Location of Remote Rural Express Station Based on Multiple Models

Yunkai Li*

Internet of Things, Beijing University of Posts and Telecommunication, Beijing, China

Liyunkai0811@163.com

Abstract – In this paper, starting from the distribution issue of rural express, the use of mobile phone positioning data, analysis of data obtained in point and OD traffic lines, dynamic traffic OD path evaluation model is set up, clear line number, eliminate the remote redundancy road, through looking for the heat point buffer model, with branch and bound method optimizing choosing six steps, this paper seeks for a measurement method to provide a reference for the site of express. Different from the traditional extensive addressing mode, this paper starts from the analysis of users' whereabouts data, comprehensively considers the objective factors such as traffic and terrain, and carries out special dissimilation analysis for all users within the scope of self-pickup cabinet coverage, so as to determine the optimal solution for site selection.

1. INTRODUCTION

China's express industry has entered the 3.0 era, which has developed rapidly in recent years. The 3.0 era of express takes data and connection as the core, which will put forward more requirements for industry management and service upgrading, and bring a better prospect for the express industry.

In 2017-2022 China's express service industry market development present situation and the investment outlook report shows that China's agriculture is in a critical period of transforming traditional agriculture to modern agriculture, the contradiction between small-scale production and big market has become the bottleneck of constraints, the rural economy development, establish perfect logistics system in the countryside become the urgent need to solve the problem.

At present, the rural courier also exist some problems, such as inflating charges, express "second rate" at the end, in some rural and remote areas, in particular, because express enterprise network radiation to the villages and towns, only some couriers to rural only by the villages and towns express delivery agents may be responsible for the delivery, thus, end node shortage has become a main reason. As an important node of the logistics network, the express station plays a role of adjustment and connection. It is an essential facility to strengthen the construction of the rural express industry. The site selection of the express station will play an important role in determining the efficiency of Courier delivery, the efficiency of villagers' picking up and the user experience. Therefore, the location of express delivery stations in remote rural areas based on big data has important theoretical and practical significance.
2. PROJECT STEPS

2.1 Collect data by mobile phone positioning

According to a survey on the communication consumption patterns of rural consumers published by Ericsson Consumer Research Office, the penetration rate of fixed-line phones in rural areas of China has dropped to 46%, while the penetration rate of mobile phones has risen to 90%. In addition, according to the guidance of the Ministry of Industry and Information Technology, 4G coverage in administrative villages across the country will be more than 98% by 2020. At present, most of the villages have communication base stations and the coverage rate is higher.

It can be seen that except for special areas, most rural areas in China have extremely high mobile phone penetration and communication base station coverage rate. In terms of cost, accuracy and practicability, the following technologies are considered: GPS, A-GPS, E-OTD, Cell-ID, AOA, TOA, CDMA positioning technology, etc. Therefore, A-GPS positioning technology was selected as the method for us to adopt positioning data in rural areas. A-GPS positioning technology is accurate, time consuming, relatively low cost, suitable for open areas such as rural areas. In the mountain villages with weak GPS satellite signals, Cell-ID (community identification code) positioning method can still realize positioning services by relying on the base station in the village, while Cell-ID is quite suitable in the rural areas with low density, but the accuracy is low due to the large area of GSM communities. In addition, the A-GPS positioning technology in the market on the smart machine has been almost completely popular. Therefore, the implementation cost is low and additional hardware equipment is not needed. Therefore, the accuracy and feasibility of using mobile phones and A-GPS positioning technology to collect positioning data in rural areas are guaranteed to a certain extent.

2.2 Obtain resident points and traffic OD lines by analyzing data

In order to find similar users in a given data set, this paper uses hierarchical clustering (HAC).

Hierarchical clustering is a kind of clustering method to build layered clusters. There are two primary implementations of this: the first is aggregation, a "bottom-up" approach. Each data point first becomes a class by itself, and then begins the interclass fusion from bottom to top. The second is a "top down" approach. All the data points start in one class, and then top down they start splitting into classes. For this selection of aggregation, we take the target function in Ward method, which is as follows:

\[ \sqrt{\frac{2|A||B|}{|A|+|B|} \cdot ||\tilde{c}_A - \tilde{c}_B||_2} \]

Using the difference formula above, users with similar activity characteristics can be grouped together.

The coordinate data of the active body within the set time period is obtained according to the previous GPS situation. The coordinate data includes the coordinates of each coordinate point and the corresponding coordinate time. This time, we choose to adopt the simulation method for parameter estimation, and use the Monte Carlo Markov chain method (MCMC) to estimate the parameter of multinomial distribution. MCMC provides a method for predicting high-dimensional parameters through numerical prediction.

