Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/132/2024.20705

Research on Optimal Layout of Express Delivery Stations in Beijing

Shiyu Zou

Chongqing DEPU Foreign Language School, Chongqing, 400000, China

zsy651000@icloud.com

Abstract. Combining DBSCAN clustering algorithm and genetic algorithm, this paper presents a method to optimize the layout of express delivery sites in Beijing. Firstly, DBSCAN algorithm is used to identify high-density areas in large-scale data and extract the preliminary express site location. Genetic algorithms are then used to further optimize the location and number of these sites to minimize total delivery costs. In problem 1, DBSCAN algorithm is used to solve the problem of identifying high-density areas, and the results show that it can effectively identify areas with intensive express demand. In problem 2, a genetic algorithm is used to optimize the preliminary clustering results, and the optimal express site layout scheme is obtained, which significantly reduces the total delivery cost. In the third problem, the Mercator projection formula is combined to convert the projection coordinates to geographical coordinates to ensure the accuracy and practicability of the calculation results. The model in this paper is efficient, adaptable, and practical and has a wide application prospect.

Keyword: DBSCAN; genetic algorithm ;express site optimization, Mercator

1. Introduction

1.1. Project Background

Express network refers to the logistics nodes set up by express enterprises in the corresponding areas of the city to provide efficient distribution services and independently arrange the spatial distance between the sender and the recipient to realize the rapid pick-up of the piece in the blank [1]. The express network, in the usual sense, includes a standard operation network and an intelligent self-pickup terminal [2]. Due to the vigorous development of express delivery services, express delivery outlets are usually close to end customers and dispersed at the entrance of cities or residential areas [2]. In recent years, with the expansion of the number and scale of express delivery enterprises, the construction of express delivery outlets has faced many challenges.

On the one hand, express delivery points are gradually growing, and the liquidity of outlets is growing. Compared with meager profits, the fixed cycle of express outlet construction is long, and it is difficult to face the risk of transfer or flow of the logistics market brought about by the temporary trend [3]. On the other hand, express outlets generally have low resource utilization. In order to compete for market share, enterprises blindly increase the network density, resulting in a low utilization rate of logistics resources [4].

In 2018, the General Office of the State Council issued the *Opinions on Promoting the Coordinated Development of E-commerce and Express Logistics*, proposing to scientifically guide the construction of express logistics infrastructure, optimize the layout of express outlets, and encourage the use of public resources and facilities to plan and arrange intelligent express boxes, taking into account factors such as regional location, functional positioning, and development level [5]. Integrated express terminal services are included in the relevant planning of public service facilities.

In the actual logistics distribution, the terminal distribution network serving residents' life is responsible for improving people's livelihood. If the scientific layout and construction of the express delivery network can play a dual role in improving the economic benefits of enterprises and the quality of life of residents [6]. At present, Beijing and other big cities are gradually promoting the intelligent development of outlets, and the construction and optimization of express delivery outlets play a crucial role in the development of logistics distribution systems and the improvement of residents' quality of life [7].

Therefore, the construction and optimization of express outlets is the key to the efficient operation of express services, but it faces many challenges such as low resource utilization, fierce market competition, and a long construction cycle. Through scientific and reasonable planning and layout, it can improve the utilization efficiency of logistics resources, enhance the economic benefits of enterprises, and improve residents' quality of life.

1.2. Research Purpose and Significance

Based on the actual logistics network layout in Beijing, this study intends to determine the rational relationship between express network layout and other service facilities in Beijing. Big data analysis technology and genetic algorithm were used to collect data on Beijing's entrance, commercial network layout, and traffic conditions that may affect network layout, analyze the correlation coefficient between the number of express network outlets in a certain region and other data and establish a multiple regression model to accurately determine the quantitative relationship between the number of logistics network outlets and various factors. A genetic algorithm is used to optimize the number and layout of the express network, and the best express network layout scheme is obtained. This study has the following significance:

For express delivery enterprises, according to the model and conclusion obtained in this study, the approximate number of express outlets in a region can be known in advance, which avoids the waste of resources caused by blindly increasing the number of outlets in areas with dense outlets and saves the investment cost of the construction of express outlets.

the perspective of urban planning by the government, the models and data obtained in this study provide a decision-making basis for the reasonable planning of express delivery outlets and various types in metropolitan areas of Beijing and are of great significance for the fine implementation of the coordinated development of Beijing-Tianjin-Hebei.

For academic research, the genetic algorithm proposed in this paper is applied to the study of the layout of express outlets to realize the intelligent development of the prediction and optimization of the number of regional express outlets. The modeling idea provides a method for reference for analyzing the layout of express outlets or logistics service facilities in other cities.

