Computers in Industry 99 (2018) 294–302

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Road surface temperature prediction based on gradient extreme learning machine boosting

Bo Liu[a,b,*], Shuo Yan[b], Huanling You[c,d], Yan Dong[d], Yong Li[b], Jianlei Lang[e,f], Rentao Gu[g]

[a] Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing 100124, China
[b] School of Software Engineering, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[c] Institute of Urban Meteorology, China Meteorological Administration, Beijing 100089, China
[d] Beijing Meteorological Service Center, Beijing 100089, China
[e] Key Laboratory of Beijing on Regional Air Pollution Control, Beijing University of Technology, Beijing 100124, China
[f] College of Environmental & Energy Engineering, Beijing University of Technology, Beijing 100124, China
[g] Beijing Laboratory of Advanced Information Networks, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Keywords:
Road surface temperature
Prediction
Neural networks
Gradient boosting

ABSTRACT

The expressway is extremely important to transportation, but high road-surface temperatures (RST) can cause many traffic accidents. Most of the hourly RST prediction models are based on numerical methods, but the parameters are difficult to determine. Statistical methods cannot achieve the desired accuracy. To address these problems, this paper proposes a machine learning algorithm that utilizes gradient-boosting to assemble a ReLU (rectified linear unit)/softplus Extreme Learning Machine (ELM). By using historical data from the airport and Badaling expressways collected between November 2012 and September 2014, sigmoid ELM, ReLU ELM, softplus ELM, ReLU gradient ELM boosting (GBELM) and softplus GBELM were applied for RST forecasting, RMSE (root mean squared error), PCC (Pearson Correlation Coefficient), and the accuracy of these methods were analyzed. The experimental results show that ReLU/softplus can improve the performance of traditional ELM, and gradient boosting can further improve its performance. Thus, we obtain a more accurate model that utilizes GBELM with ReLU/softplus to forecast RST. For the airport expressway, our proposed model achieves an RMSE within 3 °C, an accuracy of 81.8% and a PCC of 0.954. For the Badaling expressway, our model achieves an RMSE within 2 °C, an accuracy of 87.4% and a PCC of 0.949.

1. Introduction

Expressways have been extremely important to the transportation industry, but bad road conditions often cause many accidents. One of the most serious problems is high road-surface temperature (RST), which can make tires easily explode. In this case, it is easy for the driver to operate the vehicle improperly, which results in the occurrence of traffic accidents. High RST also causes the asphalt in the road to swell, pit and be otherwise damaged; any increase in vehicle traffic can also cause large areas of intense damage. Thus, a high RST not only leads to accidents but also damages roads. Therefore, forecasting RST is a significant method to prevent traffic accidents and road damage. With the use of big data, expressway RST forecasting can be determined more easily than using a traditional way. Meteorological institutions have accumulated large amounts of road data in past decades. By applying data mining algorithms to these data, we can build a more accurate model.

RST is being studied all over the world. European researchers began researching RST earlier and established a complete road monitoring and RST forecasting system. For example, Germany [1] can forecast road weather for the next 1–3 days. The United Kingdom [2] uses road radar to monitor the road conditions. China started late but also established a road monitoring system; for example, Beijing has built many road monitoring stations, which can use the BJ-RUC (Beijing rapidly updating cycle) to forecast the weather of the roads in Beijing. So far, there are two methods to forecast the road temperature: numerical and statistical.

Numerical methods use a combination of physics and math to establish an equation that can forecast the RST. Chapman [3] built a model based on GIS (Geographic Information System). Bouilloud et al. [4] established a model that can forecast RST and snow depth in France. Sokol et al. [5] used a numerical method to forecast RST and used an ensemble method to eliminate uncertainty. Liu et al. [6] built a model

* Corresponding author.
E-mail addresses: boliu@bjut.edu.cn (B. Liu), yanshuok@163.com (S. Yan), you_hl@sohu.com (H. You), dongyancandice@163.com (Y. Dong), li.yong@bjut.edu.cn (Y. Li), jllang@bjut.edu.cn (J. Lang), rentaogu@bupt.edu.cn (R. Gu).

https://doi.org/10.1016/j.compind.2018.03.026
Received 1 June 2017; Received in revised form 26 February 2018; Accepted 20 March 2018
Available online 10 April 2018
0166-3615/ © 2018 Elsevier B.V. All rights reserved.

based on the conservation of energy; the model can forecast up to a range of 24 h. Jia et al. [7] utilized basic principles of heat conduction and built a model that can forecast RST for 4 moments per day. Feng et al. [8] utilized conservation of energy and built an hourly RST forecasting model. Meng et al. [9] combined the numerical simulation products Common Land Mode (CoLM) [10] and BJ-RUC to build a model that can forecast up to a range of 3–24 h. Yang et al. [11] established a numerical model in Korea. Other researchers, such as Han et al. [12] and Gan et al. [13], also used a numerical method to develop a forecasting model.

Statistical methods build a model based on historical data and are often easier to implement than numerical methods. Diefenderfer et al. [14] and Qu et al. [15] used linear regression to build a daily highest and lowest RST prediction model for multiple areas. Recently, researchers have been able to obtain higher-quality data; Li et al. [16] Ma et al. [17] used linear regression to build an hourly model and achieved good results. Additionally, Lukanen et al. [18], Hua Tian et al. [19], and Wu et al. [20] also built models based on statistical methods.

Numerical methods that are based on physics and that can simulate the various factors of RST have strong universality. The numerical method does not require observational data; in the early exploration of a system, when there is a lack of data, the numerical method is the best choice. The parameters of the numerical method are difficult to obtain; therefore, depending on the historical data and the limited amount of experimental information, the parameters of the model will be reduced since there is no effective method to determine them. Thus, any parameters the model does have are low quality [21]. Additionally, numerical methods are also very difficult to solve because they are based on a system of equations.

Among statistical methods, multiple linear regression (MLR) is the most commonly used approach. Statistical methods are easier than numerical methods and only obtain the statistical relationship between various factors and RST. Because statistical methods only consider the influence of the environment on the RST, they have low accuracy. The integrity and accuracy of statistical data can also affect the results. However, obtaining parameters are easy to find and implement [22]. Moreover, many researchers work on statistical method to enhance accuracy, generalization, and impact on the system; therefore, statistical methods are becoming increasingly popular [21].

Parameters of most numerical methods are difficult to obtain. Building an accurate, hourly forecast model based on traditional statistical methods is difficult but frequently done. In the big data era, many problems have been solved by using big data in many domains, such as health care [23], medicine [24], recommendation systems [25] and so on. Machine learning (ML) plays a very important role in the big data research field. Like statistical methods, ML forecasts RSTs based on historical data, but the method can approximate more complex functions. To get good results, statistical methods need a large amount of data and accurate feature selection; however, ML can also achieve good results with only limited feature selection and minimal data. The parameters of ML are easy to obtain via an optimization algorithm.

