

Research Article

Probabilistic Prediction of Unsafe Event in Air Traffic Control Department Based on the Improved Backpropagation Neural Network

Yong Liao , Zhiyang Miao , and Changqi Yang 

College of Air Traffic Management, Civil Aviation Flight University of China, Sichuan, Guanghan 618307, China

Correspondence should be addressed to Yong Liao; liaoyong_1234@163.com

Received 27 March 2021; Revised 27 July 2021; Accepted 13 August 2021; Published 27 August 2021

Academic Editor: Samuel Yousefi

Copyright © 2021 Yong Liao et al. 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.

Air traffic control is an important tool to ensure the safety of civil aviation. For the departments that do the work of air traffic control, reducing the percentage of unsafe event is the core task of safety management. If the relationship between the percentage of unsafe event and their influencing factors can be effectively clarified, then the probability of unsafe event in some control department can be predicted. So, it is of great importance to improve the level of safety management. To quantitatively estimate the probability of unsafe event, a three-layer BP neural network model is introduced in this paper. First, a probabilistic representation of unsafe event related to air traffic control department is made, and then, the probability of different classes of unsafe events and safe events is taken as the outputs of the BP neural network, the factors influencing occurrence of unsafe event connected with air traffic control is taken as inputs, and the sigmoid function is chosen as activation function for the hidden layer. Based on the error function of neural network, it is proved that the general BP neural network has two drawbacks when used for the training of small probability events, which are as follows: the pattern does not ensure that the sum of probability of all events is equal to one and the relative error between the actual outputs and desired outputs is very large after the training of neural network. The reason proved in this paper is that the occurrence rate of the unsafe event is much smaller than that of the safe event, resulting in each weight in the hidden layer being subjected to the desired outputs of the safe event when using the gradient descent method for network training. To address this issue, a new mapping method is put forward to reduce the large difference of the desired outputs between the safe event and unsafe event. It is theoretically proved that the mapping method proposed in this paper can not only improve the training accuracy but also ensure that the sum of probability is equal to one. Finally, a numeric example is given to demonstrate that the method proposed in this paper is effective and feasible.

1. Introduction

China has a large population, vast geographical area, and uneven distribution of natural resources. To promote economic development and improve people's life, the exchange of people and goods among different regions is very frequent. Transportation is the tool to realize the exchange of people and goods, and different transportation modes have different characteristics. Due to the large population and uneven distribution of the population in China, the transportation demands are diverse, and to meet different people's demand, the Chinese government has been committed to building a diversified comprehensive transportation

system, creating a comprehensive transportation network integrating railroads, highways, waterways, and civil aviation [1–5].

With the popularization of China's civil aviation from high-end passengers to the public, civil aviation is becoming more and more important in China's comprehensive transportation system. In 2019, China civil aviation completed a total of 129.27 billion ton-kms of freight turnover, 660 million person-times of passengers, and 752.6 million tons of cargo and mail, with growth rates of 7.1%, 7.9%, and 1.9% year-on-year, respectively. With the advantages of safety, speedability, and convenience, civil aviation is winning more and more people's choice, and the share of civil

aviation in passenger transportation reached 32.8%, with a growth rate of 1.5% year-on-year. The number of flights is rising steadily, and Chinese transport airlines completed 4,966,200 takeoffs, with an increase of 5.8% over the previous year [6].

Safety is the primary concern for every mode of transportation. With the increasing in the number of flights, the air routes and airports are becoming more and more congested, and the controllers' workload is constantly increasing, which brings serious challenges to the safety of civil aviation. Air traffic management is an important part to ensure the safety of civil aviation, and the prerequisite for the safety of air traffic management is the safety of air traffic control (ATC), which makes the operation of aircrafts efficient, orderly, and safe. ATC is a service provided by ground-based controllers who direct aircraft on the ground and through controlled airspace, providing advisory services to aircraft in noncontrolled airspace [7]. ATC department is becoming more and more difficult for controllers to direct aircrafts, and the increase of flight volume leads to an increasing possibility of unsafe event and an increasing risk for the operation of the ATC department.

Safe operation has always been the goal pursued by civil aviation and is also the focus of the whole society [8]. There are many factors that influence safety in the ATC department. Air traffic controllers and pilots are crucial in achieving high levels of safety in air traffic operations [9]. The increasing volume of flights and the expanding route network have led to a dramatic increase in the workload of air traffic management, and a gradual increase in the number of ATC unsafe event has had a serious impact on the development of China's civil aviation [10]. Air traffic management systems are typically highly interactive, highly distributed, and complex systems [11]. The changing objective conditions lead to uncertainty, and the issue of ATC safety comes along with various uncertainties in the process of civil aviation transportation [10]. The increase in flight has led to a dramatic increase in demand for airspace, so that various unsafe factors also increase [12]. The contradiction between the capacity of air transport services and demand for transport is becoming more and more prominent. With this come the security risks in air traffic control operations [13]. The limited airspace, manpower, and equipment resources have increased the pressure on air traffic controllers and the probability of unsafe event in air transport has increased [14]. Due to the dynamic and real-time nature of the ATC control operation process, its risk level is higher than other systems, and control factors are more likely to lead to an unsafe event [15]. The risk factors of organization and management in the air traffic control system have a complex influence on flight safety [16].

To improve the safety level of air traffic control, different approaches have been taken in various ways. Safety integrity system is of great significance for modern management, and it is helpful to establish a safety integrity system for ATC [17]. Global networks of satellites for communications, navigation, and surveillance are longer-term solutions to air traffic management [18]. Automation removes some existing sources of human errors, and it can prevent some accident

[19]. The construction of a safety culture in air traffic control can protect air traffic controllers [20]. Among the many methods, safety assessment for air traffic control is the commonly used [19, 21, 22].

The Chinese government attaches great importance to civil aviation safety. According to "Rules for Safety Management of Air Traffic Management Operation Department of Civil Aviation" published by the Civil Aviation Administration of China, the operation department of air traffic control of civil aviation shall establish a safety assessment mechanism. However, this document does not put forward specific implementation plan or evaluation method. The reason is that there is no universal risk assessment methodology that can be applied to any situation. So, in the "Regulations for Safety Assessment of Air Traffic management of Civil Aviation" published by the Civil Aviation Administration of China, the Civil Aviation Administration encourages and supports research and innovation on the methods and techniques of safety assessment to make the safety assessment scientific and normative. This document also points out that the safety assessment should choose reasonable methods, based on the conditions, characteristics, and needs of the actual circumstances.

Therefore, the safety assessment in air traffic control has always been an important research topic in civil aviation, and many experts and scholars have made outstanding contributions in this field. For example, Wang and Yao [23] presented a fuzzy Petri net method to assess the risk of air traffic control. Wan and Zhang [24] dealt with the certainty and uncertainty of the assessment system as a whole and establish a risk assessment model based on game theory and set pair analysis (SPA). Yuan et al. [25] believed that there are a lot of uncertainties in air traffic control safety evaluation, such as randomness, imprecision, and ambiguity. They introduced the Dempster combination rule to improve it and proposed an ATC safety assessment method based on evidence theory. Yao et al. [26] adopted a fuzzy Petri net and introduced the risk level threshold and analytic hierarchy process to reduce the complexity of the fuzzy Petri net used for ATC safety assessment. Wang and Sun [27] used system theory process analysis (STPA) to identify potential unsafe behaviors of ATC operation system and then used first-order linear temporal logic (LTL) to normalize the identified unsafe behaviors. Finally, they proposed a safety assessment method for the unsafe behaviors. Liao et al. [28], from the perspective of probability theory, proposed a safety probability evaluation method for air traffic control based on Bayesian analysis.

Through a comprehensive analysis about these literature studies, it can be found that most of the current research on unsafe event in the ATC department is focused on safety assessment or risk assessment. ATC safety assessment is mostly based on comprehensive evaluation, that is, taking the whole ATC safety as the research object, establishing corresponding evaluation index, setting the weight of each evaluation index, scoring each evaluation index, and judging which safety level the ATC safety status belongs to by combining the weight of each index and the score of the index on the premise of the given safety level. At present,

most methods of the ATC safety assessment are qualitative assessments on the safety level in the air traffic control department, and there are few quantitative assessments on the probability of unsafe event for some air traffic control department. However, in the field of ground transportation, the application of quantitative methods has been very extensive. For example, Lin et al. used hybrid deep learning model and generative adversarial networks to traffic incident detection [29, 30]. In the field of ground traffic incident detection, the use of various quantitative research methods is very common [31–33]. For the departments that do the work of air traffic control, reducing the percentage of unsafe event in their departments is the core task of safety management. If the relationship between the percentage of unsafe event and their influencing factors can be effectively clarified, then, the probability of unsafe event in some control department can be predicted. In order to quantitatively estimate the probability of unsafe event, a three-layer backpropagation neural network model is introduced in this paper. Considering the training accuracy of the BP neural network is not good when the outputs are small probability event, this paper introduces a corresponding model to improve it.

2. Materials and Methods

2.1. Artificial Neural Networks. An artificial neural network (ANN) is modeled on the brain where neurons are connected in complex patterns to process data from the senses, establish memories, and control the body. Artificial neural networks (ANNs) process data and exhibit some intelligence. It is the piece of a computing system designed to simulate the way the human brain analyzes and processes information [34]. It is the foundation of artificial intelligence (AI) and solves problems that would prove impossible or difficult by human or statistical standards. ANNs have self-learning capabilities that enable them to produce better results as more data become available. Warren McCulloch and Walter Pitts presented the first simple systems, which are the origins of artificial neural networks (ANNs) in the 1940s. They proved that an ANN can learn any arithmetic or logical function [35]. Artificial neural networks have been widely used in various industries and have achieved excellent results [36, 37].

