2017 16th IEEE International Conference on Machine Learning and Applications

An Ensembled RBF Extreme Learning Machine to Forecast Road Surface Temperature

Bo Liu[1,2], Shuo Yan[2], Huanling You[3], Yan Dong[4], Jianqiang Li[2], Yong Li[2], Jianlei Lang[5,6], Rentao Gu[7],
[1]Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing 100124, China
[2]School of Software Engineering, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[3]Institute of Urban Meteorology, China Meteorological Administration, Beijing 100089, China
[4]Beijing Meteorological Service Center, Beijing 100089, China
[5]Key Laboratory of Beijing on Regional Air Pollution Control, Beijing University of Technology, Beijing 100124, China
[6]College of Environmental & Energy Engineering, Beijing University of Technology, Beijing 100124, China
[7]Beijing Laboratory of Advanced Information Networks, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
boliu@bjut.edu.cn, yanshuok@163.com, you_hl@sohu.com, dongyancandice@163.com, lijianqiang@bjut.edu.cn, li.yong@bjut.edu.cn, jllang@bjut.edu.cn, rentaogu@bupt.edu.cn

***Abstract*—At present, high road surface temperature (RST) is threatening the safety of expressway transportation. It can lead to accidents and damages to road, accordingly, people have paid more attention to RST forecasting. Numerical methods on RST prediction are often hard to obtain precise parameters, whereas statistical methods cannot achieve desired accuracy. To address these problems, this paper proposes GBELM-RBF method that utilizes gradient boosting to ensemble Radial Basis Function Extreme Learning Machine. To evaluate the performance of the proposed method, GBELM-RBF is compared with other ELM algorithms on the datasets of airport expressway and Badaling expressway during November 2012 and September 2014. The root mean squared error (RMSE), accuracy and Pearson Correlation Coefficient (PCC) of these methods are analyzed. The experimental results show that GBELM-RBF has the best performance. For airport expressway dataset, the RMSE is less than 3, the accuracy is 78.8% and PCC is 0.94. For Badaling expressway dataset, the RMSE is less than 3, the accuracy is 81.2% and PCC is 0.921.**

Keywords—road surface temperature; prediction; neural networks; gradient boosting

I. Introduction

Nowadays, expressway has been more and more important for transportation. Transportation is particularly affected by disastrous weather because of large traffic flow and high vehicle speed. To ensure the safety of transportation, people have paid more attention to obtaining the information of disastrous weather in expressway. Among the many meteorological factors influencing the expressway, the high road surface temperature (RST) is one of the most important factors. High RST can make tire easy to explode and leads to accident and it also leads the asphalt road to swell, pits and other phenomena, and with long-time rolling by vehicles can cause a large area of intense damage. The high RST will not only lead to an accident but also damage the road. Therefore, forecasting RST is significant to prevent traffic accidents and damages. With the arrival of big data era, it is easier to solve expressway RST forecasting by machine learning than traditional ways. Meteorological institution has accumulated plenty of road data in past decades. By applying data mining algorithms on these data, we can build more accurate models.

Worldwide researchers have contributed a lot to the study of RST forecasting. Existing methods are mainly divided into two parts: numerical methods and statistical ones. Numerical methods use physics and math to establish an equation for forecasting RST. Barber [1] regarded roads as a semi-infinite mass with uniform texture and built a model to predict the highest temperature. Sass [2] established a model that can forecast up to a range of at least 3 hours; this model is based on the equation of heat. Feng et al. [3] utilized conservation of energy and built an hourly RST forecasting model. Meng et al. [4] combined numerical simulation product Common Land Mode (CoLM) [5] and BJ-RUC (Beijing-rapidly update cycle) [6] and established a model which could forecast up to a range of 3 to 24 hours. Statistical methods build a model based on historical data and are often easier than numerical methods. Diefenderfer et al. [7] , Qu et al. [8] and Li et al. [9] used linear regression to build a daily highest and lowest RST prediction model for different areas.

