Predicting the Reasons of Employee Turnover Based on BP Neural Network Model

Ying Dai*

College of Information Engineering, Liaoning University of Transitional Chinese Medicine, Shenyang, China, 110847

* Corresponding Author Email: dy030407@163.com

Abstract. Employee turnover is related to the survival and development of enterprise team, which increases the cost of human resources and causes management difficulties and economic losses. Bp neural network model can deal with nonlinear problems, large-scale data and complex pattern recognition tasks, and plays an important role in the field of prediction. In this paper, BP neural network model is used to deeply analyze the causes of employee turnover, and deeply predict the causes of complex employee turnover with economic development and post-epidemic era. Through training and learning of relevant data of employee turnover, BP neural network can extract potential patterns and rules of employee turnover, so as to predict the trend of employee turnover in the future. So that enterprises can better understand the demands and expectations of employees, formulate more reasonable policies and plans, improve the job satisfaction and loyalty of employees, and reduce the employee turnover rate. Provide guidance for enterprise management's decision-making and rectification.

Keywords: BP neural Network Model, Prediction Model, Employee Turnover.

1. Introduction

Recently, China Salary network released the latest national graduate turnover trend chart in 2023, revealing the phenomenon of high turnover willingness of young people in all walks of life. In fact, as early as 2019, China's employee turnover rate has reached a staggering 18.9 percent, a figure that is quite rare in the world. Even in Japan, where social pressure is high, the turnover rate is only 11.3 percent. The fact that the turnover rate is still so high in the face of a growing aging population undoubtedly reveals many deep-seated problems. Therefore, how to reduce employee turnover and retain professional talents has become an important issue that enterprises need to solve urgently.

From a macro point of view, factors such as industrial structure adjustment, increasing competitive pressure, employee turnover and the imbalance of labor supply and demand are all driving the increase of turnover rate. Therefore, it has become urgent to pay attention to China's talent training, promote employment and revitalize the economy.

The impact of employee turnover on enterprises is profound. First of all, it will bring a huge financial burden to the enterprise, because the enterprise needs to devote a lot of resources to training the new employees and bear the adjustment period for them in the new position. Secondly, a large turnover of employees may lead to the disruption of the original work schedule and the forced shelving or postponing of projects, which will undoubtedly reduce work efficiency. What's worse, the employees who leave may take important project information or customer data with them, which is a great loss to the enterprise. In addition, a large number of employees leaving will also have a negative impact on the corporate culture, destroying team morale and teamwork ability, and reducing the management efficiency and brand value of the company.

Due to the subjective nature for HR and expert analysis, they often cannot accurately judge the main reasons for employee turnover. Based on the above background, this paper will use BP neural network to deeply analyze the reasons of employee turnover and build a prediction model of employee turnover. Through this model, the management of enterprises can adjust the strategic measures early, including providing benefits to employees, reducing work pressure, improving the working environment and management system. These measures can help enterprises retain talents and lay a solid foundation for future development.

2. Literature Review

2.1. Application fields of bp neural network model

Bp neural network is applied in many fields, and it can play an obvious advantage in solving and predicting problems. Problems[1][2] surrounding the water environment. Other literature compares LSTM model with bp neural network model to construct daily flow[1] rate for the next 30 days. In order to build a model for predicting water environmental carrying capacity, time series prediction and bp neural network are organically combined[2]. The advantage is that the improved model solves the problems of nonlinearity, randomness and time variation in environmental statistical characteristics. The disadvantage is that the input value of bp neural network should not be too much, which will increase the deviation. bp neural network has also solved many problems[3][4][5] in the aspect of prediction. Predecessors used k-fold cross-validation to train the original data, momentum method and ADAM algorithm to optimize and improve the bp neural network model, all of which have greatly improved the accuracy of bp neural network training. Some literatures worth exploring all use the momentum method[4][5], but the final selection algorithm is different, indicating that the research topic should be discussed in detail. Different[6] from the previous literature, it does not optimize the bp neural network model to solve the problem, but hierarchical optimization of its topology. Grey model, genetic algorithm and ant colony algorithm are used to optimize the input layer, hidden layer and output layer respectively to reduce the error of prediction. However, it also brings the problem that the weight and threshold of the neural network are not optimized. To sum up, bp neural network is an important tool for researchers to make nonlinear predictions, with many types of applications and wide applicability.

