Simulation of Logistics Frequent Path Data Mining Based on Statistical Density

Fengju Hou*
Zibo Vocational Institute, Shandong, China

***Abstract*—Sharp increases and rapid development followed the effects of a novel coronavirus outbreak on online sales and the real economy. The e-commerce mode on the Internet has attracted much attention, and users' purchases on the Internet has never been done before. However, among the many express companies, as the ones closest to consumers, they can still provide high-quality products in the face of huge market demand. Urban terminal logistics refers to the purpose of express services to meet the needs of terminal customers under the requirements of logistics centralization and customer diversification. However, the geographical distribution of logistics services in China is comprehensive, and customers' requirements are also complex. Practical problems in logistics enterprises in China significantly restrict the quality of logistics services. The final kilometer of distribution is composed of many links, and it is a very cumbersome enlace; it contains the determination of distribution scope, loading goods, arrangement of distribution sequence, arrangement of vehicles or personnel scheduling, and planning of distribution routes. A Genetic Algorithm (GA) with local search method fusion is proposed for fast logistics data modeling and mining simulation analysis. Practical examples and literature data prove the method's accuracy.**

Keywords—Statistical density; logistics; the path; the simulation data

I. Introduction

Since transportation costs have such a significant impact on the whole process, it is essential to optimize coordination routes in order to minimize transportation time and costs. This method increases efficiency while decreasing overall transportation costs. From the perspective of transportation constraints, Cheng Nan constructed a terminal transportation route containing transportation constraints in order to reduce transportation costs. Li Kunyenga proposed a basic mathematical model based on linearization by linking logistics' terminal characteristics with customers' needs. Based on this model, he proposed a new branch and truncation method, which included a new method to decompose the efficiency inequalities. The study [1] proposes using vehicle routing planning, joint distribution, and various forms to optimize rural distribution with low-density, long lines. The Genetic Algorithm (GA) is applied to optimize "at the end of the line" and maximize vehicle capacity and distribution costs. Using quadratic programming, [2] proposed a logistics distribution model that optimizes warehouse-to-distribution-center direction, customers, and express transport routes. Implementing a multigroup GA, its effectiveness was validated and confirmed. Urban terminal logistics distribution in China still has these issues due to complex and changing work.

The cost of transportation remains exceptionally high. For now, logistics costs remain high among end-city logistics companies. The current problems in China are as follows: In the tip area of China, because the customer group is widely distributed, it needs a lot of human and material resources investment in infrastructure, equipment, and other aspects to increase investment, such as the establishment of the automatic warehouse and intelligent automatic suitcase, not only to invest heavily but also to increase the cost of project operation and maintenance [3–4]. Ineffective logistics is due to a lack of resources because the requirements of each user for logistics are very different, in the whole logistics system; no one can realize the coordination process, which reduces the efficiency of the whole logistics system. There is a delay in the delivery of the goods. In the end, the logistics, according to their actual situation, were to conduct the distribution. Our country has many subjective reasons, such as imperfect traffic routes. In addition, to some areas, the scale of logistics enterprises is small, and they often need to stay in the distribution center for a short time before distribution, which leads to an extended delivery time. Most express companies have not developed a set of perfect reward and punishment mechanisms in the process of terminal distribution, which leads to problems of lax work attitudes among their staff, product loss, damage, and delayed delivery, resulting in low customer satisfaction [5–6].

According to the current research results, various intelligent algorithms are applied to logistics path planning to achieve efficient and intelligent path optimization. It is the most commonly used path planning method. With the increasing intelligence of the express delivery industry in China, many people have begun to discuss how to establish an efficient online delivery system from the perspectives of intelligent self-lifting boxes, distribution center location, and distribution path selection. Zhang Shabo put forward the cost model, pricing model, and robust model according to the characteristics of intelligent express, analyzed the case with some examples, and put forward some reference opinions on how to construct a terminal distribution network in the construction and operation. Wu Maupye et al. analyzed the key problems faced by terminal logistics in property-free communities from two perspectives of the introduction and investment of express boxes, as well as optimization strategies such as terminal logistics collaborative distribution, standardized logistics management in property-free communities, optimized pickup resource allocation, and

*Corresponding Author.

omnichannel inventory operation management in order to provide decision-making reference for the intelligent express cabinet to solve the "last kilometer" terminal logistics in the non-realty community. Lee's action is based on data mining, to solve the problem of terminal distribution mode selection, through this technology, build a terminal distribution mode selection, and solve the data instance, finally using cluster algorithm to solve the problem of the terminal distribution system, through the case study, Finally, a better result than the existing distribution system is obtained according to the "delivery to home" and "customer pick up" characteristics.