Numerical methods have better universality, but their parameters are hard to obtain. Traditional statistical methods are based on multiple linear regression (MLR). Its parameters are easy to be obtained, but its accuracy is not satisfied. In big data era, machine learning has been a very popular method for forecasting especially Artificial Neural Networks (ANNs). Extreme learning machine (ELM) [10] is a kind of ANNs which has high training speed and good generalization performance. In this paper, we combined ELM with BJ-RUC to build models. However, ELM cannot meet the needs of RST forecasting, so we utilized Radial basis function (RBF) to improve the generalization performance. RBF extracts abstract features from input features and achieves better generalization performance. ELM with RBF (ELM-RBF) [11] has been proved that RBF can improve ELM a lot. Because the parameters of RBF are randomly generated and the performance of ELM-RBF is highly depended on these parameters, ELM-RBF is unstable. Ensembling is a good way to stabilize an algorithm and obtain a robust model. Gradient boosting [12] is a very po

0-7695-6321-X/17/31.00 ©2017 IEEE
DOI 10.1109/ICMLA.2017.00-26

method in data mining contest, and it always performs well. L Singh et al. [13] has applied gradient boosting to ELM and got a good result. We used ELM-RBF as base estimator of gradient boosting and extended it to multi-output cases.

II. METHODOLOGY

A. ELM and ELM-RBF

ELM is a type of single-hidden layer feedforward neural networks (SLFNs). In ELM, the weight and bias between input layer and hidden layer are randomly assigned. Once weight and bias are determined, the activation H can be calculated, so the weight β between hidden layer and output layer is calculated by least square.

RBF neural network [14] is a type of SLFNs and the sigmoid function is replaced by RBF kernel. RBF neural networks are universal estimators [15] and possess the best approximation property [16], so they have good generalization performance and accuracy. Applying RBF kernel to ELM gives ELM better generalization performance and higher accuracy. In ELM-RBF, the parameters of RBFs are assigned arbitrarily. The activation H of hidden layer in ELM-RBF is defined as:

$$H = \exp(-\gamma \| x - c \|^2) \tag{1}$$

where x is the input vector. c is the center vector. If $y = \sigma^{-2}$, the kernel is known as the Gaussian kernel of variance σ^2 . In ELM-RBF $\gamma = \sigma^{-1}$ and vector c are uniformly randomly distributed in the interval [0, 1]. σ can be calculated like Gaussian kernel and β can be calculated like ELM.

B. Gradient ELM-RBF Boosting

Gradient boosting is a ML algorithm that produces a robust estimator in the form of an ensemble of weak estimators. Gradient boosting optimizes a loss function over estimator space iteratively choosing an estimator that points in the negative gradient direction. In this paper, we use ELM-RBF as the base estimator of the gradient boosting, this is called gradient ELM-RBF boosting (GBELM-RBF). The least square loss function is chosen as the loss function of ELM-RBF. Gradient boosting with least square loss function considers additive models of the following form:

$$F(x) = \sum_{m=1}^{M} \alpha h_m(x) \tag{2}$$

where $F(x)$ is the final model, α is learning rate, M is the number of weak estimators, $h_m(x)$ is the base estimator, in this paper, $h_m(x)$ is ELM-RBF. Gradient boosting builds additive model in a forward stagewise fashion:

$$F_m(x) = F_{m-1}(x) + \alpha h_m(x) \tag{3}$$

At each stage, an ELM is chosen to minimize the loss function L given current model $F_{m-1}(x)$:

$$F_m(x) = F_{m-1}(x) + \arg\min_h \sum_{i=1}^{n} L(y_i, F_m(x)) \tag{4}$$

where n represents the samples, y is the target value. For least square loss function, the initial model F_0 usually chooses the mean of target values. For multi-output case, F_0 choose the mean of target values of each output. The GBELM-RBF with least squares loss function was shown as below:

Algorithm 1 Gradient ELM Boosting Algorithm

Input: data x , target y , learning rate α, number of estimators M

Output: Ensembled ELM-RBF $F_m(x)$

1. Initialize $F_0(x)$ as the mean of each output
2. For m =1 to M :

(a) For i = 1, 2, 3, ……, N, Compute the negative gradient r

$$r = -[\frac{\partial L(y, F(x_i))}{\partial F(x_i)}]_{F=F_{m-1}}$$

(b) Initialize a RBF-ELM $h(x)$, Arbitrary assigned the centers c of RBF kernels and calculate the σ of RBF kernels.

(c) ELM-RBF is base estimator $h(x)$ and is used to fit r

(f) $F_m(x) = F_{m-1}(x) + \alpha h_m(x)$

3. Output $F_m(x)$

The full training algorithm of GBELM-RBF was shown in Figure 1.