2.2. General Backpropagation Neural Network. There are two types of artificial neural networks, shallow neural network and deep neural network. A shallow neural network has only one hidden layer of neurons that processes inputs and generates outputs. A deep neural network has two or more hidden layers of neurons that process inputs. According to experts [38], shallow neural networks can tackle equally complex problems. So, we use the shallow neural network to solve the problem of predicting probability of unsafe event in the air traffic control department.

Backpropagation is an algorithm that backpropagates the errors from the output nodes to the input nodes. It is the essence of neural net training. It is the method of fine-tuning the weights of a neural network based on the error rate

obtained in the previous iteration. It is a standard method of training artificial neural networks. This method helps to calculate the gradient of a loss function with respect to all the weights in the network. BP neural network is described as follows [39–41]. Given a shallow neural network has three layers of neurons that process inputs and generate outputs. this neural network has M inputs in input layer, N outputs in output layer, and K neurons (or nodes) in hidden layer. The input vector in input layer is $X = (x_1, x_m, \dots, x_M)$, the input vector in hidden layer is $HI = (hi_1, hi_k, \dots, hi_K)$, the output vector in hidden layer is $HO = (ho_1, ho_k, \dots, ho_K)$, the input vector in output layer is $YI = (yi_1, yi_n, \dots, yi_N)$, the output vector in output layer is $YO = (yo_1, yo_n, \dots, yo_N)$, the desired output vector is $t = (t_1, t_n, \dots, t_N)$, the connection weight from the m th node in the input layer to the k th node in the hidden layer is w_{mk} , and the connection weight from the k th node in the hidden layer to the n th node in the output layer is w_{kn} . The bias of each neuron in the hidden layer is b_k , the bias of each neuron in the output layer is b_n , the activation function of the hidden layer is $f_1(\cdot)$, and the activation function of the output layer is $f_2(\cdot)$.

The BP neural network uses the backpropagation algorithm to train networks. Backpropagation algorithm is commonly used in the training of artificial neural networks. The basic idea of the backpropagation algorithm is that the learning process consists of two processes: the forward propagation of the signal and the backward propagation of the error. In the process of forward propagation, the input signals pass from the input layer to the output layer after being handled in the hidden layer. If the actual outputs do not match the desired outputs, then, the process will turn into the backpropagation of error. The model of the BP neural network with one hidden layer is shown in Figure 1.

The input of the k th neuron in the hidden layer is as follows:

$$hi_k = \sum_{m=1}^M x_m w_{mk} + b_k. \quad (1)$$

The output of the k th neuron in the hidden layer is as follows:

$$ho_k = f_1(hi_k). \quad (2)$$

The input of the n th neuron in the output layer is as follows:

$$yi_n = \sum_{k=1}^N ho_k w_{kn} + b_n. \quad (3)$$

The output of the n th neuron in the output layer is as follows:

$$yo_n = f_2(yi_n). \quad (4)$$

The loss function E between the network outputs and the desired outputs is as follows:

$$E = \frac{1}{2} \sum_{n=1}^N (t_n - yo_n)^2. \quad (5)$$

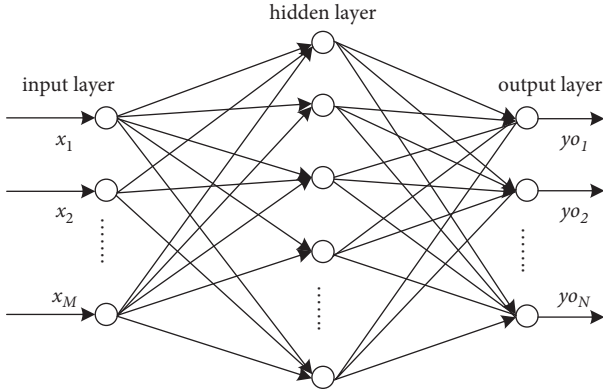


FIGURE 1: Top structure of the BP neural network.

During network training at a time, the change value of weight in each layer is obtained by gradient descent algorithm as follows:

$$\Delta w_{kn} = \eta h o_n f_2'(y i_n) (t_n - y o_n), \quad (6)$$

$$\Delta w_{mk} = \eta x_m f_1'(h i_k) \sum_{n=1}^N (t_n - y o_n) f_2'(y i_n) w_{kn}, \quad (7)$$

where Δw_{mk} represents the change value of weight from the m th node in the input layer to the k th node in the hidden layer. Δw_{kn} represents the change value of weight from the k th node in the hidden layer to the n th node in the output layer. η represents learning rate.

During network training at a time, the change value of biases obtained by gradient descent algorithm is as follows:

$$\Delta b_n = \eta f_2'(y i_n) (t_n - y o_n), \quad (8)$$

$$\Delta b_k = \eta f_1'(h i_k) \sum_{n=1}^N (t_n - y o_n) f_2'(y i_n) w_{kn}, \quad (9)$$

where Δb_n represents the change value of the n th bias in the output layer. Δb_k represents the change value of the k th bias in the hidden layer.

BP neural network is widely used in all walks of life because of its strong adaptability, including nonlinear mapping ability, self-learning ability, adaptive ability, generalization ability, fault tolerance ability, and other advantages [39, 42, 43]. At the same time, many scholars have improved the BP neural network considering its shortcomings and deficiencies, which improves the accuracy of the model [44–46]. This paper introduces the BP neural network into the probabilistic prediction of unsafe event in air traffic control department. For the problem of insufficient prediction accuracy of the general BP neural network, this paper proposes an improvement method.

2.3. Modeling Based on the Backpropagation Neural Network.

Three problems need to be solved in calculating the probability of ATC unsafe event by using BP neural network: determining the inputs, outputs, and activation function.

2.3.1. Probabilistic Representation of the ATC Event. According to the air traffic management rules of civil aviation in China, ATC unsafe event can be classified into five levels with respect to their severity: accident, serious incident, general incident, serious error, and general error. Each class of unsafe event is strictly defined in this document.

According to the probability theory, this article defines the elementary event as follows. Firstly, the elementary event e_0 is defined as safe event, e_1 represents accident event, e_2 represents the event of serious incident, e_3 represents the event of general incident, e_4 represents serious error event, and e_5 represents general error event. The probability space composed by the elementary event is $\Omega = \{e_0, e_1, e_2, \dots, e_5\}$. Ψ represents the set of all subsets of Ω , there are a total of 2^6 elements, Ψ is a σ domain, and each subset A in set Ψ is an event. The probability of A is $P(A)$, and then, the probability space of the ATC event can be expressed as (Ω, Ψ, P) . Defining the single-valued real function $X(e_i) = i$ (i is one of integer form 0 to 6 standing for different ATC event) in the probability space (Ω, Ψ, P) , then, X is the random variable about the ATC event. It is known that X is a discrete random variable, which can be seen from the value of X . The probability of different values of X can be expressed in the distribution law. In the set of elementary event defined in this paper, there is a special elementary event e_0 , which represents safe event. One important characteristic of probability is the normalization of probabilities, which means that the sum of the probability of all events is equal to one. If we do not consider the safe event e_0 , the normalization of the probability will not be satisfied, which will have a bad impact on the accuracy of the prediction.

2.3.2. Outputs of the Backpropagation Neural Network.

The outputs are usually the data that the model builder cares about. From the previously mentioned analysis, we are concerned with the probability of an unsafe event occurring in an ATC department. Therefore, the probability of unsafe event should be used as the outputs of the network. In addition to this, as previously mentioned, ensuring the normalization of the probability can improve the prediction accuracy, so the probability of safe event is also included in the outputs. Therefore, using the set of elementary event $\{e_0, e_1, e_2, \dots, e_5\}$ as the outputs of the network is the most straightforward way. Of course, the outputs can also be adjusted according to the needs of the research purpose. For example, if we only want to know the overall probability of all unsafe events and do not care about which class of unsafe event occurs, we can define event A as $\{e_1, e_2, \dots, e_5\}$. The outputs of the network are the probability of the event A and the probability of the safe event e_0 . That is to say, depending on the actual needs of the research problem, the outputs of the network can be any subset of the set Ω . It should be noted that, in order to satisfy the normalization of probability, the network's output event must be a partition of the set Ω .

2.3.3. Inputs of the Backpropagation Neural Network.

The inputs of BP neural network are usually some influencing factors that have an impact on the safety of air traffic

control, including the technical level of the controller, number of conflict points in the controlled airspace, and so on. The essence of determining the inputs is to establish the indicator system that affects the safety of ATC department. The establishment of input indicator system is relatively flexible, which usually needs to meet the following principles: the quantifiability of indicators, moderate number of indicators, and convenient data collection. There are many factors affecting the safety of ATC department, and the relationship among various influencing factors is intricate and complex, so it is very difficult to find the same influencing factors applicable to all different ATC departments. In the safety management of ATC, the analysis method most commonly used and widely recognized by the experts is SHELL model recommend by ICAO [23, 47], as described in Doc 9859 called Safety Management Manual. In this method, the factors affecting ATC safety are divided into four elements: hardware, liveware, environment, and management. The SHELL model provides a holistic overview of the influencing factors affecting ATC department. Therefore, in this paper, the hardware, liveware, environment, and management in the SHELL model are used as the inputs of the neural network. The next question is how to quantify these four inputs. There is no doubt that unsafe acts of human, equipment failures, management loophole, and objective problems in the environment may all lead to unsafe event. So, the number of unsafe acts of human in the ATC department can be used as input data for the liveware. The number of equipment failures in the ATC department can be used as input data for the hardware. The number of problems with management in the ATC department can be used as input data for the management. The number of problems with environment in the ATC department can be used as input data for the environment. Of course, the previously mentioned four inputs can be divided in depth according to the needs of the research. For example, the environment can be divided into the number of hazardous weather and the volume of flights. In short, the inputs of the neural network can be subdivided based on the previously mentioned four indicators in combination with the practical problems.