On the choice of the distribution center, Yin Xiaoqing 8 and others, based on travel time reliability and a complex network structure, built a city distribution station at the end of the cold chain planning model and conducted the simulation calculation. An example that confirms the feasibility and effectiveness of the method is the assurance of utmost timeliness. Hi-ping Ren proposed the establishment of a network in the midst of the sorting center with end customers, with the aim of resolving the intricate issue of "end" distribution for retail enterprises. The network functions as a link between the customer and the sorting center, arranging the distribution sequence from the distribution center to the client's terminal within the classification and distribution network. Terminals could be office buildings, residential areas, or districts. The existence of a hierarchical sorting mode makes the business between employees and customers less, minimizing the consumption of time, but also reduces the logistics company to the user's delivery demand and the overall layout of the delivery mode, so that in the case of meeting consumer needs more actively for consumers and the company to bring a win-win situation. Zhao Xuedong for all kinds of fresh food, all kinds of logistics center layout and route optimization, set up a balanced carbon dioxide emissions and customers at the lowest transport demand stratified programming model, and under different fresh demand, the selection of logistics center and transportation route setting has a significant effect. An example has proved the correctness of this method. Based on providing efficient, safe, and fast distribution services for e-commerce customers, John Fernier et al. set up a regional distribution center to meet customers' high-quality service requirements and, on this basis, reduce logistics costs [7-9].

The paper is organized as follows: Section II presents the research method, Section III presents the proposed methods and result analysis of this work, and Section IV concludes the work.

II. Research Method

A. Overview and Application of Statistical Density Methods

In mathematics and technology, the most significant logistics problem is the optimization of statistical density. The frequent logistics routing problem is a hybrid optimization problem with various representative problems. As soon as the research results of statistical density are published, it has attracted the attention of many researchers. A passenger starts from one place, sells things in another, and returns to the origin. There is only one visit to a place. When the distance between the two places is known, he can decide his optimal way of walking according to his situation. Before the logistics enterprise is discussed in detail, this article will elaborate on the basic knowledge of the logistics enterprise problem - the basic idea of the authorization map. In graph theory, the frequent problem of logistics routing is the minimum Hamiltonian cycle method, and the Hamiltonian cycle is a kind of node without a branch. A weighted graph is an undirected graph with a weight attached to each edge. Denoted as a weighted graph, it is a set of vertices, a set of edges, and represents the distance between vertices, Eq. (1)

(1)

Let denote the sub path from to, which takes the value 1 if it exists in the travelling salesman path and 0 otherwise, and the expression is given below, Eq. (2) and Eq. (3).

$$X_{ij} ij(i \to j) X_{ij} \tag{2}$$

$$X_{ij} = \begin{cases} 1, (i, j) \in L \\ 0, (i, j) \in L \end{cases} \tag{3}$$

where is the solution sequence, then the objective function of the travelling salesman problem is expressed as a formula, and the travelling salesman's total distance through the path must be minimized, Eq. (4) and Eq. (5).

$$L(i, j) \in E \tag{4}$$

$$mZ = \sum_{i \neq j} d_{i,j} X_{i,j}$$

$$\begin{aligned} & mZ = \sum_{i \neq j} d_{i,j} X_{i,j} \\ & \sum_{i \in v \neq i \neq j; x_{ij}} = 1 \\ & \sum_{j \in v \neq j \neq i; x_{ji}} = 1 \\ & \sum_{i,j \in V; x_{ij}} = |L|, L \in E \end{aligned} \tag{5}$$

It is a solution and then the objective function of the frequent logistics routing problem is expressed by the equation above, which requires the shortest total distance of the route through which the logistics need to pass.