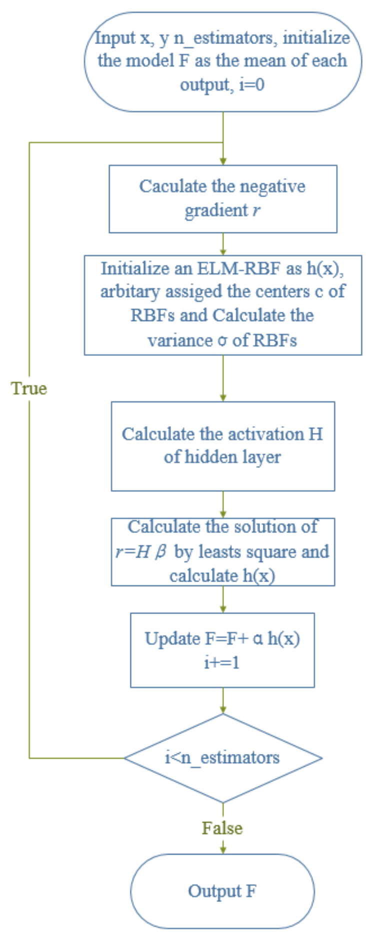

Fig. 1. The full training algorithm of GBELM-RBF

III. EXPERIMENT SETTINGS

A. Dataset

In this paper, we use the data of BJ-RUC (Beijing Rapid Update Cycle) and the data of Beijing road monitor station to forecast the RST. BJ-RUC forecasts more than 200 thousand sites of height fields, upward longwave radiation, ground surface pressure, humidity, downward shortwave radiation, 2-meter temperature, longitude 10-meter wind, altitude 10-meter wind and hourly accumulated precipitation. In this paper, we have selected the data of several nearest BJ-RUC forecast sites from the road monitoring station. Road monitor stations produce data per hour. In this paper, A1027 airport expressway and A1412 Badaling expressway were selected to analyze, because they have heavy traffic flow and complete data. We selected RST, total precipitation, precipitation intensity, road conditions, temperature as input features.

B. Evaluation

In this paper, we chose tree evaluation methods that were usually used in RST forecasting to evaluate the model.

- Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2} \tag{5}$$

- Pearson correlation coefficient (PCC):

$$PCC = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{O_i - \bar{O}}{s_O}\right)\left(\frac{P_i - \bar{P}}{s_P}\right) \tag{6}$$

Where O_i denotes the observed RST, P_i denotes the predicted RST, N denotes the number of evaluation samples, S denotes the standard deviation.

- Accuracy:

If the absolute error between observed value and predicted value is within ± 3 ℃, then the forecast is accurate.

IV. RESULT AND DISCUSSION

We selected the data of BJ-RUC and road monitor stations in past 72 hours, the data of BJ-RUC in future 24 hours. We made use of the data from November 2012 to August 2014 as training data, the data of September 2014 as testing data. We calculated the Pearson Correlation Coefficient of the features with each output and selected the features whose correlations with each output exceeded 0.5 as the input features. All the parameters were selected by gridsearch and 3-folders cross-validation.

A. Selection of the Number of BJ-RUC forecasting Sites

TABLE I. RMSE OF DIFFERENT METHODS AND RUC SITES

Stations	Methods	1	2	3	4	5
A1027	ELM	**3.916**	4.525	4.484	5.103	6.056
	ELM-RBF	**2.754**	3.131	3.177	3.502	3.670
	GBELM-RBF	**2.601**	2.855	2.994	3.056	3.190
A1412	ELM	**2.770**	3.980	4.160	5.016	6.197
	ELM-RBF	**2.553**	2.700	2.736	2.857	3.002
	GBELM-RBF	**2.326**	2.484	2.326	2.422	2.513

TABLE II. ACCURACY OF DIFFERENT METHODS AND RUC SITES

Stations	Methods	1	2	3	4	5
A1027	ELM	**57.5%**	52.5%	53.1%	48.6%	36.7%
	ELM-RBF	**75.5%**	71.9%	73.1%	65.9%	64.5%
	GBELM-RBF	**78.9%**	75.3%	76.3%	74.2%	72.1%
A1412	ELM	**60.7%**	55.2%	56.9%	47.9%	39.8%
	ELM-RBF	**76.4%**	74.1%	76.0%	73.2%	72.9%
	GBELM-RBF	**81.2%**	78.3%	80.8%	80.6%	79.0%

TABLE III. PCC OF DIFFERENT METHODS AND RUC SITES

Stations	Methods	1	2	3	4	5
A1027	ELM	**0.861**	0.811	0.808	0.747	0.710
	ELM-RBF	**0.932**	0.911	0.914	0.897	0.885
	GBELM-RBF	**0.940**	0.927	0.922	0.921	0.914
A1412	ELM	**0.816**	0.749	0.714	0.636	0.412
	ELM-RBF	**0.908**	0.898	0.891	0.878	0.862
	GBELM-RBF	**0.914**	0.913	0.922	0.917	0.910