2.2. Analysis of reasons for staff turnover

The problem of employee turnover is deeply concerned by the management of the company. How to retain employees in an enterprise and what factors cause the turnover of employees are all problems worth thinking about. Some literatures have proposed support vector machine (a binary classification model[7][8][9], the basic model of which is a linear classifier with the largest interval defined in the feature space, and the interval is the most important difference between it and perceptron), random forest (combining multiple decision trees, each data set is randomly selected and some features are randomly selected as input), XGBoost (is a kind of addition model based on boosting ensemble idea, which adopts forward distribution algorithm for greedy learning when training, learning a CART tree each iteration to fit the residual difference between the previous t-1 tree prediction results and the true value of the training sample). Logistic and a coupled model (ARMI-ANN) that combines the neural network model (ANN) with the traditional autoregressive single integral shift translation model (ARIMA) to analyze employee turnover factors. Through multiple comparison of various methods and secondary optimization, higher prediction accuracy is found, which makes the collected data set not wasted. From the psychological point of view, some literatures discuss the mediating role of job satisfaction and the moderating role[10] of employer brand. It is pointed out that there are few contents in the existing literature, which have a positive impact on employee career satisfaction, which is one of the important factors of employee turnover. However, these data are cross-sectional data, which may be biased from the actual data in order to avoid homology bias of the research data. Therefore, the current mainstream research focuses on the analysis of data sets, focusing on the employees themselves, without judging their impact from the perspective of the enterprise. There are more changes and prospects worth predicting in the future.

3. Data introduction and pre-processing

3.1. Data Introduction

In this paper, the HR Analytics Dataset of Kaggle data platform was used for data analysis, and the factors leading to employee turnover were identified from multiple perspectives. There are 34 sets of variables in this dataset, including 13 sets of quantitative variables and 21 sets of qualitative variables. There are many nonlinear data variables, and the input and output are a set of nonlinear relations.

3.2. Data Preprocessing

The first step is to delete variables that are not relevant to the research question. EmployeeID is a unique employee number and has no practical significance for prediction. Age indicates the employee's age, which has been grouped in the dataset, so it is a redundant variable. MonthlyIncome monthly income and SalarySlab payroll repeat with MonthlyRate monthly salary, retaining one; StandardHours is fixed and does not help in the prediction. Delete it.

The second step is to deal with continuous variables. For example, DistanceFromHome ranges from 1 to 29, group them (1 to 10) close together; (11-20), medium distance; (21-29), long distance. The AgeGroup is divided into the youth group AgeGroup (18-25), the young and middle-aged group AgeGroup1 (26-35), the prime age group AgeGroup2 (36-45), the middle age group AgeGroup3 (46-55) and the middle and old age group AgeGroup4 (55+). The processing of continuous variables can effectively eliminate the error caused by large data values.

The third step is to deal with discrete variables. Some ordered variables, such as whether the employee has left the company or not, and the gender of the employee, are dichotomized. Disordered variables such as work environment satisfaction, job engagement and relationship satisfaction are transformed into rank variables for quantitative analysis is shown as Table. 1.

Table 1. String encoding rule table

Variable Name	Encoding rules
Attrition	Yes:1; No:0
DistanceFromHome	Near:1; Medium:2; Remote:3
Gender	Male:1; Female:2
RelationshipSatisfaction	Dissatisfied:1; NotVerySatisfied:2; QuiteSatisfied:3; Satisfied:4
JobInvolvement	LowParticipation:1; NotHighPaticipation:2; QuiteHighPaticipation:3; HighPaticipation:4
JobSatisfaction	Dissatisfied:1; NotVerySatisfied:2; QuiteSatisfied:3; Satisfied:4
EnvironmentSatisfaction	Dissatisfied:1; NotVerySatisfied:2; QuiteSatisfied:3; Satisfied:4