2.3.4. Activation Function of the Backpropagation Neural Network. BP neural network used in this paper is a three-layer structure, which needs two activation functions. Under normal circumstances, the choice of activation function needs to be determined by the actual problem. When used in probabilistic prediction, it has its particularity; that is, the outputs of the output layer must be in interval (0, 1). Therefore, the output layer chooses the asymmetric sigmoid function as activation function as follows, whose range is also in interval (0, 1):

$$f_2(x) = \frac{1}{(1 + e^{-x})}. \quad (10)$$

Considering that the domain of input in output layer may contain negative number, if the activation function of the hidden layer chooses the asymmetric sigmoid function, its output used as the input of the output layer will not cover

the domain. Therefore, the symmetric sigmoid function should be a better choice for the hidden layer as follows, whose range is in interval (-1, 1):

$$f_1(x) = \frac{(1 - e^{-x})}{(1 + e^{-x})}. \quad (11)$$

2.4. Improvement of the BP Neural Network. BP neural network has a good fault-tolerant ability and is widely used in all walks of life. The probability of unsafe event in air traffic control department has its particularity; if we directly apply the BP neural network to probabilistic prediction, it will have a large error in accuracy. So, it needs to be improved.

2.4.1. Problems of the General Backpropagation Neural Network. There are several problems in using general BP neural network directly to predict the probability of unsafe event in the ATC department.

During the training of the network, the outputs do not satisfy the normalization of the probability, which is that the sum of the probability of all network outputs is not equal to one. When the trained network is used for prediction, prediction accuracy may be affected.

There is more than one reason for the previously mentioned problem. For example, the inputs do not fully cover all the influencing factors, the quantization of the inputs is insufficient, and the number of training data is lacking. A few data in the training set is inaccurate, and so on. In short, BP neural network can only approximate a certain function, and errors are inevitable. So, if a constraint is introduced to ensure that the sum of all outputs is equal to one in each training epoch, the problem can be solved.

ATC unsafe events are small probability events. That is, the probability of an unsafe event in the ATC department is very low compared with the probability of safe event. When the connected weight of BP neural network is adjusted, the change of weight is mainly affected by the probability of safe event. This leads to the fact that the training of the whole network is dominated by the error of safe event, which makes the relative error of unsafe event very large. Next, we prove this by starting with the principle of BP neural network.

For the convenience of proof, the case is considered, in which only two outputs are in BP neural network. γo_n is used to represent output, $n = 2$ stands for safe event, and $n = 1$ stands for unsafe event. Since the unsafe event is a small probability event, this means that the desired probability of unsafe event is far less than the desired probability of safe event. For example, in 2020, the rate of unsafe event was 0.0056 per 10,000 flights in Chinese civil aviation. In 2019, the rate of unsafe event was 0.0043 per 10,000 flights in Chinese civil aviation. It is shown as follows:

$$t_1 \ll t_2, \quad (12)$$

where t_1 represents the desired probability of unsafe event and t_2 represents the desired probability of safe event.

The following is obtained from (5):

$$E = \frac{1}{2} \sum_{n=1}^2 (t_n - y_{o_n})^2 \quad (13)$$

$$= \frac{1}{2} [t_1^2 + t_2^2 + y_{o_1}^2 + y_{o_2}^2 - 2(t_1 y_{o_1} + t_2 y_{o_2})].$$

The following are obtained from (12):

$$t_1 \ll t_2 \Rightarrow t_1 t_1 \ll t_1 t_2, \quad (14)$$

$$t_1 \ll t_2 \Rightarrow t_2 t_1 \ll t_2 t_2. \quad (15)$$

The following is obtained from (14) and (15):

$$t_1^2 \ll t_2^2. \quad (16)$$

In the training process, when the actual outputs of BP neural network y_o are close to the target t , there is $y_{o_1} \ll y_{o_2}$, and the following can be proved, like (16):

$$y_{o_1}^2 \ll y_{o_2}^2. \quad (17)$$

The following apparently hold true:

$$t_1 \ll t_2 \Rightarrow t_1 y_{o_1} \ll t_2 y_{o_1}, \quad (18)$$

$$y_{o_1} \ll y_{o_2} \Rightarrow y_{o_1} t_2 \ll y_{o_2} t_2. \quad (19)$$

The following is obtained from (18) and (19):

$$t_1 y_{o_1} \ll y_{o_2} t_2. \quad (20)$$

The following can be obtained from (13), (16), (17), and (20):

$$E \approx \frac{1}{2} [t_2^2 - 2t_2 y_{o_2} + y_{o_2}^2] \quad (21)$$

$$\approx \frac{1}{2} (t_2 - y_{o_2})^2.$$

Bringing (21) into (7) and (9), the following can be obtained:

$$\Delta w_{mk} = \eta x_m f_1'(hi_k) (t_2 - y_{o_2}) f_2'(yi_2) w_{k2}, \quad (22)$$

$$\Delta b_k = \eta f_1'(hi_k) (t_2 - y_{o_2}) f_2'(yi_2) w_{k2}. \quad (23)$$

From (22) and (23), it is easy to find out that the change of weights from the input layer to the hidden layer is mainly determined by the probability of safe event. Therefore, when weight change is based on error, the result for safe event is better, but for unsafe event with small probability is worse.

The general processing method for this situation is to normalize the inputs and outputs, which map the inputs and outputs to the same interval through a function. This method can reduce the difference among the inputs and outputs, reduce errors, and improve the accuracy. After the prediction is made, the outputs can be probability through reverse normalization. However, this method still has following problems.

- (1) It still does not guarantee the normalization of probability, and the sum of the all outputs may not be equal to 1.
- (2) Different normalized function is going to produce different results. It is hard to say which normalized function is better. Different normalized functions are selected for the same model. This may result in some difference in accuracy. So, it is unclear which normalized function can be trusted.
- (3) Normalization destroys the mapping relationship between the original inputs and outputs to some extent. Even if the network training results are good, there may be some loss of precision when facing the new inputs.

For these problems, this paper proposes an improved method of BP neural network, which can not only meet the normalization of probability, but also improve the accuracy of the training of BP neural network.

2.4.2. Improvement of the Backpropagation Neural Network.

For a certain input vector, it is assumed that the q th output of the neural network is u_q . Since there is some error in the output, it is assumed that the error of the q th output is ε_q . Then, the true value of the q th output is $u_q + \varepsilon_q$. According to random utility theory [48–50], the probability of the q th output belonging to the q th class of event is as follows:

$$p_q = \text{Prob}(u_q + \varepsilon_q > u_z + \varepsilon_z, \forall z \neq q) \quad (24)$$

$$= \text{Prob}(\varepsilon_z < \varepsilon_q + u_z - u_q, \forall z \neq q),$$

where u_z stands for the z th output of the neural network, and ε_z stands for the error of the z th network output.

Assume that the error ε obeys an independent Gumbel distribution, and its distribution function is shown as follows:

$$F_n(\varepsilon_n) = e^{-e^{-\varepsilon_n}}. \quad (25)$$

Then, its joint distribution $F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ is as follows:

$$F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \prod_{n=1}^N e^{-e^{-\varepsilon_n}}. \quad (26)$$

Find the partial derivative of the q th random variable ε_q as follows:

$$\frac{\partial F}{\partial \varepsilon_q} = e^{-\varepsilon_q} \prod_{n=1}^N e^{-e^{-\varepsilon_n}}. \quad (27)$$

On substituting (24) into (27) to find its definite integral, the following can be obtained [51–53]:

$$\int_{-\infty}^{+\infty} e^{-\varepsilon_q} \prod_{n=1}^N e^{-e^{-(\varepsilon_n + u_z - u_n)}} d\varepsilon_q = \frac{\exp(u_q)}{\sum_{n=1}^N \exp(u_n)}. \quad (28)$$

According to the random utility theory, the q th desired output t_q^i corresponding to the i th input vector can be taken as the desired probability of ATC event. So, the following is valid:

$$\frac{\exp(u_q^i)}{\sum_{n=1}^N \exp(u_n^i)} = t_q^i. \quad (29)$$

If there are N outputs, then, there are N equations like (29), which form an equation set. u_q^i can be obtained by solving the equation set. t_q^i is replaced with u_q^i used for the desired outputs in the training of BP neural network. In the prediction, the network output u is brought into (29), and then, the probability of ATC event can be obtained.

Considering any two network outputs u_p^i and u_q^i corresponds to the i th input vector. According to (29), the following can be obtained:

$$\frac{\exp(u_p^i)}{\sum_{n=1}^N \exp(u_n^i)} = t_p^i, \quad (30)$$

$$\frac{\exp(u_q^i)}{\sum_{n=1}^N \exp(u_n^i)} = t_q^i. \quad (31)$$

Equation (30) divided by (31) yields the following:

$$\frac{\exp(u_p^i)}{\exp(u_q^i)} = \frac{t_p^i}{t_q^i}. \quad (32)$$

After transformation, the following is obtained:

$$\ln \left[\frac{\exp(u_p^i)}{\exp(u_q^i)} \right] = \ln \left(\frac{t_p^i}{t_q^i} \right). \quad (33)$$

Finally, we get the following:

$$u_p^i - u_q^i = \ln \left(\frac{t_p^i}{t_q^i} \right). \quad (34)$$

According to the properties of logarithmic function, when the difference between t_p^i and t_q^i is large, that is, the value of t_p^i/t_q^i is very large or very small, the value of $\ln(yo_p^i/yo_q^i)$ will not change too much, and the difference of $|u_p^i - u_q^i|$ will become smaller. For example, if $t_p^i/t_q^i = 100000$, then $\ln(100000) = 11.51293$. Therefore, according to the analysis in Section 2.4.1, replacing t^i with u^i in the training of the network will be more reasonable.