TABLE IV. TIME OF DIFFERENT METHODS

Stations	Methods	Modeling time(s)	Forecasting time(s)
A1027	ELM	**0.1253**	**0.0004**
	ELM-RBF	0.3194	0.0135
	GBELM-RBF	17.1345	0.5530
A1412	ELM	**0.1419**	**0.0005**
	ELM-RBF	0.3810	0.0150
	GBELM-RBF	16.9862	0.5140

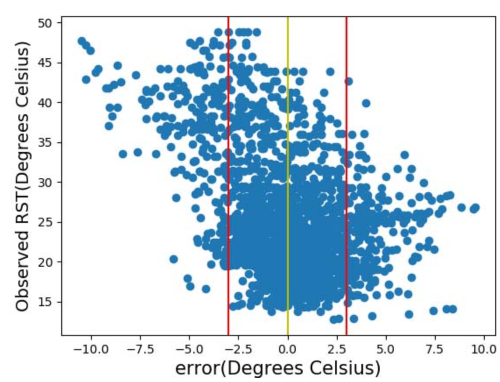

Fig. 2. Result of A1027 using GBELM-RBF with 100 hidden units, 100 estimators, and the learning rate is 0.5. .

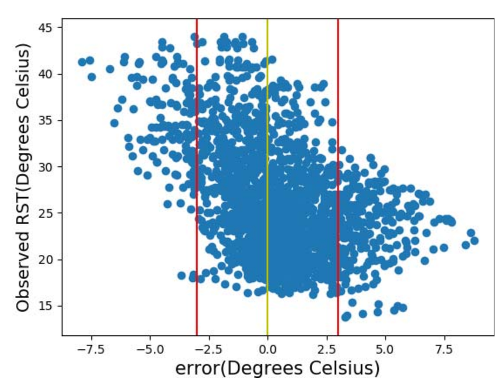

Fig. 3. Result of A1412 using GBELM-RBF with 100 hidden units, 100 estimators, and the learning rate is 0.5.

Tables I-III show the RMSE, PCC, accuracy of different methods and different RUC sites. We can conclude that adding BJ-RUC forecast sites does not reduce the RMSE, accuracy and PCC of any method. Moreover, more input features mean that the training time becomes longer. Therefore, one BJ-RUC forecasting site is the most appropriate.

By comparison, it is easy to find that GBELM-RBF has the best performance whether on station A1027 or station A1412. For station A1027, RBF kernel improves the accuracy by about 20% and reduces RMSE to less than 3℃. Besides, PCC is improved to more than 0.9. Gradient boosting further increases the accuracy to 78.8%. The RMSE and PCC are also improved. For station A1412, RBF kernel also works well on this monitoring station and improves the accuracy by more than 15%. Gradient boosting further improves the accuracy to more than 80%. RBF kernel improves generalization performance of ELM and achieves higher accuracy, because RBF kernel is capable of universal approximation. Gradient boosting stabilizes ELM and obtains a robust model. Figures 2-3 show the results of GBELM-RBF on station A1027 and A1412, respectively. The samples between the two red lines means it is accurate. From these figures, we can see that GBELM-RBF is on the lower side for high RST forecasting, but most data are forecasted accurately.

Compared with Ref. [4] who built a model on Badaling expressway and the PCC of his model is 0.9. The PCC of our model is 0.922. Ref. [9] built a model on statistical method and its accuracy in summer and autumn is 64.4% and 76.7%, respectively. In A1027 monitoring station, the accuracy of our model is 78.8%. In A1412 monitoring station, our model has the accuracy of more than 80%. Therefore, Combined GBELM-RBF with BJ-RUC can obtain a more accurate model than numerical method and traditional statistical method.

Table IV shows the modeling time and forecasting time of the models with 1 RUC sits. Sigmoid ELM has the shortest modeling time, it can complete modeling within 0.1 second. ELM-RBF completed modeling within 0.1 second too. People cannot feel an obvious difference. Because training GBELM-RBF needs to train lots of ELM-RBF, it takes much longer time than that of the other two methods. However, training GBELM-RBF only takes less than 20 seconds and it is acceptable. All these methods can forecast RST within 1 second, so people will not feel significant differences.

