**Optimization of LMBP High-speed railway Wheel Size Prediction Algorithm Based on Improved Adaptive Differential Evolution Algorithm**

Yu Zhang 1, Jiawen Zhang 1, Lin Luo1\* , Xiaorong Gao1

**Abstract**

It is beneficial for maintenance department to make maintenance strategy and reduce maintenance cost to forecast the hidden danger index value. Based on the analysis of the research status of wheel-to-life prediction at home and abroad and the repair of wheel-set wear and tear, this paper designs and implements an adaptive DE-LMBP wheel-set size prediction model. Aiming at the shortcomings of BP neural network, it is easy to fall into local extreme value. The BP algorithm is improved by Levenberg-Marquardt (LM) numerical optimization algorithm. Aiming at the shortcomings of BP neural network algorithm for randomly initializing connection weights and thresholds to fall into local extreme value , the differential evolution algorithm (DE) is used to optimize the initial connection weights and thresholds between the layers of the neural network. In order to speed up the search of the optimal initial weights and thresholds of the DE-LMBP neural network, the initial values are further optimized, and an adaptive DE-LMBP wheel-set size prediction model is designed and implemented. Compared with different algorithms, the experimental results show that the proposed adaptive DELMBP algorithm is effective and significantly improves the prediction accuracy.

**Keywords**

Wheel-set size prediction, neural network model, Levenberg-Marquardt algorithm, Differential Evolution algorithms

**Introduction**

With the increasing mileage of high-speed railways in China, railway train safety inspection is becoming more and more important. Any minor fault on the train components may affect the safety of high-speed trains and even cause major safety accidents. The wheel pair is called one of the three major consumable parts of the train, and the wear condition of the rim and the tread is a key factor affecting the safe and stable operation of the train, the ride comfort and the service life of the wheel track.1-2 As the train causes continuous wear of the wheel-set during the operation, the related expenses of maintenance and repair such as repairing or replacing it in time are one of the important components of the train maintenance cost.3 Therefore, on the basis of mastering the historical state of the wheel-set, predicting the hidden danger index of the wheel-set is beneficial to the railway

management department to formulate effective maintenance measures in time to improve the operational safety, reliability and economy of the train.

In the railway transportation and development, the wheel-set wear phenomenon of railway vehicles is widely existed in various transportation sites, which is also one of the main research directions in recent years. At present, the research on the modeling of railway wheel-set state prediction at home and abroad is mainly

1School of Physical Science and Technology，Southwest Jiaotong University, China

**Corresponding author:**

Lin Luo, School of Physical Science and Technology，Southwest Jiaotong University, Chengdu, Sichuan 610031, China

Email: [happyluolin@vip.163.com](mailto:happyluolin@vip.163.com)

divided into two categories. The first type is based on the vehicle track system dynamics model, the wheel-rail local contact model and the local wear analysis mechanism model of the wheel-rail material, involving physical quantities such as wear and creep distance, normal force, material hardness, etc. Then the numerical simulation analysis of the train wheel wear profile information is carried out.4-7 The other is to analyze and predict the historical wear profile information obtained from the statistics, and calculate the remaining service life of the wheel-set to propose a maintenance strategy, such as time series model8, support vector machine9, gray prediction algorithm10, Bayesian algorithm11 and so on. With the advent of the era of data explosion, big data technology and numerical analysis methods have gradually been applied to various device security predictions. Among them, neural network algorithms are widely used. Gebraeel12 developed an experimental device to perform accelerated bearing testing to obtain bearing vibration data samples, and based on BP neural network to establish a bearing residual life prediction model and verify its effectiveness. Wei13 established a three-layer BP neural network for multi-stress accelerated life test, which can effectively predict the failure time of normal stress level and obtain the prediction curve of reliability function. However, the traditional BP neural network converges slowly and is easy to fall into local optimum. Therefore, scholars will improve the BP algorithm or use some other algorithms combined with BP neural network to improve the shortcomings of BP. In terms of algorithm combination, Lixin14 used a combination of time series analysis and BP neural network to predict the remaining life of the cooling fan and improve the prediction accuracy. He15 proposed a combination of principal component analysis (PCA) and BP neural network to provide a good reference for the prediction of phosphorus content in BOF endpoint. In terms of algorithm improvement, Zhang16 used genetic algorithm to globally optimize the weight of BP neural network. The results of tool residual life show that the prediction effect is better than single BP neural network method. Zhaoyang17 optimized the BP neural network with LM algorithm to predict the power consumption of buildings. The results show that the LM BP neural network is more accurate and stable than the BP neural network. Huaixian18 used the particle swarm optimization algorithm to optimize the BP network to establish a prediction model for the axle-to-axle box failure of the urban rail train bogie, which is better than the BP neural network.