After the training of the network is completed, the i th sample is fed into the neural network and the p th output U_p^i is obtained. On bringing U_p^i into (29), the actual output of the i th input is obtained by the following:

$$T_q^i = \frac{\exp(U_q^i)}{\sum_{n=1}^N \exp(U_n^i)}. \quad (35)$$

Summing all the equations in (35) yields the following:

$$\sum_{q=1}^N T_q^i = \sum_{q=1}^N \frac{\exp(u_q^i)}{\sum_{n=1}^N \exp(u_n^i)} = \frac{\sum_{q=1}^N \exp(u_q^i)}{\sum_{n=1}^N \exp(u_n^i)} = 1. \quad (36)$$

So, the model can satisfy the normalization of probability.

In summary, the outputs of neural network contain all the event which is the partition of probability space, so (36) holds without question.

$$\sum_{n=1}^N T_n^i = 1. \quad (37)$$

The equation set formed by (29) has a redundant equation, which can be seen from (36) and (37). N unknown variables cannot be solved by $N - 1$ equations. Just let the one u_p^i be zero, and this problem can be solved.

It is difficult to solve the equation set formed by (29) directly. The method of element changing is introduced. Let $x_q^i = \exp(u_q^i)$, and then, (29) is transformed into the following:

$$\frac{x_q^i}{\sum_{n=1}^N x_n^i} = t_q^i. \quad (38)$$

If let $u_N^i = 0$, so $x_N^i = 1$. Equation (38) is transformed to the following:

$$\frac{x_q^i}{(\sum_{n=1}^{N-1} x_n^i + 1)} = t_q^i. \quad (39)$$

The equation set formed by (39) is the linear equations set. Solving it is very simple. After obtaining the solution to the equation set, the u_q^i can be obtained by the following:

$$u_q^i = \ln(x_q^i). \quad (40)$$

2.4.3. Adjustment of the Improved Backpropagation Neural Network Structure. Since the network output value in the improved BP neural network u_p^i does not need to be limited in this interval (0, 1), the activation function of the output layer $f_2(\cdot)$ can use the linear transfer function as follows to simplify the problem:

$$f_2(x) = x. \quad (41)$$

Then, the change value of weight in the output and hidden layers is obtained by gradient descent algorithm is as follows:

$$\Delta w_{kn} = \eta(t_n - yo_n)ho_n, \quad (42)$$

$$\Delta v_{mk} = \eta x_m f_1'(hi_k) \sum_{n=1}^N (t_n - yo_n) w_{kn}. \quad (43)$$

Similarly, during network training at a time, the change value of biases in the output and hidden layers obtained by the gradient descent algorithm is as follows:

$$\Delta b_n = \eta(t_n - yo_n), \quad (44)$$

$$\Delta b_k = \eta f_1'(hi_k) \sum_{n=1}^N (t_n - yo_n) w_{kn}. \quad (45)$$

2.5. Scope of Application of the Model. One of the core tasks of air traffic control in civil aviation is to ensure the safety of aircraft and avoid unsafe event. Unfortunately, Murphy's law tells us that absolute security does not exist. That is, it is difficult to succeed in completely avoiding unsafe event. Since it cannot be completely avoided, it must be scientifically confronted. For the department performing air traffic control work, it is very meaningful for air traffic safety management if it can predict the possibility of some unsafe event occurring. In this paper, BP neural network model is introduced into the probability prediction of unsafe event in air traffic control department. There are at least three practical applications:

- (1) The improved model in this paper provides a theoretical approach to quantify the probability of unsafe event in ATC department.

As mentioned previously, expert assessment method is mostly used in air traffic control safety research in China at present. In this class of safety assessment method, experts assess the safety situation of air traffic control department based on their own knowledge and experience. There is no doubt that this assessment method plays a positive role in the safety management of air traffic control. However, this kind of evaluation method is mainly based on the qualitative evaluation of experts but has no enough quantitative analysis. At the same time, it is based on the subjective evaluation of experts and has no enough objectivity. In this paper, the neural network model is introduced into the prediction of unsafe event in the ATC department. The historical data are used as the basis to make the evaluation result more objective. The neural network is also a quantitative mathematical model, which is more objective than the qualitative evaluation.

- (2) It can be used for the trend analysis of ATC safety development.

For the air traffic control department, the development trend of its risk has a significant impact on air traffic safety management. If the safety risk is increasing, the ATC department needs to invest more manpower, material, and financial resources into the safety management. If the safety risk is decreasing, the investment in safety management can be appropriately reduced according to the actual situation for reducing the cost. The changing trend of the rate of unsafe event reflects the development trend of safety risks to a certain extent. After predicting the probability of unsafe event in the future and combining the past and present rate of unsafe event, the safety risk trend of the ATC department can be judged, and the decision for the safety management of the ATC department can be made.

- (3) It provides a method for calculating probabilities for risk assessment.

Risk assessment is a common method for civil aviation safety management. The "Safety Management

Manual" prepared and published by International Civil Aviation Organization clearly defines risk as the product of the probability of an unsafe event and the consequences of the unsafe event. The Air Traffic Management Bureau of the Civil Aviation Administration of China has included this method in "Rules for Safety Management of Air Traffic Management Operation Department of Civil Aviation," However, neither the "Safety Management Manual" issued by the International Civil Aviation Organization nor "Rules for Safety Management of Air Traffic Management Operation Department of Civil Aviation," issued by the Civil Aviation Administration of China, give a specific method for calculating the probability of unsafe event. The model proposed in this paper provides a probability calculation method for the risk assessment for ATC department.

3. Numerical Example

The following is a numerical example to illustrate the advantages of the improved BP neural network proposed in this paper compared with other neural networks in the network training about unsafe event in air traffic control department.

3.1. Data. Suppose an air traffic control department wants to know the probability of unsafe event occurring in the next year. As analyzed in Section 2.3.3, four indicators of hardware, liveware, environment, and management related to SHELL model are used as inputs in BP neural network. Taking the unsafe event and safe event as outputs, the number of problems with management in the ATC department is used to quantify the management indicator. The number of problems with environment in the ATC department is used to quantify the environment indicator. The number of unsafe acts of human in the ATC department is used to quantify the liveware indicator. The number of equipment failures in the ATC department is used to quantify the hardware indicator. The occurrence rate of unsafe event is used to quantify the desired output of unsafe event. The occurrence rate of safe event is used to quantify the desired output of safe event. The historical data is as shown in Table 1, in which A_1 represents the safe event, $t(A_1)$ is the desired output of safe event, A_2 represents unsafe event, and $t(A_2)$ is the desired output of unsafe event. The historical data for numerical example is shown in Table 1.

3.2. Modeling. To prove that the model proposed in this paper can improve the prediction accuracy, three types of BP neural network are applied, which are called benchmark model and improved BP neural network. The benchmark model includes general BP neural network and normalized BP neural network. If the accuracy of the improved neural network is higher than the accuracy of the benchmark model used for comparison, it means that the improved neural network is meaningful. Therefore, we need to train the three

TABLE 1: Historical data for the numerical example.

No.	Liveware	Software	Hardware	Environment	$t(A_1)$	$t(A_2)$	$t(A_1) + t(A_2)$
1	2	8	8	8	0.999095	0.000905	1
2	2	1	7	10	0.999328	0.000672	1
3	1	6	7	10	0.999206	0.000794	1
4	4	10	4	2	0.999151	0.000849	1
5	8	5	10	10	0.999601	0.000399	1
6	9	1	3	5	0.999639	0.000361	1
7	5	3	2	7	0.999534	0.000466	1
8	5	8	3	4	0.999025	0.000975	1
9	9	4	5	5	0.999839	0.000161	1

different neural networks separately and compare the accuracy of the training results.

3.2.1. Benchmark Models

(1) *General BP Neural Network.* The principle of general BP neural network is shown in Section 2.2. The inputs and outputs are directly used to network training without any processing in the general BP neural network. To make the presentation clearer, the structure of the general BP neural network is described as follows. The prediction of probability of unsafe event in control department is essentially data fitting, by finding out inherent and opaque connections between the inputs and outputs through historical data. There is a consensus that the three-layer neural networks can fit most of the problems, so this model uses three-layer neural networks, which contain the input, hidden, and output layers:

(1) The input layer

Any safety-related influences can be used as inputs to the neural network, but these influences must be able to be quantified, meaning that these input data can be collected in practice. For a specific ATC department, as detailed in Section 2.3.3, the inputs of neural network can be found from the perspective of equipment, environment, personnel, and management according to SHEL model. At the same time, considering the collectability of the data, in this case, the following four indicators are used as inputs to the neural network, which are the number of equipment failures in the ATC department, number of equipment failures in the ATC department, number of problems related to management in the ATC department, and number of unsafe acts of human in the ATC department. It needs to be stressed again that, for the purpose of comparison, the input and output data are used directly to network training without any preprocessing.

(2) The hidden layer

The hidden layer needs to solve two problems: one is to determine the activation function and the other is to determine the number of nodes in the hidden layer. Considering that the domain of input in output layer may contain negative number, if the activation function of the hidden layer chooses the asymmetric sigmoid function, its output also used as the input of

the output layer will not cover its domain. Therefore, the symmetric sigmoid function should be a better choice for the hidden layer as follows, whose range is in interval $(-1, 1)$:

$$f_1(x) = \frac{(1 - e^{-x})}{(1 + e^{-x})}. \quad (46)$$

There is no method that everyone agrees on for figuring out the number of hidden layer nodes. Some experts have introduced an empirical formula as follows [54]:

$$K = \sqrt{M + N} + \alpha, \quad (47)$$

where K stands for the number in the hidden layer, M stands for the number in the input layer, and N stands for the number in the output layer, $\alpha \in [0, 10]$.

