

# Comparison of algorithms for road surface temperature prediction

Bo Liu

*Beijing Advanced Innovation Center for Future Internet Technology,  
Beijing University of Technology, Beijing, China and  
School of Software Engineering, Faculty of Information Technology,  
Beijing University of Technology, Beijing, China*

Libin Shen

*School of Software Engineering, Faculty of Information Technology,  
Beijing University of Technology, Beijing, China*

Huanling You

*China Meteorological Administration, Beijing, China*

Yan Dong

*Beijing Meteorological Service Center, Beijing, China, and*

Jianqiang Li and Yong Li

*School of Software Engineering, Faculty of Information Technology,  
Beijing University of Technology, Beijing, China*

## Abstract

**Purpose** – The influence of road surface temperature (RST) on vehicles is becoming more and more obvious. Accurate prediction of RST is distinctly meaningful. At present, however, the prediction accuracy of RST is not satisfied with physical methods or statistical learning methods. To find an effective prediction method, this paper selects five representative algorithms to predict the road surface temperature separately.

**Design/methodology/approach** – Multiple linear regressions, least absolute shrinkage and selection operator, random forest and gradient boosting regression tree (GBRT) and neural network are chosen to be representative predictors.

**Findings** – The experimental results show that for temperature data set of this experiment, the prediction effect of GBRT in the ensemble algorithm is the best compared with the other four algorithms.

---

© Bo Liu, Libin Shen, Huanling You, Yan Dong, Jianqiang Li, and Yong Li. Published in *International Journal of Crowd Science*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

This work is supported by Beijing Natural Science Foundation (4174082), National Natural Science Foundation of China (61702021), General Program of Science and Technology Plans of Beijing Education Committee (SQKM201710005021), 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).



---

**Originality/value** – This paper compares different kinds of machine learning algorithms, observes the road surface temperature data from different angles, and finds the most suitable prediction method.

**Keywords** Neural network, Gradient boosting regression tree, Random Forest, Road surface temperature

**Paper type** Research paper

## 1. Introduction

Nowadays, the demand of high-speed traffic is increasing. The expressway meets the speed requirement of vehicles; however, it also brings a greater amount of traffic accidents. Because of the heavy traffic volume and fast speed of the expressway, the road temperature is often high, which can result in some damage to the car tires and even lead to a puncture and affect the life quality and property of the people and the traffic order seriously. This paper uses historical data to predict the road surface temperature (RST) of the expressway, which can give a fair warning to traffic management departments and drivers to reduce the accident rate and ensure the normal operation of the expressway.

Many researchers all over the world have contributed a lot to the study of RST forecasting. Existing methods consist of two parts: numerical method and statistical method. Math and physics are the tools of numerical methods to establish an equation for forecasting RST (Liu *et al.*, 2017). Barber (Edward, 1957) thought roads as a semi-infinite mass with uniform texture and built a model to predict the highest temperature. Sass (1997) established a model that can forecast up to a range of at least 3 h; this model is based on the equation of heat. Feng and Feng (2012) used conservation of energy and built an hourly RST forecasting model. Meng and Liu (2009) combined numerical simulation product Common Land Mode (CoLM) (Dai *et al.*, 2003) and BJ-RUC (Wei *et al.*, 2010) and established a model which could forecast up to a range of 3-24 h.

Ensemble learning is a machine-learning paradigm where multiple learners can be trained to solve the same problems (Zhou, 2009). The first application of ensemble learning was led by Hansen and Salamon (1990) in the late 1980s. They demonstrated that the integration of multiple learners is better than that of a single learner. There are two typical strategies in the ensemble algorithm Boosting and Bagging. Boosting learns multiple classifiers by changing the weights of the training samples (Li, 2012), and linearly combines these classifiers to improve the performance of the classifier and reduce the bias of the model. Bagging is based on bootstrap sampling and trains multiple base learners. If there is a classified problem, it will adopt a voting strategy. If there is a regressive problem, a simple average method will be used. Bagging helps to reduce the variance of the model (Zhou, 2016). The random forest and gradient boosting regression tree (RF and GBRT) base learners used in this paper are decision trees.

In statistics and machine learning, least absolute shrinkage and selection operator (LASSO) is a regression analysis method that performs both variable selection and regularization to enhance the predictive accuracy and interpretability of the statistical model it produces. Robert Tibshirani introduced it in 1996 based on Leo Breiman's nonnegative garrote (Robert, 1996).

