A Novel Traffic Flow prediction model based on the Improved Extreme Learning Machine

Licheng Cui\textsuperscript{a}, Yingfei Zhan\textsuperscript{a, b}, Genning Zhang\textsuperscript{c}, Huawei Zhai\textsuperscript{d, r}

\textsuperscript{1}Police Information Department
Liaoning Police College
Dalian, Liaoning, China

\textsuperscript{2}Police Information Department
Liaoning Police College
Dalian, Liaoning, China

\textsuperscript{3}Police Information Department
Liaoning Police College
Dalian, Liaoning, China

\textsuperscript{4}Information Science and Technology
Dalian Maritime University
Dalian, Liaoning, China

\textsuperscript{a}clelicheng@163.com \textsuperscript{b}1106571236@qq.com \textsuperscript{c}529049197@qq.com
\textsuperscript{d}zhw@dlmu.edu.cn

ABSTRACT: The changes of urban traffic flow is affected by many factors, and during predicting the traffic flow, it should consider each factor. The paper considers the characteristic between time and traffic flow, and introduces the time factor as a separated input parameter. Besides, some unexpected situations have large impacts on the predicting results, so, the unexpected factor is introduced as an improved factor to the input samples. Based on these, a novel prediction model is proposed, called at-ELM, which determines the hidden node number by using PCA, and reduces the impacts of the hidden node numbers on the predicting results. After performance tests and analysis, the novel model is a little improved in prediction accuracy and completely acceptable.

1. INTRODUCTION

Extreme learning machine\textsuperscript{[1]}, ELM for short, is a special learning algorithm for the single-hidden feed-forward neural network. After selected the parameters of the input layer and hidden layer randomly, it only need to adjust the output weight of hidden layer.

Comparing with traditional learning algorithms, ELM is simple and high efficient, but there are some points needing to be solved, such as the hidden nodes number, the input information and so on. So, introducing wavelet transform and genetic algorithm, some improved ELM algorithms are proposed\textsuperscript{[1-3]}. Chen Haipeng\textsuperscript{[4]} proposed a combination predicting model, after introducing the wavelet transforming function, it used the ELM to predict the decomposed data respectively, and finally, it is to reconstitute the predicting results. The novel method didn’t modify the ELM itself, so the performance and characteristics of ELM cannot be effectively improved. Du Zhanlong and Li Fanjun
[3, 5] introduced the concept of sensitivity, proposed a novel self-adaptive ELM algorithm, dynamically removing hidden nodes based on its sensitivity, and strictly managed the hidden node number, these methods reduced its computational complexity, greatly improved the adaptive ability and generalization performance, but lack of new hidden node adding. Yang L[6] introduced the deep learning method, proposed a combination predicting method, it used the deep learning method to reduce the dimensionality of the measurement data, which made it have good separability in low-dimensionality space, and then, introduce the differential evolution algorithm into the artificial bee colony, to optimize the output weight and other parameters of ELM, in order to improve the final precision. The sequential learning algorithm is common in the support vector machine and RBF neural network, by referencing its idea, Li G[7], Alencar A S C[8] and Shao Z[9] proposed the novel ELM algorithms, which dynamically add and remove hidden nodes, strictly control the hidden node number, improve its flexibility and adaptability.

Based on the analysis above, after deeply researching the time series data of the traffic flow, the input dimension of ELM is extended, the time point parameter and the accident factor parameters are introduced, and an improved ELM is proposed, at-ELM, which can effectively improve the accuracy of traffic flow prediction at different times.

2. EXTREME LEARNING MACHINE

ELM is a special kind of single hidden layer feed-forward neural network, proposed by Guangbin Huang in 2006.

For giving training sample \((x_i, y_i), i = 1, \ldots, N, x_i \in \mathbb{R}^n\) is the input vector, \(y_i \in \mathbb{R}^m\) is the expected output vector. For the \(i\) th training sample, the algorithm output is \(y_i^* = [y_{i1}, y_{i2}, \ldots, y_{im}]^T\), \(y^*_i = \sum_{j=1}^{m} \beta_j G(a_j, b_j, x_i), i = 1, \ldots, N\).

