Highway traffic state prediction method considering risky driving behavior

Hao Zhiqiang¹, Niu Shuyun², Li Jiang Tao³, Su Fei³,⁴, a⁴, Su Fei³,⁴, b
¹Shanxi Transportation Holdings Group Co., Ltd., Taiyuan, China
²Shanxi Province Expressway Group Limited Liability Company., Taiyuan, China
³China Transport Informatics National Engineering Laboratory Co., Ltd.
⁴China Transport Telecommunications & Information Center, Beijing, China
⁷740607427@qq.com
⁸SUFEI@CTTIC.CN

Abstract—In recent years, with the continuous development of the national economy, the state and people have a higher demand for the service level of the highway, which also requires highway managers to have a more accurate prediction of the running state of the highway. So this paper proposes an Elman neural network highway state prediction method considering the risky driving behavior. In this method, expert scoring method and analytic hierarchy process are used to calculate the weight of highway risk driving behavior, and then Elman neural network is used to establish the highway state recognition model based on the weight and traffic flow data. Finally, the data of expressways within 4 days is used as the sample training model, and the data of the fifth day is used as the test sample for testing. The results show that the model has a high accuracy for predicting the future traffic status, and it can also provide reference and support for the active traffic early warning and management.

1. Introduction
In recent years, with the continuous development of my country's economy, the people have higher demand for travel activities such as tourism and shopping, and the subsequent logistics and travel pressures continue to place higher requirements on the expressway service system. Therefore, more accurate identification and prediction of expressway operating status is extremely important for improving expressway service level.

The evaluation of expressway operating status at home and abroad mainly uses indicators such as flow rate, traffic flow speed, travel time, environment, and congestion level.

Zhao Fengbo[1] divided the traffic state into four categories based on the fuzzy C-means clustering algorithm, and used heuristic methods to determine the fuzzy index. Finally, the method was verified using urban intersection traffic data, which proved that the proposed method can effectively identify roads traffic status. Liu Henan[2] classifies the functions of urban roads, uses multi-level fuzzy evaluation to quantify the evaluation indicators, and divides the traffic operation state into four levels: smooth, stable, congested and forced flow, and finally obtains a comprehensive evaluation of urban road operation conditions. But this method cannot be applied to urban networks in real time and on a large scale. D'Este[3] uses GPS to collect traffic flow speed and travel time to develop new indicators in urban congestion.
Salanova[4] analyzed, screened and merged the data collected by the traffic flow detector, based on the Greek urban traffic evaluation index, carried out the operation state evaluation based on the specific route, but this method is only suitable for the traffic state recognition under static data, and the scope of application is small.

Shi[5] proposed a traffic state prediction model based on a hybrid intelligent method, using simulated annealing algorithm and PSO algorithm to improve, and finally through the optimized BP neural network to determine the weight of evaluation indicators and traffic state threshold. Doulamis, A.D.[6] built an adaptive neural network based on video data. Kumar and Kranti[7] constructed a short-term traffic flow prediction model for non-urban roads using artificial neural networks. Zhao Jian yu[8] is based on RBF neural network and uses particle swarm algorithm to optimize the basis function center, variance and weight. Wang Yan[9] optimized the artificial neural network by genetic algorithm, which effectively improved the prediction accuracy and efficiency. Xu Rong[10] proposed an adaptive particle swarm neural network traffic flow prediction model, and used an improved Sigmoid function to dynamically adjust the original fixed weights to speed up the convergence of the model. Li Haiqiong[11] studied the relationship between traffic flow and driver driving behavior through simulated driving experiments on 21 randomly sampled drivers.

At present, most of the researches mainly analyze the road operating status through road detector data, GPS, etc., and can only predict the road operating status on a macro-scale. At the same time, the traffic flow forecasting trend in the past two years has been developed in the direction of machine learning and combined models. This type of model can complement each other's advantages to achieve better prediction results.

The prediction of highway status caused by emergencies has serious lag, so this paper adopts the highway vehicle data containing highway risk driving behavior further evaluates the highway traffic state.