In this paper, we utilize an ML algorithm called extreme learning machine (ELM) [26] to forecast RST. ELM is well known because it has a fast training speed and good generalized performance, but it also has some disadvantages. The sigmoid function is a very important activation function used in traditional ELM training. Recently, ReLU [27] and softplus [27] activation functions are very popular in deep learning because they have sparsity limits that are seldom used in ELM. Although some researchers [28,29] have applied ReLU and softplus to ELM classification problems and have demonstrated that ReLU and softplus improved the performance of ELM, little research attention is given to regression problems such as RST forecasting. In this work, we replaced the sigmoid function with ReLU or softplus. Then, we applied ELM with ReLU and softplus for RST forecasting. Because the weights and bias between the input layer and hidden layer are randomly assigned in ELM, the performance of ELM is a little random. To reduce the randomness of ELM, applying ensemble methods is a good idea. Among the ensemble methods, gradient boosting [30] is very popular in data mining research and always gets good results. Thus, we utilized gradient boosting to ELM to reduce randomness effectively. To build an accurate hourly forecasting model, an accurate weather forecasting system is necessary. We have BJ-RUC data, which is an accurate weather forecasting system. Therefore, we combined the proposed ML method with BJ-RUC and achieved good results.

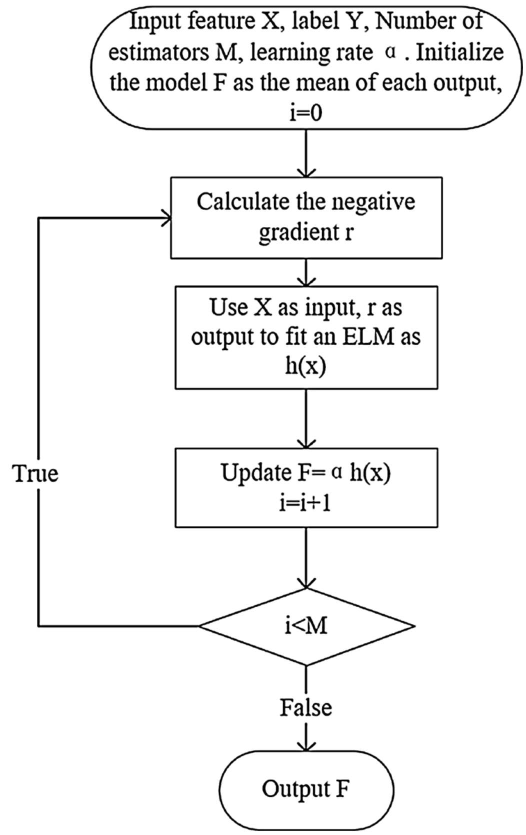

Fig. 1. The basic Gradient Extreme Learning Machine Boosting process.

2. Methodology

The main idea of the method described herein is to apply gradient boosting to ensemble ELM via an ReLU/softplus activation function. In this paper, gradient boosting was used to optimize a cost function over the function space by iteratively choosing an ELM that points in the negative gradient direction. The basic GBELM process is shown in Fig. 1.

2.1. Extreme learning machine

ELM is a type of SLFNs created by Guangbin Huang. Huang thought that training SLFNs based on the BP algorithm and gradient descent is not only inefficient but can also easily get stuck in a locally optimal solution. If an input matrix X, and an activation function g(x), are given, then an SLFNs can be represented as follows:

$$T= g(W^{*}X+b)\beta \tag{1}$$

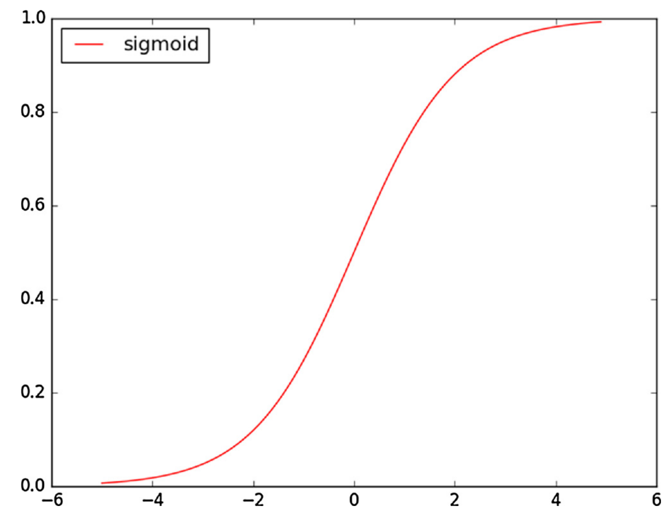

Fig. 2. Sigmoid function.

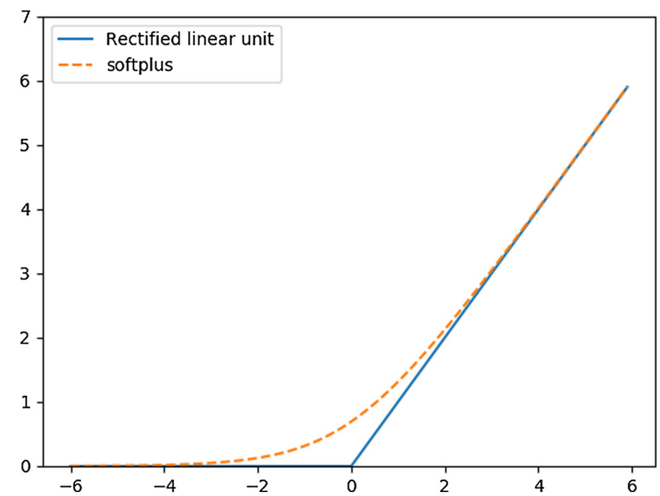

Fig. 3. ReLU and softplus functions.