For giving training sample, the output of ELM algorithm is

\[
H \beta = Y
\]

in which, \(H = [h_1 \cdots h_N]\) is the output matrix of the hidden layer, \(h_i = [G(a_1, b_1, x_i) \cdots G(a_n, b_n, x_i)]^T\), \(\beta\) is the connecting weight of the output layer, \(\beta = [\beta_1 \cdots \beta_m]\), \(Y = [y_1^* \cdots y_n^*]^T\).

ELM theory analysis believes that when the hidden node number is equal to the sample number, \(N = M\), it is only to adjust the output weight, the output error can reach zero, \(\|H \beta - Y\| = 0\), and the optimal solution of output weight \(\beta\) is

\[
\beta = H^T Y
\]

From the perspective of practical application, the impact of training samples on the ELM model increases as the amount of data increasing. In the context of big data, for achieving zero output error, it needs large amount of resources, which make the model complexity more and more. But, in real application, it is hard to achieve zero error, the error is small enough to be accepted. So, it can be to reduce the impact of training sample size on the model within the allowable error range.

Besides, in case of limited or small sample size, the value \(M\) has a relatively little impact on the computational complexity of the model. But, as the amount of data increasing, hidden nodes \(M\) could not be unlimitedly increased in order to achieve zero training error. So, it needs to control the hidden node number \(M\), the ELM algorithm believes that when \(N < M\), it can be guaranteed that any size training error can be obtained.

Based on the analysis above, the optimal solution of the output weight \(\beta\) is

\[
\beta = H^T \tilde{Y}
\]

in which, \(\tilde{Y}\) is Moore–Penrose generalized inverse of matrix \(H\). if \(\text{rank}(H) = M\), the equation (3) can be transformed to

\[
\beta = (H^T H)^{-1} H^T Y
\]

2
3. IMPROVED ELM

3.1 Time Factor
Urban traffic flow data is the time series data, which is discrete in time, and the time is an important parameter, in each time, there is a corresponding sample data, so the traffic flow at this moment reflects the nature of the traffic at that moment. For example, 7:00-8:00 is the traffic peak time, in the period, sampling data in the time \( t_{1}, t_{2}, \ldots, t_{n} \) reflects the characteristics at this moment, it is traffic peak hour, traffic flow is very large, and traffic status is red warning. 9:00-10:00 is the normal traffic time, in which the traffic is relatively stable, the sampling data value is a little small comparing with the one in peak time, and there will no or very few traffic flow surges.

After analyzing the traffic flow data in the same sampling time for consecutive days, the traffic flow is roughly the same, or it is fluctuating within a small range from a fixed value, so, these data reflect the changing characteristics of traffic in the time.

So, analyzing the relationship between the sampling time and sampling data on different time scales, it is known that the sampling time is a very important parameter, which can describe the changing characteristics of traffic flow.

Normally, the sample time is not used as a input parameter in the analyzing, which needs to be transformed, it is

\[
t_{i} = \left[ g(t) \times 60 \right] / itv
\]

in which, \( g() \) is the transformed operation, which is used to transform the time to time ID, it is the mapping from 2:30 to 2.5, \( itv \) is the sampling time interval.

3.2 Unexpected Factor
Commonly, urban traffic is affected by unexpected factors, such as weather, large event, holiday and so on. In this case, the traffic flow in a certain section or area will suddenly increase, it will lead to the changes of the traffic status, and finally, it will result in the traffic congestion and other problems, which bring some difficulties to the management of traffic departments. So, it need to predict the traffic flow changes under the influence of unexpected factors instantly and quickly, before the difficulties occur, timely and effectively take measures to avoid or reduce the impacts on the traffic flow.

Usually, unexpected factors appear separately, such as raining day, a large amount of water on the pavement and so on. But, in some cases, some unexpected factors appear at the same time, such as raining in holiday, large event in holiday, hosting large event on raining day in holiday, so in this case, these unexpected factors need to be analyzed comprehensively.