2. Data processing
Since the data used is the original data returned by the vehicle, it needs to be preprocessed. The original data is the information returned by the on-board equipment. Each piece of data is encrypted to encrypt the license plate information data to prevent the leakage of private information; each piece of information contains GPS data records containing the space (latitude and longitude) and Time data, instantaneous vehicle speed, vehicle direction angle (the instantaneous angle between the direction of the vehicle at that moment and the true north direction), and risky driving behavior ID (including whether the driver has dangerous driving behavior). Table I is sample data:

| Vehicle ID | Time                | Latitude and longitude (°) | Speed (Km/h) | Direction angle (°) | Warning ID |
|-----------|---------------------|-----------------------------|--------------|---------------------|------------|
| 5b19e6c52058b6327545512b | 2018-06-08 02:15:32.972 | 121.1892700,31.30887794 | 60.0         | 310                 | 206        |
| 5b19e6c52058b6327545512b | 2016-01-01 00:02:56.903 | 0.0,0.0                     | 60.0         | 0.00                | 202        |
| …          | …                   | …                           | …            | …                   | …          |
| 5b19e7652058b63275455781 | 2018-06-08 02:18:13.207 | 121.1992700,31.31887794 | 60.0         | 220                 | 206        |
| 5b19e7792058b63275455848 | 2018-06-08 02:18:33.343 | 121.1792700,31.31887794 | 60.0         | 341                 | 206        |
| …          | …                   | …                           | …            | …                   | …          |

As shown in Table I, the original data generally has problems such as data duplication, missing fields, and data drift. The preprocessing of the original data is the basis of the research work. Therefore, this paper uses the vehicle number to sort out the vehicle trajectory first, eliminates duplicate values, and then
uses interpolation to fill in missing fields and correct drift data[12]. Finally, based on the cleaned data, the corresponding time traffic data is obtained, and the next step is to analyze.

3. Model construction

3.1 Weight calibration of risky driving behavior indicators

The pre-processed vehicle early warning information, based on the relevant traffic accident standards and specifications[13], establishes a hierarchical structure model as shown in the figure 1:

Figure 1. The judgment matrixes corresponding to criterion layer

3.1.1 Construct judgment matrix $A-B_i \left( \chi_{vi} \right)$, where $A$ is the target layer element (highway risky driving behavior), $B_i$ ($B_1, B_2, B_3$) is the criterion layer element (speed control behavior, bad driving behaviour and driving following a car); $\beta_1, \beta_2, \ldots$ are the elements of the scheme layer; $\chi_{vi}$ represents the comparison value of the importance of $B_i$ to $B_{vp}$ for $A$, the specific values are shown in Table II:

|   | $A$ | $B_1$ | $B_2$ | $B_3$ |
|---|----|------|------|------|
| $B_1$ | 1  | 3    | 5    |
| $B_2$ | 1/3| 1    | 3    |
| $B_3$ | 1/5| 1/3  | 1    |

3.1.2 Calculate the judgment matrices for the elements of the criterion layer:

|   | Speed Control | Risky Behavior | Car-following behavior |
|---|---------------|---------------|-----------------------|
| $B_1$ | $\beta_1$ | $\beta_2$ | 1/7                |
| $\beta_1$ | 1 | 1/7 | |
| $\beta_2$ | 7 | 1 | |

|   | Risky behavior |
|---|----------------|
| $B_2$ | $\beta_1$ | $\beta_2$ |
| $\beta_1$ | 1 | 5 |
| $\beta_2$ | 1/5 | 1 |

|   | Car-following behavior |
|---|------------------------|
| $B_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| $\beta_1$ |  |  |  |
| $\beta_2$ |  |  |  |
| $\beta_3$ |  |  |  |
### 3.1.3 Level list sorting and consistency check

Single-level ranking refers to the calculation according to the judgment matrix. For an element of the upper layer, the weight of the importance order of each element related to this level is obtained through the judgment matrix to find the characteristic root and characteristic vector, and then calculate the consistency index CI and the random consistency ratio CR according to the following formulas:

\[
\beta_1 = \frac{3}{2} \\
\beta_2 = \frac{2}{3} \\
\beta_3 = 2
\]

\[
\beta_1 \beta_2 = \frac{3}{2} \times \frac{2}{3} = \frac{1}{2} \\
\beta_2 \beta_3 = \frac{2}{3} \times 2 = \frac{3}{1}
\]

\[\text{Consistency Index } CI = \frac{\mu_\beta - \mu_\alpha}{\mu_\alpha - 1} \quad (1)\]

\[\text{Random Consistency Ratio } CR = \frac{CI}{RI} \quad (2)\]

In the formula: RI is the average random consistency index (n=3, RI=0.52; n=4, RI=0.89; n=5, RI=1.12). When the ratio CR of the consistency index CI of the judgment matrix to the average random consistency index RI is less than 0.1, the judgment matrix has satisfactory consistency.