(3) The output layer

For the output layer, two problems need to be solved: one is to determine the output indicator and the other is the need to determine the activation function. There are some limitations on the output. First, the range of each output is in interval $[0, 1]$. Second, the outputs have to be the partition of the probability space. In order to meet the previously mentioned conditions, and taking into account the actual context of this case, there are two outputs, which are the probability of unsafe event and safe event. Since it is impossible to know the probability of unsafe event in advance, the occurrence rate of historical event can be used as the probability of ATC event for network training. The output layer chooses the asymmetric sigmoid function as activation function in the following, whose range is also in interval $(0, 1)$, which is also the interval of the probability:

$$f_2(x) = \frac{1}{(1 + e^{-x})}. \quad (48)$$

(2) *Normalized Neural Network.* The inputs, outputs, and number of nodes in the hidden layer of the normalized neural network are the same as those of the general neural network. In other words, the structure of the two networks is the same; the difference is that the normalized neural network needs to preprocess the input and output data before training.

The normalized neural network uses the following to normalize the inputs and outputs to the interval $[-1, +1]$ based on the general BP neural network:

$$y = \frac{2(x - x_{\min})}{(x_{\max} - x_{\min})} - 1, \quad (49)$$

where y stands for the data used for training after normalization, x_{\max} stands for the maximum value of an input or a output vector, and x_{\min} stands for the minimum value of an input or a output vector. The normalized input and desired output data are shown in Table 2.

Because the outputs are normalized to the interval $[-1, +1]$, the activation function of the output layer for the normalized neural network cannot use the sigmoid function whose range is in interval $(0, 1)$. So, the linear function is used for activation function of the output layer as follows:

$$f_2(\cdot) = x. \quad (50)$$

In the normalized neural network model, the data used for training are normalized, and reverse normalization is required when making predictions or testing training accuracy. The reverse normalization formula is shown as follows:

$$x = \frac{(y + 1) \times (x_{\max} - x_{\min})}{2} + x_{\min}. \quad (51)$$

The parameters in (51) can be referred to (49).

3.2.2. Improved BP Neural Network. The principle of improved BP neural network is shown in Section 2.4.3. In order to make the use of the improved neural network more clearly, the features of the improved neural network are summarized as follows:

(1) Defining the network structure

The purpose of defining the network structure is mainly to determine the number of inputs, outputs, and hidden layers of the network. For this case, the network structure of the three models is defined to be the same to make the comparison results more fairly. Details of the network structure can be found in Section 3.2.1.

(2) Output data preprocessing

First, let $u_2 = 0$. According to (38)–(40), u_1 can be calculated. See the third column data in Table 3. The network is trained with u_2 and u_1 as the target outputs of the neural network. Of course, to improve the speed of network training, we can also normalize the input and output of the network again for improved BP neural network.

(3) Network training

Network training involves the adjustment of weights and biases. The adjustment of weights and bias in the improved neural network are shown in (42) to (45).

(4) Probability of obtaining output

In the training process of the network, the original output data is not used. After the network training is

TABLE 2: Normalized input and desired output data.

Liveware	Software	Hardware	Environment	$t'(A_1)$	$t'(A_2)$
-0.75	0.556	0.5	0.5	-0.828	0.828
-0.75	-1	0.25	1	-0.256	0.256
-1	0.111	0.25	1	-0.555	0.555
-0.25	1	-0.5	-1	-0.69	0.69
0.75	-0.111	1	1	0.415	-0.415
1	-1	-0.75	-0.25	0.509	-0.509
0	-0.556	-1	0.25	0.251	-0.251
0	0.556	-0.75	-0.5	-1	1
1	-0.333	-0.25	-0.25	1	-1

completed, either to make predictions or calculate the accuracy of the training results, the network output needs to be converted to probability, and the conversion formula is shown in (29).

3.2.3. Model Calculation Process. The raw data can be divided into training set, validation set, and test set in the training of neural network. The training set is used for network training, the validation set is used for checking if the training process is overfitting, and the test set is used to compare. In this paper, because the samples are not many, all the samples are used for training. At the same time, the training set also is used for comparing which model is better.

There is the question of why not collect more samples. The reason is that the quantitative data on the factors affecting unsafe event are confidential to the public for air traffic control department. For example, no one will make public the number of unsafe acts caused by ATC controllers in their department. The reason is simple things like that is not honorable. So, it is very difficult to get many operation data. Taking the actual situation mentioned previously into account, the data used in this numerical example is not many. Therefore, all samples are put in the training set.

Another question is how to validate the network and compare the network, if there is no validation set and test set? Test set is used to compare the accuracy of different neural networks, after the network is trained. In the absence of a test set, the training set can be used to calculate the accuracy of the network, which may have some impact in actual use, but the purpose in this paper is to compare the accuracy of different networks and determine which neural network model is more accurate. We only need to care about the order of the different network accuracies, not too much about what the exact accuracy of the network is, so doing this will not have much impact. The validation set is mainly used to avoid overfitting; when the overfitting is detected, the network will stop training. In addition to overfitting, the conditions for stopping training the network can be the maximum number of epochs to train, performance goal, minimum gradient, maximum time to train, and so on. So, the network can train without validation set. Of course, overfitting cannot be ignored. To avoid the effect of overfitting on model comparison, this paper makes a detailed explanation in Sections 3.3 and 3.4.

TABLE 3: Analysis results of the improved BP neural network.

$t(A_1)$	$t(A_2)$	u_1	$p(A_1)$	$p(A_2)$	$e(A_1) \times 10^{-5}$	$e(A_2) \times 10^{-5}$	$\text{Re}(A_1) \times 10^{-5}$	$\text{Re}(A_2) \times 10^{-2}$	$p(A)$
0.999095	0.000905	7.00667	0.99913018	0.00086982	3.52	3.52	3.52	3.89	1
0.999328	0.000672	7.30458	0.99932105	0.00067895	0.69	0.69	0.69	1.03	1
0.999206	0.000794	7.137632	0.99917990	0.00082010	2.61	2.61	2.61	3.29	1
0.999151	0.000849	7.070602	0.99914621	0.00085379	0.48	0.48	0.48	0.56	1
0.999601	0.000399	7.82615	0.99960591	0.00039409	0.49	0.49	0.49	1.23	1
0.999639	0.000361	7.926272	0.99962773	0.00037227	1.13	1.13	1.13	3.12	1
0.999534	0.000466	7.670859	0.99955517	0.00044483	2.12	2.12	2.12	4.54	1
0.999025	0.000975	6.932098	0.99901470	0.00098530	1.03	1.03	1.03	1.06	1
0.999839	0.000161	8.733945	0.99983823	0.00016177	0.08	0.08	0.08	0.48	1

The whole process is classified into four stages:

Step 1. The training set is brought into the different models as the inputs and desired outputs for network training.

Step 2. After the training is completed, all inputs are brought into the trained network to calculate the network outputs. For the normalized neural network, use the normalized input data as inputs.

Step 3. The network outputs are the actual outputs for the general BP neural network, because the inputs and outputs are not transformed in any way in this neural network. The actual outputs for the normalized neural network are obtained, after the network outputs are processed by the reverse normalization. The actual outputs for the improved BP neural network are obtained by bringing the network outputs into (29).

Step 4. At last, the absolute error and the relative error between the actual outputs and the desired outputs is calculated. The advantages and disadvantages can be derived by comparing the errors of different kinds of BP neural network. For the normalized neural network, use the normalized input data as inputs.

For the fairness of comparison, the computing parameters of different models are set to be the same. That is, the maximum number of iterations is 1000, the learning rate is 0.01, and the number of nodes in hidden layer is $K=10$ according to (47), considering $M=4$, $N=2$, and $\alpha=7$.

3.3. Results and Analysis

3.3.1. Detailed Analysis through One Training Result. Tables 3–5 show the results of different BP neural network, where $p(\cdot)$ is the probability of the different event, which also is the actual outputs of neural network. It is obtained by feeding the inputs into the trained network and is then transformed by reverse normalization in improved BP neural network and normalized BP neural network. $e(\cdot)$ represents the absolute error between the actual outputs and desired outputs, and $\text{Re}(\cdot)$ represents the relative error between actual outputs and desired outputs. $p(A)$ is the sum of the probability of different event, which is used to check whether the normalization of probability is satisfied. The calculation formula of each parameter in the tables is as follows:

$$e(A_n) = |p(A_n) - t(A_n)|,$$

$$\text{Re}(A_n) = \frac{e(A_n)}{t(A_n)}, \quad (52)$$

$$p(A) = p(A_1) + p(A_2),$$

where A_1 stands for safe event, and A_2 stands for unsafe event.

The analysis result of the general BP neural network is shown in Table 4. The analysis result of the normalized BP neural network is shown in Table 5. The analysis result of the improved BP neural network is shown in Table 3:

- (1) From Table 3, it is easy to see that the sum of the actual outputs also used as the probability of ATC event is equal to one in the improved BP neural network. From Tables 4 and 5, it can be seen that whether outputs or inputs are normalized or not, general BP neural network and normalized BP neural network cannot guarantee that the sum of actual outputs is equal to one.
- (2) By comparing the error of the three models, it is easy to see that the general BP neural network model has the worst training precision, especially that the precision of unsafe event is very poor, and there are some relative errors reaching dozens or hundreds of times. Obviously, such model cannot be applied to application. It is easy to see that the improved model has the best precision after comparing the errors of different models.
- (3) In order to compare the accuracy of different models quantitatively on the whole, the mean square error is used to calculate the absolute total error AE and the relative total error RE of different models which is calculated as follows (the calculation results are shown in Table 6):

$$\text{AE} = \sqrt{\frac{\sum_{i=1}^9 \sum_{n=1}^2 [e^i(A_n)]^2}{18}}, \quad (53)$$

$$\text{RE} = \sqrt{\frac{\sum_{i=1}^9 \sum_{n=1}^2 [\text{Re}^i(A_n)]^2}{18}}. \quad (54)$$

TABLE 4: Analysis results of the general BP neural network.