In total, before predicting traffic flow, it needs to fully consider the impacts of these unexpected factors. So, the unexpected factor \( af \) is introduced, which is the analyzing result of multivariable function, \( af \geq 0 \), \( af \) is described as equation (6),

\[
af = F(\lambda, S, W, H, A, TA, RS, EX)
\]

in which, \( \lambda \) is the transform parameter, \( \lambda = [\lambda_{s}, \lambda_{w}, \lambda_{h}, \lambda_{a}, \lambda_{r}, \lambda_{e}] \), \( \lambda \in [0,1] \), is used to show the extent to which the unexpected factors affect the traffic flow. when \( \lambda_{i} = 0 \), it means that the \( i^{th} \) unexpected factor has little impacts on the traffic flow, when \( \lambda_{i} = 1 \), it means that the factor has large impacts on the traffic flow, in a leading role. Parameter \( S \) is the season factor, \( W \) is the weather factor, \( H \) is the holiday factor, \( A \) is the large event factor, \( TA \) is the traffic accident factor, \( RS \) is the road condition factor, and \( EX \) is the expansion factor.

3.3 Improved Algorithm
Based on the analysis above, the time factor and unexpected factor are introduced to the elm, the time factor is as a separate input factor, and the unexpected factor is used to correct each input data, showing in Fig 1
The description of the novel model: for given input vector \( (X,t,\lambda) \), \( X = (x_1, x_2, \ldots, x_n) \), \( t = (t_1, t_2, \ldots, t_n) \), \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \), the optimal solution of the output weight \( \beta \), \( \beta = (H^T H)^{-1} H^T Y \), in which, \( H = [h_1, \ldots, h_n] \), \( h_i = [G(a,b,\lambda_i x_i) \ G(a,b,t_i)]^T \).

Notes:
N1. Before determine the unexpected factor \( \lambda \), it should be known that \( \lambda \) is not a constant, it is a dynamical value, which is changing over time. So, \( \lambda \) is a time function value.
N2. Hidden node number does not need to determine randomly, too many or too few hidden nodes will lead to be bad predicting results. So in at-ELM, the principal component analysis is introduced, to determine the hidden node number by the principal component number.

4. PERFORMANCE TEST
For deeply analyzing the performances of at-ELM, the performance tests are carrying out comparing with ELM and TFPBCM. The traffic flow data from a section in Dalian is obtained to do the test, and its total amount is 10080, in which, the 6720 data is the training samples, the rest is the testing samples, and all the tests are going under the Matlab R2016a.

![Fig.2 Traffic flow prediction result of a day](image)

Table 1. Error evaluation

| Day    | mAE  | MAE  | MARE  |
|--------|------|------|-------|
| Monday | 16   | 8.2245 | 0.0689 |
| Tuesday| 18   | 8.4015 | 0.0637 |
| Wednesday| 23 | 9.0528 | 0.0691 |
| Thursday| 16   | 9.7322 | 0.0813 |
| Friday | 22   | 8.2928 | 0.0719 |
The prediction error of each day is in Table 1, the evaluating indexes are maximum mean absolute error (mAE), mean absolute error (MAE) and mean absolute relative error (MARE)[10, 11]. After deeply analyzing, the mAE is a little larger in Wednesday, Friday and Saturday, its value is more than 20, but considering the situation in that three days, it is normal. Besides, the largest MAE is 9.7322 in Thursday, and also in the normal range. MARE is fluctuating around 0.07, and its max is less than 0.09, this mean the prediction precision is more than 90%, and fully meets the real needs.

Selecting the ELM[1, 12] and TFPBCM[13] to do the comparing test, the comparing equation is

\[
    z = \left( \frac{1}{i} \sum_{i} \pi_i - \frac{1}{i} \sum_{i} \hat{a}_i \right) / \left( \frac{1}{i} \sum_{i} \hat{a}_i \right)
\]

in which, \( \pi_i \) is the predicting result in \( i^{th} \) day of the first prediction model, \( \hat{a}_i \) it the predicting result in \( i^{th} \) day of the second prediction model, and \( z \) is the improved rate of the first model versus the second one.