Calculate the weight of the elements in each level to the overall goal, and the final result is shown in the figure:

![Figure 2. The weight of risky behavior to accident probability](image)

After calculation, the consistency evaluation index CR of each judgment matrix is all less than 0.1, which meets the requirements of matrix consistency. Therefore, the judgment matrix is reasonable.

It can be seen from Figure 2 that factors such as driver fatigue and Inattentiveness during driving have the largest weight; Over Speed has second largest weight; Consecutive vehicle problem, vehicle slip, lane departure, emergency braking have the smallest weight. These weights are basically consistent with the results of road safety accident statistical analysis[14] and it can be considered that the weights of early warning information obtained in this paper are effective.

### 3.2 Model construction

Neural network has the ability to accurately approximate any nonlinear function, so it has received extensive attention in traffic state recognition[15]. Elman neural network is a global feedforward local recurrent network model, which occupies a set of context nodes to store internal state and contains memory. Therefore, it has better dynamic characteristics than static neural networks such as BP and RBF, and is particularly suitable for traffic state identification and discrimination[18]. This paper uses the original traffic data containing highway risk driving behavior to establish a macroscopic traffic operation status recognition model based on Elman recurrent neural network.
Figure 3 shows the Elman recurrent neural network with three layers of neurons. The input layer is composed of external input neurons and internal input neurons (also called context units). The feedback matrix from the hidden layer to the context unit is \( w^h \). The output of the context unit and the external input neuron is fed back to the hidden neuron. Context units are called memory units because they store the previous output of hidden neurons. The following equation describing the Elman network can be obtained.

\[
\hat{y}(k+1) = G[w^h h(k)] \\
\eta(k) = \Phi[\omega^\chi \chi(k) + \omega^\psi u(k)] \\
\chi(k) = \omega^\eta \eta(k-1)
\]

Among them, \( F \) and \( G \) are the activation functions of hidden layer and output layer neurons, and \( \omega^\chi \), \( \omega^\psi \) and \( \omega^\eta \) are the weight matrix. After the Elman network structure is established and selected, an improved backpropagation algorithm is used to train it.

Elman neural network has the advantages of less training times, small error, strong generalization ability, etc. This paper takes the risky driving behavior information and traffic flow of a certain highway within 4 days as input, and the corresponding expressway traffic state after 10 minutes within 4 days is used as training data. Elman recurrent neural network is trained; the warning information and traffic flow at different times on the fifth day are used as test data, and the highway traffic state is used as test data for verification. The model is shown in Figure 4:
4. Results
Input the test set data into the trained model to get the corresponding prediction result, namely the highway traffic state value calculated by the model. The predicted label is compared with the test label to obtain the accuracy of the model classification. The results obtained by Python are shown in Figure 5. The results show that the prediction of highway traffic state is more consistent with the actual situation.

![Figure 5. Comparison of prediction algorithm results](image)

| Serial number | Test sample       | Consider the risky driving behavior model | Ordinary ELMAN model |
|---------------|-------------------|------------------------------------------|----------------------|
| 1             | stable flow       | stable flow                              | stable flow          |
| 2             | approaching unstable flow | stable flow | stable flow |
| 3             | approaching unstable flow | approaching unstable flow | stable flow |
| 4             | approaching unstable flow | approaching unstable flow | approaching unstable flow |
| 5             | approaching unstable flow | approaching unstable flow | approaching unstable flow |
| 6             | approaching unstable flow | approaching unstable flow | approaching unstable flow |
| 7             | approaching unstable flow | approaching unstable flow | approaching unstable flow |
| 8             | stable flow       | stable flow                              | stable flow          |
| 9             | approaching unstable flow | approaching unstable flow | approaching unstable flow |
| 10            | stable flow       | stable flow                              | stable flow          |

After calculation, the applicability evaluation of the model is shown in Table V. The applicability of the model listed in the table—root mean square error RMSE, average absolute error MAE, and certainty coefficient R2, it can be seen that the applicability of the model is high and can fully meet the forecasting requirements.

| model                                          | RMSE | MAE  | R2  |
|------------------------------------------------|------|------|-----|
| Consider the risky driving behavior model      | 0.151| 0.11 | 0.97|
| Ordinary ELMAN model                           | 0.312| 0.2575| 0.93|
From the prediction results, it can be seen that the Elman neural network highway state prediction method constructed in this paper that considers risky driving behaviors has high accuracy for highway traffic state recognition; at the same time, the model can use the newly generated data to continuously iterate and further improve the accuracy rate. Therefore, it can be considered that the real-time highway traffic state identification method established in this paper is feasible, and the recognition accuracy is high, and it can provide real-time and accurate traffic information for travelers and highway managers.