In algorithmic terms, a publicly recognized data set tests and evaluates the computational power to solve a problem. It contains the best test data for a particular problem and the best solution for a problem [10-14].

B. Simulation of Logistics Frequent Path Data based on Statistical Density

This paper proposes a standard model based on statistical density simulation interval LIB for the frequent logistics path problem. The first statistical density simulation interval LIB found originates in some previous research articles by Dantzig, Fulkerson, Johnson Held and Karl, Karr, and Thompson et al. GerhardReinelt collected many practical problems, for example, in industry and geography, and gradually accumulated, eventually resulting in preliminary experimental data. LIB's statistical density simulation interval currently includes more than 100 examples, covering a wide range, from 14 small to 85,900, strongly supporting solving logistics enterprise problems. However, some of these examples are fabricated purely to solve the frequent person problem of logistics routing, such as the statistical concentration modulus simulation value of sample 225, the correction of the LIB of the statistical concentration simulation has been stopped, and the results of a large number of statistical simulation experiments.

The statistical concentration simulation section LB includes three test data:

1) The symmetry of traveller's.
2) Hamiltonian ring problem.
3) Asymmetric logistics path frequent person problem.

On this basis, in the statistical density simulation region of LB, the format of the test data set that can be used to assess the method's performance is not uniform.

Location of the space: In this format, each point is identified with a number from 0 to the largest, and the coordinates of this point in the coordinate system are given according to the number. The initial test data is the spacing between all points in this format. In this format, the initial test data are the upper triangles of the spacing matrix between points. It is divided into diagonal, nondiagonal, row-first, and column-first, according to the order of precedence of whether there is a diagonal or not and row-first [15].

In the swarm, each person is an artificial intelligence that randomly and progressively constructs a solution to a class of optimal problems. Statistical density, a traditional algorithm, has been used more in road problems because its efficiency has always been the focus of attention, Eq. (6).

$$p_k(i,j) = q\left\{\frac{[\tau(i,j)]^{\alpha}[\eta(i,i)]^{\beta}}{\sum_{\tau\in J_k(k)}[\tau(i,j)]^{\alpha}[\eta(i,i)]^{\beta}]}; j \in j_k(i)\right. \quad (6)$$

Then, as shown in the above formula, under statistical concentration will, calculate the migration of a city's risk, and then according to the "roulette gambling" method to identify the migration of the next city, in this process, the number of forbidden areas will be more, until every city covered by the taboo table, will stop for detection of the region.

Among these factors, the pheromone has the most significant coefficient of importance, and the route taken by the ants is randomized as the data changes. The lower the selected value is, the randomness of the stochastic optimization can be improved, but at the same time, the algorithm's convergence rate is reduced. As a key factor of the heuristic function, if its value is too large, it will be randomized so that it is easy to enter the optimal. In the case of abbreviated time, the randomization performance of the algorithm is improved, but the difficulty of the solution is also increased.

Then, the pheromone changes. After getting all the practical solutions, the global pheromone update is conducted according to the scheme. However, when searching, we adjust the signal to the pheromone change to obtain a new expression for the time series pheromone. Let us say that we want to minimize the error of the objective function, expressed as Eq. (7).

$$K(C1, C2, ..Ck)ZZ$$
$$Z = \sum_{i=1}^{k}\sum_{x\in c_i} \| x - u_i \|^2 \quad (7)$$

First, given the initial statistical density center, the algorithm will be iterated in the following two steps: a) Partition: The formula indicates that each node is divided into clusters such that the sum of squares within the cluster is minimized, Eq. (8)

$$m_1^{(0)}, m_2^{(0)}, ...m_k^{(0)} \quad (8)$$

The statistical density method is known as the average method. Its idea is quite simple. A sample set is divided into several small groups according to the distance of the sampling interval to ensure that the point groups are concentrated as far as possible and that the group is divided as clearly as possible. This paper proposes a method based on time series to solve the large-scale probabilistic simulation problem. It can divide an uneven urban area into several groups. The calculation accuracy is low, and the result is unstable when calculating the simulation area with less statistical density.