4. Establishment of the model

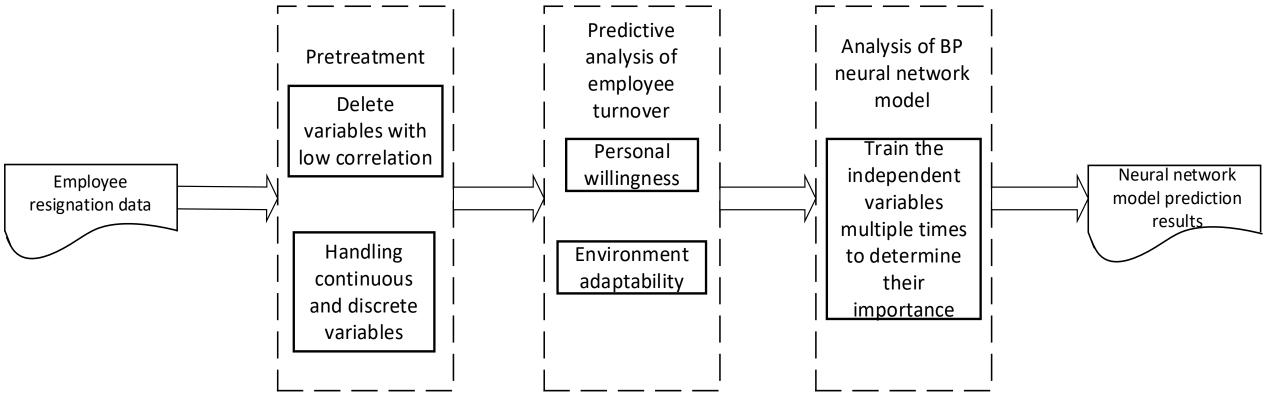

Figure 1. Training process of bp neural network model based on employee turnover

4.1. Introduction to the basic concepts of bp neural network

The BP neural network, which was introduced in 1986, is a multi-layer feedforward neural network. It utilizes an error reverse propagation algorithm for training purposes, where the weights of the network are adjusted in the opposite direction of the gradient to minimize errors. This type of neural network finds extensive applications across various fields such as function approximation, model recognition and classification, data compression, and time series prediction.

One distinguishing characteristic of the BP neural network lies in its ability to propagate signals forward and errors backward. During forward propagation, input data passes through each layer until it reaches the output layer. Neurons within the network process this input data and transmit it to subsequent layers. The final output is then compared with expected results to determine any discrepancies or errors.

In back propagation, these errors are propagated back through the network by adjusting neuron weights and thresholds accordingly. This adjustment occurs in a manner that minimizes error by moving against the gradient associated with weights and thresholds. This iterative process aims at reducing error levels to their minimum.

Another significant aspect of BP neural networks involves hidden units situated between input and output layers. Although these hidden units do not directly interact with external environments, changes in their states can greatly influence how inputs relate to outputs within the network.

Signal transmission within this type of neural network adheres strictly to rules governing both forward propagation (input movement) and reverse propagation (error movement). In essence, input data moves forward while errors move backward leading to necessary adjustments in neuron weights and thresholds is shown as Figure 1.

4.2. Establishment of bp neural network for employee turnover problem

Before the experiment, the data set was analyzed to predict and judge that the employee turnover of the enterprise has its own characteristics and environmental adaptability. First of all, we analyzed the employees' self-characteristics and grouped them by age: The AgeGroup can be divided into the youth group (18-25), the young and middle-aged group (26-35), the middle-aged group (36-45), the middle-aged group (46-55) and the old group (55+) is shown as Table. 2. The employees were grouped by their marital status as single, married and divorced. Through neural network training for these independent variables, the importance of independent variables shows that it is easier for young people aged 18-25 and single people to be distributed quit your job is shown as Table. 3.

Table 2. Importance analysis of independent variables for age distribution

Age Group	Importance	Normalization Importance
AgeGroup(18-25)	0.425	100%
AgeGroup1(26-35)	0.157	36.9%
AgeGroup2(36-45)	0.292	68.6%
AgeGroup3(46-55)	0.044	10.4%
AgeGroup4(55+)	0.082	19.2%

Table 3. Importance analysis of independent variables of marital status

MaritalStatus	Importance	Normalization Importance
Single	0.757	100%
Married	0.160	21.1%
Divorced	0.083	11.0%