where W represents the weights matrix that connects the input layer and hidden layer, b represents the bias matrix of hidden layer, β represents the connection weights between the hidden and output layer, and T is the target value matrix. This equation can be reduced as follows:

$$T = H\beta \tag{2}$$

where H represents the activation of the hidden layer. Traditionally, back-propagation was used to find W, b and β. Huang thought that W and b are not necessarily tuned, and they can be randomly selected. Once the random values are assigned to W and b, H is unique, and β can be computed as follows:

$$\beta = H^{\dagger}T \tag{3}$$

where $H^{\dagger}$ is the Moore–Penrose generalized inverse of matrix H. ELM can be regularized by adding a regularization term C, which changes the regularization form of ELM as follows:

$$\beta = H^{T}\left(\frac{1}{C} + HH^{T}\right)^{-1}T \tag{4}$$

2.2. Extreme learning machine with ReLU/softplus

Activation functions play an important role in ELM learning tasks; ELM will have good performance with a suitable activation function. The sigmoid function took an important role in traditional ELM network modeling in the past. The sigmoid function is a good threshold function. The sigmoid function is a continuous function that has a characteristic “S”-shaped curve. Fig. 2 shows the shape of the sigmoid function. The sigmoid function can be defined as follows:

$$g(x) = \frac{1}{1 + \exp(-x)} \tag{5}$$

Recently, a new activation function called ReLU (rectified linear unit) has become mainstream and has been widely used in deep learning, such as a CNN (convolutional neural network) [31] and DBN (deep belief network) [32]. Compared with the sigmoid function, ReLU is closer to the biological activation model. ReLU can regularize neural networks without pre-training; therefore, neural networks with ReLU have better generalization performance. ReLU is defined as follows:

$$g(x) = \max(0, x) \tag{6}$$

ReLU has many advantages, but it implements sparse limits by setting the activation of neurons to zero. However, mandatory sparse limits may hurt the generalization performance. Softplus was established to solve the problem above; it is a smooth approximation of ReLU. Softplus overcomes shortcomings by employing lax sparse limits and is closer to the biological activation model than ReLU. Fig. 3 shows the shape of ReLU and softplus. Softplus is defined as follows:

$$g(x) = \ln(1 + \exp(x)) \tag{7}$$

2.3. Gradient boosting extreme learning machine

Gradient boosting is an ML algorithm for both the regression and classification problems. This algorithm produces a robust estimator in the form of an ensemble of weak estimators, typically decision trees. Gradient boosting optimizes a loss function over the estimator space by iteratively choosing an estimator that points in the negative gradient direction. In this paper, we use ELM with ReLU/softplus as the base estimator of the gradient boosting, which is called gradient ELM boosting (GBELM). The least square loss function is chosen as the loss function used in gradient boosting. In this case, gradient boosting considers additive models of the following form:

$$F(x) = \sum_{m=1}^{M} \alpha h_m(x) \tag{8}$$

where F(x) is the final model, α is the learning rate, M is the number of weak estimators, and $h_m(x)$ is the base estimator. In this paper, $h_m(x)$ is ELM with ReLU/softplus. Gradient boosting builds the additive model in a forward moving fashion:

$$F_m(x) = F_{m-1}(x) + \alpha h_m(x) \tag{9}$$

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

$$F_m(x) = F_{m-1}(x) + \alpha argmin(h) \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) - h(x)) \tag{10}$$

where n represents the samples, and y is the target value. For the least square loss function, the initial model, F_0, usually chooses the mean target values. In the multi-output case, F_0 chooses the mean target values at each output.

Algorithm 1 Gradient ELM boosting Algorithm

Input: data x, target y, learning rate α, number of estimators M Output: ensemble ELM $F_m(x)$
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 an ELM as $h_m(x)$
   (c) Assign the W and b arbitrarily and calculate the activation H of hidden layer
   (d)Predict r with h_m giving x_i (i = 1, 2, 3, ……, N) and obtain $h_m(x)$
   (e)$F_m(x) = F_{m-1}(x) + \alpha h_m(x)$
3. Output $F_m(x)$

3. Experiment settings

3.1. Dataset

In this paper, weather forecasting data from the BJ-RUC [33] and from the Beijing road monitor station were used to forecast the RST. BJ-RUC is an RUC system developed for Beijing city. RUC is an internationally popular numerical forecast mode. The BJ-RUC system is activated every 3 h, 8 times per day. This paper uses the RUC numerical forecast data with 3 km resolution; Thus, in the range of BJ-RUC, every 3 km, the grid is divided into a grid each vertex of the grid is a forecasting site. BJ-RUC records height fields, upward long wave radiation, ground surface pressure, humidity, downward shortwave radiation, 2-m temperature, longitude 10-m wind, altitude 10-m wind and hourly accumulated precipitation. In this paper, we have selected several of the nearest BJ-RUC forecast site data, from the road monitoring station, at 5:00, 8:00, 17:00.

Road monitoring stations that monitor the various expressways in Beijing produce a lot of data per hour. In this paper, 2 stations with heavy traffic flow and complete data were selected to analyze the A1027 airport expressway and A1412 Badaling expressway. Road monitoring stations recorded the RST, total precipitation, precipitation intensity, road conditions, temperature, height of snow and rain.

The data from November 2012 to April 2014 and the data from June 2014 to August 2014 were used as training data. Data from September 2014 were used as test data.

3.2. Data preprocessing

Data pre-processing is mainly divided into 2 parts: missing value processing and data normalization.

Since the BJ-RUC itself has prediction ability, using the forecast value of BJ-RUC to replace the missing value is the best way. Thus, we used the forecasting data of the previous moment to replace the missing value of the current moment. However, the method is not suitable for a large range of missing values. Therefore, we drop the large range of missing values directly. Normalization can scale the data to a small range, which facilitates neural network convergence. In this paper, we normalized our data in the interval [0,1].

3.3. Experiment settings

For BJ-RUC, select height fields, upward long wave radiation, ground surface pressure, humidity, downward shortwave radiation, 2-m temperature, longitude 10-m wind, altitude 10-m wind and hourly accumulated precipitation. Past 72 h and future 24 h, totally. For road monitor station, select RST, total precipitation, precipitation intensity, road condition, temperature. In this paper, we analyze the correlation between the input features and output. We also select the features whose correlations with each output exceed 0.5 as the final input. There are many BJ-RUC forecasting sites that surround road monitoring stations. To determine how many BJ-RUC forecasting sites surround the road monitoring stations that yields the best result, we tested a different number of BJ-RUC forecasting sites and selected the number with the best performance. Because the model forecasts the RST of the upcoming 24 h, the output layer of all the ELM we used is 24 units. Grid search and K-Fold cross-validation were used to choose the parameters of the models.

In this paper, we choose four evaluation methods that are usually used in RST forecasting to evaluate the model.

1) *Root mean squared error (RMSE)*:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2} \tag{11}$$

where O_i denotes the observed RST, P_i denotes the predicted RST, and N denotes the number of evaluation samples.

2) *Accuracy*:

If the difference between the maximum temperature of the road and its calculated value is that it is within ± 3 °C, then the forecast is accurate; otherwise it is wrong.

