

## Research Article

# Evaluation and Analysis of Intelligent Logistics Distribution Using the Expectation-Maximization Algorithm Calculation Model

Zhifeng Wang 

*College of Business Administration, Xuzhou College of Industrial Technology, Xuzhou 221000, China*

Correspondence should be addressed to Zhifeng Wang; wangzf@mail.xzcit.cn

Received 25 March 2022; Accepted 25 May 2022; Published 28 June 2022

Academic Editor: Xiantao Jiang

Copyright © 2022 Zhifeng Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of this article is to solve the problem that the accuracy of logistics distribution path planning is affected by the lack of data in the process of traditional logistics distribution planning and management. This exploration innovatively applies an effective data addition algorithm expectation-maximization (EM) algorithm to the intelligent logistics distribution system to improve logistics distribution's overall efficiency and management quality. First, the concept of intelligent logistics and the composition and main functions of the intelligent logistics system are introduced. Then, the core idea of the EM algorithm and its applications in intelligent logistics are described. The logistics distribution of a chain company is taken as an example. Finally, the advantages and disadvantages of the intelligent logistics system based on the EM algorithm are compared with those of the traditional intelligent logistics systems based on variable neighborhood search (VNS), Tabu search (TS), and ant colony optimization (ACO). The performance test results show that the EM algorithm's optimal solution times are 7 times. Its convergence speed is slightly lower than that of the ACO, but there is no obvious difference. The intelligent logistics distribution system based on the EM algorithm has faster order processing speed and higher efficiency in the actual case application. The average processing time of each order is 1.78 min, which is 0.237 min less than that of VNS and only 0.022 min more than that of ACO. It reveals that the intelligent logistics distribution system based on the EM algorithm is more efficient. The study provides a new idea for the efficient distribution of enterprise logistics.

## 1. Introduction

With the upgrading of Internet technology and the vigorous development of the logistics industry, a new logistics management mode is favored by more and more logistics enterprises, that is, intelligent logistics. It is a new logistics operation mode promoted by the development of science and technology and e-commerce. It adds an "intelligent system" based on traditional logistics, which can ensure the efficient operation of logistics and reduce the operation cost [1].

Ding et al. pointed out that intelligent logistics was an effective way to deal with the challenges of rapidly changing customer expectations, seize the opportunities brought by new technologies, and promote the development of new business models [2]. With the continuous innovation of sensing technology, communication technology, and computer technology, the Internet of things technology has also been applied to more infrastructure such as environmental

monitoring, biomedicine, and intelligent wear. Song et al. emphasized that the Internet of things could create a data ocean with the assistance of various mathematical analysis technologies and explore the complex relationship among transactions represented by these data. These characteristics help to promote the development of intelligent logistics [3]. Humayun et al. proposed a layered framework based on Internet of things and blockchain for intelligent logistics, providing intelligent logistics and transportation systems. The advantages of the Internet of things and blockchain in logistics and transportation were highlighted through real case studies [4]. The most important thing in the intelligent logistics system is to improve distribution efficiency, so distribution path planning is quite crucial. The main algorithms used in logistics distribution planning are variable neighborhood search (VNS), Tabu search (TS), and ant colony optimization (ACO). Li et al. transformed the logistics scheduling problem into a mixed-integer linear programming problem, proposed a special coding method

suitable for small-scale problems, and used the variable neighborhood search algorithm framework to generate the approximate optimal solution of the problem. The important parameters were calibrated through experiments and the algorithm's robustness was analyzed. Experimental results show that the algorithm is effective [5]. Temucin and Tuzkaya minimized the total logistics cost and total delay and maximized the total average capacity utilization through the metaheuristic method based on TS. Numerical analysis shows the effectiveness of the proposed method [6]. Calabrò et al. adopted an ACO algorithm to solve the vehicle routing problem of inbound logistics. The effectiveness of this method in cost reduction and scheduling was verified by actual data, which provided useful suggestions for the large-scale operation of freight services [7].

In conclusion, it reveals that worldwide research on intelligent logistics and distribution path has shown significant breakthroughs and achievements, but most of these problems and research results are confined to the form of traditional logistics. In the traditional logistics system, there may be missing data in the observation data due to observation conditions, instrument faults, human errors, accidents, improper downloading and uploading, improper storage process, and other factors. Various logistics distribution planning algorithms are highly dependent on data, and the amount and quality of data directly determine the algorithm's accuracy. Expectation-maximization (EM) is an effective data addition algorithm, which can provide important support for solving the problem of data missing in logistics management and improving the quality of logistics distribution management. Based on this, first, the relevant theories of intelligent logistics distribution and EM algorithm are summarized. Next, the innovative logistics algorithm based on EM is put forward. Finally, the effectiveness of this method is verified by an actual case. This exploration provides important support for consolidating the market position of logistics enterprises, enhancing their core competitiveness, and enhancing the brand influence.

## 2. Intelligent Logistics Distribution and Application of the EM Algorithm

*2.1. Intelligent Logistics Distribution.* Intelligent logistics distribution is to use integrated intelligent technologies such as RFID and sensors to make the logistics distribution system have the ability to think like people, solve some basic problems in the process of distribution, and ensure the normal progress of distribution. It means the ability to analyze and make decisions according to the relevant information provided by the logistics information platform. Intelligent logistics emphasizes the management of logistics activities with the help of dynamic information of machinery and equipment and the Internet. Intelligent logistics matches informatization and thing-thing interconnection. Logistics distribution is more automatic, advanced, and intelligent through more refined management. Its purpose is to make the operation activities of logistics more efficient,

use resources more fully, and enrich the forms of value expression to improve the overall level of the logistics industry [8–10]. Figure 1 shows its main distribution functions.

In Figure 1, the first is the stocking and out-of-warehouse function. The intelligent logistics uses an intelligent storage system to realize the automatic picking and stock out of goods, which avoids the error of manual operation and improves efficiency. The second is the delivery function. The global positioning system is used to realize the real-time monitoring of vehicles and goods. The third is the delivery terminal service function. On the basis of fulfilling their basic responsibilities, enterprises also need to provide convenient value-added services for customers to enhance their goodwill toward enterprises. The last is the information processing function. The information is collected and processed through the computer terminal, and big data are adopted to analyze customer preferences and push relevant demand information.

Figure 2 shows the flow of the intelligent logistics distribution management system.

Figure 2 shows that the intelligent logistics distribution management system is a management platform serving logistics distribution enterprises based on the geographic information system (GIS), global positioning system (GPS), and Internet of things. Its functions include real-time monitoring, two-way communication, vehicle scheduling, real-time information query of goods, and planning distribution route. First, mobile phones and fixed-line broadband receive real-time information through the Internet, optical fiber network, and wireless modem. Then, the command and control center obtains the position, speed, and cargo information of the vehicle through the GPS satellite; the internal distribution system replans the route, and then, the global system for mobile communications' (GSM) wireless communication network transmits the scheme to the on-board GPS terminal. Finally, the GPS terminal will intelligently prompt the next driving path of the vehicle [11–13].

Figure 3 shows the main functions of the intelligent logistics scheduling platform.

Figure 3 suggests that the intelligent logistics scheduling platform has the following functions, such as the graphical display of goods information; order information; road information and network nodes; the distribution of customer points and the attributes; number and order number of goods required; the online query of vehicle speed and location, environment, batch, and quantity of goods; and the start and end time and sequence number of services received by each customer. Moreover, the platform can also calculate and display the optimization path according to the goods information clicked by a given customer, feedback real-time road conditions, collect real-time traffic information and capture orders, and collect other information related to travel vehicles.