Deep Learning (DL) is one of the newest trends in Machine Learning and Artificial Intelligence Research. The term DL was first introduced to machine learning (ML) in 1986, and later used for artificial neural networks (ANN) in 2000. Deep learning methods are composed of multiple layers to learn features of data with multiple levels of abstraction (LeCun *et al.*, 2015). To learn complicated functions, deep architectures are used with multiple levels of abstractions, that is, non-linear operations; for example, ANNs with many hidden layers.

The five algorithms that we use represent different learning strategies. We can discover the characteristics of the data from different perspectives and compare the performance of the five strategies.

## 2. Algorithm

### 2.1 Gradient boosting regression tree

GBRT is a member of the boosting family, which can promote weak learners to be strong learners (Friedman, 2001). It uses the steepest descent approximation method. The key is to use the negative gradient of the loss function in the current model value as an approximation of the residuals in the regressive problem to fit a regression tree. Following Equation (1), after many iterations and updates, we finally got Equation (2):

$$-\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right] f(x) = f_{m-1}(x) \quad (1)$$

$$\hat{f}(x) = f_M(x) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (2)$$

### 2.2 Random Forest

L. Breiman (2001) proposed RF, which is a powerful performance of the multi-purpose classification and regression algorithm. RF is composed of multiple random trees, and the average value of output of the random trees is used as the predictive result. The random tree is a variant of decision tree, that is, in the process of decision tree construction, introducing the random nature: selecting k features from all features randomly as feature set in the decision tree, and then select an optimal feature from this subset for partitioning (Breiman, 1996).

### 2.3 Least absolute shrinkage and selection operator

The main idea of the LASSO regression method (Tibshirani, 2011) is to minimize the sum of squared residuals under the constraint that the sum of the absolute values of the regression coefficients is less than a normal number, so that variables with small or zero regression coefficients can be filtered out and effectively solve the problem of multicollinearity. It has the advantage of subset selection, while at the same time it can perform variable selection and unknown parameter estimation.

As usually (Robert, 1996), there is a data set  $(h^t, t_i), i = 1, \dots, N$ , and  $h^t$  is a predictive value.  $t_i$  is a real value. The estimated amount  $(\alpha, \beta)$  of LASSO can be defined as:

$$(\alpha, \beta) = \operatorname{argmin} \left\{ \sum_{i=1}^N \left( y_i - \alpha - \sum_j \beta_j x_{ij} \right)^2 \right. \quad (3)$$

$$\left. \text{s.t. } \sum_j |\beta_j| \leq t \quad (4) \right.$$

in which  $t$  denotes a training parameter.

### 2.4 Multiple linear regression

MLR is the simplest model to study the correlation between a dependent variable and multiple independent variables. The usual multiple linear regressions' model shows as:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon \quad (5) \quad \text{Comparison of algorithms}$$

Among them,  $\beta_0, \dots, \beta_m$  are regression coefficients,  $m$  represents the number of independent variables and  $\varepsilon$  stands for random error. It is generally assumed that  $\varepsilon$  is a Gaussian distribution with a mean of zero and a variance of  $\delta$  two (Liu, 2005).

### 2.5 Neural network

The term NN has evolved to encompass a large class of models and learning methods. Here we describe the most widely used “vanilla” neural net, sometimes called the single hidden layer back-propagation network, or single layer perception (Hastie, 2009).

We have built a single hidden layer neural network, which shows as:

$$y(x, w) = \sigma \left( \sum_{j=1}^M w_{kj}^{(2)} h \left( \sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^1 \right) + w_{k0}^{(2)} \right) \quad (6)$$

And the activation Equation  $h(\cdot)$  is a logistic sigmoid function (7):

$$h(a) = \frac{1}{1 + \exp(-a)} \quad (7)$$

## 3. Experiments

### 3.1 Data processing

This paper uses the data of BJ-RUC (Beijing-rapidly update cycle) and the data of Beijing pavement inspection station to conduct experiments. BJ-RUC is an RUC system developed for Beijing and is an internationally popular numerical forecasting model. The RUC recorded upward long-wave radiation, surface pressure, humidity, downward short-wave radiation, 2 m temperature, longitudinal 10-m wind, latitudinal 10-m wind and hourly cumulative rainfall. We chose the data at #121107 monitoring station for experiments.

The pavement monitoring stations are located in multiple expressways in Beijing and record data every hour. This paper selects a monitoring station with large traffic volume and relatively complete data for analysis, that is, A1412 Badaling Expressway.