3) *Pearson correlation coefficient (PCC)*:

$$PCC = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{O_i - \overline{O}}{S_o}\right)\left(\frac{P_i - \overline{P}}{S_p}\right)$$

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

4. Results and discussion

First, we explored the number of BJ-RUC forecast sites for RST forecasting. Then, we made use of the data from November 2012 to August 2014 and the data from June 2014 to August 2014 to build models, and we compared these models to find the advantage of each. The September 2014 data were used to evaluate the model. All the experiments were carried out on a machine with Intel I7 6770HQ, memory 16 GB, 64-bit operate system, and Anaconda 4.2.0. All the parameters were selected by grid search and 3-folders cross-validation. In grid search, the range of hidden units is 10, 20, 30……300; the range of the number of estimators is the same as the hidden units number; the range of learning rate is 0.001, 0.005, 0.01, 0.05, 0.1, 0.5; the range of the regularization coefficient is the same as the learning rate.

4.1. Selection of the number of BJ-RUC forecasting sites

To determine how many of the nearest BJ-RUC forecasting sites from the road monitor station are the most appropriate, we used sigmoid ELM, ReLU ELM, softplus ELM, ReLU GBELM and softplus GBELM to model the data from June to August 2014 via a different number of BJ-RUC forecasting sites.

From Figs. 4–9, we can see that no methods can be improved by adding more BJ-RUC forecasting sites. The performance of ReLU ELM and softplus ELM declined less than the sigmoid ELM, because ReLU

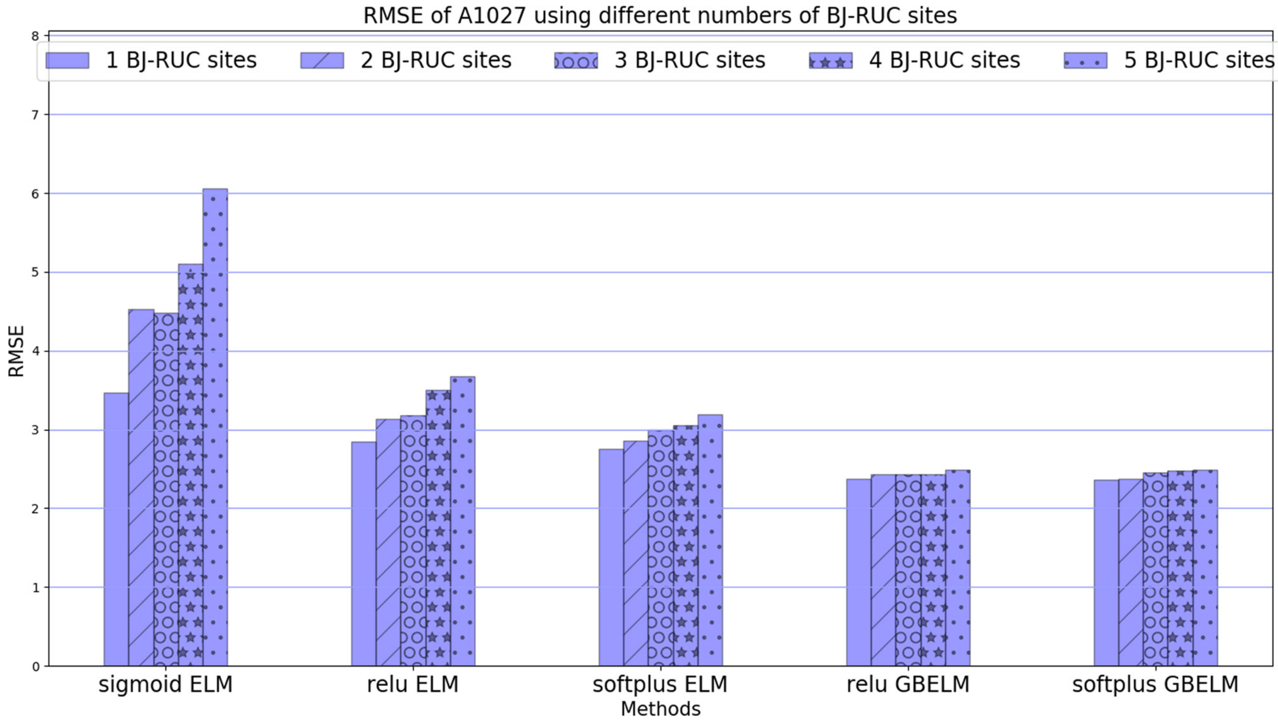

Fig. 4. The RMSE of five methods using different numbers of nearest BJ-RUC forecasting sites in A1027 road monitor station.

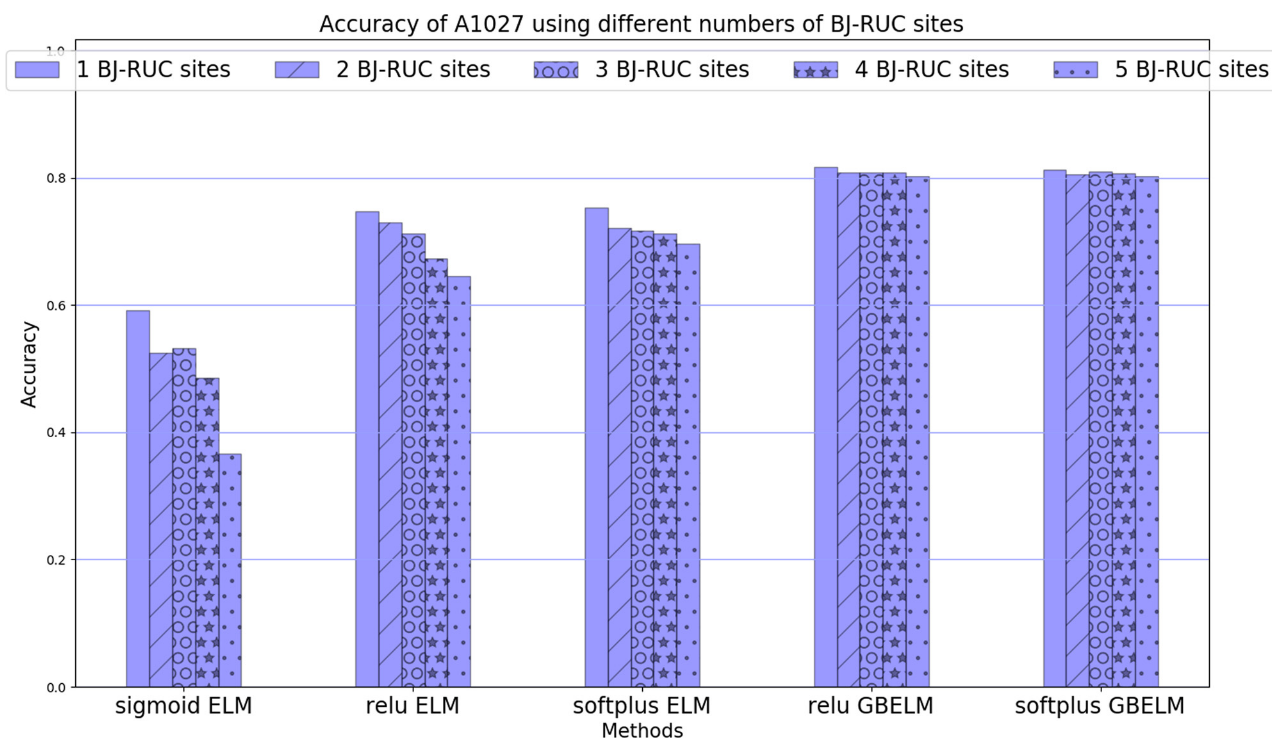

Fig. 5. The accuracy of five methods using different numbers of nearest BJ-RUC forecasting sites in A1027 road monitor station.

and softplus have sparsity limitations to solve any redundant features, while the sigmoid ELM, with regularization term, has no such powerful sparsity limitations. The performance of GBLEM is almost same with the different BJ-RUC forecasting sites, indicating that GBELM reduces the feature selection requirements.

