

## Research on the Express Delivery Delay Prediction Based on Neural Network in the Background of Big Data

Bin Ye<sup>1, a \*</sup>, Jia Zuo<sup>2, b</sup>, Xin Zhao<sup>3, c</sup> and Lin Luo<sup>4, d</sup>

<sup>1, 2, 3, 4</sup>Chengdu Neusoft University, Chengdu, Sichuan, 611844

<sup>a</sup>yebin@nsu.edu.cn, <sup>b</sup>zuoja@nsu.edu.cn, <sup>c</sup>zhaoxin@nsu.edu.cn, <sup>d</sup>luoalin@nsu.edu.cn

**Keywords:** Big data; Neural network; Logistics; Freight; Express

**Abstract.** For the current express industry, the paper sets out to explore the causes of delay during transport-using the big data technology, collect appropriate freight data, and establish corresponding analysis indicators and corresponding training set, using Mat-lab and other tools for simulation and training, use the artificial neural networks, and build applications model to analyze the simulation results. It also implements the Logistics delay prediction for freight, and explain the Application Method in logistics decision-making and management system, and application prospects.

### Introduction

Based on the latest release of “the Twenty-seventh China Internet Network Development Situation Survey” by China Internet Network Information Center (NIC CN-) , China's Internet users has reached the scale of the Internet, online shopping has become the fastest growing Internet application, and the amount of online shopping market has reached 926 billion 230 million yuan by the end of 2014. However, the problem of express delay is becoming increasingly prominent in the online shopping process.

In express industry, *The express service standards* express the limit of delivery service time is clearly defined, the urban express delivery time is no more than 24 hours, and the suburban express delivery time limit is no more than 72 hours. Shop staffs from Taobao and courier companies are mostly able to comply with this time limit, but it is difficult for some courier companies to guarantee the time limit at the peak of the businesses. Recently, when consumers value the goods of the shops, there are also many complains about the very slow express. In online shopping, bad feedback to merchants due to express delay accounts for more than 70%, fully illustrating those businesses, courier companies, and customers highly concern the delivery delay.

In modern express enterprises, the control over the transportation time of the goods directly affects the management of customer relationship, the management of the supply chain and the transportation decision system. If you can grasp the delivery time in advance and reference an early warning mechanism. To solve the current logistics disputes, establish a good reputation of enterprises, as well as adjust the transportation strategy, it is of great significance.

This article will use a more mature neural network to explore the reasons for the delay of delivery, predict the delay time, and introduce the method of introducing into the network system.

### Analysis of the Reasons for the Delay in Express Delivery under the Background of Big Data

Many factors affect the delivery time, and these factors relate to the social, economic, traffic and individual characteristics. For different situations and research environments, the degrees of influence produced by various factors are different [1], and the main factors influencing the freight time are mainly in the following aspects:

**The Level of Regional Economic Development.** From the perspective of social development, when the economic development in the region is better, and the city planning based on the traffic infrastructure construction has a good promotion, the transportation is more abundant, and the logistics industry is relatively developed. At the same time, the probability of goods delay in the region will be smaller. On the contrary, when the regional economic development is lagging behind, the probability of goods delay will be greater.

**Regional Cultural Background.** There is a certain correlation between social culture and economic development; the people of different cultural background have different understandings of the transportation delay. Particularly, some people live and work at a relatively fast pace; therefore, they have more stringent requirements for on-time delivery, and the relationship between logistics enterprises and customers is tense; while in some areas, the people have a more tolerant attitude. Thus, regional culture has caused people to have different attitudes towards goods delay.

**Individual Difference.** Individual customers are very complex. They have great individual differences, and have different requirements for on-time delivery, including delivery time and receiving time; these differences also affect the actual arrival of freight. For example, some people cannot receive goods in time for work reasons. Although some times the delay is caused by the individuals' own reasons, not by logistics enterprises' reasons, for the logistics enterprises, loss of time is the consequence. So, we also pay close attention to the individual differences.

**The Nature of the Goods.** With the rapid change of transportation demands towards logistics enterprises, for the development of enterprises, logistics enterprises are no longer so picky about the varieties of goods. The special sizes of goods increase the difficulty of logistics transportation, and there are special requirements with respect to special objects in different stages of transportation, e.g. fragile, explosive goods or chemical products should be subject to strict inspection which is inevitable, but also causes the freight delay.