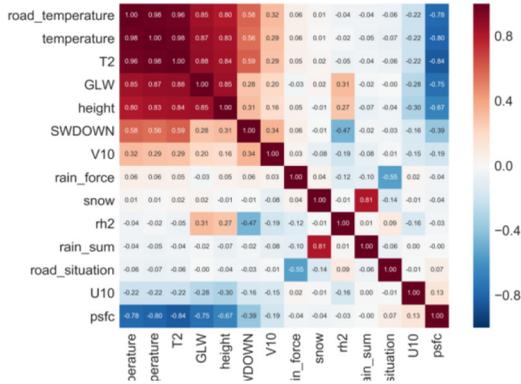
We use data from September 2012 to June 2015 as a data set. For single data missing, average values are filled in. If the data more than five fields in a day or the data more than three consecutive days are missed, we will delete the missing data. In the end, we obtained 1,347 data.

The Pearson correlation coefficient for each variable with Road Temperature can be derived from Figure 1. The correlation coefficient between T2 (The temperature of 2 m above the surface of the road) and the Temperature and the target variable is higher than 0.9, so the field of Temperature with a correlation of 0.98 is discarded. The algorithms used in this paper are MLS, LASSO, RF and GBRT, NN.

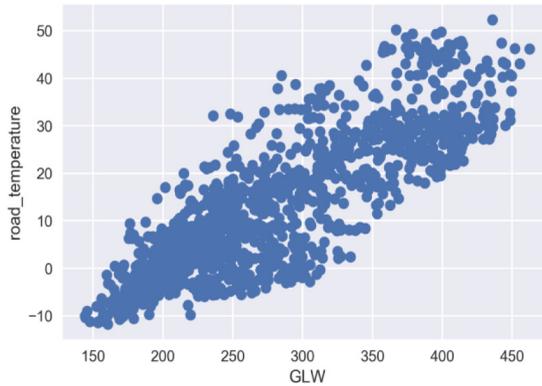
### 3.2 Feature selection

We extract the features whose absolute values of correlation coefficient are greater than 0.9 or less than 0.5. The results are shown in Figures 2 and 3.

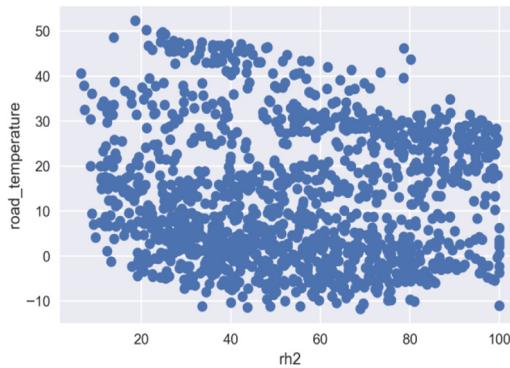
**Figure 1.**  
Pearson correlation coefficient between each variable



**Figure 2.**  
Scatter plot of upward long wave radiation (GLW) and road\_temperature



**Figure 3.**  
Scatter plot of humidity (rh2) and road\_temperature



The results show that while the absolute value of the correlation coefficient is less than 0.5, there is no good correlation. This paper selects the variable with an absolute value of the correlation coefficient greater than 0.5 as features.

After we remove the variables that are less relevant to the target variables, we will get the variables that we need in the experiment, that is, the features. We combine the eigenvalues of every day into an eight-dimensional vector, because we have 1,347 pieces of data totally, so the resulting input matrix dimension is 1347\*8.

### 3.3 Model training

We use five machine learning models, where RF, GBRT and NN require tuning parameters. The base learner of the RF is a decision tree, and the number of base learners is 500. The RF uses a Bagging strategy. It can reduce the variance of models, so the depth of the tree can be relatively large. We set the depth of the tree to 13. The GBRT base learner is also decision tree with the number of base learners 700. Because GBRT uses a Boosting strategy, it can reduce the deviation, and the depth of the tree can be small. In this article, it is set to 3. There are many hyper parameters needed to be adjusted in NN. The number of neurons in hidden layer is 500, and the activation function is logistic sigmoid function. In this paper, the “hold-out” method is used to divide the data set into two mutually exclusive sets. One set is used as the training set S and the other is used as the test set T. After training the model at S, we use T to evaluate the test error as an estimate of the generalized error. In this experiment, data sets were randomly divided into training sets and test sets, of which the training set accounted for 70 per cent and the test set accounted for 30 per cent.