$t(A_1)$	$t(A_2)$	$p(A_1)$	$p(A_2)$	$e(A_1) \times 10^{-2}$	$e(A_2) \times 10^{-2}$	$Re(A_1) \times 10^{-2}$	$Re(A_2) \times 10^0$	$p(A)$
0.999095	0.000905	0.97353244	0.024643473	-2.56	2.37	-0.03	26.23	0.99818
0.999328	0.000672	0.931128293	0.008308042	-6.82	0.76	-0.07	11.36	0.93944
0.999206	0.000794	0.963675363	0.017205835	-3.55	1.64	-0.04	20.67	0.98088
0.999151	0.000849	0.886911194	0.12358584	-11.22	12.27	-0.11	144.57	1.01050
0.999601	0.000399	0.968466684	0.015835696	-3.11	1.54	-0.03	38.69	0.98430
0.999639	0.000361	0.981705286	0.116324163	-1.79	11.60	-0.02	321.23	1.09803
0.999534	0.000466	0.936218516	0.0111588	-6.33	1.07	-0.06	22.95	0.94738
0.999025	0.000975	0.933914479	0.031134188	-6.51	3.02	-0.07	30.93	0.96505
0.999839	0.000161	0.940056061	0.009269256	-5.98	0.91	-0.06	56.57	0.94933

TABLE 5: Analysis results of the normalized BP neural network.

$t(A_1)$	$t(A_2)$	$p(A_1)$	$p(A_2)$	$e(A_1) \times 10^{-5}$	$e(A_2) \times 10^{-5}$	$Re(A_1) \times 10^{-5}$	$Re(A_2) \times 10^{-2}$	$p(A)$
0.999095	0.000905	0.99900973	0.00096991	8.53	6.49	8.53	7.17	0.999980
0.999328	0.000672	0.99931329	0.00065971	1.47	1.23	1.47	1.83	0.999973
0.999206	0.000794	0.99930812	0.00073200	10.21	6.20	10.22	7.81	1.000040
0.999151	0.000849	0.99908764	0.00085421	6.34	0.52	6.34	0.61	0.999942
0.999601	0.000399	0.99965091	0.00034770	4.99	5.13	4.99	12.86	0.999999
0.999639	0.000361	0.99976458	0.00024216	12.56	11.88	12.56	32.92	1.000007
0.999534	0.000466	0.99938241	0.00059731	15.16	13.13	15.17	28.18	0.999980
0.999025	0.000975	0.99926498	0.00081474	24.00	16.03	24.02	16.44	1.000080
0.999839	0.000161	0.99962508	0.00036466	21.39	20.37	21.40	126.50	0.999990

It is not difficult to see from Table 6 that the improved neural network model has the highest precision, because both the absolute total error and the relative total error of this model are minimal.

3.3.2. Comprehensive Analysis through 100 Training Sessions.

The previously mentioned analysis is based on the training results of each neural network at one time, and we can visualize the difference in the training accuracy of different neural networks from a numerical perspective. It is well known that the training process of neural networks is an optimal optimization process, in which the output error is minimized as the objective function and the network weights and bias are the decision variables. To prevent the neural network from falling into local optimum, when initializing the weights and bias, random initialization is used and multiple training will be performed to take the best one for the actual application. Therefore, the one-time training results cannot fully demonstrate the advantages of the improved model proposed in this paper in terms of training accuracy. To address this issue, we made 100 training sessions for each network and calculated their average error from 1 to 100 times, as shown in Figure 2. The black curve is the change of the average error of the improved neural network with the training sessions. The red curve is the change of the average error of the normalized neural network with the training sessions. The blue curve is the change of the average error of the general neural network with the training sessions. It can be easily seen from the figure that the average error of the improved neural network is the best at any time. It is enough to prove that the higher training accuracy of the improved neural network is not accidental.

TABLE 6: Precision analysis of different models.

General BP	Absolute total error (AE)	4.234648×10^{-2}
	Relative total error (RE)	8.547633×10^{-0}
Normalized BP	Absolute total error (AE)	1.238086×10^{-4}
	Relative total error (RE)	3.199931×10^{-1}
Improved BP	Absolute total error (AE)	1.730846×10^{-5}
	Relative total error (RE)	1.834107×10^{-2}

3.3.3. Excluding the Effect of Overfitting on Accuracy.

Overfitting is a potential pitfall in neural network training. Since the samples collected in this paper are not many, all samples are used for training. Inevitably, a question arises whether the improved neural network in this paper is overfit to make the accuracy higher than the accuracy of other benchmark models. To illustrate this question, first, we are clear that overfitting may be caused by an increase in the number of training epochs. Based on this consensus, we compare the accuracy of the three models under different number of training epochs one by one. At the early stage of training, the number of training epochs is low, and the possibility of overfitting is extremely low, at which time we compare the training accuracy, and the higher accuracy of the improved neural network due to overfitting can be excluded. To improve the fairness of comparison, the training parameters of the three different neural network models are set to be the same, the maximum epochs of training is 1000, the learning rate is 0.01, and the same gradient descent algorithm is used. At the same time, to avoid the influence of falling to the local optimum on the training results, all three models are initialized by randomly assigning weights and bias to the network. Considering the randomness of the single training results, the average error is

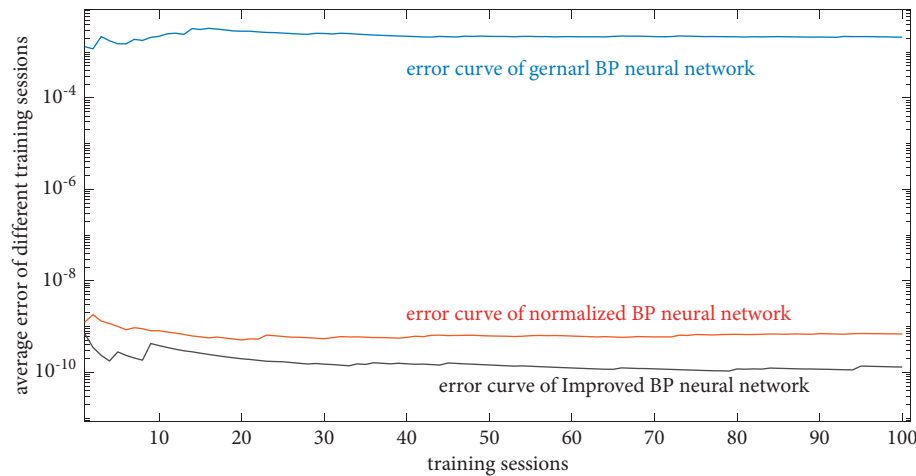


FIGURE 2: Average error of different networks with different training times.

taken for comparison after 100 independent training sessions. The training results are shown in Figure 3. The horizontal axis represents the 1000 training epochs, and the vertical axis represents the average error of 100 training sessions at each training epochs. The black curve is the change of the average error of the improved neural network after 100 training sessions with the number of training epochs. The red curve is the change of the average error of the normalized neural network after 100 training sessions with the training epochs. The blue curve is the change of the average error of the general neural network after 100 training sessions with the training epochs. It is not difficult to see from the figure that, at every training epoch, the accuracy of the improved neural network proposed in this paper is higher than that of the benchmark model, which excludes the possibility that the improved neural network has high accuracy due to overfitting.

3.4. Application of the Model. The trained BP neural network can be used for prediction of unsafe event in ATC department, by feeding new inputs to the network. For example, the safety managers in an air traffic control department have formulated the safety management objective in the next year, which requires that the number of unsafe acts of the controllers should not exceed five, the number of equipment failures should not exceed four, the number of management problems that cannot be solved in time should not exceed six, and the number of environmental problems should not exceed four. The question is what is the probability of an unsafe event, given that the above objectives can be achieved? The question can be answered using the model proposed in this paper. The solution process is described as follows.

In this paper, the entire samples are used for network training because of the limited samples, which is feasible for comparing the accuracy of different networks. However, when using the network for practical application, this approach may result in overfitting and lead to reduction in the generalization ability of the model. For addressing this

problem, the early stopping method can be used to avoid overfitting. The general practice is to first divide the data into three subsets. The first subset is the training set, which is used for computing the gradient and updating the network weights and biases. The second subset is the validation set. The error on the validation set is monitored during the training process. The validation error normally decreases during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error on the validation set typically begins to rise. The training stops when the error of the validation set is found to grow continuously in the training. The network weights and biases are saved at the minimum of the validation set error. The test set error is not used during training, but it is used to compare different model. Of course, the prerequisite is that there is enough data to ensure that there are enough samples in the different sets.

To illustrate the use of the network, the samples in Table 1 are divided into three sets, the training set, the validation set, and the test set. The training set contains five samples, whose number is 2, 6, 7, 8, and 9 in Table 1. The validation set contains two samples, whose number is 4 and 5 in Table 1. The test set contains two samples, whose number is 1 and 3 in Table 1. There are two points to clarify: one is that how the samples in the set are assigned is random, and the samples in the previously mentioned set is the outcome after the authors have divided samples to different set and trained the network many times to achieve the best result. The other is that this process is only to demonstrate the use of the model, since the samples are not many. The training results may not be very good. This is the reason why all the samples are input to training without dividing the sample into different set, when the network comparison was made in Section 3.3.

The training parameters are as follows: the learning rate is 0.01, the maximum number of epochs to train is 1000, the performance goal of training set is zero, and the maximum validation failures of validation set is six. After performing many training sessions, a network with better results was selected for illustration. The performance curves for different sets of improved BP neural network are shown in Figure 4.

