to fit with it. We used the three drugs delivery centers above to illustrate the effectiveness of our drugs delivery system in optimizing the eradication of Ebola. We simulate a 26-day experiment on these 41 cities and the trend of SIR model is shown as below. Since the quantity of susceptible people is quite large, it's not shown in the picture. And the statistics of SIR comes from the Internet.[4] we generate the initial value of λ and μ using Maximum Likelihood Estimation② and set λ to $0.4*10^{-6}$ and set μ to 0.3

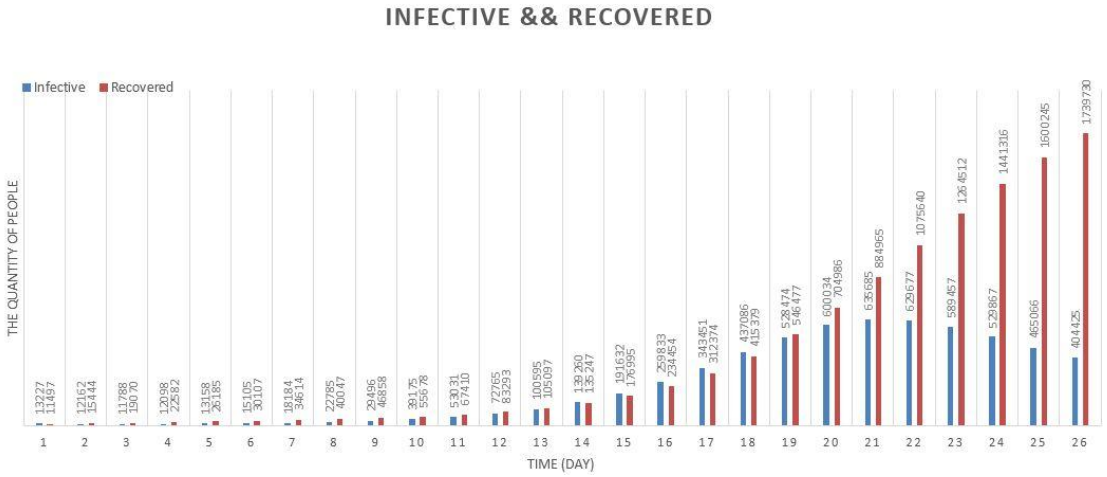


Figure 6-3 SIR model without using our delivery system

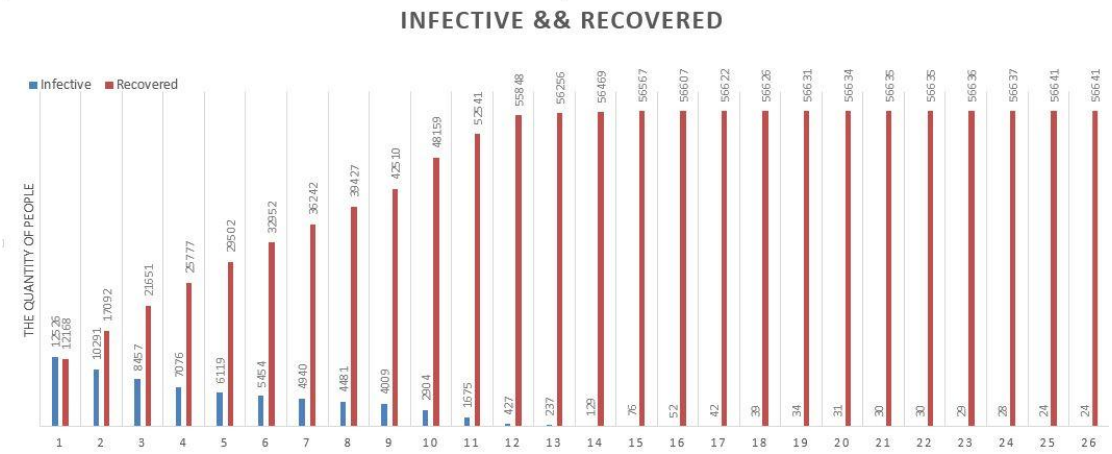


Figure 6-4 SIR model using our delivery system

This 26-day simulate experiments show that without applying our drug delivery system, the quantity of infective keeps increasing for some day and starts to decrease. But while using our delivery system, the situation becomes better quickly. In a conclusion, our drugs delivery system performs well and is effective in optimizing the eradication of Ebola.

[4] http://maps.who.int/SimpleViewer_WHO/?appid=3ada31510f2046d0939f0a1f362b241f

http://globserver.cn/en/node/38348

http://globserver.cn/en/node/41523

7 Sensitivity Analysis

Of course, there're some uncertainties in our model because we use heuristic algorithm like K-Means and Genetic Algorithm. Here we consider several important parameters that have some influence on the output of our model.