*2.2. EM Algorithm.* The EM algorithm is a data addition algorithm. It mainly adds some potential data based on observed data sequence and makes up the lost data through

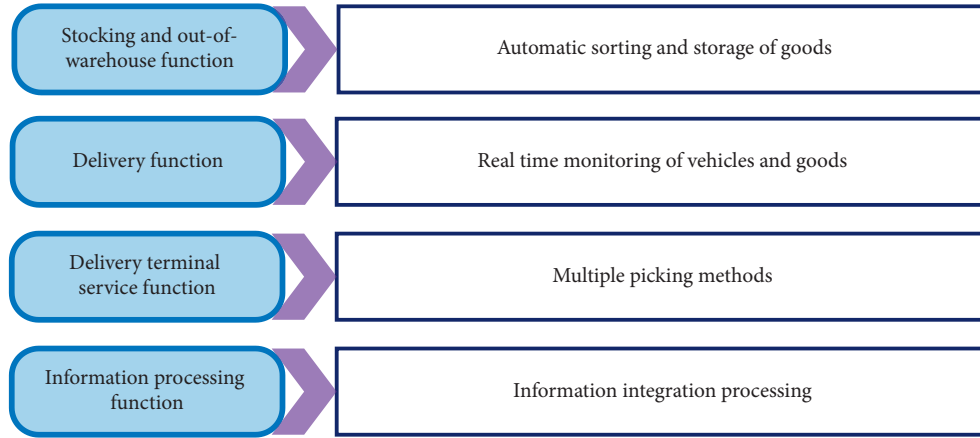


FIGURE 1: Main functions of intelligent logistics distribution.

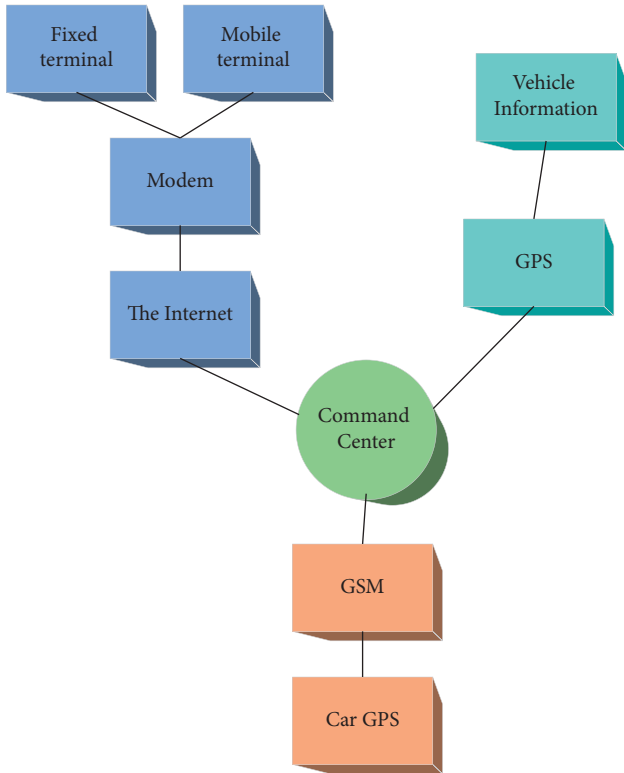


FIGURE 2: Flow of intelligent logistics distribution management system.

mathematical methods to change incomplete data into complete data. It means that realizing data conversion processing is the biggest feature of the EM algorithm [14].

2.2.1. *Common EM Acceleration Algorithms.* Parameter-expanded expectation-maximization (PX-EM) algorithm: PX-EM accelerates the convergence speed of the algorithm. The original model can be obtained again by integrating the added parameter  $\alpha$  and appropriate functions into a large-scale model. This process should be realized by selecting a special value  $\alpha_0$  of  $\alpha$  [15]. In the original function sequence, if the parameter  $\theta$  needs to be estimated, the

matching model parameter is  $\varphi = (\theta^*, \alpha)$ .  $\theta^*$  and  $\theta$  have the same dimensions. For a known transformation  $R$ ,  $\theta = (\theta^*, \alpha)$ . When  $\alpha = \alpha_0$ ,  $\theta^* = \theta$ . The parameters of the model are adjusted so that the information about  $\alpha$  cannot exist in the observed data  $Y$ , that is,

$$f_x(Y_0|\theta^*, \alpha) = f_x(Y_0|\theta^*, \alpha'). \quad (1)$$

where  $f_x$  is the density function of the corresponding extended model under any  $\alpha'$ , and the complete data  $Y = Y_0$  can be screened out. The PX-EM algorithm is a simple improvement of the EM algorithm, which is realized by improving the  $t$ -th iteration:

The PX-E step is calculated as follows:

$$Q(\phi|\phi^t) = E(\log_{f_x}(Y|\phi)|Y_0, \phi^t). \quad (2)$$

The PX-M step is calculated as follows:

$$\phi^{t+1} = \max Q(\phi|\phi^t). \quad (3)$$

Each iteration of PX-EM will increase the value of  $f_x = (Y_0|\theta, \alpha)$ , and its convergence property is consistent with the standard.

2.2.2. *Data Loss Mode and Mechanism.* At present, some methods to deal with missing data are limited to some specific patterns, which have some limitations. Therefore, it is necessary to understand the missing mode of data, which is generally divided into a single value missing mode and arbitrary missing mode. If all missing values are the same attribute, this mode is a single value missing mode. This situation is relatively simple, but it is rare in most complex data. If the missing values belong to different attributes, the data filling method of this missing type is complex, and specific problems need to be analyzed, which is considered as the arbitrary missing mode. Generally, missing data are divided into three categories: missing at random, missing completely at random, and missing not at random. Figure 4 shows a specific description.

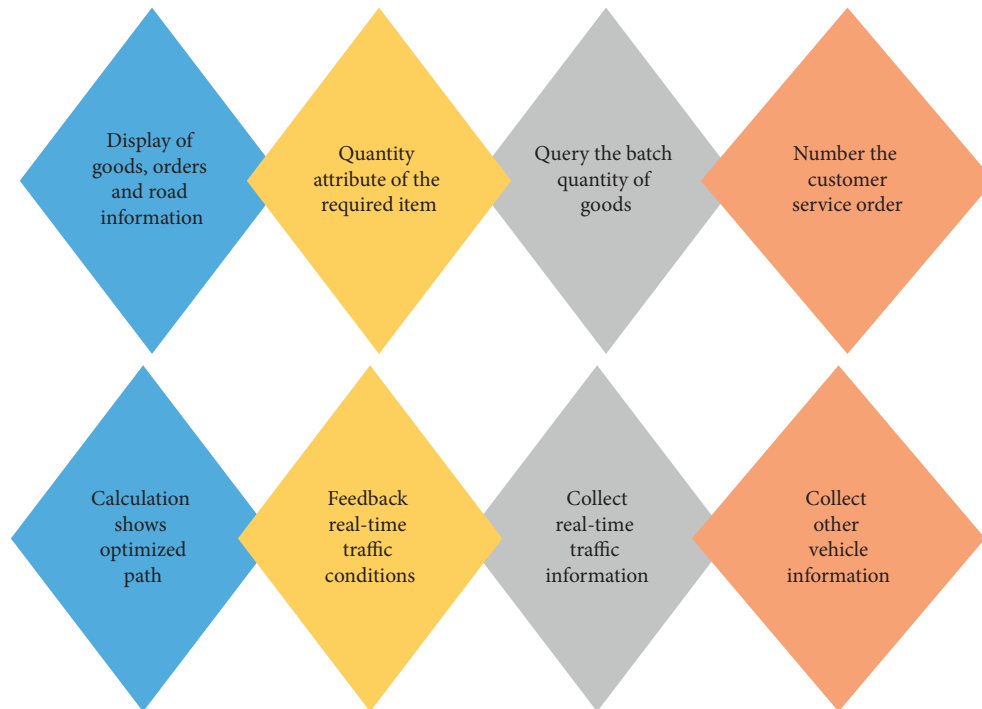


FIGURE 3: Functions of the intelligent logistics scheduling platform.