III. Analysis of Results

A. Large-scale Statistical Density Analysis of Logistics Routing Data

After clustering in a large-scale statistical density simulation interval, the best route is determined based on the probability distribution of each simulated area. The pheromone is the key factor of the research object, and its change is related to the effect of the whole method. Under the traditional computational density, the starting concentration of each channel is the same and constant because the length of the channel will increase, so the subject will be attracted to the higher concentration so that the least one can be found in the screening repeatedly. However, this method has a drawback: it can lead to finding an incorrect route under uncertainty, resulting in a longer, time-consuming search.

Therefore, this paper presents a correction method based on point distance to correct the pheromone density of adjacent sides. We select a minimal path from a line with a start and end and multiple paths. In this line, the distance of the path is represented by the straight-line distance from the start point to the endpoint. The initial pheromone concentration between nodes can be defined as follows: Eq. (9) and Eq. (10)

$$d_{AC}D_{Aj} + d_{jc}A - j - C_{ij} \quad (9)$$

$$\tau_{ij}(0) = \frac{d_{AC}}{d_{Aj} + d_{jc}} \quad (10)$$

According to the property that pheromone is inversely proportional to path length, the above equations indicate that when the node is closer to the edge, that is, the closer it is to the shortest straight line distance, the value of is closer to 1, and the statistical density will be biased towards such node movement.

This method guides the primary research target to avoid blind searches to improve the solution rate. After determining the new, improved method, the operation flow parameters of the algorithm are shown in Table I.

TABLE I. SIMULATION PARAMETERS OF LOGISTICS DATA WITH STATISTICAL DENSITY

Name	Value
P	50
Nc max	500
α	1
β	5
ρ	0.5

1) In terms of statistical concentration, other methods are used to generate a certain amount of pheromone at the starting point, and then the statistical density method is used to calculate, which improves the computational efficiency and accuracy.

2) The particle swarm optimization (PSO) algorithm is introduced to give each iteration a statistical concentration. Then, the solution generated by the statistical density system of each generation is used as the initial solution, and then the solution of other algorithms is repeated to speed up the calculation speed and achieve better results.

3) In the aspect of statistical density, the traditional selection method is based on experience, but if the selection is improper, it will reduce the operation speed, so other algorithms can be used to conduct.

4) The algorithm has the shortcomings of a premature algorithm, cannot adapt to the global optimal, is time-consuming, and so on. The statistical density algorithm is used to optimize the initial value and combined with other algorithms, such as crossover and mutation operation, to generate the optimal population.

B. Influence of Statistical Density on Data Mining of Logistics Frequent Paths

In the fundamental statistical density analysis, a pheromone is a positive feedback, and its change degree directly affects the degree of optimization of the research object. Therefore, this paper proposes a correction method based on positive and negative feedback to avoid excessive accumulation of pheromones.

In the research object's second iteration, its average path is calculated, and the updated pheromone standard formula is modified to make it move in a better direction, kf_{arg}. The statistical density is expressed by the average of the paths, as shown in the following Eq (11).

$$f_{avg} = \frac{1}{m} \bullet \sum_{k=1}^{m} L_i^k \tag{11}$$

The calculation formula in the following formula expresses the pheromone dynamic adjustment operator. Its function is to adjust the pheromone concentration on the path in each iteration, Eq. (12).

$$\Delta_v = \sum_{i=1}^{m} \frac{L_i^k - f_{wrg}}{L_{best}^k} \bullet L_i^k \tag{12}$$

L_{best}^k It represents the shortest path and the path length passed by the current particle. The pheromones update Eq. (13) after adding the regulation operator is as follows: L_i^k

$$\tau_v(t+1) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij} + \Delta_v \tag{13}$$