7.1 Uncertainties in K-Means

As for the Improve K-means algorithm, there are three essential parts which can make an influence to the output.

The first part is the number of central points. It directly determines the number of clusters, which represents the number of medicine delivery center in our module. The following two pictures show how output changes with it.

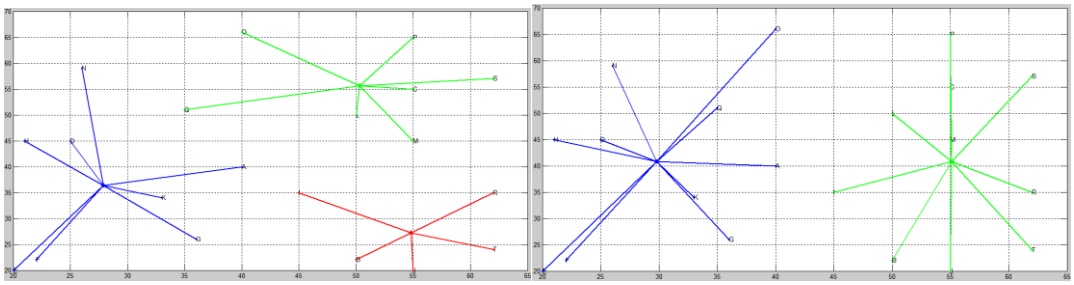
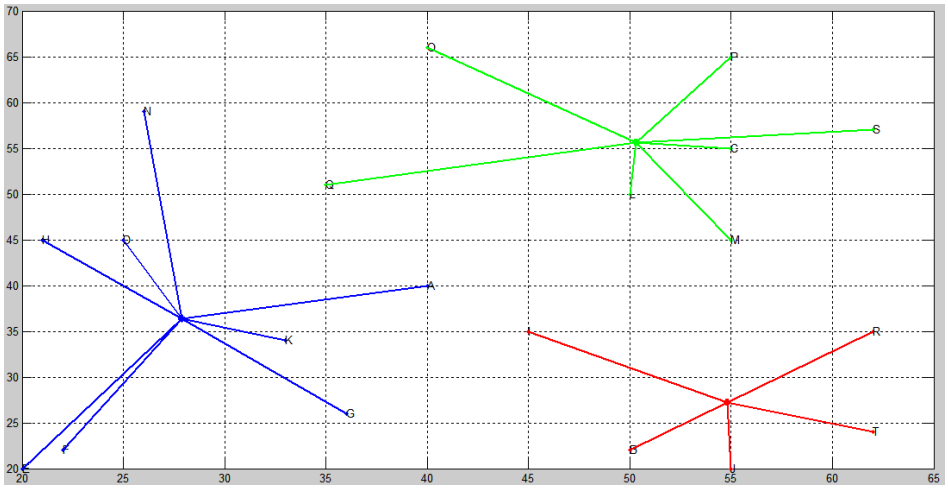


Figure 7-1

The second part is the initial values assigned to central points. Because the values of central points decide which clusters the given point belongs to, and keep affecting during the process of iterations. Different initial values may cause different output. See the following two pictures, with the same given points and different initial central points, the output becomes different.