### 3.4 Performance metrics

To evaluate the generalization performance of the learners, it requires not only an effective and feasible experimental estimation method, but also an evaluation standard that measures the generalization ability of the model. Performance metrics reflect the task requirements. While comparing the capabilities of different models, using different performance metrics often leads to different evaluation results. The essence of the task is a regression problem. The evaluation metrics include Mean Squared Error (MSE), Mean Absolute Error (MAE) and  $R^2$ :

$$\text{MSE} : E(f; D) = \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 \tag{8}$$

$$R^2 : E(f; D) = 1 - \frac{\sum_{i=0}^n (f(x_i) - y_i)^2}{\sum_{i=0}^m (f(x_i) - y_i^*)^2} \tag{9}$$

$$\text{MAE} : E(f; D) = \frac{1}{m} \sum_{i=1}^m |f(x_i) - y_i| \tag{10}$$

where  $y_i$  denotes the observed RST,  $f(x_i)$  denotes the predicted RST,  $m$  denotes the number of evaluation samples.

**4. Result and discussion**

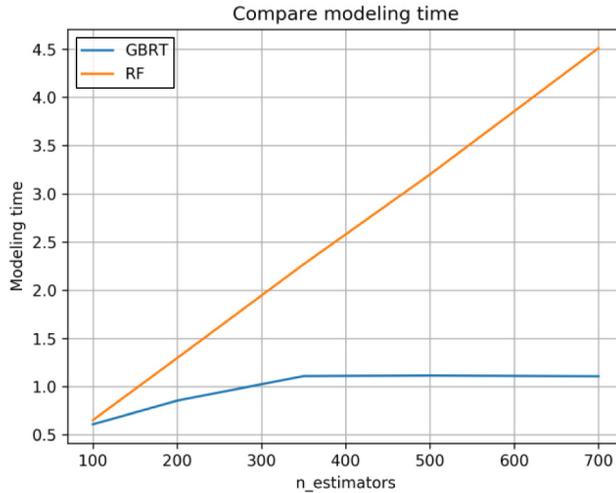
The [Table I](#) shows that GBRT has the best generalization performance. Compared with GBRT, RF has a similar generalization capability, but the modeling time of RF is six times that of GBRT. [Figure 4](#) and [Table II](#) show that with the increase of the number of base learners, the difference in the modeling time of different ensemble strategies is significant. The reason for this phenomenon is that the depth of the base learner is different. Although the modeling time for MLR and LASSO is short, the predictive accuracy is poor. The NN is not very effective without overfitting. We need more data and features and deeper networks.

We can see the performance comparison of the five algorithms directly from [Figures 5](#) to [14](#).

**Table I.**  
The performance of different methods

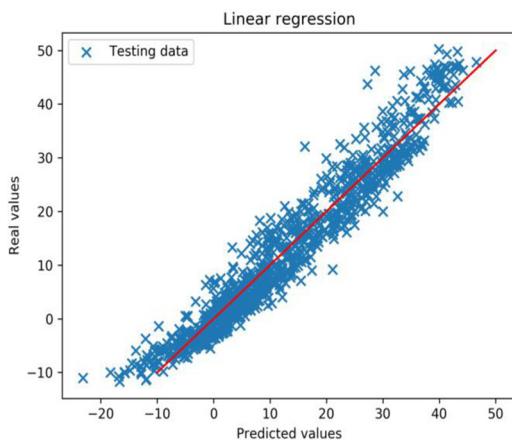
Methods	MSE	MAE	R2	Modeling time
MLR	14.8045	3.1006	0.9313	0.0052
LASSO	14.0944	3.0123	0.9311	0.0030
RF	9.1697	2.2572	0.9574	5.7513
GBRT	8.4899	2.1295	0.9606	0.8670
NN	17.5930	3.0189	0.9132	0.0650