At present, the method for predicting wheel wear based on dynamic model is mature, but most of them are based on simulation numerical analysis, and field data is not used. Based on the historical surface information data, there are few researches on the method of wheel wear. Although the field data is used, the prediction accuracy is not high, and the algorithm combined with the global optimization and training process is less. In this paper, based on the idea that the wheel diameter value changes with time to find useful information, the problem of wheel-set life is studied. The corresponding prediction model is established to predict the trend of wheel diameter change, which is used as an auxiliary method for the maintenance department to formulate maintenance strategy. Aiming at the shortcomings of BP neural network which is it easily to fall into local extreme value, the LM numerical optimization algorithm is used to improve the problem. Aiming at the shortcomings of BP neural network algorithm for randomly initializing connection weights and thresholds to fall into local extremum, differential evolution algorithm (DE) is used to optimize initial connection weights and thresholds between layers of neural network. In order to speed up the search for the optimal initial weights and thresholds of the DE-LMBP neural network, the initial values are further optimized, and a design is implemented based on Adaptive DE-LMBP wheel-set size prediction model.

**Theory**

**LMBP neural network algorithm**

Backpropagation (BP) neural network is a multilayer feed forward neural network, which consists of input layer, single layer or multi-layer hidden layer and output layer. For single hidden layer BP neural network, the input layer contains  nodes, the hidden layer contains  nodes, and the output layer is 1 node, as shown in Figure 1.

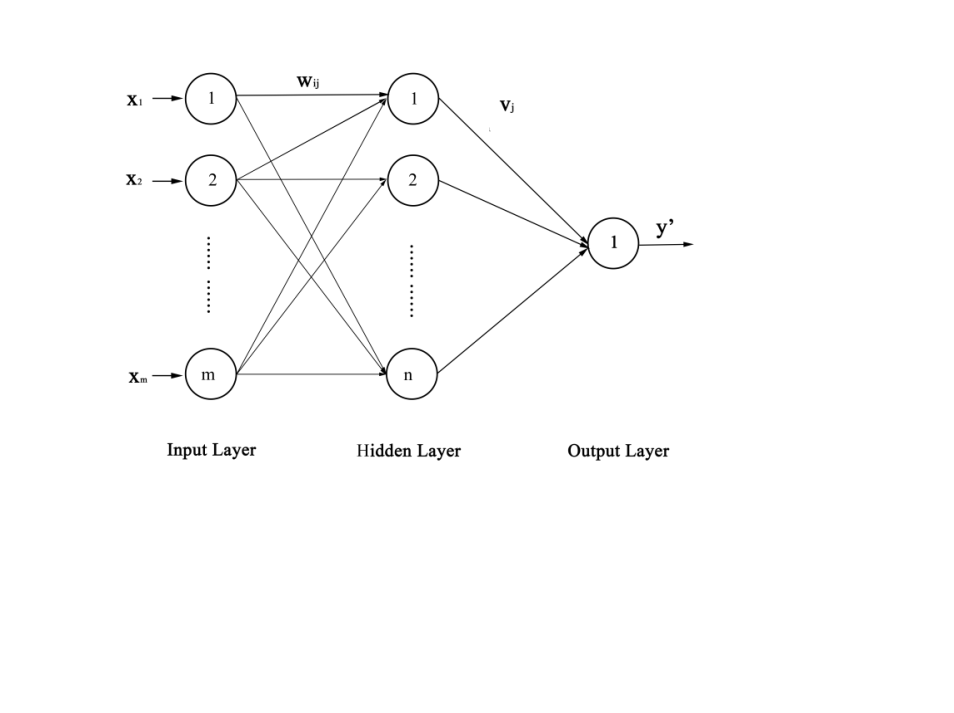


Fig.1Topological structure of single hidden layer BP network

The weight matrix from the input layer to the hidden layer of BP network is marked as , the weight matrix from the hidden layer to the output layer is marked as , the threshold matrix of the hidden layer is marked as , and the threshold of the output layer is marked as . The expression is as follows:







The standard BP algorithm minimizes the sum of squares of errors between the expected output vectors of training samples and the actual output vectors of the network by adjusting the weight vectors and thresholds between the connecting layers. The sum of squares of errors is the objective function that Levenberg-Marquardt (LM) algorithm needs to optimize.

LM algorithm is used to adjust the weights and thresholds of BP neural network. The formula is as follows:

 (1)

Where  is the weight vector of iteration,  is the constant (adjustment factor) greater than zero in the LM algorithm, which is used to control the iteration of LM algorithm,  is the unit matrix,  is the Jacobian matrix of the error to the weight differential.

In the process of network training, with the increasing number of iterations, when  approaches zero, LM algorithm approaches Gauss-Newton method, which converges faster than BP algorithm based on gradient descent method. And it has the advantage of faster calculation speed and higher accuracy when the error is closer to the minimum value. The algorithm provides a compromise between the speed of Newton's method and the gradient descent method, which guarantees convergence.

**Differential evolution algorithm (DE)**

The essence of differential evolution algorithm (DE) 19 is a greedy algorithm based on real coding with the idea of preserving the best. The basic principle of the algorithm is to randomly select two individuals in the population to generate difference vectors and sum them with the third individual to generate new individuals (variant individuals), cross-operate the parent and the corresponding variant individuals, select the individuals with better fitness between the parent and the offspring individuals, and these individuals with better fitness were selected as offspring.

DE algorithm mainly includes population initialization, mutation operation, crossover operation and selection operation:

(1) Population initialization

The three matrices , ,  and threshold  of BP network are mapped to the chromosome strings of the difference algorithm. The mapping relationship is as follows:



Let , in which  is the i’th chromosome of the  generation, ,, ,  is the population size,  is the largest evolutionary algebra, and  is the chromosome length.

(2) Mutation operation

The mutation operation is based on individual vector difference. Assuming that the current evolutionary individual is , three chromosomes  are randomly selected from the population of this generation,  is the variation factor. The difference between the two individual variables is taken to obtain :

 (2)

(3) Cross-operation

The  of the variant individual and the current evolutionary individual operate in a discrete crossover manner to generate the crossover  to increase the diversity of the population.

 (3)

Where  is a random number between ,  is a crossover factor and  is a random integer of . This crossover strategy ensures that at least one component of  is contributed by  corresponding components.

(4) Fitness function

The fitness evaluation is carried out by using the square error measure in the following form:

 (4)

Where  is the number of training samples,  is the actual output of the network, and  is the expected output.

(5) Selection operation

Selecting  and  as the crossover individuals and the current evolutionary individuals to select the best one according to greedy way by comparing their fitness, that is:

 (5)

Repeat (2-5) until conditions are met.

**Adaptive DE-LMBP algorithm**

In order to speed up the search for the optimal initial weight and threshold of DE-LMBP neural network and further optimize the initial value, an adaptive DE-LMBP neural network model is proposed. The difference between the adaptive DE-LMBP neural network model and the DE-LMBP neural network model is that the crossover probability and mutation probability of the DE-LMBP neural network model are fixed values. The crossover probability and mutation probability of the adaptive DE-LM BP neural network model are adjusted with the individual fitness.