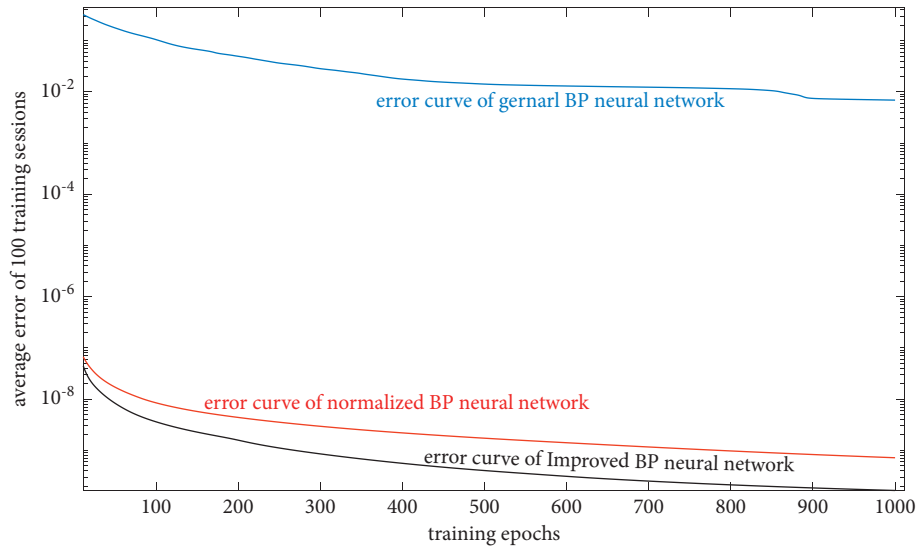


FIGURE 3: Average error of 100 training sessions of different neural network regarding training epochs.

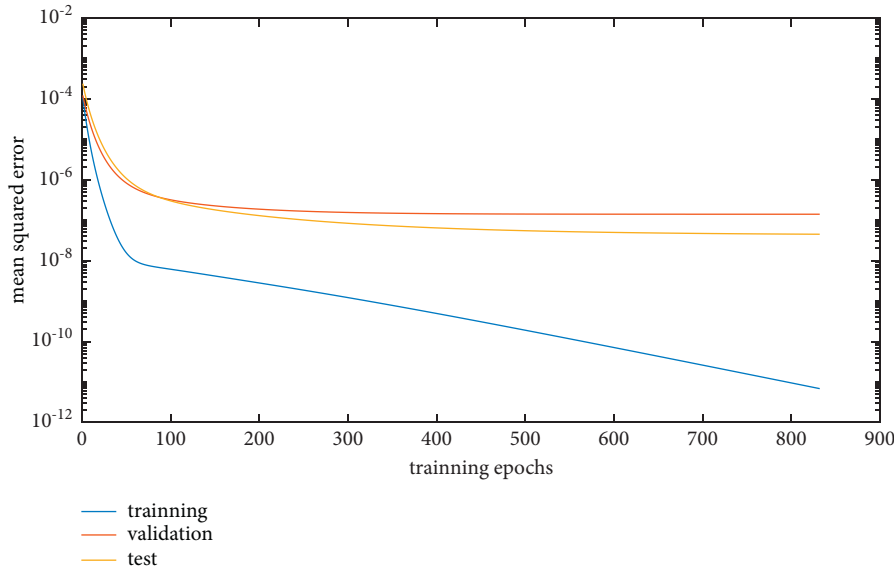


FIGURE 4: Performance curves for different sets of improved BP neural network.

In this training process, with the increase of training epochs, the errors of different sets are decreasing, and the decrease rate is smaller and smaller, which is fully consistent with the properties of gradient descent method. When the training epoch reaches 831, the error in the validation set was increasing for six consecutive times (the increase is small and the curve in Figure 4 is not clear) and reaches the maximum validation failure set by user. At this time, if the training continues, overfitting will occur, so the network stopped training. Obviously, the optimal network is obtained at 825 training epoch, so the network at this time can be used for practical applications.

Once the network has been trained, the question proposed in the first paragraph in this section on how to apply the trained network and known conditions to predict the probability of an unsafe event should be the

focus. In fact, when the network has been trained, the problem becomes very simple. The probability can be predicted by feeding this set of data [5, 4, 6, 4] into the network. The predicted result is that the probability of unsafe event is 0.0006093, and the probability of safe event is 0.9993907.

3.5. Summary

- (1) With its powerful ability of function fitting, BP neural network can be used to effectively predict the probability of unsafe event in air traffic control department, after reasonably designing the network structure and collecting historical data for training the neural network.

- (2) The general BP neural network cannot guarantee the accuracy of output, if it is directly applied to predict the probability of unsafe event in ATC department, and it cannot ensure normalization of probability. However, after normalization of inputs and outputs before training the network with the help of normalized BP neural network, the prediction accuracy is improved, but the normalization of probability for actual outputs is still difficult to satisfy.
- (3) The improved BP neural network proposed in this paper has high accuracy for the training of small probability event and can guarantee that the sum of the probability of all events is equal to one.

4. Conclusions

As an important means to ensure the safety of civil aviation, air traffic control plays a pivotal role in the protection of passengers' lives and property. To improve the level of safety management of the department which performs air traffic control work, the authors introduce artificial neural network to predict the probability of unsafe event for the air traffic control department. A three-layer neural network containing an input layer, hidden layer, and output layer is designed to solve the problem. The influence that affects ATC safety but can be quantified and collected is used as inputs according to SHEL model. The probability of unsafe event is used as the outputs of the network. The general BP neural network cannot be used for the network training of small probability event, which is proved theoretically in this paper, because the general BP neural network does not ensure that the sum of probability of all the outputs is equal to one and the error between the actual outputs and desired outputs is very large after the training of neural network. To address this issue, a new mapping method is put forward from the probabilistic viewpoint in this paper. It is theoretically proved that the mapping method proposed in this paper can not only improve the training accuracy, but also ensure that the sum of probability is equal to one. Finally, a case study demonstrates that the improved BP neural network model in this paper has a higher accuracy in predicting the probability of unsafe event in the air traffic control department.

However, one should note that this study also has two limitations, which should be improved in future research endeavors. The first limitation is that our current analysis is based on Chinese civil aviation. In particular, the outputs of the neural network are built on the air traffic management rules of civil aviation in China. So, if the model is used outside of China, the outputs of the network need to be modified according to the actual situation in other countries. The second limitation is that, like other artificial neural networks, the number of the nodes in the hidden layer has a great influence on the accuracy of the network, which is the common fault of artificial neural network models. In the actual application, the number of the nodes in the hidden layer needs to be adjusted several times to achieve good results.

BP neural network needs data for training. If a few data in the dataset are wrong, these wrong data will have side effects on network training. When a large amount of data is collected, the data should be preprocessed to remove the unreasonable data, which is very helpful to improve the accuracy of the network. This paper does not study the method of filtering data, and how to filter the data to remove some unreasonable data is the direction of future research.

Data Availability

Data used to support this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of the paper.

Acknowledgments

This research was supported by the General Project of Science and Technology Department of Sichuan Province (no. 2020YJ0500), General Project of Civil Aviation Flight University of China (nos. J2015-53, J2016-22, and J2019-105), Special Fund Project of Education and Teaching Reform of China (no. E2020028-5), and Open Fund of Key Laboratory of Flight Techniques and Flight Safety (grant no. FZ2020KF08).

References

- [1] P. Zhao, D. Lu, H. Hu et al., "Population-development oriented comprehensive modern transport system in China," *Acta Geographica Sinica*, vol. 75, no. 12, pp. 2699–2715, 2020.
- [2] C. Fu, "Evaluation on integrated transportation capacity and coordination in the urban agglomeration," *Journal of Transportation Systems Engineering and Information Technology*, vol. 17, no. 2, pp. 21–27, 2017.
- [3] M. Shang and L. Zhang, "Study on the optimization of comprehensive transportation system in China under the perspective of total factor efficiency," in *Proceedings of the 2nd International Conference on Advances in Energy Resources and Environment Engineering*, Sapporo, Japan, August 2017.
- [4] J. Jia and Y. Fang, "Underground space development in comprehensive transport hubs in China," in *Proceedings of the 15th International Scientific Conference Underground Urbanisation as a Prerequisite for Sustainable Development*, St. Petersburg, Russia, September 2016.
- [5] N. Wang and W. Li, "Thoughts on China's comprehensive transportation system under financial crisis," in *Proceedings of the 5th Advanced Forum on Transportation of China*, Beijing, China, October 2009.
- [6] Carnoc, "The work of China's civil aviation in 2019," 2020, <http://news.carnoc.com/list/518/518087.html>.
- [7] S. A. Gómez, A. Goron, A. Groza, and I. A. Letia, "Assuring safety in air traffic control systems with argumentation and model checking," *Expert Systems with Applications*, vol. 44, pp. 367–385, 2016.
- [8] J. Tang and B. Dong, "Safety risk evaluation of air traffic management system based on combination weights method