**Transportation Distance.** The transportation distance is an important basis for logistics enterprises to charge cargo transportation fees, but not the farther the distance is, the longer the delivery time is, because there are more vehicles to choose from. It is not absolute that the delay time is proportional to the transportation distance, but the transportation distance is one of the main causes of delay.

**The Differences of Express Enterprises.** Express enterprises make logistics decision-making, and control the process of logistics transportation. Implementation of enterprises' decision-making system, management, as well as transportation personnel's basic business abilities and occupational ethics will affect the delivery of goods. To reduce the number of freight delay and improve the transportation conditions of enterprises, the fundamental method is to establish a good corporate reputation.

**Natural Causes.** natural factors; In the process of freight transportation, no matter what means of transportation is used; people will inevitably deal with the nature. Important factors and natural factors may cause freight delay, such as force majeure, including earthquakes, tsunamis, landslides, wind, and rain and so on. Similarly, normal/good weather and environment have positive effects on logistics transportation, and will produce less delay.

Most of traditional analyses are based on past experience. However, in the face of changing markets and large customers, the traditional analyses are inadequate, and the intelligent early warning cannot play its due role. The fundamental reason is obvious, i.e. sufficient data are not collected effectively to comprehensively analyze the logistics activities involving various factors. In today's age of increasing data quantity, data collection, data processing and analysis provide a chance to use the traditional data processing method, and big data technologies are used to establish a really useful intelligent early warning system for logistics. The principle is shown in Fig. 1 below:

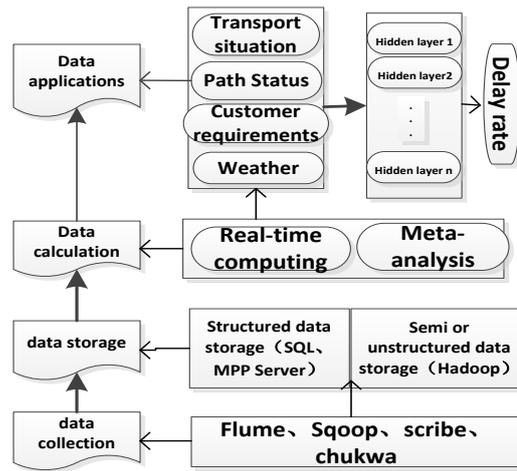


Figure 1. Data collection, calculation and processing process

### The Establishment of the 3-Prediction-Model Method and Process

**The Basic Principle of BP Neural Network.** The BP artificial neural network is McClelland, and Rmenlhart research and design, based on the error back propagation algorithm (Back-propagation) of a multilayer feed forward neural network. In the development history of the artificial neural network, there is no effective algorithm to find the connection weights of the hidden layer for a long time. Until the error back propagation algorithm (BP algorithm) is proposed, it successfully solves the problem of the weight adjustment of multilayer feed forward neural network for solving the nonlinear continuous function.

The learning rule is to use the steepest descent method, through back propagation to constantly adjust the weights and thresholds of the network. The network error square and the minimum of the BP neural network model topology consists of an input layer (input), a hidden layer and an output layer (output layer), as shown in Fig. 2.

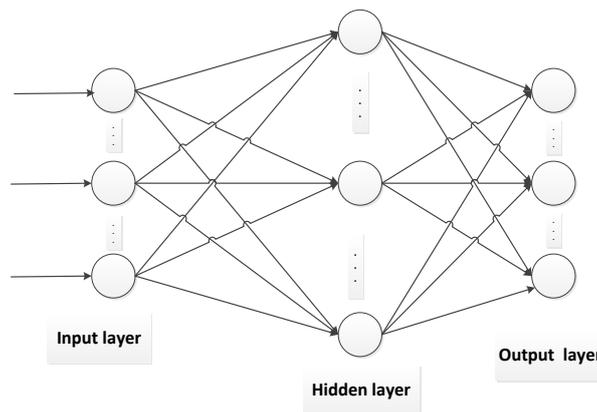


Figure 2. BP neural network structure diagram

The main process of BP algorithm is inputting learning samples and using the back-propagation algorithm of the network weights and biases are repeated adjustment and training. To make the output vectors and expected vector as close as possible to, when the square error of the network output layer and the error is less than the specified training is complete, then save the network weights and biases.