After parameter estimation, the group with the highest probability is selected as the target group G and the probability set FHU corresponding to the candidate position is obtained. After that, the probability vector P needs to be predicted according to observations. The method of maximum likelihood probability is adopted here. For unknown parameters P = (P_1, P_2, ..., P_x) can be expressed as:

\[ I(p) = I(p_1, ..., p_k) \prod_{l=1}^{k} p_{n_l}^{n_l} \]

\[ g_h = g_i, v_i = \max(p) \]

\[ F_{n_i} = \{ f_i | x_i \in g_h \} \]
Finally, select the largest candidate position in the set frequency and the final permanent position $X_t$ in $gh$.

$$x_i = x_{i_1}, \quad f_i = \max (F_{i_1})$$

If more than one candidate location meets the conditions, the candidate location closest to the geographic center of all candidate locations in $G$ is selected as the permanent location. The geographic center location is roughly obtained by calculating the average longitude and latitude of all candidate locations.

Based on the "time-space" two-layer clustering method, the mobile phone registration points were clustered to get the travel stopover points. Based on the K-means algorithm, the region is divided, and the OD matrix of the trip is calculated by combining the information of the trip stop and resident points and the results of the division. The speed, distance and time interval between two points are used to determine the state of the point. The specific steps are shown in the figure below:

2.3 Dynamic traffic OD path evaluation model

On the basis of POI interest point statistics, with the purpose of establishing a comprehensive induction benefit evaluation function model based on extended and relay-type indexes of induction device, we constructed the attribute set of dynamic OD path, namely the comprehensive evaluation index system of path, as follows:

$$X = \{x_j | j = 1,2,...,n\}$$

In combination with the particularity of rural areas, we set $n=4$, where $x_1$ is the travel distance; $X_2$ is the travel time; $X_3$ is the reliability of travel time; $X_4$ is the hot spot index (heat), which is a comprehensive evaluation value considering the number and significance of POI interest points on each section of the route.

Because the dimensions of each index are different, the dimension of each evaluation index is unified as 1, and the evaluation interval is $[0,1]$. At the same time, the evaluation index is uniformly converted into the cost type, that is, the smaller the design index is, the better. At the same time, during the operation of the traffic network, there are the subjective personality preferences of the travelers and the random variation of the road network itself. Therefore, the random variable $X$ is determined to describe the random variation of the two within a certain range:

$$x_j = \bar{x}_j + \varepsilon_j$$

Where, $\bar{x}_j$ is the estimated or measured value of the traffic system model for the path indicator $x_j$; $\varepsilon_j$ is the error,

$$x_j = [x_j^{down}, x_j^{up}]$$

For a given OD route, it is better to take travel distance and travel time as the goal of path evaluation. Here, the reliability of travel time and hot spot indexes are mainly considered.

The reliability of road section travel time refers to the probability that vehicles or pedestrians pass the road section within a specified time, namely:

$$R_a = P(t_a(x_a) \leq \theta_a t_a^0)$$

Where, $R_a$ is the reliability of travel time on section $A$; $Ta^0$ is the free outflow travel time on section $A$; $\theta_a$ is the blocking coefficient of $a$, which reflects the traffic congestion level. $\theta_a$ should be greater than 1, because when the traffic capacity of a section changes randomly, it is always less than its maximum capacity, so the travel time is always greater than the free flow time.

The index of interest points should consider both the interest index and the significance. Interest index is used to represent the relative number of POI interest points on both sides of a road section. This
index can reflect the heat of a single road section or the whole path without dimension. The interest index is a continuous variable, and its value range is defined as 0~5. The size of the index value represents different levels of heat.

The section interest index (\(H_{\text{link}}\)) and path interest index (\(H_{\text{path}}\)) can reflect a certain section respectively. When the interest index is transitioning from a section to a path, weighting method is adopted for integration. The theoretical calculation formula of \(H_{\text{link}}\)

\[
H_{\text{path}} = \frac{\sum_{i=1}^{n} A_i \times H_{\text{link}}}{\sum_{i=1}^{n} A_i}
\]

Where: \(H_{\text{link}}\) represents the interest index of each section in the path; \(H_{\text{path}}\) represents the interest index of the path; \(A_i\) represents the weight of the importance of each road section; \(N\) represents the total number of sections on the path.

The significance (Sig.) of POI is mainly affected by public recognition (Cog.), spatial distribution (Cen.) and individual characteristics (Char.). Here, a POI significance measurement model composed of these indicators is proposed:

\[\text{Sig.} = c_1 \times \text{Cog.} + c_2 \times \text{Cen.} + c_3 \times \text{Char.}\]

In the formula, the sum of weight coefficients \(c_1\), \(c_2\), \(c_3\) of each vector is 1. It is easy to know that the data format after normalization processing is dimensionless and the value interval is between 0 and 1. The determination of the weight coefficient is determined according to the way of the difference of public cognition. The preset value of the weight coefficient of experts is modified according to the survey of the level of public cognition. This method can be transplanted to different regions in a low way.

