Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 6
ISET 2024

Research on cargo volume prediction and personnel scheduling optimization of sorting center based on BP neural network and improved genetic algorithm

Wenzhuang Liu[1, #, *], Dehui Xu[2, #], Xinyu Wang[1, #], Ruiqian Sun[3, #]

[1]Department of Architectural, North China University of Science and Technology, Tangshan, China, 063210

[2]Department of Applied Physics, North China University of Science and Technology, Tangshan, China, 063210

[3]Department of New Energy Metallurgy and Energy, North China University of Science and Technology, Tangshan, China, 063210

*Corresponding author: 15613823128@163.com

#These authors contributed equally.

Abstract. With the development of Internet technology, e-commerce has become the primary choice for consumers to shop, and the huge order transaction volume has generated logistics problems. As an important part of the logistics network, the sorting platform is essential to improve its management efficiency. To enhance the efficiency of the sorting center, the topological relationship diagram is used to show the logistics network, and the cargo volume of the sorting center in the next 30 days and every hour is predicted based on the BP neural network model. Based on the mixed integer linear programming model and the improved genetic algorithm SCAGA model, the total man-days are minimized to ensure that the cargo volume is processed. At the same time, to ensure that the actual hourly efficiency is as balanced as possible, the optimization problem of personnel scheduling is studied. To maintain the efficient operation of the logistics industry, the accurate prediction of the number of goods and the reasonable scheduling of personnel, to provide a set of efficient and economical personnel scheduling programs for the logistics company, thus improving the overall operational efficiency and service quality.

Keywords: Logistics Network, Topological Relationship Diagram, BP Neural Network, Personnel Scheduling Optimization.

1. Introduction

To maintain the efficient operation of the logistics industry, personnel scheduling based on cargo volume prediction is the top priority to improve the efficiency of the sorting center. Accurate prediction of cargo volume and reasonable scheduling of personnel provide logistics companies with a set of efficient and economical personnel scheduling programs, thereby improving the overall operational efficiency and service quality. To maintain the efficient operation of the logistics industry, personnel scheduling based on cargo volume prediction is the top priority to improve the efficiency of the sorting center. Accurate prediction of cargo volume and reasonable scheduling of personnel provide logistics companies with a set of efficient and economical personnel scheduling programs, thereby improving the overall operational efficiency and service quality.

E-commerce increases productivity and reduces operating costs. The development of e-commerce cannot be separated from the support of the logistics industry. The development of e-commerce has put forward higher requirements for logistics transportation [1-3]. Ahmad Alqatawna et al. combined SARIMAX, ARIMA, AR, and long Short-term memory (LSTM) models to provide comprehensive analysis and insights for the order volume prediction of logistics companies [4]. Zhou Cheng et al. developed a new combinatorial forecasting model for predicting China's logistics freight volume and determined the combinatorial weights through the POS-BP neural network [5]. Ning Xu et al. made a short-term prediction of China's logistics industry through the adaptive gray prediction model [6]. Ling Liu et al. divided the predicted value into a dual neural network architecture with a linear part and a nonlinear part, improved the hybrid dual Arima &BP model, and predicted the freight flow [7]. Li Guangru et al. applied Elman neural network to the cargo throughput prediction of Zhoushan Port of Ningbo, and analyzed and compared the prediction results with BP neural network and RBF neural network. The results show that the Elman neural network can better adapt to the characteristics of throughput data changing with time, and its prediction performance is better [8]. Hu Xiuwu et al. optimized the scheduling of call centers by combining the heuristic algorithm with the neighborhood search algorithm of the simulated annealing mechanism [9]. Guo Wei et al. optimized the resources and scheduling of pilots based on AHP and fuzzy comprehensive evaluation [10].

Although models such as SARIMAX, ARIMA and LSTM have been applied, there may be a lack of scalability for different regions or types of logistics operations. Each model may perform well in a particular environment, but it may not be universally effective.Beyond that, optimization strategies may not be dynamic enough to adapt to changing workloads and operational needs. Real-time adjustments to staffing and scheduling based on current and projected needs are necessary.