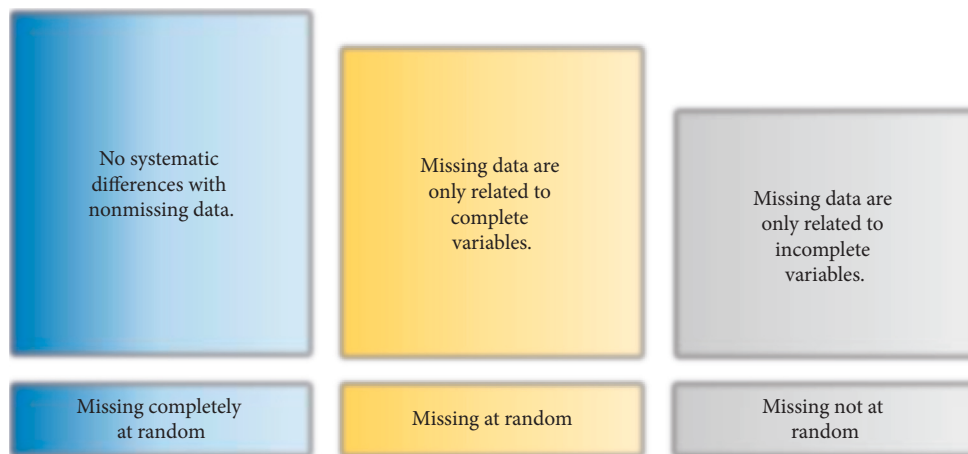


FIGURE 4: Classification of missing data.

2.2.3. Missing Data Processing Methods Commonly Used in the Measurement

(1) *Filling Method:* According to the auxiliary information or potential information, some mathematical method is adopted to determine the reasonable estimated value to replace the missing value of the data to make the data more complete. Then, the whole dataset is processed by conventional methods. The filling method should be different according to different data information. The filling method can generally be divided into two categories according to the different number of missing values to construct estimates. One is the single imputation and the other is the multiple imputation [16]. Figure 5 shows the common types of single imputation.

Multiple imputation is developed based on single imputation, using a series of possible values to replace each missing value. It mainly finds the data law and randomly generates a value that can replace the missing data, rather than replacing the missing value with a single value. The estimation of uncertain information is more accurate than single imputation through the parameter distribution and connection variables between missing data. The multiple imputation method has high reliability, fully considers the uncertain information contained in the missing data, and greatly reduces the amount of data calculation. Therefore, it has become the most widely used data filling method at present. According to the mode and variable type, the multiple imputation method can use trend scoring, random

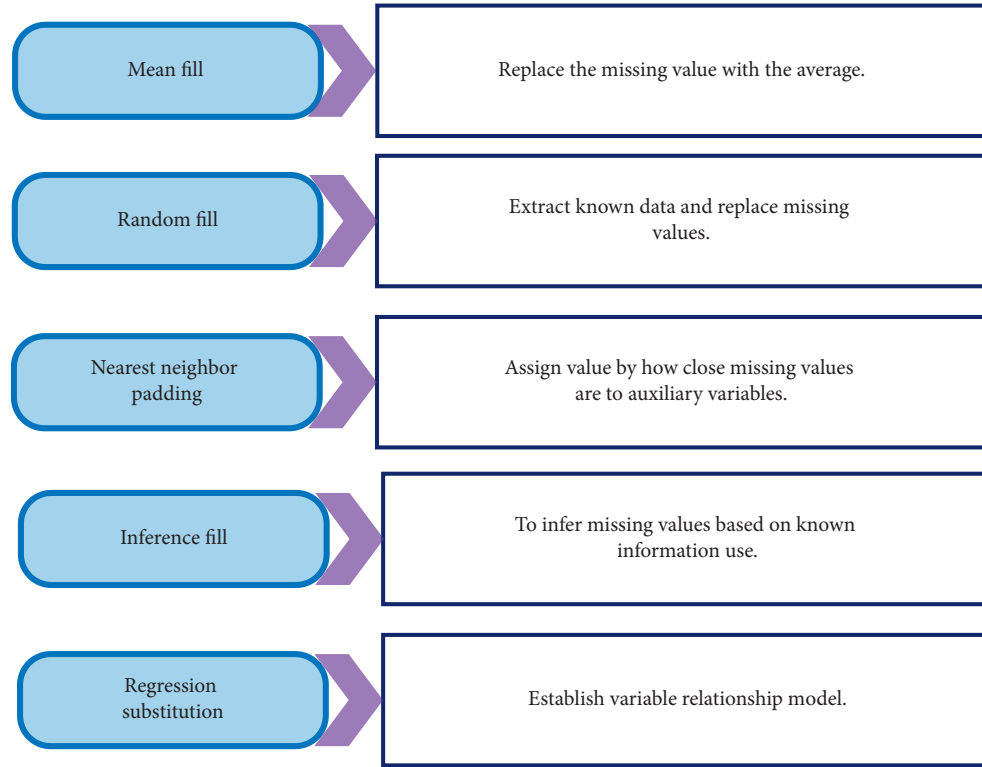


FIGURE 5: Five common filling methods of the single imputation method.

regression filling, and Markov chain Monte Carlo (MCMC) models to fill the missing data [17–19].

The trend scoring method mainly uses the self-service method to fill the missing value of each group of data and then divides the observation data into several subsequences. The processing of incomplete data is carried out according to the following steps:

The first is to fit the logistic regression model equation:

$$\text{logit}(p_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{i-1} X_{i-1}. \quad (4)$$

$p_i = \text{Pr}(R_i = 0 | X_1, X_2, \dots, X_{i-1})$ .  $(\beta_0, \beta_1, \beta_2, \dots, \beta_{i-1})$  is the model's parameter.

The trend score for each missing data on variable  $X_i$  is calculated. According to the trend score, the observation data are grouped according to the fixed number of groups and the number of groups is determined according to the number of observation measurements. Finally, the missing data in each group are estimated and filled by the approximate Bayesian method. The above steps are repeated until  $X_i$  is filled.

MCMC is a method based on the Bayesian inference. This method has two-step cycles, namely, filling and a posteriori. The data are corrected in real time and updated to fill the missing data [20]. For missing information, the posterior probability density of the parameter is as follows:

$$p(\theta|y) = \frac{p(\theta|y)p(\theta)}{\int p(\theta|y)p(\theta)d\theta}. \quad (5)$$

Equation (5) is the posterior probability density under complete observation data. The posterior probability  $y_{obs}$  of the missing data can be obtained only when the observed data  $y_{obs}$  are completely filled. Similarly, the posterior probability  $p(\theta|y_{obs})$  of observation data cannot be obtained directly. Only after inference and simulation can the incomplete observation data be supplemented and then be estimated.