Optimising the statistical concentration in the iteration is easy because its calculation method is highly dependent on the amount of information. The particle swarm optimization algorithm calculation method uses the maximum value of Best and Pest to realize the particle location. The concept of Best and Pest is applied to the statistical concentration so that the research object has the characteristics of particles in the iteration, not only through the pheromone transmission but also through the optimization of the individual and the overall optimization. In the statistical density, the particles' adaptive crossover and mutation strategy prevents the method from falling into the maximum value. Firstly, the primary statistical concentration was used in the initial optimization. After passing through each channel, the FAG of the average road surface was calculated by using the statistical concentration, and the particles with fitness values below F were selected for cross-operation. When the fitness of path selection is not high, the position of the initial particle is replaced by the intersection point. Otherwise, it cannot be replaced. The crossover method randomly selects two locations, R and R, in a city and then makes a second crossover. Using local map technology, the contradiction between numbers can be effectively handled. Based on this, two arbitrary particles are exchanged to evaluate their adaptability, and when the particles' fit reaches an appropriate level, they will be replaced.

C. Effect of Statistical Density Algorithm on Data Simulation of Logistics Frequent Paths

Because the statistical density in this chapter is based on the minimization of the frequent person problem for logistics routing, it contradicts the up-and-down stochastic optimization principle of the benchmark function. The purpose of this paper is to evaluate the performance of the fusion algorithm in a specific city. The comparison method uses the last improved ACO and PSO and conducts ten separate experiments on different data to reduce the interference of random factors due to instability and record the best and average results. Table II illustrates the three methods for calculating the mean and the accuracy of the calculation, which are an example of computational density simulation.

For the "one-Kilometer" logistics problem, describe it as a multiple transportation point, that is, M mailers, N delivery points; this delivery point can be a user's address, can also be a smart ATM or a Courier, in this mode, all the information is considered as a job.

TABLE II. OPERATION TIME OF LOGISTICS DATA ALGORITHM BASED ON STATISTICAL DENSITY

Dataset	ACO	PSO	PSO-IACO
chn 31	22	17	10
ei1101	113	79	43
ch 130	253	189	78
rand 200	396	248	111

In the real situation, the "one-meter" distribution mode, under the premise of ensuring high quality and ensuring the delivery needs of customers, as far as possible to reduce the company's operating expenses. Therefore, applying the composite method to solve the terminal logistics problem still needs to conduct the division of the target area and route planning. Due to the current logistics terminal logistics environment in China, there are many kinds of couriers, and the target geographical distribution is wide; therefore, in the actual logistics process, it is inevitable to produce some problems such as going the wrong way, taking a long way, and repeating the road. If we cannot give an effective solution to this problem, there will be a lot of problems, such as unbalanced work, low efficiency, and repeated delivery in the end distribution process. Therefore, in the transportation of the "final kilometer", the "minimum route" is the primary optimization purpose, and then other restrictions are changed or added according to the actual situation to maximize the interests of customers and the company.

D. Optimizing Frequent Logistics Routes based on Statistical Density

Due to the use of multi-traveller mode, there will be energy limitations, capacity limitations, time limitations, and other issues. Energy limitation refers to the consumption of logistics and the continuous life of other intelligent vehicles. In the logistics process, the capacity limitation is the core of the vehicle loading capacity. In the case of delayed or early check-in by customers, there are also certain requirements on the delivery time of goods, namely the arrival sequence of the target. The "one kilometer of the city" problem to be dealt with in this paper should be modified or explained according to the actual situation to make the following interpretation:

1) Description of the distance on the actual case: The point of specific distribution and transport path between the plan because the curve of the road makes the actual transport distance cannot simply use coordinates to determine, in order to ensure the accuracy of the test and practical value, the point - point line spacing in the coordinate system to replace the actual distance of the target.

2) Description of delivery personnel: The delivery company is equipped with M delivery personnel, which can be assigned according to the task's scale.

3) Description of time constraints: Time constraints are added to each route. When the average speed of each train is the same, there is a positive relationship between the time and the total transportation distance. Therefore, in the future modelling process, in order to ensure real-time and balance, the maximum length of each route is limited.