2. Model

2.1. Topological diagram

A topological diagram is a mathematical model used to represent and describe the connections between objects (nodes). In a topology diagram, objects are connected by edges, which can be directed or undirected.

Topological diagrams can be used to analyze the flow of goods between different sorting centers, as well as to identify key nodes and routes with large volumes of goods. This type of topology is very useful for logistics and supply chain management and can help decision-makers optimize transportation routes, improve efficiency, and reduce costs.

The topology diagram is used to show the logistics network, in which the nodes represent the sorting center, the edges represent the freight transportation lines, and the weights represent the quantity of freight transportation. The visualization results are shown in Figure 1.

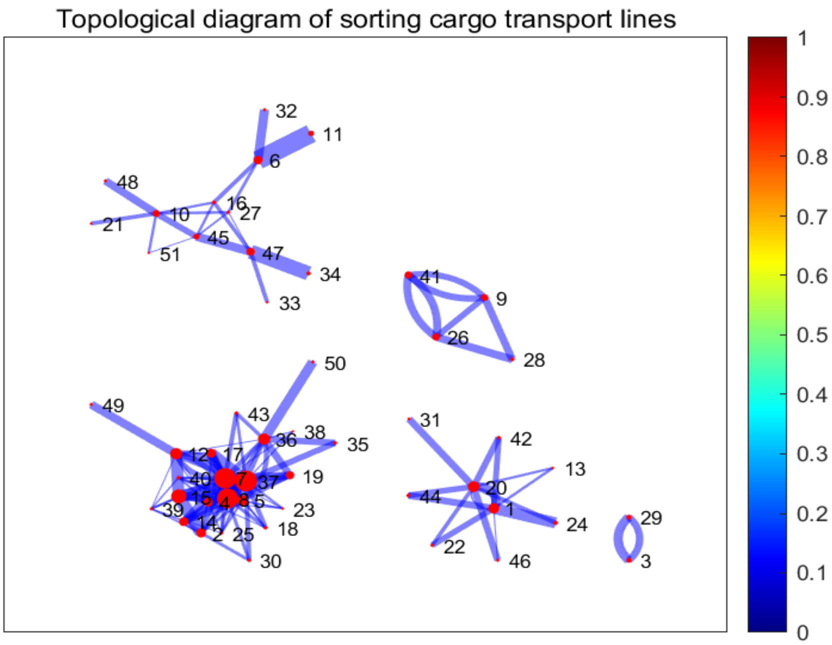

Figure 1. Topological diagram of sorting cargo transport lines

2.2. The basic fundamental of BP neural network

The BP neural network is a type of multi-layer feedforward neural network that is trained using the error backpropagation algorithm. It consists of an input layer, a hidden layer, and an output layer. This network has the advantage of a flexible network structure and a good nonlinear mapping function, making it uniquely suited for solving uncertainty problems. The network structure is illustrated in Figure 2.

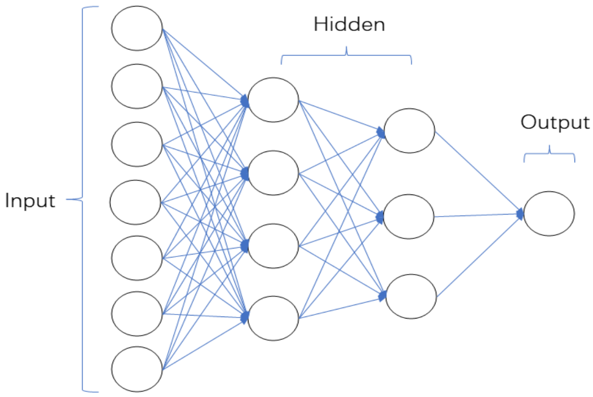

Figure 2. Neural network structure

During the forward propagation process, information enters the network through the input layer and then passes through each layer to produce the final output. If the output does not match the expected result, the error is sent to the backpropagation stage. In the backpropagation stage, the signal difference is calculated in reverse, using methods such as gradient descent to adjust the thresholds and weights of neurons in each network layer. It's important to set a maximum number of learning iterations in advance for this process.