4.2. Result of different methods

In this paper, we use the data of monitoring stations A1027 and A1412 as test data to compare 5 prediction methods: sigmoid ELM, ReLU ELM, softplus ELM, ReLU GBELM and softplus GBELM.

Tables 1 and 2 show the RMSE, accuracy, PCC, modeling time and

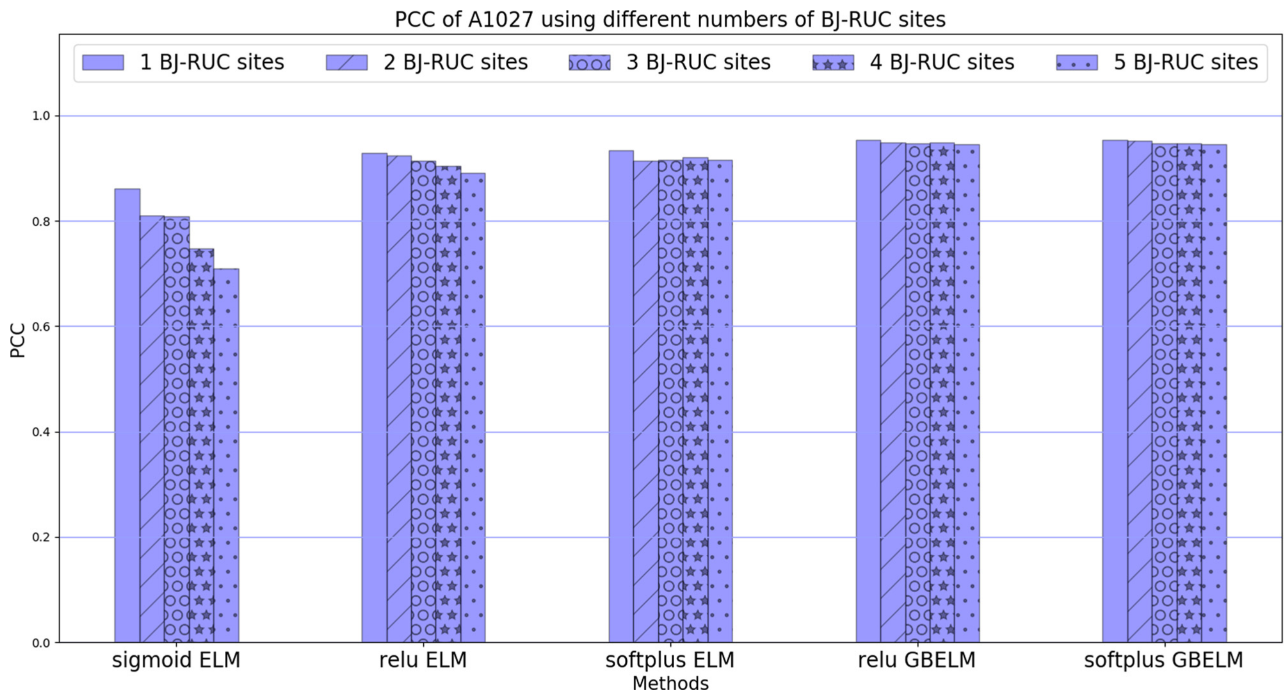

Fig. 6. The PCC of five methods using different numbers of nearest BJ-RUC forecasting sites in A1027 road monitor station.

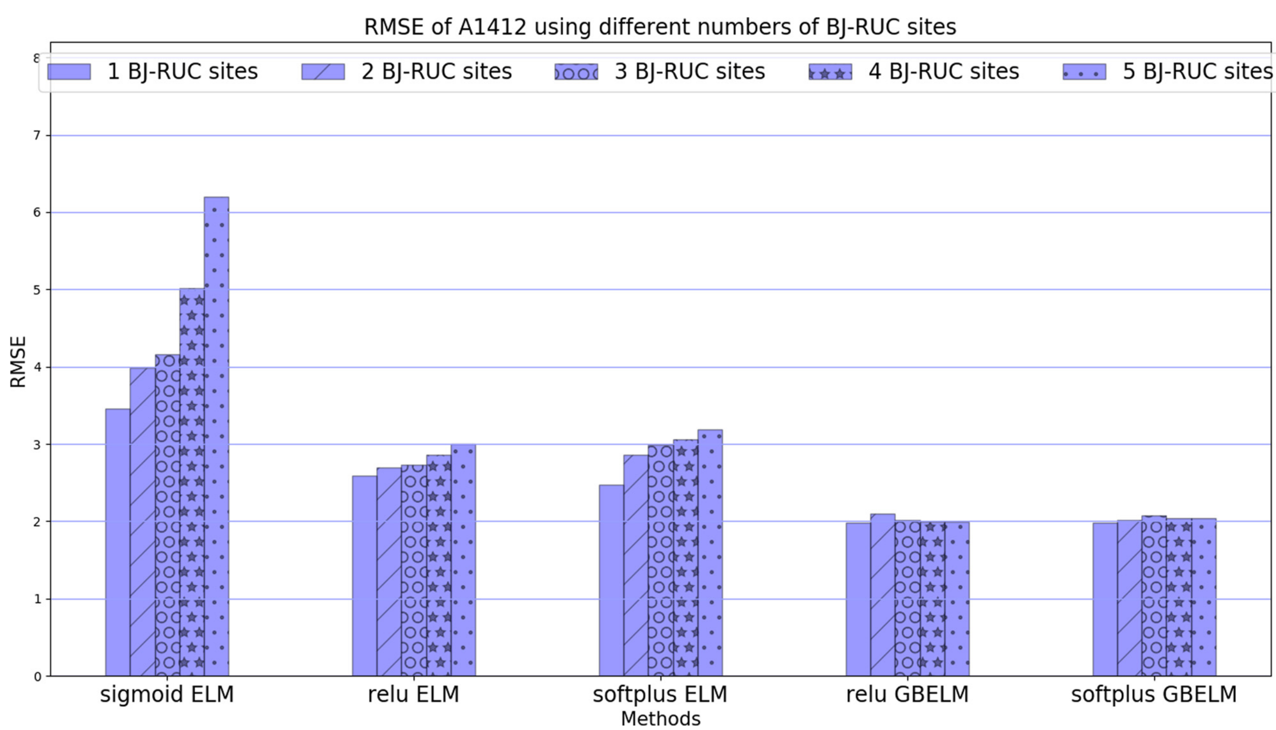

Fig. 7. The RMSE of five methods using different numbers of nearest BJ-RUC forecasting sites in A1412 road monitor station.