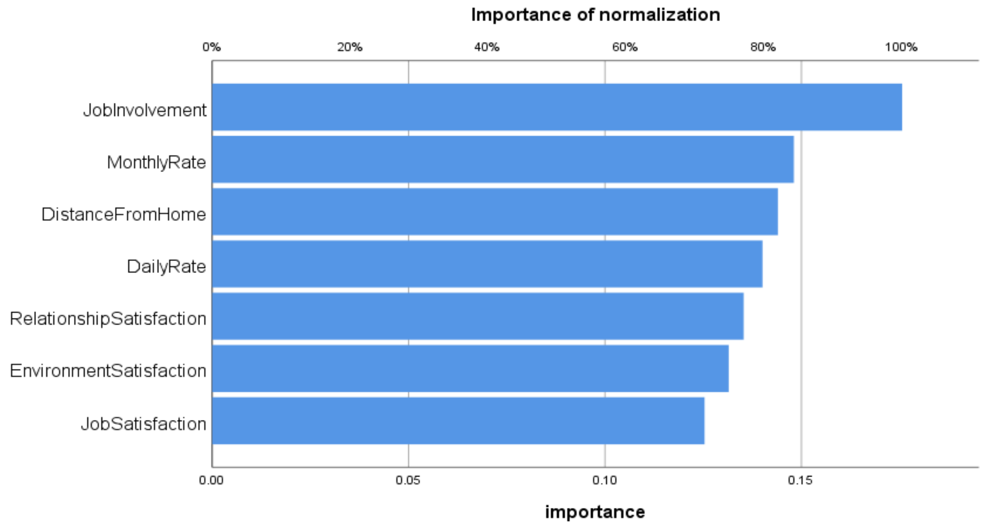

Figure 2. Normalized importance analysis of environmental treatment

Secondly, in order to judge the impact of the working environment and treatment on employee dimension, we selected seven key variables, namely daily salary, monthly salary, work environment satisfaction, job participation, job satisfaction, relationship satisfaction, and commuting distance, to predict employee dimension. From the predicted results shown as Figure 2, it can be seen that job participation satisfaction, commuting distance and monthly salary are the main factors for employee turnover.

4.2.1 Listing of structural details of bp neural network model

In the training of neural networks, a crucial foundational step involves computing the input and output for each layer of neurons. This necessitates a deep comprehension of the architecture and connection weights employed in these networks. At the onset of training, it becomes necessary to compute the input for each neuron by multiplying the previous layer's output with the current layer's weight, along with an added bias term. This process continues until reaching the output layer, resulting in a predicted network output is shown as figure 3.

Once we obtain this predicted network output, it can be compared to the desired output to calculate an error function that quantifies their disparity. The error function's partial derivative provides insights into how this error is distributed among neurons within the output layer. By leveraging these partial derivatives, we gain an understanding of each neuron's contribution to overall error and subsequently adjust connection weights accordingly.

In addition to focusing on adjustments within the output layer, it is also essential to calculate partial derivatives of the error function for every neuron residing in hidden layers. Armed with this information, we can modify hidden layer weights effectively and reduce errors. Throughout training, precise adjustment of connection weights holds significant importance as it directly impacts prediction accuracy within neural networks. Through continuous weight modifications, networks gradually acquire more accurate and valuable feature representations.

In addition to adjusting the weights, it is essential to utilize the output of neurons for optimizing network performance. Specifically, we can make use of the output from each neuron in both the hidden and output layers to rectify connection weights. This adjustment aids in faster convergence during learning and enhances prediction accuracy. Furthermore, leveraging input from the hidden layer and input layer further optimizes network learning by accurately capturing input data characteristics.

Throughout training, it is crucial to calculate global errors as a means of evaluating overall network performance and learning progress. Global error serves as a comprehensive indicator reflecting the network's forecasting ability holistically. By comparing global errors after different training iterations, we can assess whether there is gradual improvement within the network.

Ultimately, it becomes necessary to determine if the network error meets desired requirements. If the current network achieves satisfactory predictive accuracy, then we can conclude our algorithm. Otherwise, we proceed with selecting subsequent learning samples along with their corresponding expected outputs for further rounds of learning. This iterative process continues until optimal performance is attained within practical applications is shown as figure 4.