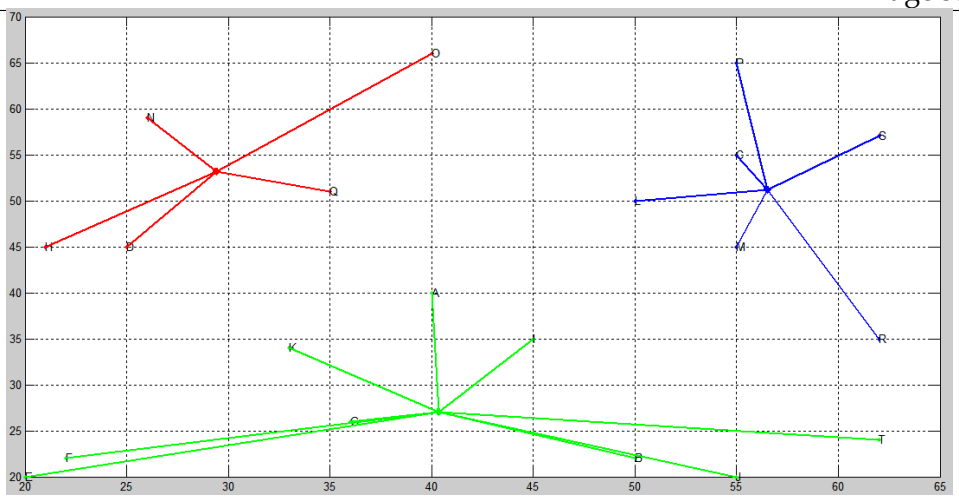
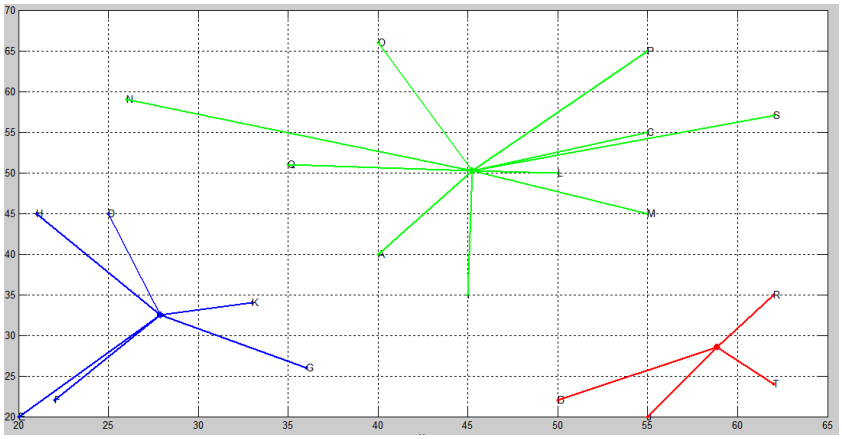


Figure 7-2

The third part is the constant k in the formula

$$w(\mathrm{i}) = k * \mathrm{NumberOfPatients}$$

Through the method, we can find that with the increase of the value of k, the difference of weights among epidemic areas will become larger, which finally results in the distance the central point move towards the epidemic area with larger weight. Assume points A, K, R are with the largest weight in their clusters. Now set k as 0.5 or 0.8 and the result is shown in the following figure. You can easily find out that when k is 0.8, central points are closer to the epidemic areas with larger weight.



7.2 Uncertainties in MDVRP model

As we mentioned above, we add three penalties P_1, P_2 and P_3 onto the distance in order to balance influence of the cost of the distance and the seriousness of disaster and the multiplier W_1, W_2, W_3 may affect our output seriously. Fortunately we set W_1, W_2, W_3 to 1 at first and it performs quite well. See one of the good solutions at Figure x1 in our experiment at someday.

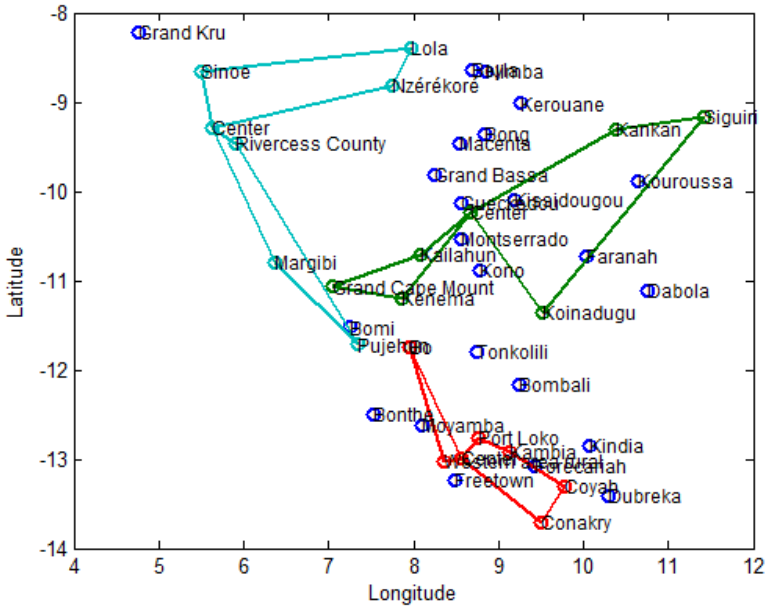


Figure 7-3 a relatively good solution

But when we change W_1, W_2, W_3 into a larger number, for example, 5 or more, the solutions become relatively worse obviously. See as figure 7-4. The SIR value of epidemic areas in the picture is the same to the figure 7-3.