The key to the adaptive algorithm lies in the variation and crossover operation of the algorithm, that is, the dynamic adjustment of the cross-factor  and the variation factor  according to the individual fitness value. The basic idea is: let  is the highest fitness of a certain generation group, and  is the average fitness of the generation group. The difference between the maximum fitness and the average fitness of indicates the stability of the population to a certain extent. The smaller the difference, the smaller the individual fitness difference in the population, and the greater the possibility that the population reaches precocity. On the contrary, the larger the difference is, the greater the individual fitness difference and the divergence of individual characteristics. Therefore, when  is small, the values of  and  should be increased, conversely, when  is larger, the values of  and  should be reduced. The calculation formulas for CR and F are as follows:

 (6)

Analysis of the above formula shows that the values of  and  do not depend on the fitness of any individual, and all individuals have the same crossover and mutation probability, which is not conducive to the convergence of the algorithm to the global optimal solution. In addition, when the population approaches the global optimal solution,  decreases, the values of  and  increase, the optimal individuals are easily destroyed, and the population may not converge to the global optimal solution. To this end, the above formula is adjusted to:

 (7)

Where  is the larger fitness value of the two individuals to be crossed, and  is the fitness value of the individual to be mutated.

When the crossover and mutation factors are adjusted according to formula (7), the probability of crossover and mutation is close to or equal to zero for individuals whose fitness is close to or equal to the maximum fitness. In this way, in the early stage of evolution, the good individuals are almost in a state of unchanged, but at this time, the good individuals are not necessarily the global optimal solution. The final crossover and mutation functions are obtained by the following adjustment:

 (8)

Where  are between zero to one. The method increases the crossover probability and mutation probability of the individuals whose fitness is close to and equal to the maximum fitness in the population to  and  , and can control the degree of approaching the maximum fitness through  and , for poor individuals whose fitness is lower than the average fitness, a large crossover probability and mutation probability are uniformly adopted to make these individuals evolve toward the optimal individual.

**Evaluation index**

In this paper, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and correlation coefficient R are selected as model evaluation indicators.

 (9)

 (10)

 (11)

 (12)

Where  is the expected output value and  is the network output value.  and  are the average of the real value and the network output value, respectively, and  is the predicted sample number. The smaller the RMSE, MAE and MAPE value, the higher the prediction accuracy of the model; the closer the R value is to 1, the higher the prediction accuracy of the model.

**Experimental procedures**

The flow chart for predicting the change trend of wheel pair diameter based on improved adaptive DE-LMBP neural network is shown in Figure 2.



Fig.2 Adaptive DE-LMBP neural network prediction process

According to Figure 2, we have designed the following wheel-set dimensional change trend prediction process as shown in Algorithm 1.

|  |
| --- |
| Algorithm 1: Wheel-set Size Prediction |
| 1. Denoise the selected data and normalize it.  2. Construct sample feature points, construct the input data set according to  in the original sequence, and construct the output data set according to . And divided into training sets and test sets according to a certain proportion.  3. Use adaptive differential evolution algorithm to find the optimal initial weight and threshold.  4. In the iterative process, the LMBP algorithm continues to adjust the weight threshold to get the best results.  4. Use the model to predict wheel-set size trends.  5. Calculate model evaluation indicators. |

**Data sample preparation**

The data studied in this paper comes from the LY Series Dynamic Inspection System. As shown in Figure 3, it is installed on the operation line or entrance line of the Electric Multiple Units(EMU) will pass by. When the railway enters the detection area at a limited speed and triggers the detection sensor, the system enters the working state.



Fig.3 LY series dynamic inspection system for wheel profile

The Dynamic Inspection System for Wheel Profile uses the "light intercept image measurement technology" to measure the wheel alignment size online. The principle is shown in Figure 4. When the railway enters the detection range, the laser line source is projected from both sides of the track to the wheel surface to form an optical curve and is captured by the CCD camera. Through real-time image acquisition, processing and correction, the real contour curve is obtained, which the numerical calculation is performed to obtain the wheel size parameter. On this basis, numerical calculation is performed to obtain the wheel size parameter.



Fig.4 Optical intercept image detection principle

The research object of this paper is the China Railway High-speed (CRH) wheel-set. The tread shape of the CRH380BL model is the wear-type tread of the S1002CN. Figure 5 shows the historical wheel diameter measurement data for the CRH380BL-3539 railway for one and a half years. The detection data of the wheel pairs is susceptible to the measurement position, maintenance personnel's measurement habits and manual corrections, loads, rail conditions and other factors. Historical measurement data will have data fluctuations and obvious abnormal points and repairs. This experiment does not consider the phenomenon of repair.

