

Predictions on Seaports Freight Throughput based on the Extreme Learning Machine Neural Network

Nan Yao

China Academy of Railway Sciences, Beijing, 100081, China

Email: yaonan@rails.cn

Keywords: Extreme Learning Machine; Neural Networks; Support Vector Machine; Seaports Freight Throughput; Prediction Model

Abstract. The researches and predictions on seaports freight throughput in China have become increasingly important as the fast growth of Chinese economy and the development of “the Silk Road Economic Belt and the 21st-Century Maritime Silk Road”. Extreme Learning Machine, a relatively new neural network algorithm published in the last a few years, is creatively utilized in this paper to mine historical data of seaports freight throughput in China. A new prediction model is built to predict seaports freight throughput in China. According to test results, the model shows good performances. The researches presented in this paper may be valuable for both real applications and theoretical studies.

Introduction

China has a long sea coast and many large seaports. The aggregated seaports freight throughput in China is more than 200 million TEUs (twenty foot equivalent units) per year [1]. Seaports freight throughput can be a key factor to measure economic growth and social developments. It is usually evaluated carefully when governments make their national strategies and specific policies. Besides, these seaports are connected with immense railway networks and highway networks, and are influencing integrated developments of the whole transportation industry. Also, the related study results may provide solid supports to the initiatives of “the Silk Road Economic Belt and the 21st-Century Maritime Silk Road”. Therefore, a great deal of significance is attached to researches shown in this paper.

The seaports freight throughput is decided by many different factors and the interactions among them. In this paper it is examined and studied from a point of view of data mining. Some advanced data mining algorithms published in the last several years [2] have been taken into consideration to ensure a better usage of the historical data of seaports freight throughput in China. The goal of this paper is to build a more effective and efficient prediction model by the data and to discuss the value of such models.

Previous publications from the related fields in the academic world have been examined. The major data mining techniques have been used in these publications include polynomial regression, Support Vector Machine (SVM), BP neural network, RBF neural network, and grey model [3][4][5][6][7]. Among these algorithms, neural network algorithms and SVM perform better in the seaports freight throughput predictions [2] [8]. However, the SVM and the neural network methods used in the previous predictions may be inefficient or too complex, because complicated and multiple iterations are required to implement these algorithms. Moreover, prediction accuracy or prediction speed of all the previously used algorithms may not be satisfied enough to predict seaports freight throughput [2] [8]. The Extreme Learning Machine, a relatively new neural network algorithm, may provide a solution to the above problems.

In short, it is worthwhile to apply ELM to seaports freight throughput. Thus, a prediction model based on ELM is developed in this paper.

Prediction Model

The seaports freight throughput prediction is a supervised regression data mining task. It is suitable to apply ELM to this type of tasks [9]. This algorithm can be utilized widely to provide accurate and efficient predictions [9][10][11][12].

The fundamental principles of ELM are described below.

ELM is a relatively simple algorithm with three steps. For a given training dataset: $\text{TrainData} = \{ (x_i, t_i) \mid x_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m, i=1, \dots, N \}$, hidden node output function is $G(a, b, x)$, and hidden nodes amount is L . The three steps are as following:

Step 1

Assign values to hidden node parameters randomly: (a_i, b_i) , $i=1, \dots, L$. According to simulation results, the ELM model is good enough as long as L is larger than 800. Therefore, L is set as 1000 in this paper.

Step 2

Calculate hidden layer output matrix which is named as H .

$$\sum_{i=1}^L \beta_i G(a_i, b_i, x_j) = t_j, j = 1, \dots, N$$

This equals to $H\beta = T$. The i^{th} column in H is the output from the i^{th} hidden node, and the corresponded inputs are x_1, x_2, \dots, x_N .

$$H = \begin{bmatrix} h(x_1) \\ \mathbf{M} \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(a_1, b_1, x_1) \wedge G(a_L, b_L, x_1) \\ \mathbf{M} \\ G(a_1, b_1, x_N) \wedge G(a_L, b_L, x_N) \end{bmatrix}_{N \times L}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \mathbf{M} \\ \beta_L^T \end{bmatrix}_{L \times m}$$

$$T = \begin{bmatrix} t_1^T \\ \mathbf{M} \\ t_N^T \end{bmatrix}_{N \times m}$$

Step 3

Calculate output weights β : $\beta = H^\dagger T$

H^\dagger is the Moore-Penrose generalized inverse of hidden layer output matrix H .

In this paper, the three calculation steps above are integrated into a machine learning process. During this learning process, some historical data of seaports freight throughput in China is divided into a training dataset and a testing dataset. The ELM prediction model generated by the training dataset is utilized to produce predicted seaport freight throughput that can be compared to the real throughput in the testing dataset.

According to the methodologies and mathematical models described above, computer programs are written via Matlab. After the machine learning process, training time, testing time and prediction accuracy will be output by computer(s).

The algorithm provided above has several advantages:

First, ELM can show good performances in enhancing prediction accuracy and in reducing over-fit problems.

Second, compared to BP neural networks, RBF neural network, SVM and other advanced data mining algorithms, the ELM is relatively simple and does not require iterative tuning.

Third, the learning speed of the ELM is extremely fast [9][10][11][12].

Fourth, the hidden node parameters a_i and b_i are independent to each other, and they are independent to training data. In the ELM calculation process, the hidden node parameters can be set before seeing training data [11][12].