- and matter element model,” *Advances in Aeronautical Science and Engineering*, vol. 12, no. 1, pp. 30–38, 2021.
- [9] H. H. De Jong, S. H. Stroeve, and H. A. P. Blom, “The roles of air traffic controllers and pilots in safety risk analysis,” in *Proceedings of the European Safety and Reliability Conference 2006*, Estoril, Portugal, September 2006.
- [10] J. Guo, C. Pan, and Y. Liao, “A method for unsafe event prediction of air traffic management (ATM) based on GBO-Markov model,” *Journal of Safety and Environment*, vol. 16, no. 4, pp. 48–51, 2016.
- [11] Z. Zhang and W. Li, “Study on risk control network model of approach control system,” *China Safety Science Journal*, vol. 29, no. 2, pp. 32–37, 2019.
- [12] Y. Zhao and J. Wan, “On the operational risk assessment of the air traffic control based on the set pair analysis method,” *Journal of Safety and Environment*, vol. 18, no. 3, pp. 871–875, 2018.
- [13] Y. Zhao and J. Wan, “An operational risk assessment model for air traffic control based on the set pair extension coupling method,” *Journal of Safety and Environment*, vol. 18, no. 6, pp. 2059–2063, 2018.
- [14] Y. Zhang, X. Wang, M. Wu et al., “Evaluation of the unconventional operational risks for the air traffic control based on the fuzzy hierarchical analysis process and the cloud model,” *Journal of Safety and Environment*, vol. 16, no. 4, pp. 42–47, 2016.
- [15] F. Chen, Q. Chen, and B. Xu, “Analysis on operation risk topics of air traffic control (ATC) based on text mining,” *Journal of Safety Science and Technology*, vol. 16, no. 11, pp. 47–52, 2020.
- [16] Y. Sikirda, T. Shmelova, and D. Tkachenko, “Automated system for evaluation of the organizational risk factors influence on flight safety in air traffic control,” in *Proceedings of the 2018 IEEE 5th International Conference on Methods and Systems of Navigation and Motion Control*, Kiev, Ukraine, October 2018.
- [17] K. Zhang, R. Sun, W. Feng et al., “Discussion on construction of safety integrity system in air traffic control,” in *Proceedings of the 6th International Symposium on Project Management*, Chongqing, China, July 2018.
- [18] J. A. Adam, “Air traffic control: satellites and artificial intelligence promise improved safety and efficiency,” *IEEE Spectrum*, vol. 28, no. 2, pp. 27–32, 1991.
- [19] V. D. Hopkin, “Safety and human error in automated air traffic control,” *IEEE Conference Publication*, vol. 463, pp. 113–118, 1999.
- [20] J. Zhang, W. Wang, Y. Liu et al., “Establishment of the safety culture in civil aviation air traffic control,” in *Proceedings of the 12th COTA International Conference of Transportation Professionals*, Beijing, China, August 2012.
- [21] Z. Bingxiang and L. Jimin, “Research on safety level assessment of air traffic control,” in *Proceedings of the 2010 International Conference on Computer and Communication Technologies in Agriculture Engineering*, Chengdu, China, June 2010.
- [22] Z. Guoliang and Y. Shaohua, “Evolution of safety management in China’s air traffic control safety risk information management,” in *Proceedings of the IFA 34th International Conference, and IATA-Sharing Knowledge to Improve Safety*, Shanghai, China, August 2004.
- [23] Q. Wang and D. Yao, “A new method for evaluating air traffic control safety,” in *Proceedings of the 2017 6th International Conference on Network, Communication and Computing*, Kunming, China, December 2017.
- [24] J. Wan and M. Zhang, “Evaluation of air traffic control risk situation based on game theory and set pair analysis,” *China Safety Science Journal*, vol. 28, no. 11, pp. 142–148, 2018.
- [25] D. Yuan, Y. Lei, and N. Yang, “Approach to the aviation and air traffic and transportation safety evaluation based on the Dempster’s combination rule,” *Journal of Safety and Environment*, vol. 18, no. 3, pp. 1014–1019, 2018.
- [26] D. Yao, Q. Wang, and X. Gan, “Safety risk assessment on the air traffic control via the improved fuzzy petri net,” *Journal of Safety and Environment*, vol. 18, no. 2, pp. 413–417, 2018.
- [27] J. Wang and X. Sun, “Research on safety analysis of air traffic control system based on STPA,” *Journal of Wuhan University of Technology*, vol. 39, no. 12, pp. 49–55, 2017.
- [28] Y. Liao, J. Guo, and C. Yang, “Probability assessment of ATC safety based on Bayes analysis,” *China Sciencepaper*, vol. 10, no. 19, pp. 2271–2275, 2015.
- [29] Y. Lin, L. Li, H. Jing, B. Ran, and D. Sun, “Automated traffic incident detection with a smaller dataset based on generative adversarial networks,” *Accident Analysis & Prevention*, vol. 144, Article ID 105628, 2020.
- [30] L. Li, Y. Lin, B. Du, F. Yang, and B. Ran, “Real-time traffic incident detection based on a hybrid deep learning model,” *Transportmetrica A: Transportation Science*, vol. 16, no. 4, pp. 1–21, 2020.
- [31] P. Mercader and J. Haddad, “Automatic incident detection on freeways based on Bluetooth traffic monitoring,” *Accident Analysis & Prevention*, vol. 146, Article ID 105703, 2020.
- [32] J. D. G. Paule, Y. Sun, and Y. Moshfeghi, “On fine-grained geolocalisation of tweets and real-time traffic incident detection,” *Information Processing & Management*, vol. 56, no. 3, pp. 1119–1132, 2019.
- [33] J. Lu, S. Chen, W. Wang, and B. Ran, “Automatic traffic incident detection based on nFOIL,” *Expert Systems with Applications*, vol. 39, no. 7, pp. 6547–6556, 2012.
- [34] X. Zhou, T. Chen, and T. Qiu, “BP neural network forecast of flight “estimated final arrival time” based on Levenberg-Marquardt algorithm optimization,” in *Proceedings of the 2020 5th International Conference on Electromechanical Control Technology and Transportation*, Nanchang, China, May 2020.
- [35] W. S. McCulloch, “A heterarchy of values determined by the topology of nervous nets,” *Bulletin of Mathematical Biophysics*, vol. 7, no. 2, pp. 89–93, 1945.
- [36] Z. Wu, S. Tian, and L. Ma, “A 4D trajectory prediction model based on the BP neural network,” *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1545–1557, 2020.
- [37] Y. Yang, L. Chen, Y. Xiong, S. Li, and X. Meng, “Global sensitivity analysis based on bp neural network for thermal design parameters,” *Journal of Thermophysics and Heat Transfer*, vol. 35, no. 1, pp. 187–199, 2021.
- [38] I. J. Goodfellow, D. Erhan, P. Luc Carrier et al., “Challenges in representation learning: a report on three machine learning contests,” *Neural Networks*, vol. 64, pp. 59–63, 2015.
- [39] L. Xu, T. Quan, J. Wang et al., “GR and BP neural network-based performance prediction of dual-antenna mobile communication networks,” *Computer Networks*, vol. 172, Article ID 107172, 2020.
- [40] L. Wang and X. Bi, “Risk assessment of knowledge fusion in an innovation ecosystem based on a GA-BP neural network,” *Cognitive Systems Research*, vol. 66, pp. 201–210, 2021.
- [41] Y. Guo, “Credit risk assessment of P2P lending platform towards big data based on BP neural network,” *Journal of Visual Communication and Image Representation*, vol. 71, Article ID 102730, 2020.

- [42] M. Zou, L. Xue, H. Gai et al., "Identification of the shear parameters for lunar regolith based on a GA-BP neural network," *Journal of Terramechanics*, vol. 89, pp. 21–29, 2020.
- [43] J. Hao, Y. Lin, G. Ren et al., "Comprehensive benefit evaluation of conservation tillage based on BP neural network in the Loess Plateau," *Soil and Tillage Research*, vol. 205, Article ID 104784, 2021.
- [44] Y. Zhou and S. Li, "BP neural network modeling with sensitivity analysis on monotonicity based spearman coefficient," *Chemometrics and Intelligent Laboratory Systems*, vol. 200, Article ID 103977, 2020.
- [45] M. Mirsadeghi, M. Shalchian, S. R. Kheradpisheh et al., "STiDi-BP: spike time displacement based error back-propagation in multilayer spiking neural networks," *Neuro-computing*, vol. 427, pp. 131–140, 2021.
- [46] Z. Yu, L. Qin, Y. Chen et al., "Stock price forecasting based on LLE-BP neural network model," *Physica A: Statistical Mechanics and Its Applications*, vol. 553, Article ID 124197, 2020.
- [47] A. Oztekin, *Development of a Safety Assessment Tool for Air Traffic Control System* SAE International, Warrendale, PA, USA, 2015.
- [48] M. Peşki, "Large roommate problem with non-transferable random utility," *Journal of Economic Theory*, vol. 168, pp. 432–471, 2017.
- [49] J. L. Lusk, J. M. Crespi, B. R. Mcfadden et al., "Neural antecedents of a random utility model," *Journal of Economic Behavior & Organization*, vol. 132, pp. 93–103, 2016.
- [50] M. S. Hasnine, A. Graovac, F. Camargo et al., "A random utility maximization (RUM) based measure of accessibility to transit: accurate capturing of the first-mile issue in urban transit," *Journal of Transport Geography*, vol. 74, pp. 313–320, 2019.
- [51] A. Papola, F. Tinessa, and V. Marzano, "Application of the combination of random utility models (CoRUM) to route choice," *Transportation Research Part B: Methodological*, vol. 111, pp. 304–326, 2018.
- [52] D. Hojo, K. Murono, H. Nozawa et al., "Utility of a three-dimensional printed pelvic model for lateral pelvic lymph node dissection education: a randomized controlled trial," *Journal of the American College of Surgeons*, vol. 229, no. 6, pp. 552–559, 2019.
- [53] K. J. Ray, P. Lalitha, N. V. Prajna et al., "The utility of repeat culture in fungal corneal ulcer management: a secondary analysis of the MUTT-I randomized clinical trial," *American Journal of Ophthalmology*, vol. 178, pp. 157–162, 2017.
- [54] C. Song, S. Xie, J. Wang et al., "Research on tower crane safety based on grey relation and BP neural network," *Computer Simulation*, vol. 36, no. 10, pp. 472–476, 2019.