5. Conclusion
This paper aims at the existing highway state prediction research that does not consider the impact of the driver’s risky behavior on the traffic state. Therefore, this paper proposes an Elman neural network highway state prediction method that considers the risky driving behavior. This method calibrates the weight of the impact of risky driving behavior on the traffic state, and uses the ELMAN neural network to learn historical data to predict the highway traffic state. In order to further prove the effectiveness of this method, this paper respectively compared the experimental results of the model proposed in this paper and the traditional ELMAN neural network model. The results show that the combined model has better prediction accuracy than the traditional ELMAN neural network model.

Acknowledgements
This work was financially supported by The National Key Research and Development Program of China(2017YFC0803900).

References
[1] Zhao F B. Sun M, Traffic status recognition method based on fuzzy cluster analysis [J]. Microcomputer Applications, 2006, 022(002):9-11,23.
[2] Henan Liu. The research of highway network traffic conditions evaluation and identification [D], Chongqing Jiaotong University, 2010.
[3] D’Este G M, Zito R, Taylor M A P. Using GPS to Measure Traffic System Performance[J]. Computer-Aided Civil and Infrastructure Engineering, 2010, 14(4):255-265.
[4] Maria, S. G. J., Evangelos, M., Panagiotis, T., Iraklis, S., Luigi, S., & Georgia, A.. Multisource Data Framework for Road Traffic State Estimation[J]. Journal of Advanced Transportation, 2018;1-9.
[5] Shi Z Q, Fang Y, Chen H Z. Traffic State Index Prediction Model Based on Intelligent Methods [J]. Applied Mechanics and Materials, 2014:1508-1513.
[6] Doulamis A D, Doulamis N D, Kollias S D. An adaptable neural-network model for recursive nonlinear traffic prediction and modeling of MPEG video sources[J].IEEE Transactions on Neural Networks, 2003, 14(1):150-166.
[7] Kumar K, Parida M, Katiyar V K. Short Term Traffic Flow Prediction for a Non Urban Highway Using Artificial Neural Network[J]. Procedia Social & Behavioral Sciences, 2013, 104:755-764.
[8] Zhao J Y, Jia L, Yang L C. Traffic flow prediction of RBF neural network based on particle swarm optimization [J]. Journal of Highway and Transportation Research and Development, 2006(07):116-119.
[9] Wang Yan, Hua W, Li min X. Highway traffic prediction with neural network and genetic algorithms[C]/ Vehicular Electronics and Safety, 2005. IEEE International Conference on. IEEE, 2005.
[10] Xu R, Zhou D, Jiang S Z. Adaptive particle swarm neural network traffic flow prediction model [J]. Journal of Xi’an Jiaotong University, 2015, 49(010):103-108.
[11] Li H Q. Simulation experiment research on driving behavior difference under different traffic flow conditions [D]. Kunming University of Science and Technology, 2012.
[12] Ma L., Wang S F, Liu J Y, Deng Y Y. Research on Application Requirements and Basic Architecture of Big Data in Expressway Operation [J]. Technology of Highway and Transport, 2017(06):126-128.

[13] CHEN Chuan. Design of Expert Research and Judgment System Based on Analytic Hierarchy Progress[J]. Computer and Information Technology, 2020,(28)04,1-4

[14] Liu Z Y, Li R M, Liu Y Q. Graded Warning Methods of Dangerous Degree of Typical Traffic Violations on Highways[J]. Journal of Transport Information and Safety, 2017(4).

[15] Z. W. Luo, Z. J. Wu, Z. J. Han, “Radial basis function neural network for a traffic flow model,” Journal of Tsinghua University, vol. 41, pp. 106-110, September 2001. (in Chinese)

[16] Y. G. Sa, Y. C. Hu, J. M. Xu, “The neural network model of freeway dynamic traffic flow,” Journal of South China University of Technology, vol. 30, pp. 91-94, August 2002. (in Chinese)

[17] J. M. Xiao, T. F. Zhang, X. H. Wang, “Neural network model of freeway traffic flow based on rough sets,” Computer Engineering, vol. 31, pp. 193-196, May 2005. (in Chinese)

[18] X. R. Liang, Y. X. Wei, “Freeway traffic flow modeling based on recurrent neural network and wavelet transform,” Proceedings of the International Conference on Transportation Engineering, Chengdu, China, pp. 1088-1093, July 2007.