forecasting time of the A1027 and A1412 monitoring stations. By comparison, it is easy to find that GBELM with ReLU/softplus have the best performance for either station. For station A1027, ReLU/softplus activation function improves the accuracy by more than 10%, and reduces the RMSE to less than 3 °C. Gradient boosting further improves the accuracy to more than 80%. In general, ReLU GBELM has the best performance. For station A1412, ReLU GBELM has the best performance. ReLU/softplus also works well on this monitoring station, improving the accuracy by approximately 20%. Only when the RMSE-based ReLU GBELM and softplus GBELM are within 2 °C, the accuracy is greater than 85%. GBELM, with either an ReLU activation function or softplus activation function, can obtain a similar performance. It is hard to say which activation function is better. Experiments are necessary when selecting the activation function (e.g., ReLU v softplus). Figs. 10 and 11 show the best A1027 and A1412 models. From these figures, we can see that for over 40 °C high RST forecasting, these modeling performances are good, but the final result is a little on the higher side.

Compared with Ref. [8], their model yielded a PCC of 0.9 on the

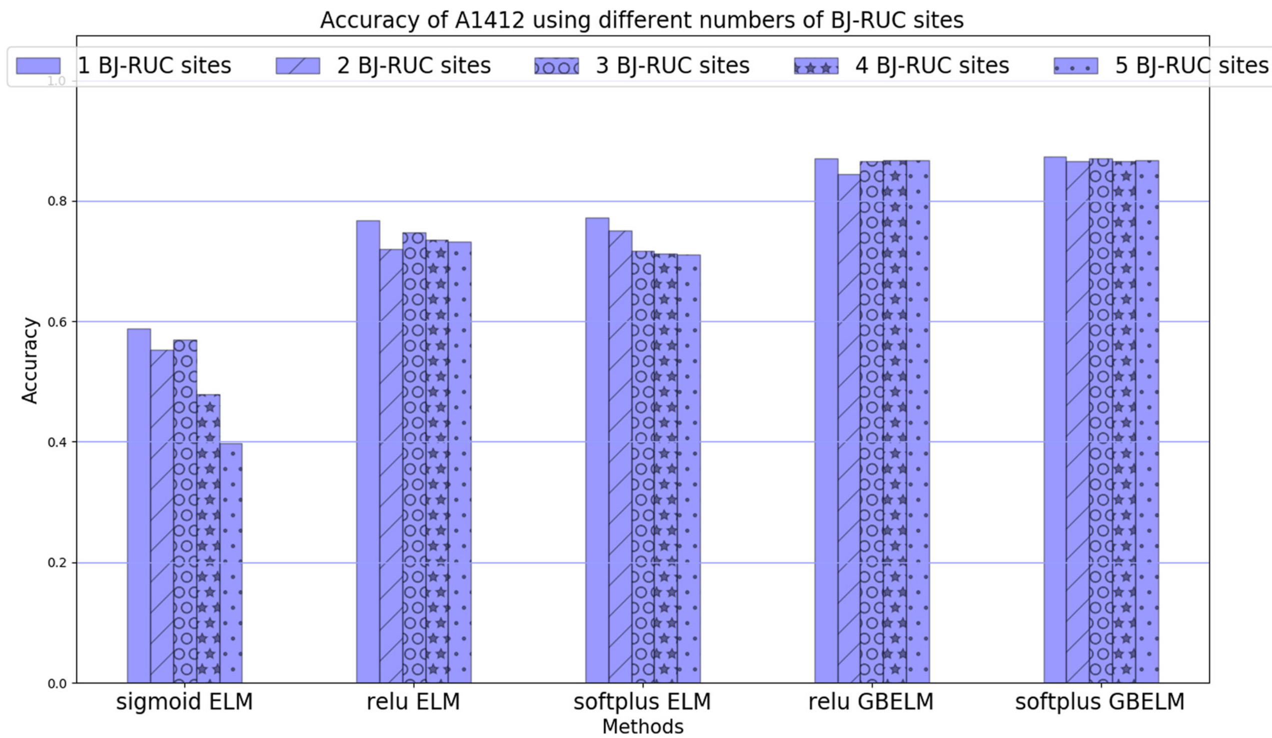

Fig. 8. The accuracy of five methods using different numbers of nearest BJ-RUC forecasting sites in A1412 road monitor station.