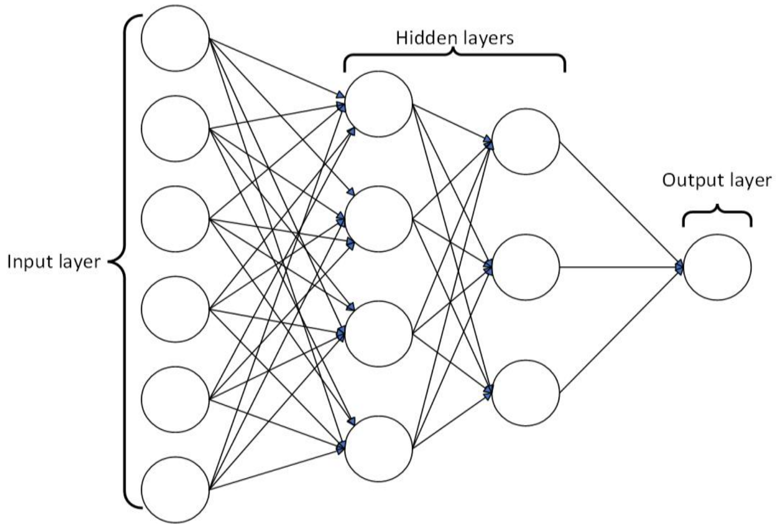

Figure 3. Topology of bp neural network

4.2.2 Algorithm flow chart

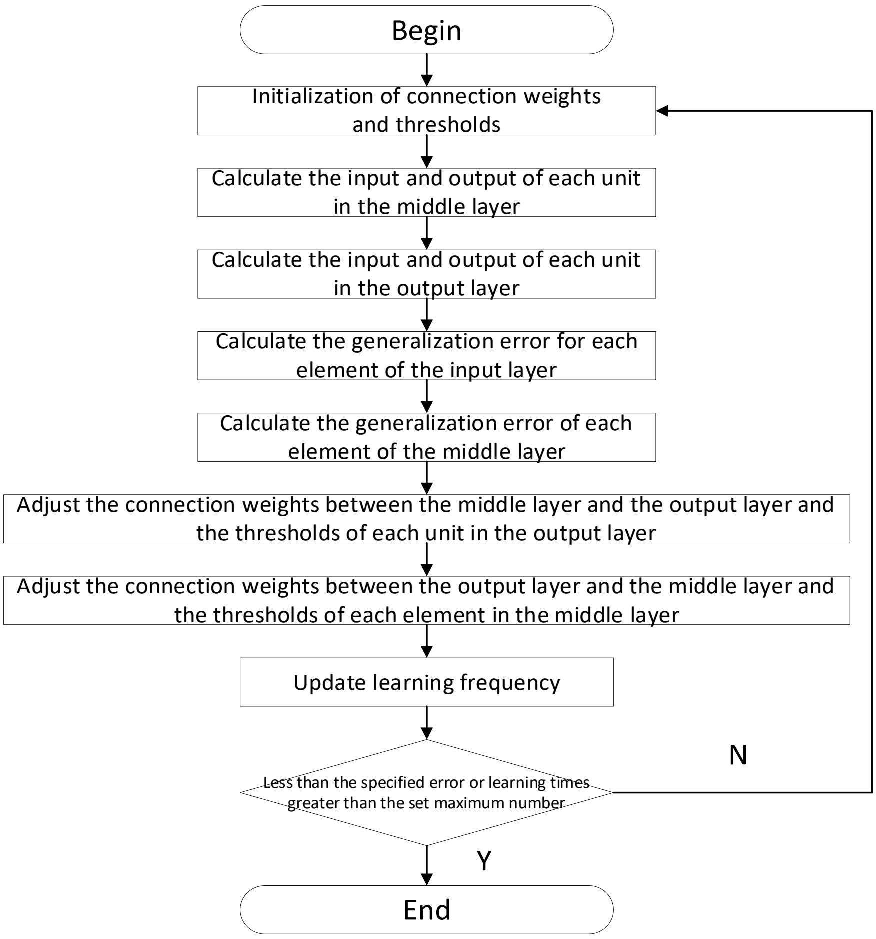

Figure 4. Algorithm flow chart of bp neural network model based on employee turnover

5. Solution of the model

5.1. Evaluation index

1. Mean Square Error (MES)

$$MSE = (1/n) * \sum(y\_i - y\_hat\_i)^{\wedge}2 \tag{1}$$

Where n is the sample size, y_i is the true value, and y_hat_i is the model prediction. MSE calculates the average of the squared difference between the true value and the predicted value. The mean square error is measured in units of the sample value, for example, if the sample value is measured in dollars, MSE is also measured in dollars. In order to better compare model performance on different data sets, MSE is sometimes converted to other forms of metrics, such as Root Mean Squared Error (RMSE) or Mean Absolute Error (MAE).