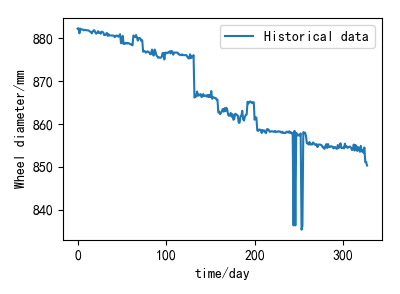


Fig.5 Historical wheel diameter measurement data

According to the wheel-set data of CRH380BL-3539 railway, the sample data of time series are constructed. The time interval of the time series is one day, and 450 data points of wheel diameter data with equal time interval are obtained. Considering that the data center has undergone several repairs and manual correction processing, the outliers in the data are corrected. The final sample data is shown in Figure 6.

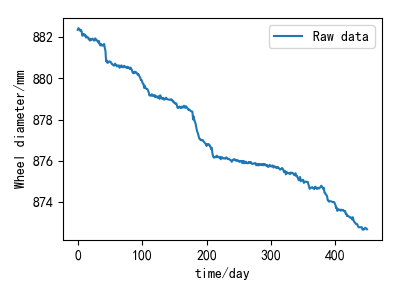


Fig.6 Data sample set

**Set network parameters**

Construct sample feature points20. Select the correlation coefficient to observe the input dimension. It can be seen from Figure 7 that the correlation coefficient R is the largest when the input dimension is 10, and the correlation coefficient is gradually decreased when it is lower or higher than the 10-dimensional input layer. Therefore, the input layer is set to 10-dimensional wheel pair diameter history data, and the output layer is the one-dimensional wheel diameter value of the next moment. Enter the wheel diameter values at the time of the wheel diameters , respectively, and the target output is the wheel diameter value at time . Using the empirical traversal method to select the hidden layer nodes, the optimal number of training effects is 10.

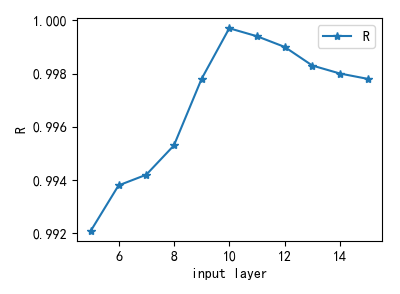


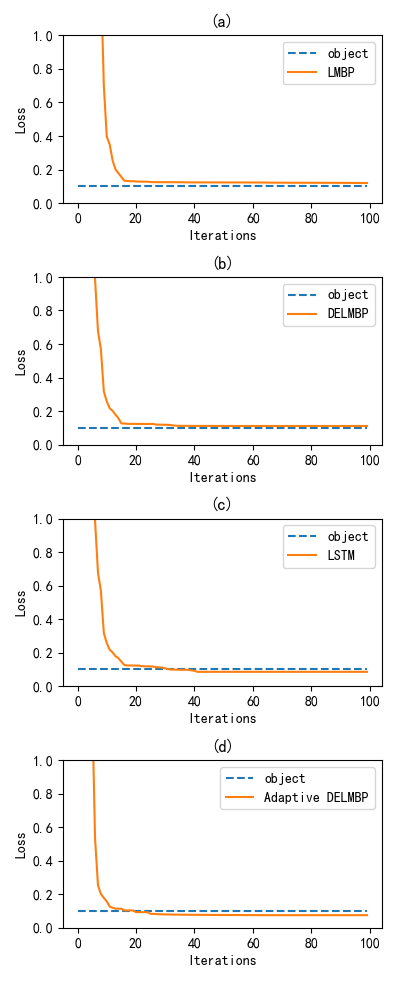
Fig.7 The relationship between the input dimension and the correlation coefficient R