$$\partial_{\mathrm{h}} = \sum_{i=1}^{d} v_{ih} x_i + \theta_h \tag{1}$$

From the output layer to the hidden layer there is the following relationship:

$$\beta_{\mathrm{j}} = \sum_{h=1}^{q} v_{hj} b_h + \theta_j \tag{2}$$

From the hidden layer to the output layer there is the following relationship:

$$E = \frac{1}{2} \sum_{\mathrm{k}=1}^{2} (y_k - T_k)^2 \tag{3}$$

2.3. The basic fundamental of MILP and SCAGA

ResultsMixed integer linear programming model is an optimization model in operations research. It has the high efficiency of linear programming and the discrete processing ability of integer programming. It can solve many problems including discrete selection and continuous assignment.

Decision variable: X_{ijk} is set as the number of regular workers in shift k of the i sorting center on day j; Y_{ijk} is the number of temporary workers on shift k on day j of the i sorting centre; Z_{ijk} is the total number of people for shift k of the i sorting center on day j.

Parameter: T_{ijk} is the amount of goods to be processed by shift k of the i sorting center on day j; Q_{jk} is the number of regular workers on shift k on day j; M is the highest number of regular workers, M=25; N is the maximum official number of temporary workers, N=20.

Objective function

$$\min z_{ijk} = \sum_{j=1}^{30}\sum_{k=1}^{6}(x_{ijk} + y_{ijk}) \tag{4}$$

Constraints: The number of shifts with completed cargo is expressed as follows:

$$s,t = \begin{cases} \min z_{ijk} = \sum_{j=1}^{30}\sum_{k=1}^{6}(x_{ijk} + y_{ijk}) \geq T_{ijk} \\ \sum_{k=1}^{6} Q_{jk} \leq 60 \\ X_{ij} \geq 0; Y_{ij} \geq 0 \end{cases} \tag{5}$$

2.4. The establishment of improved genetic algorithm SCAGA model

Genetic algorithm is a model used to simulate the genetic and evolutionary process of biological populations in the natural environment. For general biological populations, genetic algorithm is completely random, so it is often very easy to lead to unbalanced evolutionary results in the iterative process. In order to ensure that the initial individuals are evenly distributed in the solution space, this question uses the improved genetic algorithm SCAGA.

The process of model building is as follows:

Step1 According to the population size, the solution space is divided into N equal subspaces

Step2 Each subspace forms subentities in a completely random manner

Step3 The initial population is obtained by combining all the sub-individuals

The algorithm formula accompanying the gradual reduction in the iterative process is as follows:

$$x = \sin\left(\left(1 - \frac{gen}{gen_{\max}}\right) \cdot \frac{\pi}{2}\right) \tag{6}$$

where gen the representative represents the current evolutionary generation, and gen_{max} represents all evolutionary generations.

In order to adjust crossover probability and mutation probability within the effective range, the following common formula for probability update is established:

$$z = x \cdot y = \sin\left(\left(1 - \frac{gen}{gen_{\max}}\right) \cdot \frac{\pi}{2}\right) \cdot \ln(1 + Na) \tag{7}$$

$$p_c^{'} = \frac{1}{1 + e^{-x}} \tag{8}$$

where N is population size, α is percentage of fitness for a particular individual.

Therefore, the crossover probability of two individuals is:

$$p_c = \frac{p_{c1}^{'} + p_{c2}^{'}}{2} \tag{9}$$

Since the mutation operator affects the effectiveness of the algorithm to a certain extent, it becomes particularly important to determine the value of the mutation probability. In general, the value of the mutation probability becomes very important, and the way to dynamically adjust the mutation probability becomes very important.

$$p_{\mathrm{m}} = \left| \sqrt{\sin((1 - \frac{gen}{gen_{\max}}) \cdot \frac{\pi}{2}) - 0.5} \right| \tag{10}$$

where p_m is before the probability of variation.