2. Root mean Squared Error (RMSE)

$$RMSE = sqrt(MSE) \tag{2}$$

RMSE is the square root of MSE and is used to align the units of error with those of the original data.

3. Mean Absolute Error (MAE)

$$MAE = (1/n) * \sum \left| y\_i - y\_hat\_i \right| \tag{3}$$

MAE calculates the average of the absolute value of the difference between the true value and the predicted value, which is a better indication of the actual predicted value error.

5.2. Analysis of model solving

In the process of solving, the solving quality and training effect of the BP neural network model are affected by many factors, such as the setting of initial weight, the selection of learning rate, and the determination of the number of iterations. In order to improve the performance and generalization ability of the model, regularization, momentum method, adaptive learning rate and other technical means can be used to optimize the model.

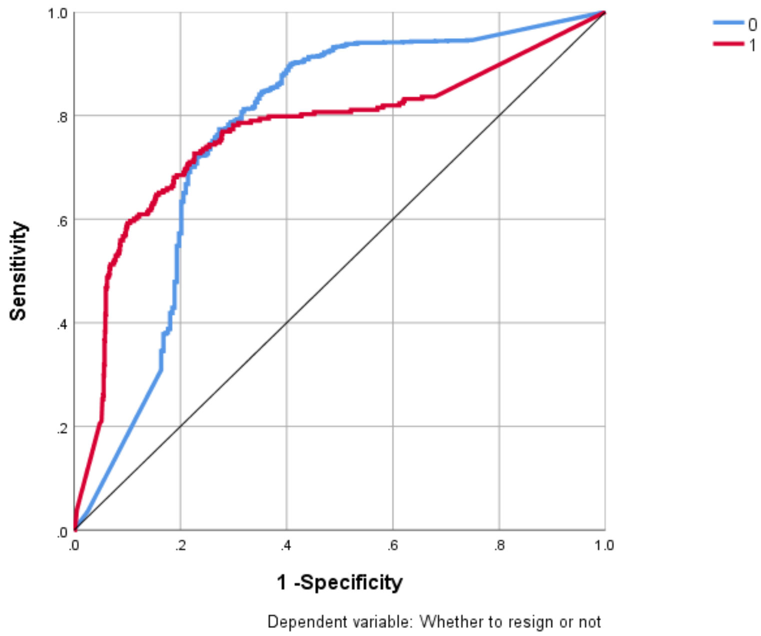

Figure 5. ROC curve

Look at the shape of the ROC curve, the closer to the top left corner, the more accurate the model is. figure 5 shows the ROC curve of employee turnover rate and environmental adaptability, which shows that the accuracy of this model is relatively high.

Table 4. classifies the sample training set and test set

sample	Actual measurement	No	Yes	Correct percentage
train	No	797	72	91.7%
	Yes	39	124	76.1%
	Overall percentage	81.0%	19.0%	89.2%
test	No	282	91	75.6%
	Yes	54	21	28.0%
	Overall percentage	75%	25%	67.6%

Performance on the training set: This is mainly assessed by looking at the accuracy of the model on the training set, loss function values, and other metrics. A good model should perform well on the training set and be able to accurately predict new data. Performance performance on the trial set: This is a key metric to evaluate the model's ability to generalize. A good model should perform well not only on the training set, but also on the test set. If the performance on the test set is poor, it may mean that the model is overfitting. This model has a good performance in both training and testing is shown as Table. 4.

5.3. Conclusion of the model experiment

First of all, BP neural network model in this model shows a high classification accuracy, targeted to solve the causes of employee dimension related problems, and gives the enterprise improvement ideas. Secondly, BP neural network has strong nonlinear mapping ability, which makes it have great advantages in dealing with complex problems. Compared with traditional classification methods based on statistical theory, BP neural network does not require normal distribution of training samples, so it has wider applicability in practical applications. However, we also note that BP neural network still has some defects. For example, it is very sensitive to the setting of initial weights and thresholds, improper Settings may cause the performance of the network to decline, and further improvement of the optimization algorithm may be needed to improve the performance of the model when facing complex problems.