In this paper, the fuzzy analysis method in fuzzy mathematics theory is used to evaluate the importance of each path, and the weight fuzzy analysis is used to solve the consistency problem of the optimal ordering problem of multiple paths. The introduction of the weight vector of the path index simplifies the complexity of judging the relative importance of the target, thus effectively solving the consistency problem in the process of optimization sequencing.

Specific evaluation steps for dynamic OD path are as follows:

a. Based on all the OD path set \(Y = \{y_i | i = 1, 2, \ldots, M\}\) and path evaluation index set \(\{x_j | j = 1, 2, \ldots, N\}\), given path \(y_i\) about evaluation index \(x_j\) attribute interval \(a_{ij} = [a_{ij}^{\text{low}}, a_{ij}^{\text{up}}]\). Forming path evaluation matrix \(A = (a_{ij})_{M \times N}\), and standardization of decision matrix \(\tilde{A} = (\tilde{a}_{ij})_{M \times N}\).

b. Determining the weight vector \(W\) of dynamic OD path evaluation index.

Solve the comprehensive cost of OD paths, and calculate the possibility degree \(p_{ij} = P(\tilde{z}_i \geq \tilde{z}_j)\) between the values of each OD path, so as to establish the possibility degree complementary matrix \(P = (p_{ij})_{M \times N}\) between each OD path.

Solve the ordering vector \(\omega\) of the complementary judgment matrix, according to the size of its component to evaluate the order of each OD route, to determine the advantages and disadvantages of all OD routes. The specific algorithm flow is shown in the figure below:

2.4 Selection of the number of routes

The selection of rural roads is particularly a more complex problem, based on the current situation of contemporary rural roads and the existing problems. Secondly, express the ark set should be considered road traffic capacity, load-point distribution, travel purpose, travel mode, travel time, travel distance, travel times, such as freight, shipped quantity, investigation, each traffic area and between each traffic area and amount of goods to and from field between, some of the basic data units, etc. (namely, passenger traffic OD and freight traffic OD) When we are faced with route selection, we take two aspects as the main consideration: dividing the number of express containers according to the resident number of rural population and dividing the line ratio in combination with traffic OD.
Specific selection strategy for classifying the country (mainly population, at the same time should also consider the way the total (basic is proportional to the population, but has not ruled out part of the special region)) : this project will be the country according to the division for very large rural population, large village, large village, medium-sized country, small villages, tiny villages. Refer to the number of roads selected and the number of express cabinets set.

The overlap part is selected by using the ratio of traffic OD line, and the number of lines is selected by considering the actual situation of the road (such as intersection situation, corresponding distance, etc.).

2.5 Road clustering
After we obtained traffic OD data and determined alternative roads, in order to simplify the complex road network in the village system, we conducted road clustering to eliminate redundant roads and remote paths. The core idea is to view the path as a node by obtaining several discrete points with a fixed distance on the road, so as to simplify the system's analysis of a large number of remote, low-weight and obviously non-target roads and improve the efficiency of calculation. Considering that the whereabouts reflected by traffic OD data often take the road section as the end point as the unit, the following analysis uniformly takes the road section as the research object.

First of all, considering that the discrete target points in this system are obtained by the interval points, we may initialize each target point of several sections cut by a road through other roads in the same cluster, then the number of clusters is equal to the number of sections in the system. For example, the "well" character system formed by four roads will be initialized into 12 roads and 12 clusters. Then, according to the distance similarity between objects, the larger clusters are divided according to a rule, and the intermediate results are combined according to the same rule. This rule is simply expressed as: reject distance and advance.

For each initial cluster, if the network distance between two adjacent objects is greater than the predetermined threshold ε, the two adjacent objects are taken as the boundary to divide the original cluster into two smaller capacity clusters. Otherwise, it is treated as a separate cluster without processing. This process ensures the compactness inside the cluster, that is, the network distance between any two adjacent objects in the cluster is not greater than ε. In addition, in order to simplify the later clustering merge process, it is also necessary to specify its effective node for each cluster in this process.

According to the distance similarity between cluster blocks, the adjacent clusters are cyclically merged to form the final clustering result. Specifically, using the effective node chain list of the cluster block, according to the threshold ε and the network distance between the cluster blocks, the adjacent cluster blocks around the nodes are merged (there is no need to merge between the cluster blocks after splitting on the same edge), and the adjacent clusters around each node are merged repeatedly until they can no longer merge.