In the iterative process, the value of pm gradually decreases. This amplification can not only avoid falling into the local optimal solution in the early morning, but also improve the convergence of SCAGA in the later stage.

3. Results

3.1. Analysis of BP Neural Network Results

The model is solved by specifying the maximum number of iterations as 10000 and the target accuracy as 0.4. Build the training set and test set, train and test the data set of BP neural network, and predict the cargo volume. The result is shown in Figure 3 and Figure 4. The results of some predicted values are shown in Table 1 and Table 2.

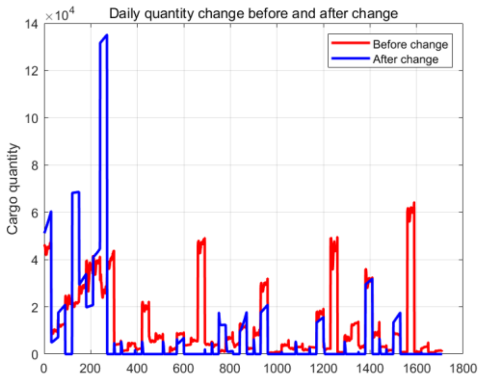

Figure 3. Daily quantity change before and after change

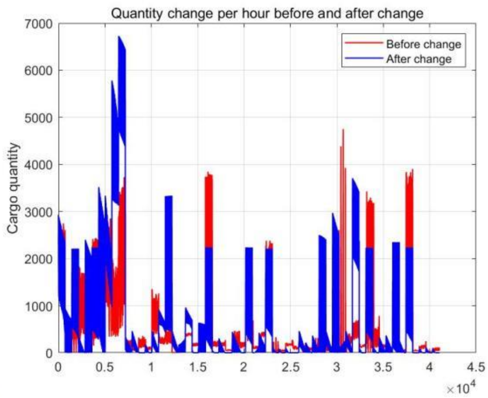

Figure 4. Quantity change per hour before and after change

Table. 1. Volume per day for the next 30 days

Sorting center	Date	Cargo quantity
SC1	2023/12/1	51153
SC2	2023/12/1	5000
SC3	2023/12/1	17572

Table. 2. Volume per hour for the next 30 days

Sorting center	Date	Hour (s)	Cargo quantity
SC1	2023/12/1	0	2918
SC1	2023/12/1	1	2929
SC1	2023/12/1	2	2924
SC1	2023/12/1	3	2908
SC1	2023/12/1	4	2787
SC1	2023/12/1	5	1786
SC1	2023/12/1	6	1241
SC1	2023/12/1	7	1209
SC1	2023/12/1	8	1233
SC1	2023/12/1	9	1313
SC1	2023/12/1	10	2918

From the results, we can see that the changed volume increased significantly on certain days and at certain times, especially on those days with higher volumes, which may mean that the change in the network led to the optimization of the volume distribution, allowing the sorting center to process the cargo more efficiently. The change in network structure seems to have had a positive impact on the daily and hourly volumes at the sorting centers, especially during peak hours. This change may be due to new network designs that distribute goods more efficiently, reduce bottlenecks, or improve resource utilization.

3.2. Analysis of MILP and SCAGA results

To better understand the values of different variables in the optimization solution, the distribution diagram of decision variables is solved first, as shown in Figure 5.