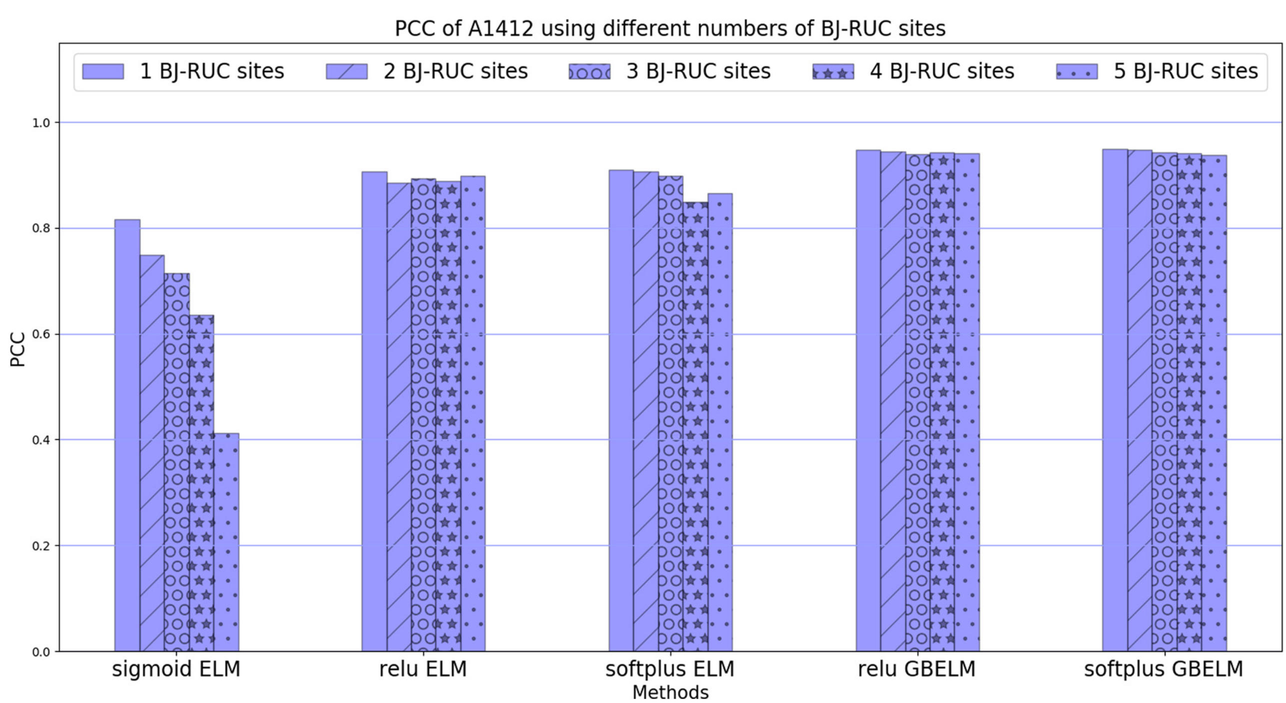

Fig. 9. The PCC of five methods using different numbers of nearest BJ-RUC forecasting sites in A1412 road monitor station.

Badaling expressway, while the PCC of our model on the Badaling expressway is 0.949. Ref. [16] uses 4 years of hourly data and a lot of analysis to build a statistical model, and their temperature distribution in autumn is similar to A1027. The accuracy of their model in autumn is 76.7%, whereas the accuracy of our model on A1027 is 81.8%; moreover, we use much fewer data and a simpler feature selection. Compared with ELM, GBELM has a much higher accuracy and a higher PCC, which means that the trend fitting process is better. Therefore, combined GBELM with BJ-RUC can obtain a more accurate hourly model than a numerical method, statistical method and ELM. We also need fewer data, and the method reduces the requirements for feature selection.

ReLU ELM has the shortest modeling time. This method can complete the modeling within 0.1 s, but ReLU ELM and softplus ELM can also complete the modeling within 0.1 s. People cannot feel an obvious difference in time. ReLU GBELM and softplus GBELM takes much longer to complete the modeling than the other three methods, but they can complete the modeling within 10 s, which is acceptable. All the

Table 1
A1027 Monitoring station Results.

	RMSE	Accuracy	PCC	Modeling time (s)	Forecasting time (s)
ELM(sigmoid)	3.4692	59.1%	0.861	0.1253	0.0022
ELM(ReLU)	2.8466	74.7%	0.929	**0.0676**	**0.0012**
ELM(softplus)	2.7539	75.3%	0.933	0.0812	0.0025
GBELM(ReLU)	2.3684	**81.8%**	**0.954**	7.2216	0.1534
GBELM (softplus)	**2.3662**	81.2%	**0.954**	7.0281	0.1854

The bold values mean the best values.

Table 2
A1412 Monitoring station Results.

	RMSE	Accuracy	PCC	Modeling time (s)	Forecasting time (s)
ELM(sigmoid)	3.4522	58.8%	0.816	0.1319	0.0019
ELM(ReLU)	2.5846	76.8%	0.907	**0.0686**	**0.0015**
ELM(softplus)	2.4725	77.3%	0.909	0.0877	0.0020
GBELM(ReLU)	**1.9839**	87.1%	0.948	8.0216	0.1231
GBELM (softplus)	**1.9839**	**87.4%**	**0.949**	7.4281	0.1324

The bold values mean the best values.

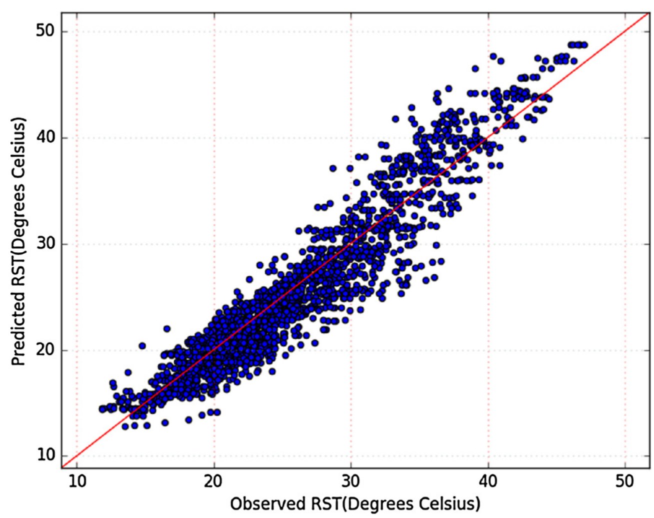

Fig. 10. A1027 results using ReLU GBELM with 100 hidden units, 130 estimators and a 0.05 learning rate.

methods can forecast RST within 1 s, so people will not feel a significant time difference.

5. Conclusions and future work

This work proposed an ensemble ML algorithm called GBLEM, which applies the ReLU and softplus functions to ELM in a regression problem. GBELM not only obtains better results than traditional methods but also reduces the requirement for feature selection. We applied GBELM to RST forecasting and built an accurate hourly model. By comparing ELM using sigmoid-based activation functions and ReLU and softplus activation functions, we found that ReLU and softplus obviously improve the performance of ELM in a regression problem. Gradient boosting helps ReLU ELM and softplus ELM improve their performance further. Softplus cannot replace ReLU as ReLU replaces the sigmoid function. When applying GBELM to a machine learning problem, experiments are necessary to select the appropriate activation function between ReLU and softplus. Few ML researchers pay attention to the RST forecasting problem. We applied an ML algorithm to RST

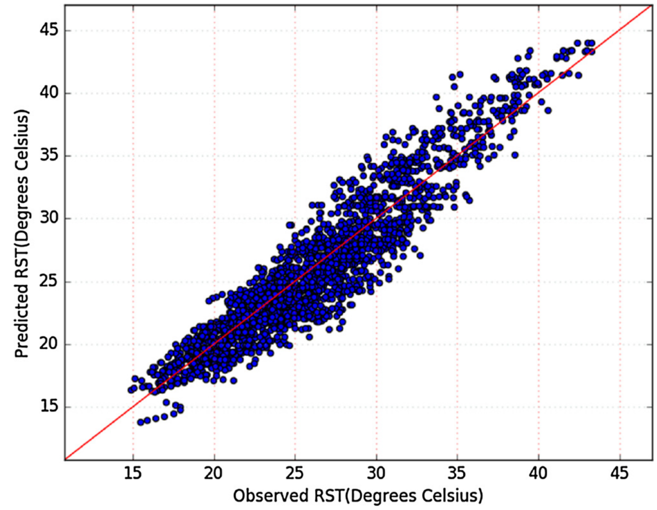

Fig. 11. A1412 result using softplus GBELM with 50 hidden units, 160 estimators, learning rate is 0.05.

forecasting and got better results than the numerical and statistical methods.