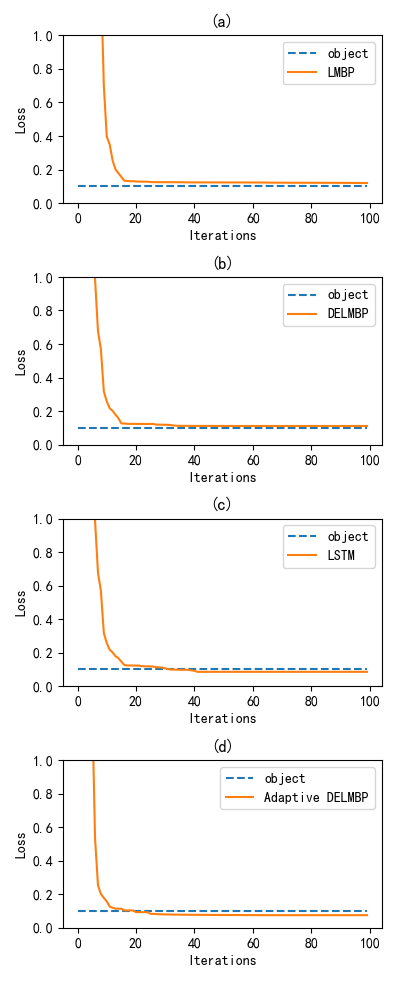
The neuron transfer function of the hidden layer and the output layer of the neural network adopts a continuous and differentiable sigmoid function. In order to avoid the saturation area of S-type function and improve the convergence speed and sensitivity of the network, the sample data are normalized before network training. In this paper, we use min-max standardized data to make the normalized data in the  interval. The ratio of data training set to test set is 8:2.

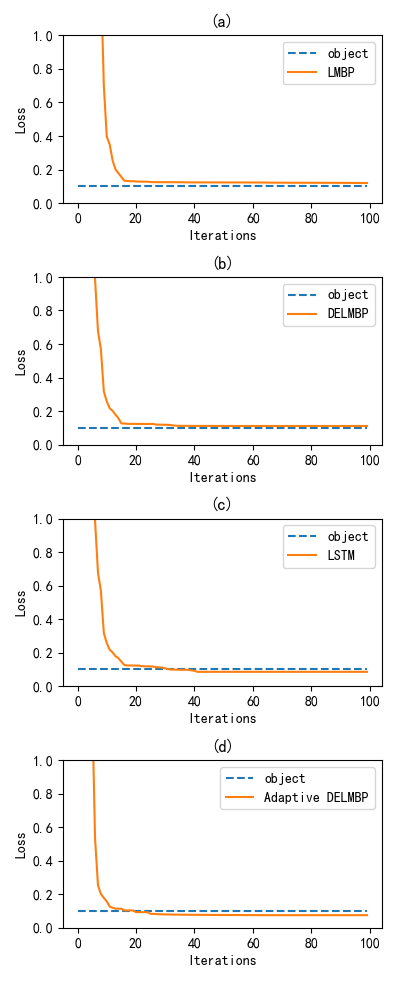
**Results and discussions**

In order to better understand the performance of the improved adaptive DELMBP neural network model, the LMBP neural network model with the same network structure and the standard DELMBP neural network model, and the Long Short-Term Memory neural network (LSTM), were established with the same data samples for comparative analysis. The LSTM neural network is a variant of the cyclic neural network (RNN), which is suitable for the analysis of time series data.

Because the problem of the neural network itself will lead to the randomness of the results, I have carried out several simulation experiments on four models to obtain the optimal convergence speed and accuracy. The convergence simulation results of LMBP, DELMBP, LSTM and adaptive DELMBP four prosperous network models are shown in Figure 8.