According to the stationary distribution  $\theta$  and  $p(Y_{mis}|Y_{obs}, \theta^{(t)})$ , the missing data filling value is as follows:

$$Y_{mis}^{(t+1)}: p(Y_{mis}|Y_{obs}, \theta^{(t)}). \quad (6)$$

$Y$  and  $\theta$  are irrelevant. Based on each missing data, the filling value is calculated as follows:

$$Y_{mis,i}^{(t+1)}: p(Y_{mis,i}|Y_{obs,i}, \theta^{(t)}). \quad (7)$$

Random regression filling method: there is a linear regression relationship among observation data  $(Y_1, Y_2, \dots, Y_3)$ . If observation data are missing, the fitting model is as follows:

$$EY_i|\beta = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \cdots + \beta_{i-1} Y_{i-1}. \quad (8)$$

$(\beta_0, \beta_1, \beta_2, \dots, \beta_{i-1})$  is the coefficient of the regression model. The final filling value is calculated as follows:

$$Y_i^* = \beta_0^* + \beta_1^* Y_1 + \beta_2^* Y_2 + \cdots + \beta_{i-1}^* Y_{i-1} + \sigma^* \varepsilon, \quad (9)$$

where  $\sigma^*$  represents the variance of the regression model,  $\varepsilon$  is the error vector in the normal random state, and  $\beta^*$  represents the filled regression coefficient after  $n$  times of

replacement. The random regression filling method reflects the uncertainty of missing data and filling value by adding an additional residual term subject to normal distribution or other distribution.

(2) *Gray Model Method*: It is to research and analyze with time series and establishes the equation with the sequence of number, that is, the model composed of a single variable first-order differential equation. It is a prediction method for modeling after transforming the original irregular sequence into a more regular generated sequence [21]. The data processing steps are as follows:

First, the order ratio of the original sequence  $x^0 = \{x^0(t) | t = 1, 2, \dots, n\}$  is calculated as follows:

$$\sigma(k) = \frac{x(k-1)}{x(k)}, k = 2, 3, \dots, n \quad (10)$$

whether  $\sigma(k)$  falls into the tolerance interval is judged as the basis for modeling;

Next, the level ratio test data are transformed. For the sequence that cannot pass the order ratio test, the data need to be processed by translation transformation, logarithmic transformation, square root transformation, and other related transformations. Then, a cumulative transformation is conducted on the qualified data to generate a new sequence as follows:

$$x^1 = \{x^1(t) | t = 1, 2, \dots, n\}, x^1(t) = \sum_{k=1}^t x^0(k), \quad (11)$$

where  $x^1$  refers to the calculation of the one-time accumulation sequence. Then, the mean sequence  $z^1(k) = 0.5(x^1(k) + x^1(k-1))$  of  $x^1(t)$  is calculated. Intermediate parameters are calculated as follows:

$$\begin{aligned} C &= \sum_{k=2}^n z^1(k), D = \sum_{k=2}^n x^0(k), E = \sum_{k=2}^n z^1(k)x^0(k), F \\ &= \sum_{k=2}^n z^1(k)^2 \end{aligned} \quad (12)$$

The obtained intermediate parameters are used to calculate the model parameters:

$$a = \frac{C D - (n-1)E}{(n-1)F - C^2}, b = \frac{DF - CE}{(n-1)F - C^2}. \quad (13)$$

The final model is as follows:

$$\hat{x}^1(k+1) = (x^0(1))e^{-ak} + \frac{b}{a}, \quad (14)$$

where  $x^0(1)$  is the calculated value of the original sequence model.

The summation principle is as follows. The first data of the original sequence are the first data of the generated column. The sum of the first and second row data is the second data of the generated column. The sum of the third data of the original sequence and the second data of the generated column are the third data of the generated column. According to this rule, the new generated column can be obtained. The accumulated restored sequence data are as follows:

$$\hat{x}^0(k+1) = \hat{x}^1(k+1) - \hat{x}^1(k). \quad (15)$$

### 2.3. Application of the EM Algorithm in Intelligent Logistics

**2.3.1. Logistics Distribution Steps.** Figure 6 shows a complete logistics distribution process.

The first step is to divide the basic delivery area. First, the customer location is systematically analyzed and divided into regions, and then each customer is assigned to the delivery area to make basic preparations for the later distribution decision-making. The second step is vehicle stowage. Due to the different attributes of distribution goods, in order to ensure the safe distribution of goods and improve the distribution efficiency, it is necessary to distinguish the goods with different attributes before distribution. In this way, the distribution methods and tools can be quickly and accurately determined after receiving the order. The third step is to arrange vehicles. The company needs to determine the type and tonnage of distribution vehicles. The fourth step is to determine the delivery order. The fifth step is to choose the distribution route. The delivery time shall be determined according to the actual factors, such as the geographical location of customers; the traffic conditions during the delivery, the route with the shortest distance and the lowest cost; and the special requirements of some customers or actual environment on the delivery time, model, and order when necessary. The last step is to deliver the goods to the customer.

**2.3.2. Case Analysis.** The distribution system between nine branches and the distribution center of a chain company is selected as an example to discuss the application effect of the EM algorithm. Considering the limitation of the hard time window, if the area division is too large, there will be order points that exceed the specified time and cannot complete the service. It has become the resistance of the team to complete the task on time and increased the company's overall operating cost [22]. Besides, it is stipulated that the fleet dispatched by the company shall complete the distribution of each area. Vehicles cannot be distributed across regions, the time window requirements should be strictly followed, and rejection is not allowed except under special circumstances. In addition, it is assumed that the driving speed of vehicles is not lower than the average driving speed of the road, the status, position, and speed of these vehicles are monitored in real-time by the central distribution system. The real-time information of any vehicle can be mastered at any time and the adjustment planning of the vehicle path can be completed at any time. Figure 7 shows the distribution of specific branches.

In Figure 7, the distribution center is set to 0, and the nine branches are set from 1 to 9.

**2.3.3. Algorithm Comparison.** As mentioned above, at present, the most commonly used algorithms in logistics distribution mainly include VNS, TS, and ACO. In order to

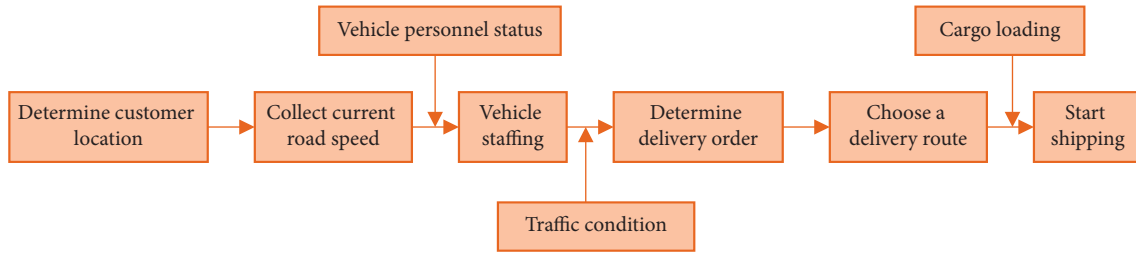


FIGURE 6: Flow chart of logistics distribution.