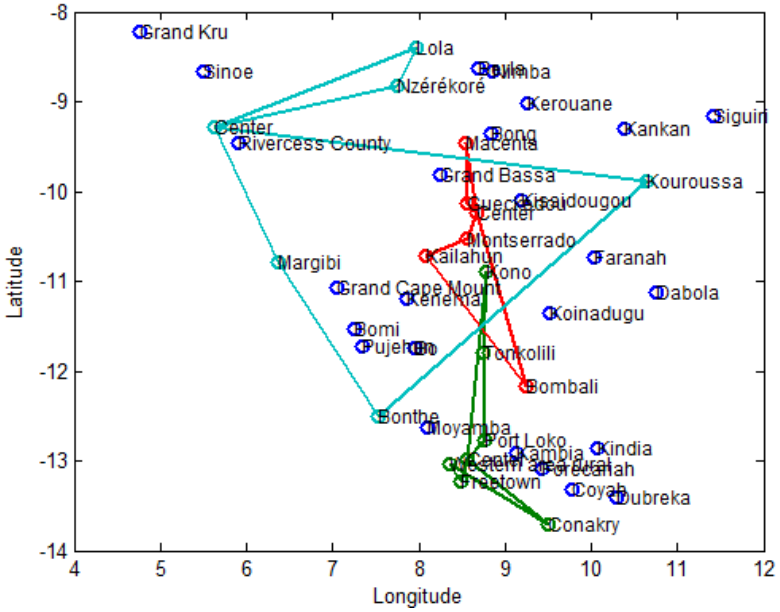


Figure 7-4 a relatively bad solution

8 Strengths and Weaknesses

8.1 Strengths

1. We use weighted K-Means algorithm to site the drugs delivery center and the location we obtain is useful in our delivery system.

2. Heuristic algorithms such as genetic algorithm combining with local search are applied in our model so when the scale of input grows large, our model will be still efficient compared to traditional mathematical method.

3. MDVRP combining with SIR balances the cost of the transportation and the severity o f the disaster.

8.2 Weaknesses

1. K-Means and its variations have a number of limitations. In particular, K-means has difficulty in detecting clusters with non-spherical shapes or widely different sizes or densities. This is because K-means is designed to look for globular clusters of similar sizes and densities, or clusters that are well separated.

2. The amount of parameters in our model is a little large and it’s not easy to determine appropriate value for them to make our system perform well.

9 Further Attempts

1. Since we use weighted K-Means algorithm to site the drugs delivery center, we need to set the number of drugs delivery center first. But sometimes it's hard for us to determine how many centers to set is better. To attain an appropriate number of centers, we can use other clustering algorithm to replace K-Means in order to gain a relatively appropriate number of centers such as ISODATA algorithm.

2. Although we consider the speed of manufacturing of drugs, the amount of vehicles in a drugs delivery center is unlimited. But in reality, the amount of vehicles in delivery center is always under limitation. If we add the constraint of the amount of vehicles, our model will be closer to real life. But in fact, it's hard to add this constraint since the split method we use is based on infinite number of vehicles.

10 Appendices

10.1 Appendix A---A Non-technology letter for World Medical Association to use in their announcement

Ebola virus disease, (EVD; also Ebola hemorrhagic fever), or simply Ebola, is a severe, often fatal disease in humans and spreads in the humans by direct contact with the blood or body fluids of an infected person. Once infected, the fatality rate is up to 50%. And considering the poor conditions of health care in West Africa, the present situation of Ebola is imminent. According to latest statistics from the WHO, there have been almost 22 500 reported confirmed, probable, and suspected cases of EVD in Guinea, Liberia and Sierra Leone, with almost 9000 reported deaths (outcomes for many cases are unknown). We World Medical Association have focused on studying the production of vaccines and drugs. Recently we eventually found a new medication to stop Ebola and cure patients whose disease is not advanced. So how to effectively deliver these vaccines and drug to control the epidemic situation becomes particularly important. We apply a special drugs delivery system to help design an efficient schedule scheme to optimize the eradication.

Considering the spread of the disease, the quantity of the medicine needed, delivery systems, locations of delivery, speed of manufacturing of the vaccine or drug, the condition of health care and so on, we establish 3 main models to simulate the delivery system in Guinea, Liberia and Sierra Leone.

Firstly, we use weighted K-Means algorithm to choose the drugs delivery centers while considering both the cost of distance and the severity of the disaster. Then we apply a modified MDVRP model with some certain constraint such as the load of the vehicle, the amount of drugs in the center, combining with the SIR model which can predict the trend of the disease to draw optimal delivery routes for the vehicles in the drugs delivery center.

Subsequently, we collected massive samples and do a great deal of experiments. The results show that our models are able to optimize the eradication of Ebola while using an efficient medicine delivery scheme at a lower cost, which take various factors into account, to help those struggling against Ebola virus disease.