**Analysis of the Establishment and Quantification of Indicators.** In order to realize the scientific prediction, improve the prediction accuracy, facilitate the practical application, the principle of establishing the index system should follow the following principles:

Firstly, to try to contain the factors mentioned above and more elements, the more manifest is its

science, not sweeping. In fact, to be combined with the reality, samples are selected from the real investigation; combination with the reality of data has application values. Second, the analysis of indicators should not be too much, because too much computation is too large and is not practicable in the actual system. Finally, the index through simulation tests calculates the error rate, by adjusting the index and number of samples and the error rate is relatively small. Therefore, for the reasons for express delay, there are man-made or natural factors. Through the questionnaire analysis and network survey, based on the sample analysis of the impact of the reasons for the delay, the main reasons include distance, weather, customers, and arrival areas. These factors have a widely coverage. Meanwhile, for reducing the analytical index, the operation is relatively less but more conducive to the application of the model. This paper selects a number of factors as candidates for the index analysis. Through simulation tests, it is found that the error is relatively small. For this, the detailed analysis is carried out on these indexes, so as to facilitate the simulation tests.

**Prediction of the Delivery Delay Model Based on the BP Neural Network.** In the analysis of indicators, there are four sub indicators, e.g. the input layer, the output layer, the unit number of the hidden layer and problems [2]. The experience formula is as the following [3]:

$$L = \frac{3\sqrt{mn}}{2} \tag{1}$$

In formula, N and N as a input and output unit number; L is a hidden layer unit number. In the actual design of the network, based on the speed of the convergence, setting precision further adjustment [4].

By determination of the empirical formula (3-1), the hidden layers can be determined [5] [6]. But in order to more accurately determine the optimal hidden layer number, the network is trained by using the neural network toolbox of matlab7.0. When the training error reaches the precision requirement, the training is stopped. By this, the training practice, the actual step number and the maximum error value of different hidden layers can be calculated. Given the learning accuracy for  $E = 0.01$ , the hidden layers tested are listed in the following Table 1.

Table 1 The error margin of different hidden layers and training steps

Hidden layer	2	3	4	5	6
Max error	0.135	0.071	0.092	0.085	0.133
training steps	7353	2705	3352	2806	3502

Through the analysis, the relative error is small and the number of training steps is the minimum when the hidden layer number is 3.

**Analysis of the Index Normalization Processing.** In the BP neural network's delay prediction model, four input indexes in the input layer should first be normalized in the training before the [0,1]. These values are only as a mark in the analysis. In order to facilitate the calculation, uniform standards are used, and each identifies the normalized value to represent the influence on the results. There is no unified standard for the quantitative value divided by 5, and the value is a normalized value of each sample, so that the quantitative value is between [0, 1]. In addition, the expected output value is the actual delay rate. The actual delay time is determined from the corresponding sample, and the delay time is also normalized. The processing method is as follows: the scheduled arrival time divided by the real time of arrival, the value is a decimal less than 1, and the value of the said anti delay rate is small. Although the delay time is longer, it also meets the requirement of normalization. T is used to express the actual transportation time, and the H is used to express the expected transit time. Then the calculation formula for the delay rate F as below:

$$F = 1 - \frac{T}{H} \quad (2)$$

In the sample processing, some do not conform to the requirements, and the samples are removed. For example, in real life, the actual time of arrival sometimes is shorter than the predetermined time, and then there is no delay. In the selection of samples, samples as these are eliminated

**Analysis of the Training Method of the BP Neural Network.** In the design of the BP neural network, the number of neurons in each layer, the initial value and the learning rate of the network should be considered. It has been proved that the three layers of the BP neural network can realize the mapping of the multidimensional unit cube  $R_m$  to  $R_n$ , which is able to approximate any rational function. Therefore, the author uses three layers of structure design, and gives priority to considering increasing the number of neurons in the hidden layer. According to the previous formula (1) the number of neurons in the hidden layer can be set to 3.

When the MATLAB neural network toolbox is used to train the network, there are two methods to realize many kinds, and the commonly used methods are the gradient descent method (traingd), the momentum gradient descent method (traingdm) and the adaptive LR gradient descent method (traingda). The author has taken into account the number of layers, as well as the number of neurons of the hidden layer. So the training method is selected to consider the training rate. The adaptive LR gradient descent method can adaptively adjust the learning rate, so as to increase the stability and improve the precision of speed. Therefore, the author employs the adaptive LR gradient descent method (traingda). And the learning rate is  $lr=0.06$ .

**Program Code and Results Analysis.** After the sample collection and the normalization process are completed, 80 groups of samples are selected and used to train the hidden layer and the accuracy of the training.