V. Conclusions and Future work

In this paper, we utilized gradient boosting to ensemble ELM-RBF. It improves the generalization performance of ELM-RBF and has higher accuracy. We test the performance of GBELM-RBF on RST forecasting and prove that combining GBELM-RBF with BJ-RUC can obtain a more accurate model than numerical methods and traditional statistical methods.

There are still lots of work to do in the future. We will try Xgboost to ensemble ELM. Pearson Correlation Coefficient was used to select features, it is a traditional way. In the future, we will try to combine deep learning with gradient ELM boosting, and this may get a better result.

Acknowledgment

This work is supported by Beijing Natural Science Foundation (4174082), Natural Science Foundation of China (61702021), General Program of Science and Technology Plans of Beijing Education Committee (SQKM201710005021), China Postdoctoral Science Foundation (2016M590027), Beijing Postdoctoral Research Foundation (2016ZZ-17, 2016ZZ-20), Fundamental Research Foundation of Beijing University of Technology (PXM2017_014204_500087), and Funds of Beijing Advanced Innovation Center for Future Internet Technology of Beijing University of Technology (BJUT). Corresponding author: Huanling You

References

[1] Barber, E.: Calculation of maximum pavement temperatures from weather reports, Highway Research Board Bulletin, 1957, (168)

[2] Sass, B.: A Numerical Forecasting System for The Prediction of Slippery Roads, Journal of Applied Meteorology, 1997, 36, (6), pp. 801-817

[3] Feng, T., and Feng, S.: A Numerical Model for Predicting Road Surface Temperature in The Highway, Procedia Engineering, 2012, 37, pp. 137-142

[4] Meng, C., and Liu, C.: Summer Road Temperature Disaster Forecast of Expressway in Beijing, J.of Institute of Disaster-Prevention Science and Technology, no. 03, 2009, pp. 26-29.

[5] Dai, Y., Zeng, X., Dickinson, R. et al.: The Common Land Model, 2003, 84, (8), pp. 1013-1023

[6] Wei, D., You, F., Fan, S., Yang, B., and Chen, M.: Evaluation and Analysis of Sounding Quality of Beijing Rapid Renewal Cycle Prediction System (BJ-RUC) Model, Journal of Meteorological, vol. 36, no. 8, 2010, pp. 72-80.

[7] Diefenderfer, B., Al-Qadi, I., Reubush, S., and Freeman, T.: Development and validation of a model to predict pavement temperature profile, in Editor (Ed.) Development and validation of a model to predict pavement temperature profile (Citeseer, 2003,edn.), pp.

[8] Qu, X., Wu, H., Zhang, Y., and Jia, J.: The Atmospheric Physical Quantity and Radar Echo Characteristic of A Strong Convective Weather, Journal of Arid Meteorology, no. 03, 2010, pp. 352-357.

[9] Li, L., Song, Y., and Han, G.: Pavement Temperature Characteristics And Prediction Model of Tai'an Expressway, Journal of Shandong Meteorology, no. 02, 2016, pp. 58-63.

[10] Huang, G., Zhu, Q., and Siew, C.: Extreme Learning Machine: Theory and Applications, Neurocomputing, 2006, 70, (1), pp. 489-501

[11] Huang, G., and Siew, C.: Extreme learning machine: RBF network case, in Editor (Ed.) Extreme learning machine: RBF network case (2004,edn.), pp. 1029-1036

[12] Natekin, A., and Knoll, A.: Gradient Boosting Machines, A Tutorial, Front Neurorobot, 2013, 7, pp. 21

[13] Singh, L., and Chetty, G.: Email Personalization and User Profiling Using RANSAC Multi Model Response Regression Based Optimized Pruning Extreme Learning Machines and Gradient Boosting Trees, in Arik, S., Huang, T., Lai, W.K., and Liu, Q. (Eds.): Lecture Notes in Computer Science (Springer Int Publishing Ag, 2015), pp. 302-309

[14] Kokshenev, I., and Braga, A.: A Multi-objective Learning Algorithm for RBF Neural Network, in Editor (Ed.) A Multi-objective Learning Algorithm for RBF Neural Network (2008,edn.), pp. 9-14

[15] Park, J., and Sandberg, I.: Universal Approximation Using Radial-Basis-Function Networks, Neural Comput, 2014, 3, (2), pp. 246-257

[16] Girosi, F., and Poggio, T.: Networks and the Best Approximation Property (Massachusetts Institute of Technology, 1989. 1989)