IV. CONCLUSION

To sum up, this paper carries out three approaches for a single logistics problem: One is for a single logistics problem, using an improved statistical density and density method is optimized to it, and then USES the method of target classification, and then adopted based on the improved pheromone correction rule of statistical strength on the cluster planning, and the connection between these clusters established a whole calculation concentration simulation spacer ring. The transportation mode of the terminal is optimized. On this basis, the interval mode based on probability simulation cannot meet the transportation demand of the final kilometer, and the interval mode based on data density simulation can better meet the transportation demand of the terminal. Then, according to the "end" transportation situation in the logistics system, a M data distribution simulation region is established. In addition, the optimal combination of two different methods for distribution, on the premise of meeting the actual needs of customers, increasing the time limit, and making sure that the number of Courier service equalization effectively overcomes the traditional way of distribution, is not reasonable, the workload imbalance problem put forward an effective method of path selection, It has a practical guiding role in improving the customer's intimacy and improving the utilization rate of logistics.

REFERENCES

[1] R. Wang, X. Li, Z. Zhang and M. Hongguang, "Modeling and simulation methods of sea clutter based on measured data", *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 12, No. 1, 2050068 (2021).

[2] J. Kearns, A. M. Ross, D. R. Walsh, *et al.*, "A blood biomarker and clinical correlation cohort study protocol to diagnose sports-related concussion and monitor recovery in elite rugby [Study Protocol]", *BMJ Open Sport & Exercise Medicine*, vol. 6, no. 1, 2020.

[3] X. Yang, R. Zhang, F. Pan, *et al.*, "Stochastic user equilibrium path planning for crowd evacuation at subway station based on social force model," Physical A: Statistical Mechanics and its Applications, vol. 594, no. 127033, 2022.

[4] B. Zhu, L. Zhang and Y. Zhang, "The Early Warning Method of Drug Adverse Reaction Monitoring Based on Data Mining Algorithm Was Studied," *Journal of Physics: Conference Series*, vol. 1852, no. 3, 2021.

[5] F. Wu, M. Liu, G. Huang, *et al.*, "Simulation of stationary non-Gaussian multivariate wind pressures based on moment-based piecewise Johnson transformation model", *Probabilistic Engineering Mechanics*, vol. 68, no. 103225, 2022.

[6] M. J. Marijnissen, C. Graczykowski and J. Rojek, "Simulation of the comminution process in a high-speed rotor mill based on the feed's macroscopic material data," *Minerals Engineering*, vol. 163, no. 106746, 2021.

[7] T. Wisniewski and R. Szymański, "Simulation-based optimization of replenishment policy in supply chains," *International Journal of Logistics Systems and Management*, vol. 38, 2021.

[8] Y. Zhao, H. Gao, S. Zhang, *et al.*, "Simulation and selection of fin stabilizers for polar cruise ships based on Computational Fluid Dynamics", *Journal of Physics: Conference Series* vol. 1, pp. 9, 2021.

[9] C. Jin, J. Xiao and C. Cao, "Concept of biomimetic mechanical foot based on muscle simulation", *Journal of Physics Conference Series*, 1885(4):042017, 2021.

[10] K. Ni and Y. Ma, "Simulation of LPG tank truck leakage and explosion accident based on CFD", *Journal of Physics: Conference Series*, vol. 2003, no. 1, 2021.

[11] X. Wang and Y. Zhang, "A Data Simulation Method of Bank Fraud Transaction Based on Flow-Based Generative Model", *Journal of Physics: Conference Series*, vol. 1631, no. 1, pp. 1-9, 2020.

[12] Q. Cheng, H. Shen, H. Chu, *et al*. "Research on Logistics Simulation and Optimization of Die Forging Production Line Based on Flexsim", *Journal of Physics Conference Series*, vol. 1624, no. 022063, 2020.

[13] Z. H. Wang and T. Y. Li, "Computer optimal path simulation planning of oil and gas storage and transportation based on similarity measurement of decision space", *Journal of Physics: Conference Series*, vol. 1533, no. 3, pp. 5, 2020.

[14] Y. Dong, L. Shao and J. Shi, "Study on Numerical Simulation of Roadway Backfill under Freeway Based on FLAC3D", *Journal of Physics: Conference Series*, vol. 1549, no. 4, pp. 6, 2020.

[15] C. Yu, S. Cen and S. Luo, 'Simulation design and optimization of production line of a cross-axis machining based on Plant Simulation", *Journal of Physics: Conference Series*, vol. 1654, no. 1, pp. 6, 2020.