2. 2. Research Methods

2.1. The Principle of Genetic Algorithm

Genetic Algorithm (GA) is an optimization algorithm based on natural selection and genetic mechanisms to search for optimal solutions by simulating biological evolution. Its basic principle includes the following aspects [8].

2.2. Encoding

The solution to the problem is represented as a string (chromosome), each string consisting of a series of genes. It usually adopts binary coding, real coding, or symbol coding [8-10].In this study, each

DOI: 10.54254/2755-2721/132/2024.20705

chromosome represents a delivery network layout scheme, where genes represent the location and number of delivery networks.

2.3. Initial Population

A certain number of chromosomes are randomly generated to form the initial population. The diversity of the population guarantees a wide coverage of the search space [8].

2.4. Fitness Function

The fitness function is used to evaluate how good or bad each chromosome is and usually corresponds to the objective function of the problem. In this study, the fitness function can comprehensively consider the coverage of outlets, service quality, operating costs, and other factors [11].

2.5. Selection

Select good individuals to enter the next generation according to their fitness value. Common Selection methods include Roulette Wheel Selection, Rank Selection, Tournament Selection, etc [12].

2.6. Crossover

A new chromosome (progeny) is created by exchanging parts of the genes of two chromosomes. Common Crossover methods include Single-point Crossover, Multi-point Crossover, Uniform Crossover, and so on. Example of a single point crossing:

Parent 1: 110010 | 110
Parent 2: 101101 | 001
Child 1: 110010 | 001
Child 2: 101101 | 110

2.7. Mutation

Random changes are made to some genes of the chromosome to increase the diversity of the population and prevent falling into local optimality. Commonly used Mutation methods include Bit Mutation, Swap Mutation, Insertion Mutation, etc. Example of bit variation (encoded in binary):

Original Chromosome: 110010110
Mutated Chromosome: 110011110

2.8. Termination Condition

Determine whether to terminate the algorithm according to the preset conditions. Common termination conditions include reaching the maximum number of iterations, convergence of fitness values, or reaching a preset threshold.

2.9. Symbol Description

There are five variables, P_{ij} presents population density of grid points, Q_{ij} presents the total demand of the grid points, SC_{ij} presents fixed cost of setting up delivery points at grid points, d_{ij} presents distance from delivery point to building, and Q_{ij} presents the amount of goods delivered per kilogram at express points set up at grid points. Then, the total cost consists of distribution costs, fixed costs, and operation costs.

2.10. Model Establishment

First, we conduct a basic modeling on a small scale. The main goal of the model is to minimize the total cost, including distribution cost and fixed construction cost, to optimize the location and number of express stations.

The minimize cost formula [13]:

$$Total\ cost = \sum_{(i,j)\,\in\,S}\left(d_{ij} \cdot \frac{1}{p_{ij}} \cdot Q_{ij} \cdot E_{ij} + C_{bij} \cdot T_{ij} + O_{bij} \cdot E_{ij}\right) \tag{1}$$

In this formula, d_{ij} represents the distance between grid point (i , j) to the center of the delivery station is calculated by the Pythagorean theorem.

Second, we set some constrains to limit the feasibility range and help to get the optimal solution of the express station location.

The distribution volume of each region must meet its total demand.

$$Q_{ij} \geq D_{ij} \quad \forall(i,j)$$

The total delivery volume of all express stations is fixed.

$$\sum_{(i,j)\in S} Q_{ij} = 10000$$

The total demand of the region does not exceed the total delivery volume of the express station:

$$\sum_{(i,j)\in S} D_{ij} \leq 10000$$

Third, we assume each region has a different value of population density, economic prosperity, terrain factors, construction costs, and operating costs. As shown in Table 1, "Population Density" represents the number of people in each square, and the higher the value, the lower the distribution cost in that area; "Economic Prosperity" indicates the economic degree of each square, the higher its value indicates the higher the distribution cost of the area; "Terrain Factor" indicates the difficulty of the terrain of each block, and the higher the value, the higher the construction cost of the area; "Construction cost" indicates the amount of RMB needed to build the express station in each block; "Operational Cost" refers to the number of RMB required per square to operate the express station.