Table 2. Improve rate (mAE)

|       | ELM  | at-ELM | TFPBCM |
|-------|------|--------|--------|
| ELM   | 0    | -0.15347 | -0.24125 |
| at-ELM| -0.15347 | 0      | -0.011316 |
| TFPBCM| -0.24125 | -0.011316 | 0      |

Table 3. Improve rate (MAE)

|       | ELM  | at-ELM | TFPBCM |
|-------|------|--------|--------|
| ELM   | 0    | -0.18451 | -0.21084 |
| at-ELM| -0.18451 | 0      | -0.011639 |
| TFPBCM| -0.21084 | -0.011639 | 0      |

Table 4. Improve rate (MAPE)

|       | ELM  | at-ELM | TFPBCM |
|-------|------|--------|--------|
| ELM   | 0    | -0.19815 | -0.28427 |
| at-ELM| -0.19815 | 0      | -0.01278 |
| TFPBCM| -0.28427 | -0.01278 | 0      |

Table 2, Table 3 and Table 4 show the comparing analyzing results of ELM, at-ELM and TFPBCM, in which, zero means the model itself is comparing, and the negative values indicate that the prediction accuracy of the model is improved. Conversely, positive values indicate that the prediction accuracy of the model is reduced. Overall analysis, the model at-ELM proposed by the paper is improved in prediction accuracy comparing with ELM and TFPBCM, the max is nearly 20%.

In summary, after the comparing test and analyzing above, at-ELM proposed by the paper is better in prediction accuracy, and fully meets the real requirements.

5. CONCLUSION

For solving the impacts of the time and unexpected factor, a novel model is proposed, called at-ELM, which introduces the time factor and unexpected factor as input parameters. In the predicting process, at-ELM considers the time and unexpected factors, such as weather, large event and so on. Besides,
the PCA is introduced to determine the hidden node number, reduce the impacts of the too many or too small hidden nodes on the predicting results. After the performance test, the predicting results of at-ELM is acceptable completely.

ACKNOWLEDGMENTS
This research is partially supported by Liaoning Provincial Natural Science Foundation of China (Grant No.20170520196), and by the National Natural Science Foundation of China (Grant No. 51479021, 51679025).

REFERENCES
[1] Huang G, Zhu Q, Siew C. Extreme learning machine: Theory and applications. Neurocomputing, 2006, 70(1): 489-501
[2] Min H, De-Cai L. An Norm 1 Regularization Term ELM Algorithm Based on Surrogate Function and Bayesian Framework. ACTA AUTOMATICA SINICA, 2011, 37(11): 1344-1350
[3] Fan-Jun L, Hong-Gui H, Jun-Fei Q. Pruning algorithm for extreme learning machine based on sensitivity analysis. Control and Decision, 2014, 29(6): 1003-1008
[4] Haipeng C, Xiaohang T, Yu W, et al. Short-Term Parking Space Prediction Based on Wavelet-ELM Neural Networks. Journal of Jilin University(Science Edition), 2017, 55(2): 388-392
[5] DU Zhan-long, Xiao-Min L, Lei-Ping X, et al. Improved sensitivity-analysis based pruning extreme learning machine. Control and Decision, 2016, 31(2): 249-255
[6] Yang L, Li Y, Li Z. Improved-ELM method for detecting false data attack in smart grid. International Journal of Electrical Power & Energy Systems, 2017, 91("":): 183-191
[7] Li G, Liu M, Dong M. A new online learning algorithm for structure-adjustable extreme learning machine. Computers & Mathematics with Applications, 2010, 60(3): 377-389
[8] Alencar A S C, Rocha Neto A R, Gomes J P P. A new pruning method for extreme learning machines via genetic algorithms. Applied Soft Computing, 2016, 44: 101-107
[9] Shao Z, Er M J. An online sequential learning algorithm for regularized Extreme Learning Machine. Neurocomputing, 2016, 173(2016): 778-788
[10] Zhang H, Zhang S, Yin Y. Online sequential ELM algorithm with forgetting factor for real applications. Neurocomputing, 2017, 261: 144-152
[11] Cui L, Zhai H, Qu Z, et al. A variable-structure sequential ELM algorithm based on the characteristics of time-varying system, 2018
[12] Xuemei J, Junwei G, Xin L, et al. Short-Term Traffic Flow Forecasting Based on Extreme Learning Machine. Journal of Qingdao University(Engineering & Technology Edition), 2015, 30(4): 58-61, 67
[13] Lanlan R, Qinning L. Short-term Traffic Flow Prediction Algorithm Based on Combined Model. Journal of Electronics & Information Technology, 2016, 5(38): 1227-1233