In this paper, we use a least square loss function for GBELM with ReLU and softplus. In the future, we will try other robust loss functions, such as Huber or Quantile loss functions to further improve the performance. Xgboost can also be applied to ensemble ELM and may obtain a more accurate model than GBELM.

Acknowledgments

This work is supported by National Natural Science Foundation of China (61702021), Beijing Natural Science Foundation (4174082, 4182040), General Program of Science and Technology Plans of Beijing Education Committee (SQKM201710005021), China Postdoctoral Science Foundation (2016M590027), Beijing Postdoctoral Research Foundation (2016ZZ-17), 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).

References

[1] X. Li, An investigation report on the meteorological information system of the German Federal Highway, J. Chin. Foreign Roads 5 (1997) 23–26.
[2] S. Meng, Management and meteorological information of foreign expressway, Meteorol. Sci. Technol. 04 (2000) 60–62.
[3] L. Chapman, J.E. Thornes, A geomatics-based road surface temperature prediction model, Sci. Total Environ. 360 (1–3) (2006) 68–80.
[4] L. Bouilloud, E. Martin, F. Habets, A. Boone, P. Le Moigne, J. Livet, M. Marchetti, A. Foidart, L. Franchisteguy, S. Morel, Road surface condition forecasting in France, J. Appl. Meteorol. Clim. 48 (12) (2009) 2513–2527.
[5] Z. Sokol, V. Bli Ňák, P. Sedlák, P. Zacharov, P. Pe Ice, M. Kuthan, Ensemble forecasts of road surface temperatures, Atmos. Res. 187 (Supplement C) (2017) 33–41.
[6] X. Liu, Y. Yu, G. Lei, Z. Liu, Using radiant balance theory to calculate concrete road-surface temperature in summer, J. Appl. Meteorol. Sci. 15 (5) (2004) 623–628.
[7] L. Jia, L. Sun, L. Huang, J. Qin, Numerical temperature prediction model for asphalt concrete pavement, J. Tongji Univ. (Nat. Sci.) 35 (8) (2007) 1039–1043.
[8] T. Feng, S. Feng, A numerical model for predicting road surface temperature in the highway, Procedia Eng. 37 (2012) 137–142.
[9] C. Meng, C. Liu, Summer road temperature disaster forecast of expressway in Beijing, J. Inst. Disaster-Prevent. Sci. Technol. 03 (2009) 26–29.
[10] Y. Dai, X. Zeng, R. Dickinson, et al., The common land model, Bull. Am. Meteorol. Soc. 84 (8) (2003) 1013–1023.
[11] C. Yang, M. Park, D. Yun, A road surface temperature prediction modeling for road weather information system, J. Korean Soc. Transp. 29 (2) (2011).
[12] D. Han, J. Zhang, J. Wei, Temperature characteristic of asphalt pavement structure in Beijing based on temperature survey of test road, J. Beijing Univ. Technol. 4 (2015) 566–572.
[13] X. Gan, X. Ji, H. Cao, Empirical prediction model of low-temperature field of asphalt pavement, J. Beijing Univ. Technol. 4 (2015) 584–589.
[14] B. Diefenderfer, I. Al-Qadi, S. Reubush, T. Freeman, Development and validation of a model to predict pavement temperature profile, Development and Validation of a

Model to Predict Pavement Temperature Profile, Citeseer, 2003.
[15] X. Qu, H. Wu, Y. Zhang, J. Jia, The atmospheric physical quantity and radar echo characteristic of a strong convective weather, J. Arid Meteorol. 03 (2010) 352–357.
[16] L. Li, Y. Song, G. Han, Pavement temperature characteristics and prediction model of Tai'an expressway, J. Shandong Meteorol. 02 (2016) 58–63.
[17] J. Ma, C. Liao, Y. Zhang, Variation characteristics of road surface temperature on highway of changsha and its prediction model, Sci. Technol. Square 02 (2017) 22–25.
[18] E. Lukanen, R. Stubstad, R.C. Briggs, Temperature predictions and adjustment factors for asphalt pavement, Temperature Predictions and Adjustment Factors for Asphalt Pavement, Federal Highway Administration, 2000 Publication No. FHWA-RD-98-085.
[19] H. Tian, H. Wu, L. Zhao, H. Chen, K. Li, X. Yang, Characteristics and statistical model of road surface temperature on Huning expressway, J. Appl. Meteorol. Sci. 06 (2009) 737–744.
[20] H. Wu, C. Ma, R. Yang, J. Zhang, Variation characteristics of road surface temperature on highway, J. Arid Meteorol. 04 (2014) 665–670.
[21] J. Qian, L. Sun, Management and meteorological information of foreign expressway, J. Chin. Foreign Roads 25 (6) (2005) 19–23.
[22] Y. Jing, G. Wang, C. Zhu, Road surface temperature analysis and extreme weather effects, J. China Sci. Technol. Inf. 08 (2009) 21–22.
[23] J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, H. Pan, Emerging information technologies for enhanced healthcare, Comput. Ind. 69 (Supplement C) (2015) 3–11.
[24] J. Li, C. Liu, B. Liu, R. Mao, Y. Wang, S. Chen, J. Yang, H. Pan, Q. Wang, Diversity-aware retrieval of medical records, Comput. Ind. 69 (Supplement C) (2015) 81–91.
[25] J. Li, J.J. Yang, Y. Zhao, B. Liu, M. Zhou, Enforcing Differential Privacy for Shared Collaborative Filtering, (2018).
[26] G. Huang, Q. Zhu, C. Siew, Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks, (2004), pp. 985–990.
[27] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, Deep Sparse Rectifier Neural Networks, (2011), p. 275.
[28] J. Peng, L. Chen, I. Ather, A. Yu, Extreme learning machine based on rectified nonlinear units, in: Z. Zeng, X. Bai (Eds.), AER-Advances in Engineering Research, Atlantis Press, 2016, pp. 1525–1530.
[29] Y. Sun, X. Yang, Y. Hu, P. Wang, ELM algorithm based on softplus activation function, J. Beijing Univ. Technol. 41 (2015) 1341–1348 0254-0037(2015) 41:9 < 1341:JYSJHH > 2.0.TX;2-P9.
[30] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Front. Neurorobot. 7 (2013) 21.
[31] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[32] G. Hinton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[33] D. Wei, F. You, S. Fan, B. Yang, M. Chen, Evaluation and analysis of sounding quality of Beijing rapid renewal cycle prediction system (BJ-RUC) model, J. Meteorol. 36 (8) (2010) 72–80.