Table 1. Parameter value of different types of regions

Area Type	Population Density/square	Economic Prosperity	Terrain Factor	Construction Cost/RMB	Operational Cost/RMB
Living quarters	700	1	1	100000	5000
Industrial area	500	1.2	0.8	120000	6000
waters	300	0.9	1.1	90000	4000
cropland	200	1.1	1.2	80000	3000
Central area	800	1.3	0.9	130000	7000
Commercial area	1000	0.8	1.3	150000	8000
Open area	400	1	1	100000	5000

Finally, we establish a basic map of small scale. As shown in Figure 1, the green part presents living quarters, blue part presents the industrial area, the light grey part presents waters, the yellow part presents cropland, pink part presents central area, brown part presents the commercial area, and darkgrey presents part open area.

DOI: 10.54254/2755-2721/132/2024.20705

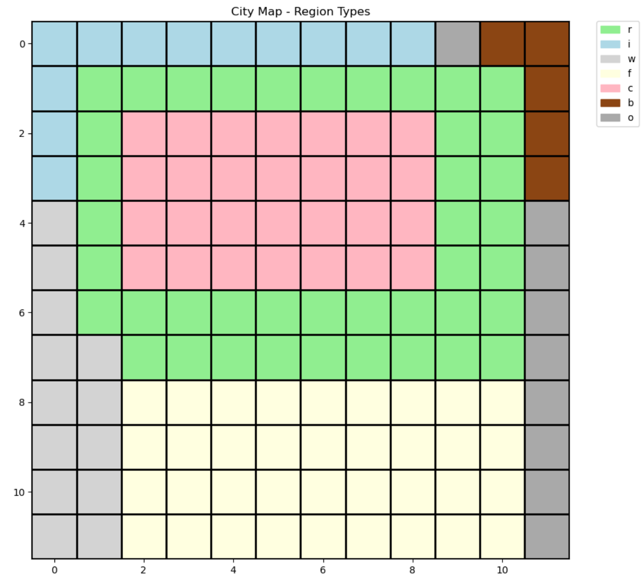

Figure 1. Map of small scale

2.11. Model Assumptions

It is postulated that the demand for express delivery in diverse regions of Beijing remains relatively stable and, consequently, will not undergo significant fluctuations during the study period. It is hypothesized that the services provided by all express outlets are homogeneous; in other words, the service capacity, quality, and type of each outlet are identical. The assumption is made that the optimization objective is singular: to minimize operating costs, provided that coverage and service quality are ensured. It is also assumed that each express outlet has a fixed service scope, meaning that an outlet can only serve customers within a certain radius. Taking Beijing as an example for optimization, we introduce the DBSCAN-GA based express site selection model (as depicted in Figure 2). In accordance with the existing data and the shape file of Beijing, we have roughly divided each region, as illustrated in Figure 2.

Figure 2. Regional division of Beijing

DOI: 10.54254/2755-2721/132/2024.20705

2.12. The Principle of DBSCAN Algorithm

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm [1]. It determines the cluster by defining a region (with ε as the radius) and a minimum number of points (min samples).

Given point P and distance threshold ϵ,If point Q meet the following conditions, then Q is P. The direct density can reach the point.

$$\text{dist. }(P, Q) \leqslant \epsilon \tag{2}$$

If point P has at least "min samples" as the core point, and the center point of the smallest min samples of the direct density reachable points is the core point. Then, the genetic algorithm is used to optimize the objective function. Then, the genetic algorithm is used to optimize, and the objective function is constructed for iteration to achieve the DBSCAN condition, that is, to end the iteration.

2.13. Mercator projection

In the calculation process, in order to calculate the distance, we use the Mercator projection of latitude and longitude. The Mercator Projection is a type of map projection first proposed by the Flemish geographer and cartographer Gerardus Mercator in 1569. The Projection is an isogonal projection, or Conformal Projection, which preserves the Angle and shape in a small area but distorts the area, especially near the poles. The Mercator projection uses a mathematical formula to convert the latitude and longitude (geographic coordinates) of the Earth's surface to coordinates on a plane (Mercator coordinates).

The Mercator projection maps points on a sphere onto a cylinder of infinite height. Mercator projection forward transformation formula (from geographic to Mercator coordinates) [14].

$$\begin{gathered} x = R \cdot \lambda \\ y = R \cdot \ln\left(\tan\left(\frac{\pi}{4} + \frac{\phi}{2}\right)\right) \end{gathered} \tag{3}$$

In this formula, R represents the radius of the Earth (an average of 6,371km is usually used), λ represents longitude in radians, ϕ represents the dimension expressed in radians, ln represents the natural logarithm. Mercator projection inverse transformation formula (from Mercator coordinates to geographic coordinates).

$$\begin{gathered} \lambda = \frac{x}{R} \\ \phi = 2 \cdot \arctan\left(e^{\frac{y}{R}}\right) - \frac{\pi}{2} \end{gathered} \tag{4}$$

In this formula, x and y are Mercator coordinates. Additionally, e, which is approximately 2.71828, is a constant of nature.