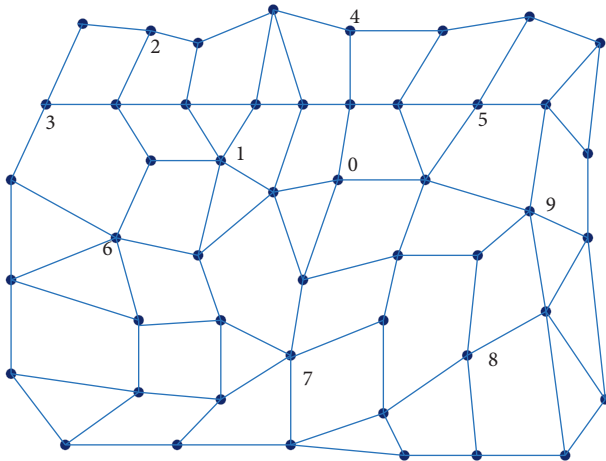


FIGURE 7: Distribution of branches of a chain company.

prove the feasibility of the proposed algorithm, the three algorithms are compared with the EM algorithm introduced. The comparative experiments are carried out without changing the parameters in convergence performance, path length, vehicle time consumption, and cost saving.

For the VNS algorithm, the initial tabu length is set to 1, the maximum interval of average solution repetition is 1, the proportion of Tabu length increase is 1.1, the proportion of Tabu length decrease is 0.9, the maximum number of iterations is 300, the maximum number of solution repetitions is 3, and the maximum number of solutions in tabu table is 6. For the TS algorithm, the initial Tabu table length is set to 4 and the neighborhood size is 10, which is gradually adjusted according to the search process. If the feasible solution is not improved after the number of moves reaches a certain level, it is possible to produce local cycles. At this time, it is necessary to increase the length of the taboo table. If all movements are prohibited, the neighborhood size needs to be increased. For the ACO algorithm, the number of ants is 30, the pheromone volatilization coefficient is 0.3, and the amount of information released by ants after completing one cycle is 50. The condition for stopping the cycle is that the difference between the optimal solutions obtained from two adjacent cycles is less than 0.01, the value of the parameter heuristic factor  $\alpha$  is 1, and the expected heuristic factor  $\beta$  is 3.

**2.3.4. Algorithm Running Environment.** The algorithm is realized by MATLAB simulation software. Table 1 shows the operating environment.

### 3. Experimental Result

**3.1. Basic Statistics of Logistics Distribution.** The average speed and real-time vehicle speed of different road types among all branches are obtained through measurement. Figure 8 presents the specific results:

Figure 8 shows that the average speed of class I road is the highest, which is 60 km/h, and the real-time vehicle speed is 45 km/h. The main sections include 3-5 and 4-0-7. The average speed of the class II road is 40 km/h and the real-time vehicle speed is 30 km/h. The main sections include 0-1, 1-9, and 5-7. All other sections are class III roads.

Figure 9 shows the distance among branch stores (km) and the demand for goods (ton):

Figure 9 displays that the branch with the farthest distribution distance is store 3, followed by store 6. Store 7 needs the most goods, followed by store 1. It suggests that there is no linear relationship between the distance between each store and the distribution center and its demand. When designing the vehicle route, priority should be given to store 3 and store 6, which are farthest away, and store 7 and store 1, which are in greatest demand.

**3.2. Performance Comparison of Four Algorithms.** Each algorithm is tested independently 30 times. Their optimal solution, the worst solution, and the frequency of the optimal solution are solved, respectively, and the convergence results of each algorithm are counted. Figure 10 displays the experimental results:

Figure 10 shows that the results of the optimal solution and the worst solution of the four algorithms are basically the same under the same scale. The index with the large difference is the number of times to obtain the optimal solution. The optimal number obtained by the VNS algorithm is only three times. The TS and ACO get the same number of optimal solutions, 6 times. The EM algorithm is the best and the number of times to get the optimal solution is 7.

12 times of optimal path operation are conducted on these four algorithms. Figure 11 shows the calculation results:

Figure 11 shows that the optimal solution of the obtained path is 67.5, in which the VNS does not obtain the optimal path, and the TS obtains the optimal path at the third time. ACO calculates 5 times to obtain the optimal path. The EM algorithm obtains the optimal path in the second operation, and in the latter operation, the number of times to obtain the optimal path is the most.

TABLE 1: Environment required for the system test.

Project	Configuration
Operating system	Windows 10
Central processing unit (CPU)	Intel core i7-7700 3.60 GHz
Random access memory (RAM)	8 GB
Hard disk	500 GB

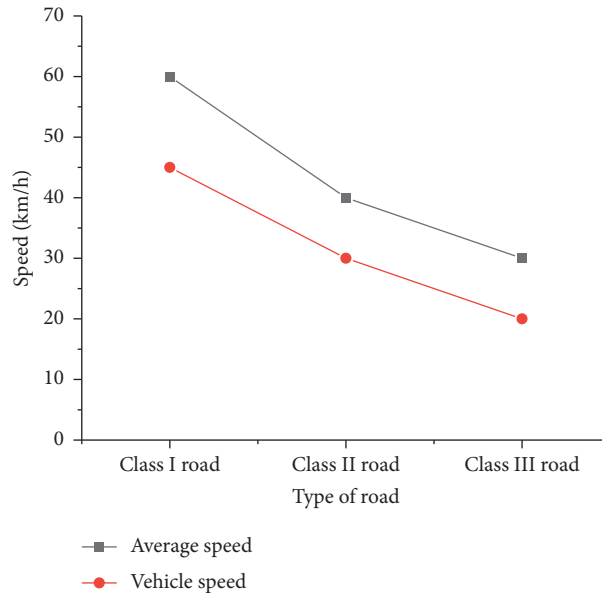


FIGURE 8: Real-time vehicle speeds for different road types.

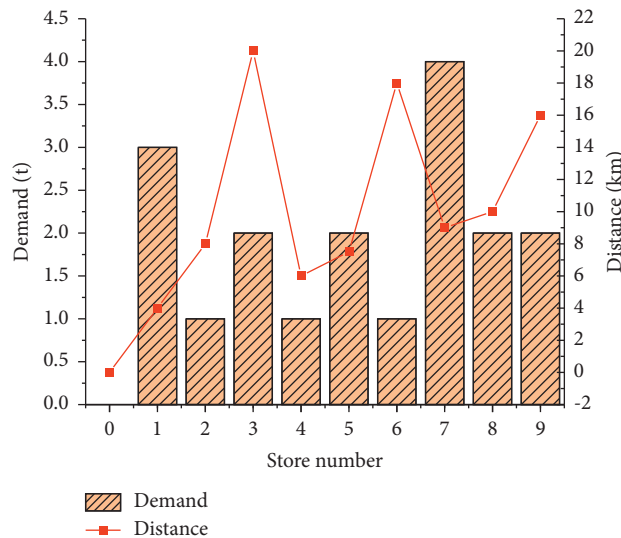


FIGURE 9: Distance between branch stores and goods demand.

3.3. Comparison of Application Examples of Four Algorithms. Without changing the parameter setting, the use time (min) and the use of vehicles of the four algorithms are compared when the number of transportation branch stores is 3, 6, and 12, respectively. Figure 12 displays the specific results.

The average number of orders that can be processed by one vehicle in the four methods is 12.963, 15, 15, and

21.17. Figure 13 displays that the average time-consuming of each order processed by the four methods is 2.017 min, 1.80 min, 1.658 min, and 1.78 min, respectively. The efficiency of the VNS algorithm is the lowest. Although the average time consumption per order of the EM algorithm is not the least, the EM algorithm is the most efficient algorithm.