After the operation of the relevant results and the following learning process is shown in Fig. 3:

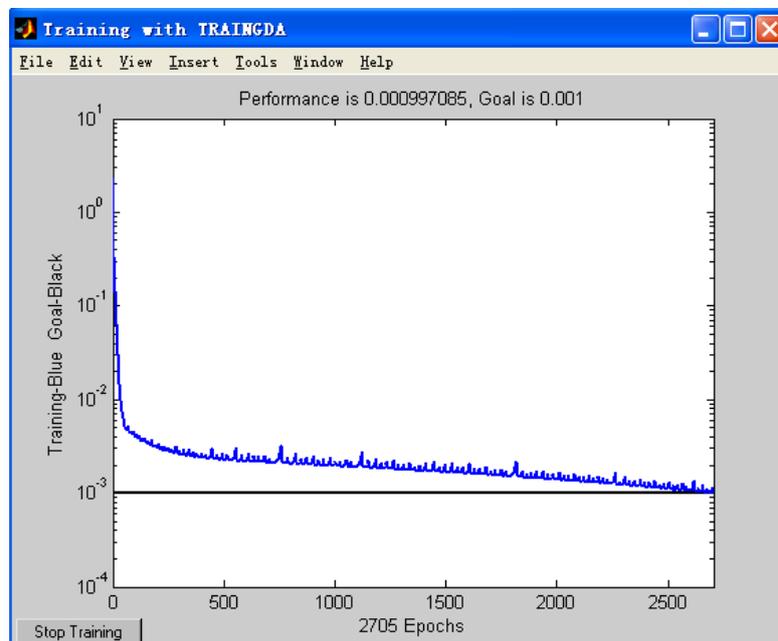


Figure 3. Learning process diagram

After the operation of the relevant results, the learning process is shown in Fig. 3. It can be seen from the chart that the network training is completed after running 2705 steps.

After the testing, the number of neurons of the hidden layer in the BP neural network is 3, the maximum error is 0.0715, and the minimum error is 0.0047, which is basically controlled between 0.4%-7%

## Summary and Research Prospects

From the training results, it can be seen that the use of the BP neural network model can be more accurate than the freight delay forecast. In practical application, a logistics decision system leads to an early warning mechanism, by which, a decision can be made ahead of time, and the customers can be notified. The logistics decision-making system is more perfect [7] [8], and can effectively alleviate the problem due to the delay caused by the tension of customers. In the network environment, the model can also be used as a strong support for an e-commerce system, so that customers and businesses, as well as the logistics companies can effectively monitor goods [9] [10]. However, the social environment and the natural environment are very complex, and there are many factors influencing the freight delay, so the analysis of indexes in different regions and different social environments can be adjusted, in order to make forecasting more scientific and more reasonable.

## References

- [1] Hong Tao. Study on the value of travel time [D]. Beijing Jiaotong University, 2011.
- [2] Shen, Hui, Liu, Zhigui, Li, Chunju. Design of traffic volume prediction based on BP neural network [J]. Journal of Southwest University of Science and Technology.2008 23 (2): 73.
- [3] Tao. Research on character recognition algorithm based on high order neural network [D]. Northeast Normal University 2010.
- [4] Zhang Jianping. Evaluation of the operation quality of air traffic control based on BP network [J]. Journal of Southwest Jiao Tong University, 2013 (6).
- [5] Quan-bing HE. Learning Capability Study on Improved RBF Neural Network [A]. Journal of Simulation (VOL.3, NO.2) [C]. 2015
- [6] WU Jin-Song. Study on neural networks in optical character recognition [A]. Proceedings of 2015 4th National Conference on Electrical, Electronics and Computer Engineering (NCEECE 2015)[C]. 2015
- [7] Jun Tang. Research on the operation mechanism of modern logistics enterprises under the low-carbon economy [A]. Proceedings of 2014 International Conference on Information Management and Management Engineering IMME 2014)[C]. 2014
- [8] Peng Liu, Yan Yan, Jianjian Zhang, Baosheng Jiang, Yajuan Qiao, Baona Sun. Research of Urban Logistics Node Layout Solution[A]. Journal of Simulation (VOL.3, NO.6)[C]. 2015
- [9] CHEN Yining. General Application of Data Mining Technology in Logistics Enterprises Management [J]. Journal of Guangxi Politics and Law Management Cadre Institute.2011 (02)
- [10]XU Peipei. HE Yue. Regional Logistics Demand Forecasting based on Self-organizing Data Mining [J].Statistics and Decision Making.2011 (6)