DOI: 10.54254/2755-2721/132/2024.20705

3. Results

3.1. the Result of Establishing Model

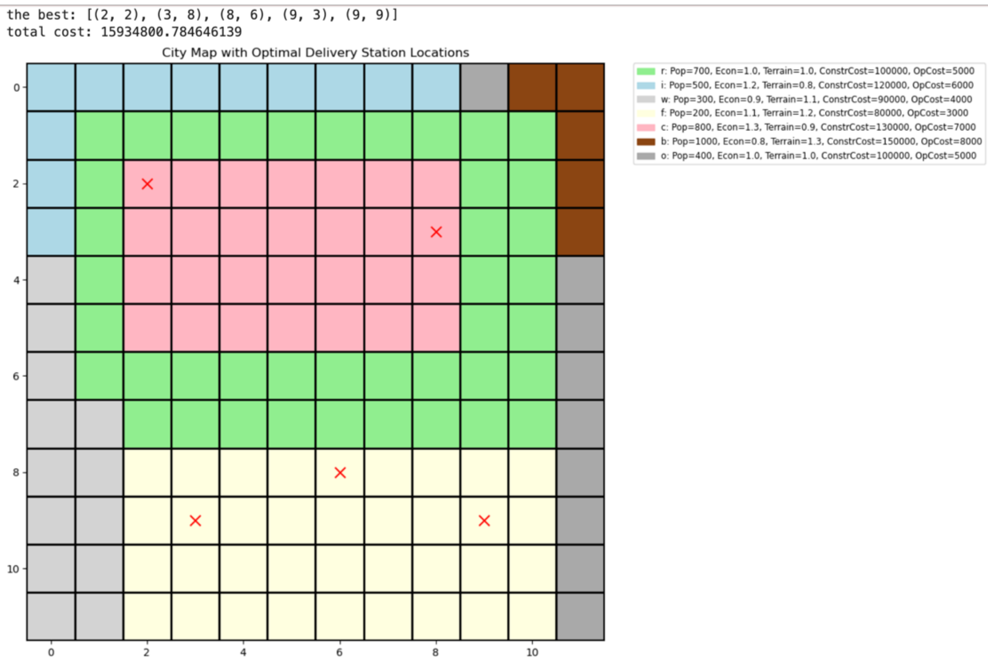

Figure 3. Optimal express station location

As shown in Figure 3, The Red Cross represents the location of the delivery station, and the box in the upper right corner describes the set values of population density (Pop), economic prosperity (Econ), terrain (Terrain), construction costs (ConstrCost) and operating costs (opCost) for each color block. At the top, “the best” represents the specific location of each of the five delivery stations, and “total cost” refers to the total cost of delivering these five delivery stations.

3.2. The Result of Beijing Map

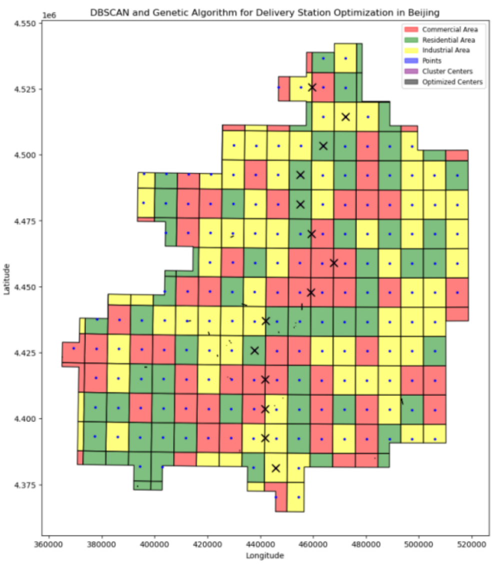

Figure 4. Selection of express delivery sites in Beijing

DOI: 10.54254/2755-2721/132/2024.20705

As shown in Figure 4, the red area represents the commercial area, the green area represents the residential area, the yellow area represents the industrial area, the blue dot is the center of each small square, and the black cross represents the location of the Courier station. The vertical line on the far left represents latitude, and the horizontal line represents longitude. Most delivery stations are located in commercial and residential areas. According to DBSCAN, we searched the following 14 coordinates for regional sub-sites, with a total cost of 4767639.4840919785.