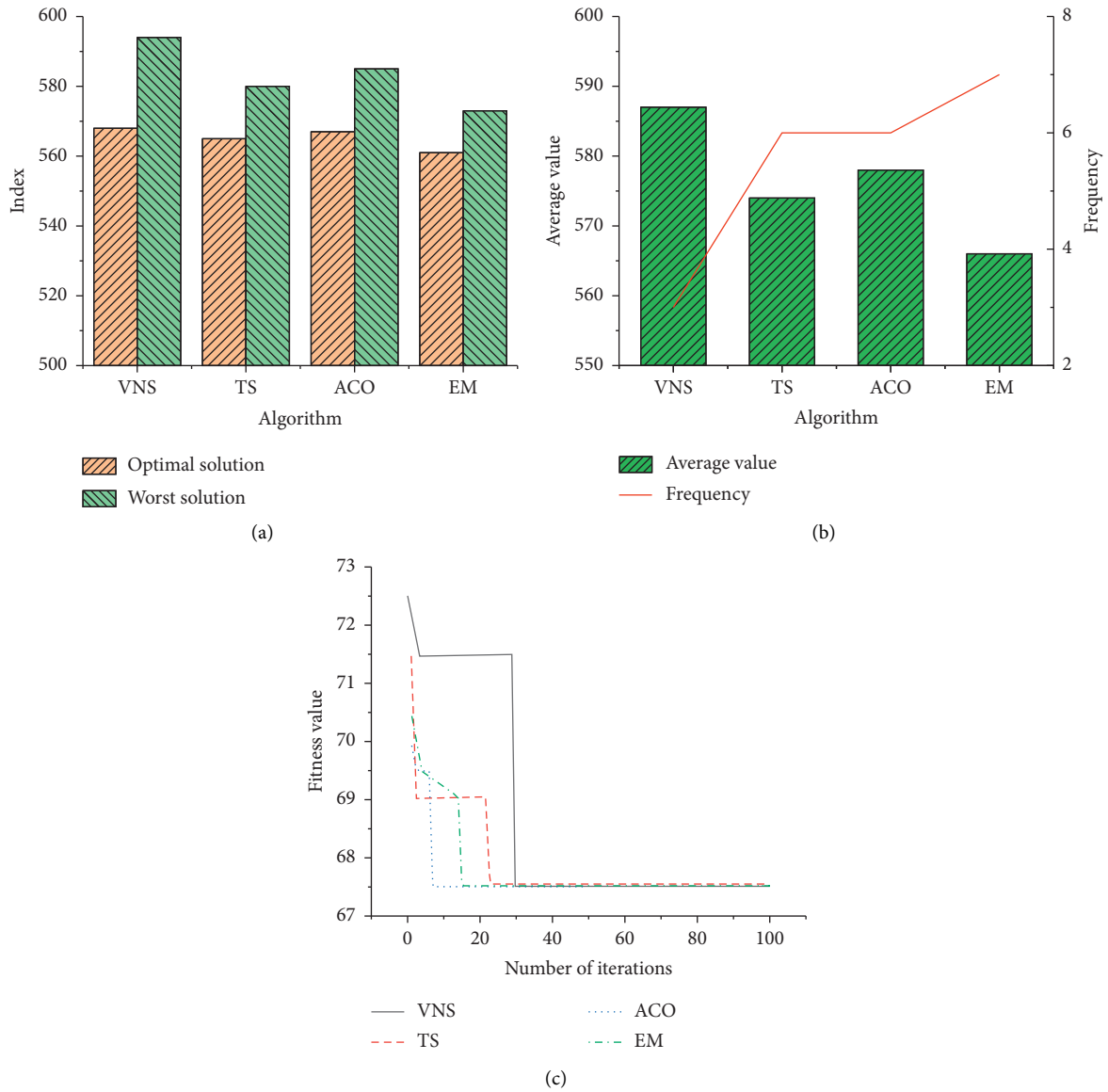


FIGURE 10: Statistics of independent test results. (a) Statistics of the optimal solution and worst solution; (b) statistics of average value and frequency of reaching optimal solution; and (c) convergence of four algorithms.

Finally, the advantages and disadvantages of the four algorithms are considered from the evaluation factors such as the saved distribution distance, the spent fuel cost and the comprehensive cost of distribution. Figure 13 shows the specific results.

As shown in Figure 13, considering the fuel cost, EM algorithm saves the most, which is 3.17 yuan. Considering the comprehensive cost of distribution, the optimized design of the other three algorithms reduces the comprehensive cost of the original scheme to varying degrees. Among them, EM algorithm saves the most cost, which is 7.81 yuan.

3.4. Results and Discussion. The EM algorithm is compared with VNS, TS, and ACO algorithms in terms of convergence

performance, path length, vehicle time consumption, and cost saving. The results are as follows: (1) the optimal solution obtained by the VNS algorithm is only three times, indicating that the algorithm is very easy to converge to the local minimum solution. The TS algorithm and ACO algorithm get the same number of optimal solutions, both of which are 6 times. The EM algorithm is the best, and the number of times to get the optimal solution is 7. The convergence shows that the ACO algorithm has the fastest convergence speed, followed by the EM algorithm, and the VNS algorithm has the slowest convergence speed. (2) The VNS algorithm does not get the optimal path. The TS algorithm gets the optimal path at the third time, but after more operations, the number of times to get the optimal path is less. The ACO algorithm obtains the optimal path after 5 operations, the EM algorithm obtains the optimal path in the second operation, and in the later operation, the

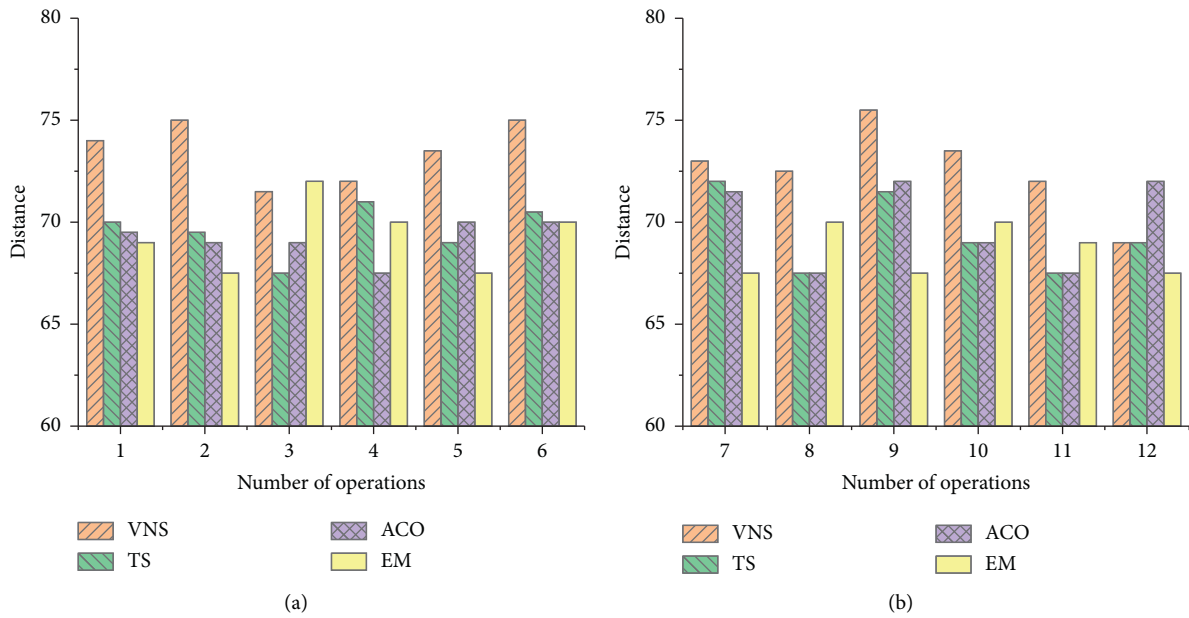


FIGURE 11: Calculation results of four algorithms; (a) statistics of the first six calculation results and (b) statistics of the last six calculation results.