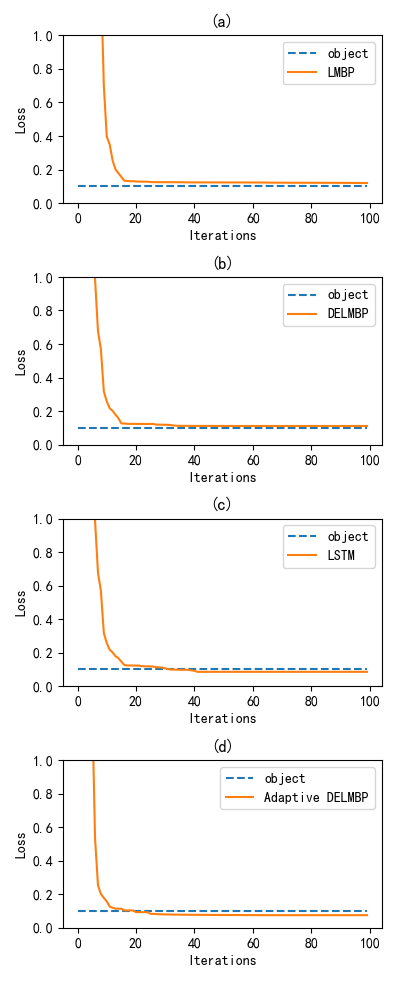


Fig.8 LMBP, DELMBP, LSTM and adaptive DELMBP convergence speed comparison

It can be seen from Figure 8 that the LMBP neural network needs more than 100 times to reach the convergence goal of 0.1. Although the convergence speed of the DELMBP neural network is faster than LMBP, it still takes 100 times to reach the convergence target of 0.1, and the LSTM neural network has excellent convergence speed. The adaptive DELMBP algorithm can achieve the convergence goal less than 20 times, and the training time is less than other models. This shows that the adaptive DELMBP converges much faster than the other three models.

The predicted effects of the three models are shown in Figure 9. Although LSTM converges slightly faster than the other two hybrid models, the prediction results are poor. The expected effects of the other three models are better, but the prediction effect of adaptive DELMBP is better than the other three models, and the stability is better.

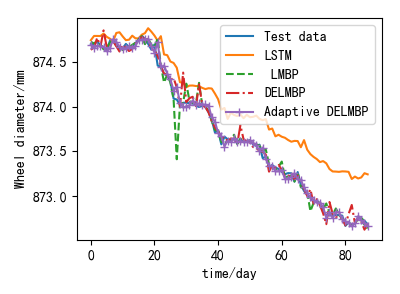


Fig.9 LMBP, DELMBP, LSTM and Adaptive DELMBP prediction results

Tab. 1 Prediction accuracy of three models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | LMBP | DELMBP | LSTM | Adaptive DELMBP |
| MSE | 0.0128 | 0.0103 | 0.1057 | 0.0038 |
| RMSE | 0.1134 | 0.1015 | 0.3251 | 0.0616 |
| MAE | 0.0882 | 0.0772 | 0.2859 | 0.0480 |
| MAPE | 0.0099 | 0.0084 | 0.0103 | 0.0054 |
| R | 0.9955 | 0.9974 | 0.9564 | 0.9997 |

It can be seen from Table 1. Although all the four models can predict the wheel diameter better, the index values of the adaptive DELMBP neural network model are better than the other three models. This proves that the improved adaptive DELMBP neural network model is more suitable for the prediction of the trend of wheel-set size change.

**Conclusion**

In this paper, the combination of DE algorithm, LM algorithm and BP neural network is used for wheel-set size prediction. The main work and contribution are as follows: In order to overcome the shortcomings of traditional BP neural network algorithm, the training process is trapped in local extreme points, and its accuracy is improved. This paper firstly uses the adaptive DE algorithm with powerful global optimization ability to perform global pre-optimization, and then uses LM algorithm for deep optimization and training BP neural network. Compared with other three prediction models, the experimental results show that the improved adaptive DELMBP algorithm proposed in this paper is effective. Compared with the LMBP neural network, the standard DE-LMBP neural network algorithm and the LSTM neural network, the prediction accuracy is significantly improved.

Therefore, in practical applications, the algorithm can be applied to the prediction of the wheel size of high-speed trains, and the maintenance department is provided with a reference according to the trend of the wheel size prediction, thereby effectively reducing the maintenance cost.

**Acknowledge**

This work was supported by the National Nature Science Foundation of China (Grant No. 61471304) and we wish to acknowledge them for their support. The authors also thank Southwest Jiaotong University Photoelectric Engineering Institute for their kind support in the experiment.

**References**

1. Wang L, Yuan H, Na WB, et al. Optimization of the re-profiling strategy and remaining useful life prediction of wheels based on a data-driven wear model. Systems Engineering-Theory&Practice,2011, 31(06):1143-1152.