4. Conclusion

This study combines the DBSCAN clustering algorithm and the genetic algorithm to optimize the layout of express delivery stations in Beijing. The DBSCAN algorithm can find places with dense express delivery demand. At the same time, the genetic algorithm can improve the initial clustering results and make the total delivery cost as low as possible. Using DBSCAN allows the model to process a lot of data when there is a lot of data, and the genetic algorithm can also make the location and number of delivery stations more accurate. This research can improve the overall efficiency, not the efficiency of a small area.

This research has many practical benefits. First, for express companies, this model can give the optimal number and location of express stations so that the overall cost is low. Second, in terms of urban planning, the results of this study can provide the government with an idea to help plan or improve urban logistics facilities. Finally, this study uses a genetic algorithm in the field of logistics, which provides a certain reference for academic research on the optimization of placement.

As for the disadvantages and improvement methods of this model, the first disadvantage is that this model is too sensitive to parameters, such as the distance between neighbors, the minimum sample size, crossover rate, and variation rate, which have a great impact on the results. If the parameters are not selected well, the results may have great errors. Another problem is that the calculation is too complicated when the model is processing a large number of data, the genetic algorithm is very difficult to calculate, it takes a long time to get the result, and it will be trapped in the local optimal solution, even if the genetic algorithm can find the optimal position everywhere in the map, but in the complex space may still be trapped in the local optimal solution, can not find the global optimal solution. We can improve the model by optimizing the parameters. We can use cross-validation or grid search to make the parameters of DBSCAN and GA better, so as to improve the stability and accuracy of the model. A hybrid algorithm can also be used to combine other optimization algorithms to make the search power in the whole space stronger and prevent getting trapped in the local optimal solution. Parallel computing can also be used to allow the calculation process of the model to be carried out at the same time, speeding up the processing of large amounts of data and making the model more efficient.

In conclusion, this study gives a complete and useful way to improve the layout of delivery stations. It can be used in logistics, urban planning and other fields. This approach can not only solve the current logistics problems but also pave the way for the future design of intelligent logistics network progress.

References

[1] C. Dülsen, Flick, T. , Ghring, T. , Wagner, W. , & Wensing, M. . (2023). A novel approach for FPGA-to-server data transmission over an ethernet-based network using the express data path technology. IOP Publishing Ltd.

[2] Yuzhao Zhang, Zhenjiang Zhang, Zhimo Jiang, Xiaorong Wang&Muchen Ye.(2024).Study on Low-carbon Transport Scheme of Express Freight Based on Space-time Network.Engineering Letters(6).

[3] Diaz, R. , Phan, C. , Golenbock, D. , & Sanford, B. . (2022). A prescriptive framework to support express delivery supply chain expansions in highly urbanized environments. Industrial management & data systems.

[4] Bing-Xin, W. , & Hong-Lian, G. .(2014). Service operation process optimization research in an outlet of Zhongtong Express. IEEE.

[5] The Central People's Government of the People's Republic of China. General Office of the CPC Central Committee. The General Office of the State Council issued the Opinions on Promoting the Coordinated Development of E-commerce and Express Delivery Logistics [EB / OL]. (2018-01-23). https://www.gov.cn/zhengce/content/2018-01/23/content_5259695.htm.
[6] Zhou, X. J. . (2013). Research on logistics value chain analysis and competitiveness construction for express enterprises. American Journal of Industrial and Business Management, 3(2), 131-135.
[7] Zhang, Y. , Su, M. , Cai, C. , Liu, Z. , Zhang, D. , & Han, L. . (2020).Research on the layout of express delivery outlets in jinan city. IOP Conference Series: Earth and Environmental Science, 526(1), 012203 (6pp).
[8] Xufa, W. . (1995). Genetic algorithm and its application. MINI-MICRO SYSTEMS.
[9] Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. Addison-Wesley.
[10] Holland, J. H. (1992). Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. MIT Press.
[11] Li, Q. , Sun, J. , & Zhou, J. . (2006). Improved algorithm for calculating ga fitness function of aeroengine pid controller parameter optimization. Journal of Nanjing University of Aeronautics & Astronautics, 38(2), 162-165.
[12] Lipowski,A.,&Lipowska, D..(2012). Roulette-wheel selection via stochastic acceptance. Physica A Statistical Mechanics & Its Applications, 391(6), 2193-2196.
[13] Ma, R. , & Yuan, J. . (2016). Online scheduling to minimize the total weighted completion time plus the rejection cost. Journal of Combinatorial Optimization.
[14] Richard Gott, J. , Mugnolo, C. , & Colley, W. N. . (2006). Map projections minimizing distance errors. Cartographica the International Journal for Geographic Information & Geovisualization, 42(3), 219-234.