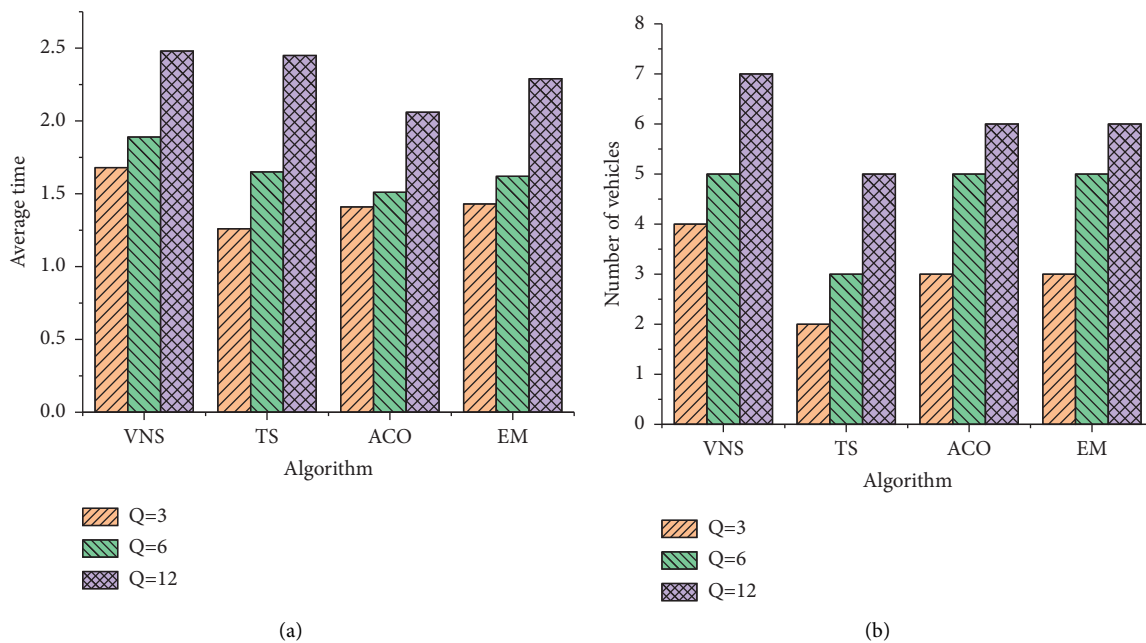


FIGURE 12: Comparison of time spent and vehicles used; (a) average time used and (b) number of vehicles used.

number of times to obtain the optimal path is the most. Therefore, the effect of EM algorithm is better. (3) The average time consumption of each order processed by the four methods are 2.017 min, 1.80 min, 1.658 min, and 1.78 min, respectively. The efficiency of the VNS algorithm is the lowest. Although the average time consumption per order of the EM algorithm is not the least, but taken together, the EM algorithm is the most efficient algorithm. (4) Considering the fuel cost, the EM algorithm can reach the optimal path faster, and its cost saving is the most, which is 3.17 yuan. The optimized design of the

distribution route reduces the company's fuel cost, energy consumption, exhaust emission, and environmental pollution. Considering the comprehensive cost of distribution, the comprehensive cost of the TS algorithm is basically the same as that of the original scheme. After the optimization design of the other three algorithms, the comprehensive cost is reduced to varying degrees compared with the original scheme. Among them, the EM algorithm saves the most cost, which is 7.81 yuan.

Combined with the actual situation, the company should not only consider one factor in choosing the distribution

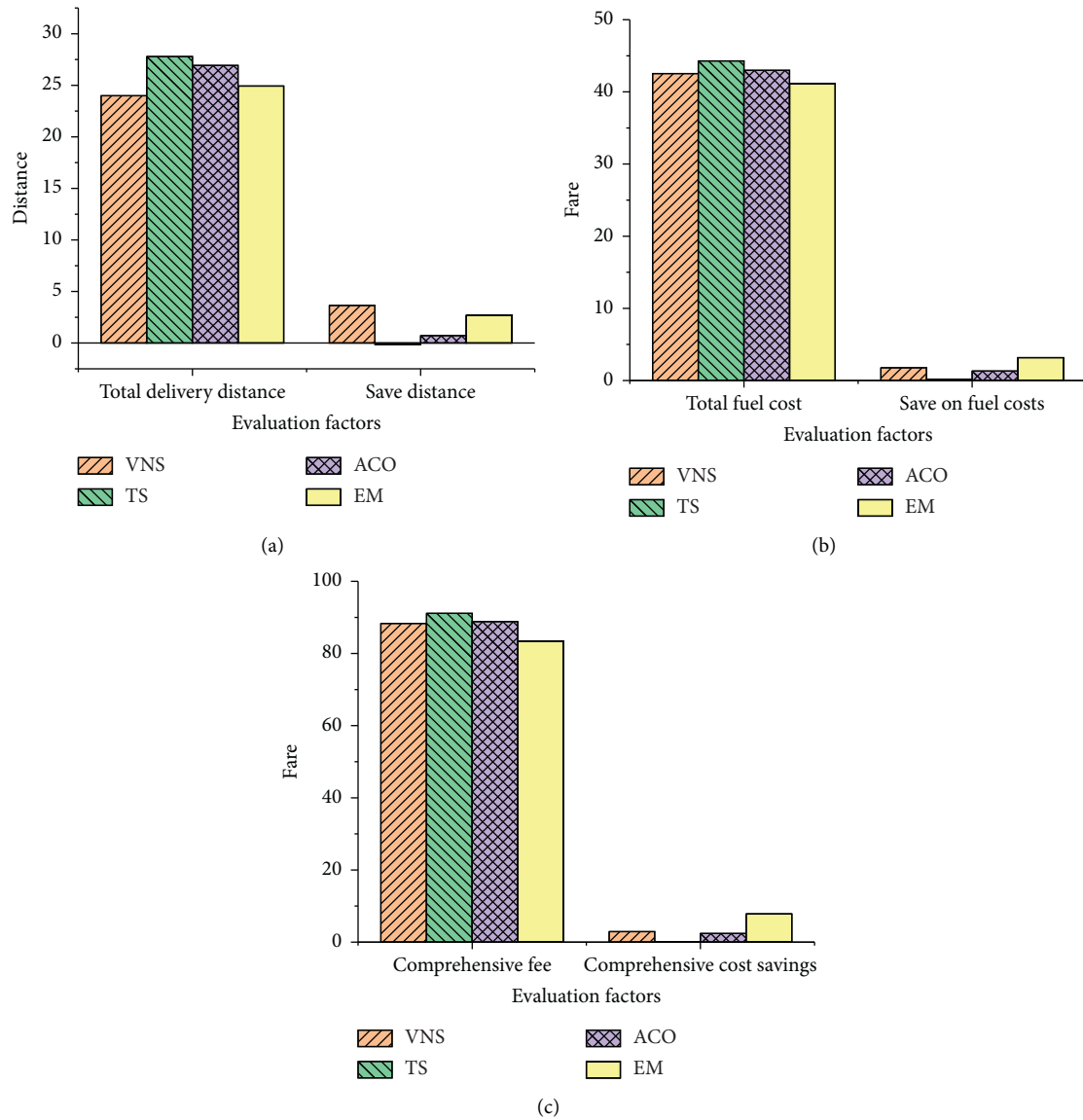


FIGURE 13: Advantages and disadvantages of four algorithms under different evaluation factors; (a) comparison of distribution distance of different algorithms; (b) comparison of fuel cost of different algorithms; and (c) comparison of comprehensive cost of different algorithms.

scheme, but comprehensively measure the factors such as vehicles, manpower, distance, and fuel cost. Hence, the comprehensive cost should be the basis for the company to decide the distribution scheme. If the operation time allows, the result obtained by EM algorithm is the best, but the convergence speed of this algorithm is slightly slower than that of ACO algorithm, but there is no significant difference.