In general, the BP neural network model has shown good performance in this experiment, but there is still some room for improvement. In future studies, we will continue to explore more effective optimization algorithms and network structures to improve the classification accuracy and generalization ability of the model.

6. Conclusions

With the intensifying competition in the talent market, employees have higher and higher requirements for salary, welfare and career development. From the prediction of this model, it can be seen that job participation satisfaction, commuting distance and monthly salary are the main factors for employee turnover, and young people aged 18-25 and single people are more likely to lose employees.

Therefore, it is crucial for organizations to prioritize the psychological well-being and job satisfaction of their employees in order to attract both young talents and experienced professionals. The extensive data on factors contributing to employee turnover serves as a foundation for developing a prediction model using the BP neural network. This study utilizes the matured BP neural network model, which excels in both theoretical understanding and performance. One notable advantage of this model lies in its exceptional ability to map nonlinear relationships and its adaptable network structure. The number of middle layers and neurons within each layer can be flexibly adjusted based on specific circumstances, resulting in varying performance outcomes depending on the chosen structure.

By improving the subjective decision-making opinions of experts and management, enterprises can better understand the demands and expectations of employees, formulate more reasonable policies and plans, improve employees' job satisfaction and loyalty, and reduce employee turnover. This contributes to the sustainable development and long-term stability of the enterprise. However, BP neural network also has the following major shortcomings. The learning speed is slow, and even a simple problem usually requires hundreds or even thousands of times of learning to converge.

References

[1] TAO Siming, LIANG Zhongmin, Chen Zaini, Qu Tian, HU Yiming. Application of long short-term memory network to medium and long term runoff forecast [J]. Journal of Wuhan University: Engineering and Technology Edition, 2021, 54(1): 21-27.

[2] CAO Ruoxin, ZHANG Kexin, Zeng Weihua, Ma Junwei, Xie Yuqian, Li Qing. Early warning of water environment carrying capacity based on BP neural network: A case study of Beihe Canal [J]. Chinese Journal of Environmental Sciences, 21,41(5): 2005-2017.

[3] Xiong Yi, Zhan Zhihong, Ke Fangchao, Zhou Qiupeng, Sun Liping, Liao Shuang, Ren Yulun, Zhou Renjun. Substation Maintenance cost prediction based on improved BP neural Network [J]. Journal of Electric Power Science and Technology, 2021, 36(4): 44-52.

[4] Tang Shao, Teng Zhaosheng, Sun Biao, Hu Qing, Pan Xifu. ADAM Improved BP Neural Network and Application of Weigh-in-motion [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(4): 127-135.

[5] ZHANG Hui 'E, LIU Dagui, ZHU Tingting, LIU Guanghui, HU Heng, XIAO Changjiang. Optimization of BP neural network for PV short-term power prediction based on similar day and momentum method [J]. Smart Electric Power, 201,49(6): 46-52.

[6] ZHU Xianhui, Yu Yue, Shi Nan, Xu Liang, Jian Youwei. Research on Hierarchical optimization of BP neural Network and its application in wind power prediction [J]. High Voltage Electrical Apparatus, 202,58(2): 158-163+170.

[7] WANG Guanpeng, Qin Shuangyan, CUI Hengjian. [7] Analysis and Prediction of Influencing Factors of Employee turnover [J]. Systems Science and Mathematics, 2022,42(6): 1616-1632.]

[8] Peng Lijun, Ma Yuelu. [8] Seasonal Problem of employee turnover and application of SARMI-ANN Model [J]. Operations Research and Management, 2021,30(2): 139-145.]

[9] Zheng Jian, Liu Renjing. Employee turnover prediction of Qinghai Electric Power Company based on interval variable RF algorithm [J]. Operations Research and Management, 2022,31(9): 210-216.]

[10] Zhu Fei, Yue Meiqi, Zhang Jiesuan. A study on the influence mechanism of consistency between organizational career management and human resource management intensity on employee turnover intention -- The mediating role of career Satisfaction and the moderating role of employer brand [J]. Journal of Central University of Finance and Economics, 2021(12): 105-118.