2. Lu JH. Analysis on the wearing of Guangzhou Metro Line 1 vehicles. Technology for Electric Locomotives, 2001(03):16-17+29.

3. Hua S. Research on Wheel Wear Predicting and Re-profiling Strategy Based on Intelligence Analysis. Nanjing University of Aeronautics and Astronautics, 2017.

4. Ignesti M, Innocenti A, Marini L, et al. A numerical procedure for the wheel profile optimisation on railway vehicles[J]. ARCHIVE Proceedings of the Institution of Mechanical Engineers Part J Journal of Engineering Tribology 1994-1996 (vols 208-210), 2014, 228(2):206 – 222.

5. Lim R, Mba D. Fault Detection and Remaining Useful Life Estimation Using Switching Kalman Filters[M]// Engineering Asset Management - Systems, Professional Practices and Certification. Springer International Publishing, 2015:53-64.

6. Son S W, Jung H S, Kwon T S, et al. Fatigue life prediction of a railway hollow axle with a tapered bore surface[J]. Engineering Failure Analysis, 2015, 58:44-55.

7. Han P, Zhang W H. A new binary wheel wear prediction model based on statistical method and the demonstration [J]. Wear, 2015, 324-325(324):90-99.

8. Hong T P, Yang B S. Estimation and forecasting of machine health condition using ARMA/GARCH model[J]. Mechanical Systems & Signal Processing, 2010, 24(2):546-558.

9. XING Zongyi, MAO Lingli, LIAO Guiling, et,al. Forecasting of wheelset size of urban rail train based on PSO-SVM model[J].Journal of Shengyang University of Technology, 2014, 36(4):411-415.

10. LI Ying. Prediction of Wheelset Tread Wear Based on Grey Theory[D]. Lanzhou Jiaotong University, 2015.

11. Lin J, Pulido J, Asplund M. Reliability analysis for preventive maintenance based on classical and Bayesian semi-parametric degradation approaches using locomotive wheel-sets as a case study[J]. Reliability Engineering & System Safety, 2015, 134:143-156

12. Gebraeel N , Lawley M , Liu R , et al. Residual Life Predictions from Vibration-Based Degradation Signals: A Neural Network Approach[J]. IEEE Transactions on Industrial Electronics, 2004, 51(3):694-700.

13. ZHANG Wei，LI Xiaoyang，JIANG Tongmin，et al. Life-prediction of multi-stress accelerated life testing based on BP algorithm of artificial neural network[J]. Acta Aeronautica et Astronautica Sinica，2009，30(9)：1691-1696.

14. Lixin W , Zhenhuan W , Yudong F , et al. Remaining life predictions of fan based on time series analysis and BP neural networks[C]// Information Technology, Networking, Electronic & Automation Control Conference, IEEE. IEEE, 2016.

15. He F, Zhang L. Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network[J]. Journal of Process Control, 2018, 66:51-58.

16. Zhang Z , Li L , Zhao W . Tool Life Prediction Model Based on GA-BP Neural Network[J]. Materials Science Forum, 2016, 836-837:256-262.

17. Zhaoyang Ye, Moon Keun Kim. Predicting electricity consumption in a building using an optimized back-propagation and Levenberg–Marquardt back-propagation neural network: Case study of a shopping mall in China[J]. Sustainable Cities and Society, Volume 42, 2018, Pages 176-183.

18. Yin HX, Wang K, Zhang TZ, et al. Fault prediction based on PSO-BP neural network about wheel and axle bogie in urban rail train. COMPLEX SYSTEMS AND COMPLEXITY SCIENCE, 2015, 12(4):97-103.

19. Hou Y , Zhao L , Lu H , et al. Fuzzy neural network optimization and network traffic forecasting based on improved differential evolution[J]. Future Generation Computer Systems, 2017, 81.

20. SHEN Zhongjie, CHEN Xuefeng, HE Zhengjia, et al. Remaining life predictions of rolling bearing based on relative features and multivariable support vector machine[J]. Journal of Mechanical Engineering, 2013, 49(2):183-189.