Considering that the threshold ε has a certain influence on clustering, the system needs to collect actual data for training, in order to find the appropriate ε value. After this value is determined, we can cluster the previously proposed sites through the above operation. Considering the equal density of discrete target points in the system (fixed length selection point), then, the corresponding sections of sites that were ultimately excluded from the maximum cluster can be eliminated.

Citing the "well" system mentioned above, the result will be the four sections represented by the "mouth" in the middle after processing according to the above rules. Of course, based on the actual situation, the selection of sections is not necessary. If all networks are included in the same cluster, the deletion operation can be skipped.

After the above processing, the number of road sections limited by traffic OD will be further reduced, and the road will be subdivided into sections, which will greatly reduce the complexity and workload of the algorithm.
2.6 Buffer finding hot point

The hot point is the key point related to this problem, and it is also the inevitable choice to further obtain the pick-up point of the express cabinet. The search for the hot point mainly needs to integrate several aspects: first, the terrain, the living location of the villagers, the relevant roads, water conservancy and a series of conditions; second, the reasonable use of traffic OD data and optimization; This paper mainly adopts ArcMap in GIS (Geographic Information System) for operation, and its main model is buffer analysis problem. Open ArcMap and call the local topographic data and DEM data (DEM data is mainly used to analyze the selection of hot points by topographic factors) to set relevant elements: Set convenient terrain, road with relatively short distance (integrated clustering algorithm), and cover more populated areas and roads with high traffic OD ratio to the maximum extent, etc. The buffer analysis adopts an accuracy of about 30m, which can be optimized to the maximum extent and reduce errors.

The ID in the buffer attribute is taken as the discriminant condition, and the values of corresponding fields in residential houses and places with large flow of people are set to 1, and the corresponding fields of the buffer in the main road network are set to -1 to distinguish the buffer levels. Take the union of five buffer regions and use the union tool in stacking analysis. Add a class field to the layer property sheet after union, as the selection basis, and add the superposition effect of each condition on it. (Tool: field calculator). Then, according to the value of the class field, it is displayed in a gradient color band. The darker the color, the better the condition of the area.

![Fig1. Arcmap analysis of sample buffer area](image)

We use the conditional function tool to extract the best position. It has been determined that only locations with a suitability value of 9 (the highest value of Suit_Areas output) are considered optimal locations. In the conditional expression, all areas with a value of 9 will retain their original value (9). Regions with values less than 9 will be changed to NODATA. Leaving raster data or constant value parameters taken when the input condition is false applies the default value. By default, if any value in the input condition raster data does not satisfy the input condition, it is assigned NoData in the output raster.

3. CONCLUSION

On the basis of solving the location problem, this project adopts the idea of optimizing the operation complexity in many places and dealing with it in the simplest way with the least error.

For example, the use of A-GPS in the selection of location information can ensure very low cost while maintaining the best accuracy. At the same time, OD uses road clustering algorithm combined with GIS analysis in the link of route selection, which more skillfully connects the complex road situation in rural areas. The final point of interest is selected by spatial analysis, GIS ArcMap. Compared with the general traditional mapping algorithm and the general algorithm for map processing, it has more efficient advantages, and can greatly reduce the time complexity. Due to the
improvement of the accuracy of the national satellite map, the use of GIS can minimize the error and cover a wider range. For the location of rural areas, it has maximum convenience and carpet coverage.

REFERENCES

[1] Chao Wang. A Dynamic OD Acquisition and Evaluation Method Based on Mobile Phone Location Data. TN929.53

[2] Zheng Menglei, Xue Xin, Chen Hui, Lu Xiane. A Dynamic Traffic OD Algorithm Based on 4G Data DOI: 10.26914 / Arthur c. nkihy. 2019.003833 classification number: U495 Regional Traffic Demand Forecasting Based on OD Data Measured by Mobile Phone Application Software, GE Mengxue, SONG Guohua, WANG Pengsen, HOU Jie, YAN Xiaoyong "Feature Analysis and Visualization Based on User Big Data

[3] Chen Jidong, Meng Xiaofeng, Lai Caifeng. Object clustering based on road network. Journal of Software, Vol.18, No.2, February 2007, pp.332–344.

[4] kaitexue the location analysis of ArcGIS | spatial analysis ] 【, 2018.10

[5] Li Yi. Study on the Branch and Boundary Method of Logistics Center Location Problem. 2012.5

[6] Xiong Wei, Operations Research (3rd edition), Machinery Industry Press, 2014:81

[7] Si Shoukui, Sun Zhaoliang. Mathematical Modeling Algorithms and Applications. 2nd Ed. : National Defense Industry Press, 2015