#### 4. Conclusion

In the traditional logistics distribution management system, there may be missing data in the observation data due to observation conditions, instrument failure, human error, accidents, improper download and upload, improper storage process, and other factors. It will have a certain adverse impact on the overall quality of logistics planning and management. Based on this, first, the relevant theories of

intelligent logistics distribution and EM algorithm are summarized. Next, the intelligent logistics distribution scheme based on the EM algorithm is proposed. Finally, the algorithm widely used in the research of logistics distribution planning is selected as the control and the effectiveness of this method is verified by an actual case. The results show that the convergence speed of this method is the fastest. Its speed and times of obtaining the optimal path are the highest. The overall efficiency is the highest when processing orders. The fuel cost and comprehensive cost saved are the most. It can be concluded that under the condition of allowable operation time, the result obtained by the EM algorithm is the best. The research deficiency is that the problems of road condition information and traffic rules and regulations are ignored in the case study. Therefore, the follow-up research should also make reasonable improvement in combination with the specific situation to further

enhance the advantages of the algorithm. Applying this method to the actual logistics distribution management system can make a certain contribution to improve the quality of enterprise logistics distribution management and enhance the core competitiveness.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This research study was sponsored by these projects: Xuzhou Industrial Vocational and Technical College Doctoral Program, project number: XGY2021EB01; Xuzhou Industrial Vocational and Technical College Service Industry Development Research Institute Industrial R&D Project: "Research on the Evaluation System of Smart Logistics Distribution from the Perspective of Customers"; Jiangsu Province 2019 "Qing Lan Project" Excellent Teaching Team Project "Practical Teaching Team of Business Professional Group."

### References

- [1] A. J. Pesurnay, "Local wisdom in a new paradigm: applying system theory to the study of local culture in Indonesia," *IOP Conference Series: Earth and Environmental Science*, vol. 175, no. 1, p. 012037, 2018.
- [2] Y. Ding, M. Jin, S. Li, and D. Feng, "Smart logistics based on the internet of things technology: an overview," *International Journal of Logistics Research and Applications*, vol. 24, no. 4, pp. 323–345, 2021.
- [3] Y. Song, F. R. Yu, L. Zhou, X. Yang, and Z. He, "Applications of the Internet of things (IoT) in smart logistics: a comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4250–4274, 2021.
- [4] M. Humayun, N. Jhanjhi, B. Hamid, and G. Ahmed, "Emerging smart logistics and transportation using IoT and blockchain," *IEEE Internet of Things Magazine*, vol. 3, no. 2, pp. 58–62, 2020.
- [5] Y. Li, X. Yuan, J. Zhu, H. Huang, and M. Wu, "Multiobjective scheduling of logistics UAVs based on variable neighborhood search," *Applied Sciences*, vol. 10, no. 10, p. 3575, 2020.
- [6] T. Temucin and G. Tuzkaya, "A multi-objective reverse logistics network design model for after-sale services and a tabu search based methodology," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 4, pp. 4139–4157, 2020.
- [7] G. Calabrò, V. Torrisi, G. Inturri, and M. Ignaccolo, "Improving inbound logistic planning for large-scale real-world routing problems: a novel ant-colony simulation-based optimization," *European Transport Research Review*, vol. 12, no. 1, p. 21, 2020.
- [8] X. Wu, J. Mai, J. Zhou, M. Jiang, K. Wang, and K. Wang, "Concept and key technologies of intelligent logistics," *Journal of Physics: Conference Series*, vol. 1646, no. 1, p. 012092, 2020.
- [9] W. Di, "An empirical study on the relationship between the capital structure and the performance of the listed intelligent logistics companies," *International Journal of Social Science and Education Research*, vol. 3, no. 7, pp. 170–175, 2020.
- [10] X. Wang, X. Le, and Q. Lu, "Analysis of China smart city upgrade and smart logistics development under the COVID-19 epidemic," *Journal of Physics: Conference Series*, vol. 1570, no. 1, p. 012066, 2020.
- [11] M. Woschank, E. Rauch, and H. Zsifkovits, "A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics," *Sustainability*, vol. 12, no. 9, p. 3760, 2020.
- [12] O. Quan, Z. Jun, and S. Wang, "Investigation of the construction of intelligent logistics system from traditional logistics model based on wireless network technology," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, 7 pages, 2019.
- [13] S. Li, Q. Sun, and W. Wu, "Benefit distribution method of coastal port intelligent logistics supply chain under cloud computing," *Journal of Coastal Research*, vol. 93, no. sp1, p. 1041, 2019.
- [14] M. Mozaffari, B. Safarinejadian, and M. Shasadeghi, "A novel mobile agent-based distributed evidential expectation maximization algorithm for uncertain sensor networks," *Transactions of the Institute of Measurement and Control*, vol. 43, no. 7, pp. 1609–1619, 2021.
- [15] T. Tottori and T. J. Kobayashi, "Forward and backward bellman equations improve the efficiency of the EM algorithm for DEC-POMDP," *Entropy*, vol. 23, no. 5, p. 551, 2021.
- [16] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu, and L. Zhu, "A missing power data filling method based on improved random forest algorithm," *Chinese Journal of Electrical Engineering*, vol. 5, no. 4, pp. 33–39, 2019.
- [17] J. H. Jang, A. K. Manatunga, C. Chang, and Q. Long, "A Bayesian multiple imputation approach to bivariate functional data with missing components," *Statistics in Medicine*, vol. 40, no. 22, pp. 4772–4793, 2021.
- [18] P. Garcia-Patos and R. Olmos, "Multiple imputation in multilevel models. A revision of the current software and usage examples for researchers," *Spanish Journal of Psychology*, vol. 23, no. 1, p. e46, 2020.
- [19] C. H. Hsu, Y. He, C. Hu, and W. Zhou, "A multiple imputation-based sensitivity analysis approach for data subject to missing not at random," *Statistics in Medicine*, vol. 39, no. 26, pp. 3756–3771, 2020.
- [20] F. F. D. Nascimento and W. D. N. Moura, "Inference using MCMC to a new conjugate prior to positive parameters, applied in environmental data," *Communications in Statistics: Case Studies, Data Analysis and Applications*, vol. 7, no. 1, pp. 36–48, 2021.
- [21] W. Zhou, X. Wu, S. Ding, X. Ji, and W. Pan, "Predictions and mitigation strategies of PM2.5 concentration in the Yangtze River Delta of China based on a novel nonlinear seasonal grey model," *Environmental Pollution*, vol. 276, no. 1, p. 116614, 2021.
- [22] Z. Zhang, Y. Liu, J. Zhang, X. Song, and X. Song, "Research on the influence of new generation of information technology on contemporary enterprise logistics management information system," *Journal of Physics: Conference Series*, vol. 1648, no. 4, p. 042039, 2020.