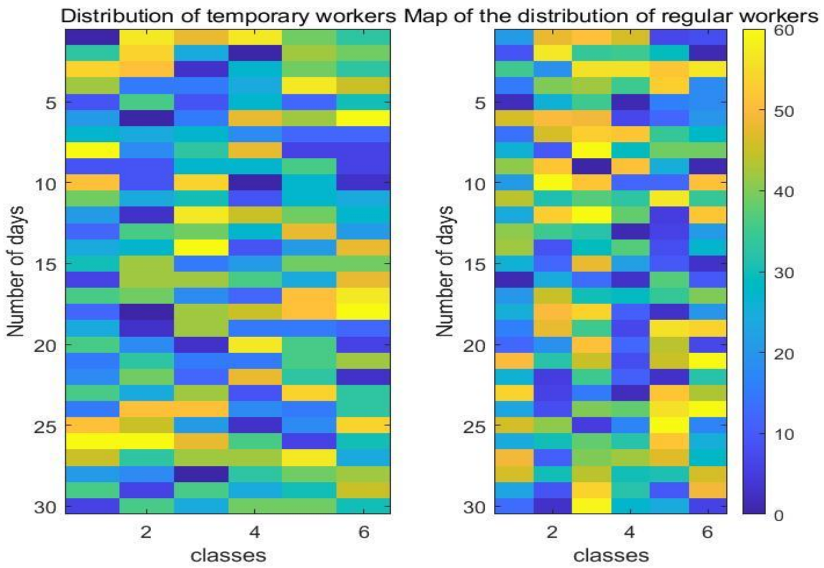

Figure 5. Distribution of decision variables

As can be seen from the map of the distribution of decision variables, the number of regular workers also varies across shifts, but looks smaller overall than the number of temporary workers, but the use of temporary workers is more concentrated in some shifts, which may mean that additional human resources are needed in these shifts to cope with work peaks. For the temporary worker scheduling distribution map, the number of temporary workers in shift 6 is the largest, and the number of other shifts is relatively small, while for the regular worker scheduling distribution map, the number of regular workers in shift 4 is the largest, and the number of regular workers in other shifts is more average.

Then, the change trend of the optimal fitness value of the algorithm in the multi-generation evolution process is calculated, as shown in Figure 6.

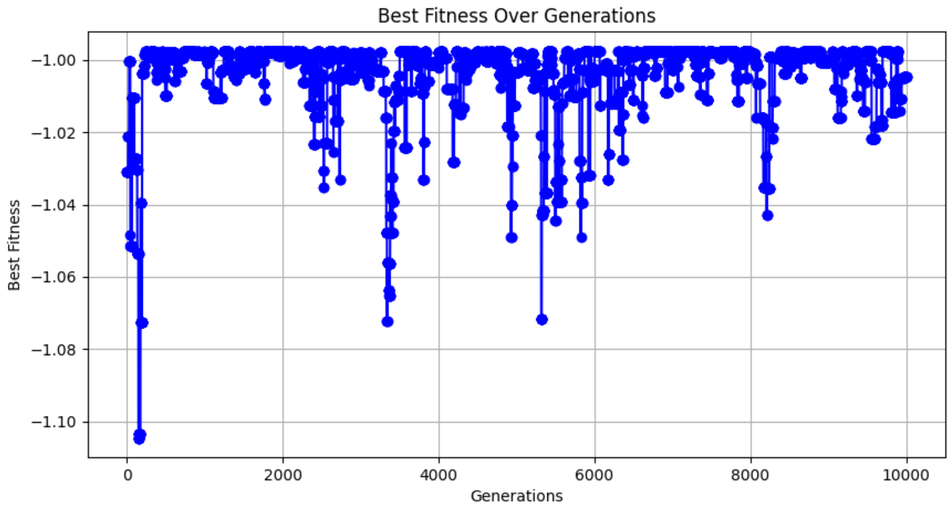

Figure 6. Variation trend of optimum fitness value during multi-generation evolution

In the initial stage (generation 0), the optimal fitness value is about -1.08, which indicates that there is still a certain gap between the optimal solution in the initial population and the goal of the problem. With the increase of algebra, the optimal fitness value gradually increases, which indicates that the algorithm constantly finds better solutions in the search process. The curve as a whole presents an upward trend, which indicates that the algorithm gradually approaches the optimal solution of the problem in the iterative process, which means that the algorithm has found a stable optimal solution or has converged to the local optimal solution.

The number of temporary and regular workers per shift at the sorting center is analyzed, and a portion of the attendance per shift at each sorting center for the next 30 days is output in Table 3.