Test Design

The historical data used in this paper is published by the Transportation Department of China. The data includes monthly seaport freight throughput in China, from February, 2001 to March, 2013. The training dataset consists of the corresponded data from February, 2001 to December, 2011. The testing dataset includes the other part of the historical data (i.e. the data from January, 2012 to March, 2013).

During a data understanding process, two special situations are noticed. First, the seaports freight throughput in Februaries shows a unique fluctuation. Second, there is a significant decrease of the seaports freight throughput in November, 2008. This may be caused by the International Financial Crisis 2008. Therefore, all the training data is further divided into three parts: the seaports freight throughput data from March, 2001 to October, 2008 but exclude Februaries, the seaports freight throughput data from November, 2008 to December, 2011 but exclude Februaries, and all the Februaries in the training dataset. It is also noticed that some data in Decembers is missing value.

Next, according to the calculation steps and the principles discussed in the preceding part of the paper, codes are input to Matlab to apply the Extreme Learning Machine algorithm to the three parts of the training datasets respectively. Then the three resultant models are integrated by time (month).

After this training process, a prediction model is built and is used to predict seaport freight throughput between January, 2012 and March, 2013. Also, the real seaports freight throughput in this period is listed. Finally, more comparisons and analysis are conducted.

Test results

The seaports freight throughput between January, 2012 and March, 2013 is as following:

Month	Real Seaports Freight Throughput (in ten thousand TEUs)	Predicted Seaports Freight Throughput (in ten thousand TEUs)
2012-1	1231.0	1313.2
2012-2	1059.5	991.6
2012-3	1266.2	1337.7
2012-4	1288.2	1349.8
2012-5	1353.3	1362.1
2012-6	1338.6	1374.9
2012-7	1352.6	1387.3
2012-8	1346.1	1399.7
2012-9	1415.3	1412.8
2012-10	1336.9	1425.2
2012-11	1350.1	1437.6
2012-12	N/A	1449.3
2013-1	1400.2	1462.5
2013-2	1075.0	1069.6
2013-3	1373.1	1485.6

Total training time: 0.0624s

Total testing time: <0.0001s

Computing environment: CPU is Core i3 3.1 GHz; with 4 GB RAM and 64 bit Windows 7

RMSE: 55.36 (unit: ten thousands TEUs)

Prediction Accuracy (1-RMSE/Average Real Seaports Freight Throughput): 95.74%

Conclusion

The conclusion of the above research is summarized as following:

First, given the complexities of the international trade environments in recent years, the ELM neural network algorithm shows a very fast learning speed and high prediction accuracy in predictions on seaports freight throughput in China.

Second, compared to other advanced algorithms such as SVM, BP neural network and RBF neural network, the ELM is easier to be understood and utilized. With the help of Matlab or other programming tools, an ELM programming package can be written, and the ELM can be implemented efficiently and easily.

Third, based on the ELM research methods discussed in this paper and related tests presented in the paper, the ELM neural network algorithm is expected to play a more important role in machine learning related areas.

References

- [1] Y.-J. Yang, M.-H. Cao. Ports Freight Throughput in China has reached more than 200 million TEUs per year [J]. *Zhujiang Water Transportation*, 2015(9):28.
- [2] G.-B. Huang, H.-M. Zhou, X.-J. Ding, and R. Zhang. Extreme Learning Machine for Regression and Multiclass Classification”[J] *IEEE Transactions on Systems, Man, and Cybernetics—PART B: Cybernetics*, 2012(2): 513-529.
- [3] Yang Y. Development of the regional freight transportation demand prediction models based on the regression analysis methods[J]. *Neurocomputing*, 2015, 158:42–47.
- [4] Pu Z, Yang L, Guo Z. Applied Research on Logistics Demand Prediction Based on Support Vector Machine of Genetic Algorithm[C]// *Fourth International Conference on Computational & Information Sciences*. IEEE Computer Society, 2011:510-513.
- [5] Huang Y, Lin Y. Freight Prediction Based on BP Neural Network Improved by Chaos Artificial Fish-Swarm Algorithm[C]// *Computer Science and Software Engineering, International Conference on*. IEEE, 2009:1287-1290.
- [6] J.-T. Li, C.-W. Ma, G.-Q. Sun. Dynamic Ports Freight Prediction Based on RBF Neural Network [J]. *Journal of Dalian Jiaotong University*, 2008, 29(4):27-32.
- [7] Y. Ou. Discussions on Ports Freight Throughput Prediction Methods [J]. *Water Transportation Engineering*, 2003(1):22-25.
- [8] Al-Deek H. Which Method Is Better for Developing Freight Planning Models at Seaports—Neural Networks or Multiple Regression?[J]. *Transportation Research Record Journal of the Transportation Research Board*, 2001, 1763(1):90-97.
- [9] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: Theory and applications. [J] *Neurocomputing*, 2006, 70: 489–501.
- [10] W. Deng, Q. Zheng, and L. Chen. Regularized extreme learning machine [C]. *Proc. IEEE Symp. CIDM*, 2009: 389–395.
- [11] G.-B. Huang, Q.-Y. Zhu, K. Z. Mao, C.-K. Siew, P. Saratchandran, and N. Sundararajan. Can threshold networks be trained directly?[J] *IEEE Trans. Circuits Syst. II*, 2006, 53: 187–191.
- [12] M.-B. Li, G.-B. Huang, P. Saratchandran, and N. Sundararajan. Fully complex extreme learning machine [J]. *Neurocomputing*, 2005, 68: 306–314.