Table. 3. The number of attendees per shift at the sorting center

Sorting center	Date	Classes	Number of regular workers	Number of temporary workers
SC1	2023/12/1	00:00-08:00	10	41
SC2	2023/12/11	05:00-13:00	5	0
SC3	2023/12/3	14:00-22:00	10	7
SC7	2023/12/28	16:00-24:00	10	10
SC28	2023/12/10	12:00-20:00	10	1
SC31	2023/12/17	14:00-22:00	10	2
SC44	2023/12/10	00:00-08:00	10	3
SC53	2023/12/15	12:00-20:00	10	20
SC57	2023/12/22	05:00-13:00	10	5
SC1	2023/12/1	00:00-08:00	10	41

4. Conclusions and outlooks

The development of e-commerce has brought a huge volume of orders, which in turn has created logistics problems. The sorting center is the central node of the logistics platform, and the cargo volume prediction of this link is an important research problem of the e-commerce logistics network. In order to improve the efficiency of the sorting center, this paper studies the cargo volume prediction of the sorting center and the optimization of personnel scheduling. The topological diagram was used to visualize the logistics network and optimize the transportation route. The BP neural network model and linear integer programming model, combined with the improved genetic algorithm SCAGA model, were used to predict the cargo volume and optimize the optimal attendance of each shift in each sorting center. Finally, the cargo volume and the optimal attendance of each sorting center were obtained. The model can be applied to different types of logistics networks, such as express delivery companies, e-commerce platforms, supply chain management, etc., to analyze the adaptability and optimization effect of the model under different business models.

The topology diagram is used to analyze the logistics network, which helps to identify the key points and routes, and provides a more intuitive perspective for optimizing the transportation routes.

Through mixed integer linear programming model and improved genetic algorithm SCAGA model, the scheduling problem is effectively solved and the rational allocation of human resources is realized. Models may not fully take into account all external factors such as extreme weather events and may need to be adapted for different logistics networks and business scenarios to maintain their effectiveness and usefulness.

References

[1] Akıl S, Ungan M C. E-commerce logistics service quality: customer satisfaction and loyalty[J]. Journal of Electronic Commerce in Organizations (JECO), 2022, 20(1): 1-19.

[2] Tsang Y P, Wu C H, Lam H Y, et al. Integrating Internet of Things and multi-temperature delivery planning for perishable food E-commerce logistics: A model and application[J]. International Journal of Production Research, 2021, 59(5): 1534-1556.

[3] Chen Guobin. Discussion on Logistics under the condition of E-commerce [J]. Journal of Fujian Executive College of Economics and Management, Fujian Administration College, 2001 (3): 53-55.

[4] Alqatawna A, Abu-Salih B, Obeid N, et al. Incorporating Time-Series Forecasting Techniques to Predict Logistics Companies' Staffing Needs and Order Volume[J]. Computation, 2023, 11(7): 141.

[5] Cheng Z, Juncheng T. Adaptive combination forecasting model for China's logistics freight volume based on an improved PSO-BP neural network[J]. Kybernetes, 2015, 44(4): 646-666.

[6] Xu N, Gong Y, Bai J. Adaptive Grey Prediction Model with Application to Demand Forecasting of Chinese Logistics Industry[J]. Journal of Grey System, 2019, 31(1).

[7] Liu L, Pan C C. Forecast of Line-level Logistics Cargo Volume Based on Hybrid Dual Neural Network Model[C]//2019 Chinese Control And Decision Conference (CCDC). IEEE, 2019: 4326-4329.

[8] Li Guangru, Zhu Qinghui. Port cargo throughput prediction based on Elman neural network [J]. Journal of Chongqing Jiaotong University (Natural Science Edition), 2020, 39(06): 8.

[9] Hu Xiuwu, WANG Ruicheng, WANG Xiuli. Call center scheduling optimization model and algorithm considering class constraints [J]. Industrial Engineering and Management, 2020, 25(2): 51-58. (in Chinese)

[10] Wang Zhenyu, Guo Wei, Yu Kai, et al. Research on Pilot Resource Optimization Allocation and Scheduling Risk Evaluation based on C# language and AHP principle [J]. Science and Technology Innovation, 2020 (31): 92-93.