**Figure 4.**  
Modeling time of GBRT and RF

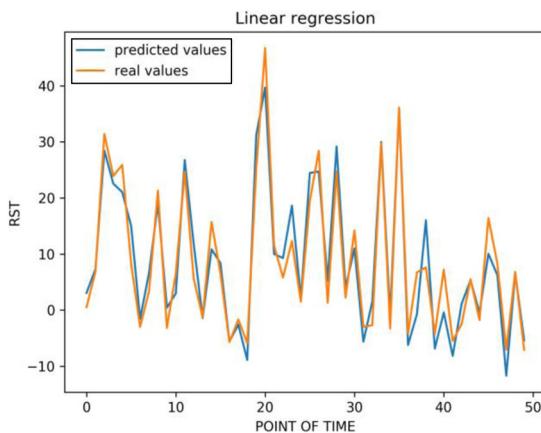


**Table II.**  
The performance of different number of base learners

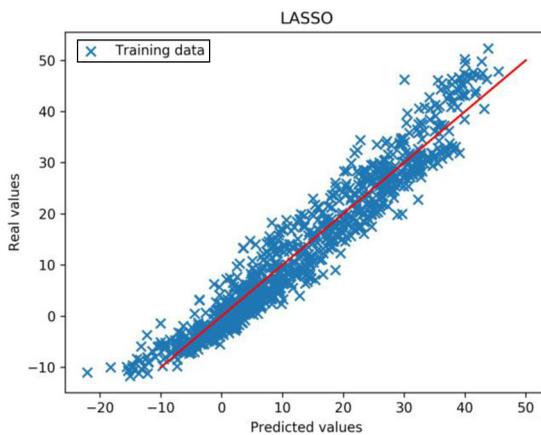
Decision trees	RM-MSE	RF-R2	GBRT-MSE	GBRT-R2
100	8.3527	0.9595	7.7539	0.9624
200	8.2295	0.9635	7.7182	0.9658
350	11.3758	0.9434	9.3037	0.9537
500	9.5780	0.9560	9.4538	0.9565
700	7.2041	0.9626	6.7853	0.9648



**Figure 5.** Regression plot for data samples with MLR

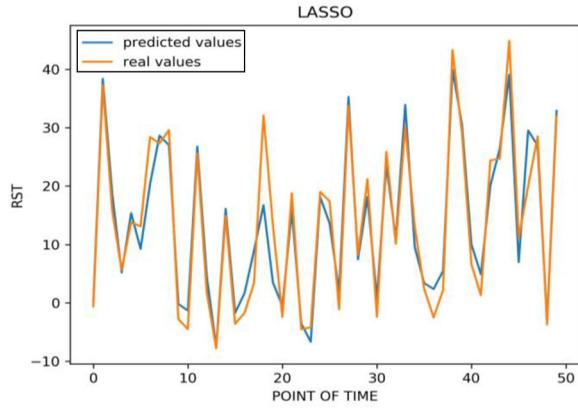


**Figure 6.** Line chart of real value and predicted value with MLR

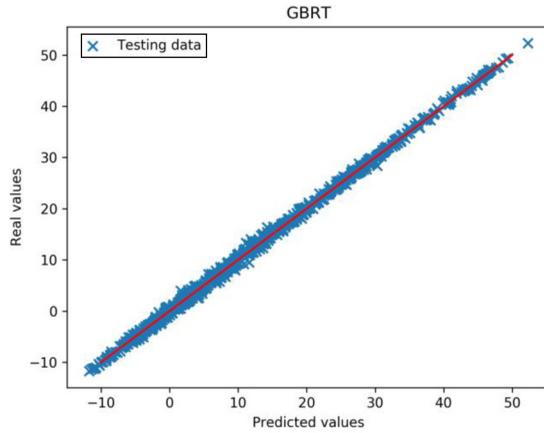


**Figure 7.** Regression plot for data samples with LASSO

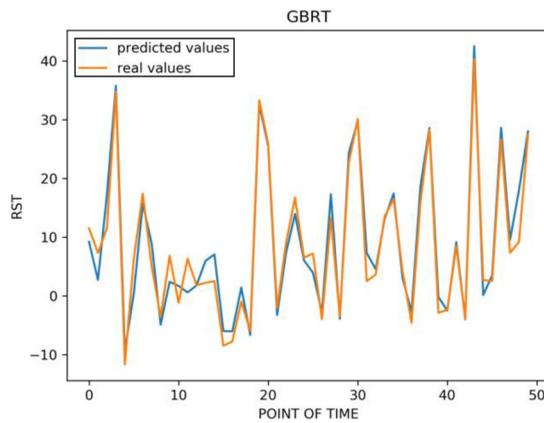
**Figure 8.**  
Line chart of real  
value and predicted  
value with Lasso



**Figure 9.**  
Regression plot for  
data samples of  
GBRT



**Figure 10.**  
Line chart of real  
value and predicted  
value with GBRT



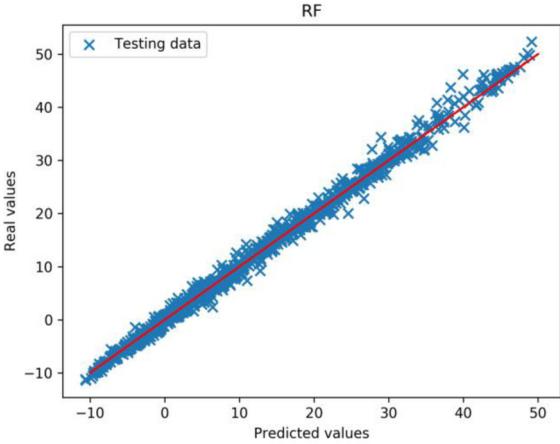


Figure 11. Regression plot for data samples with RF

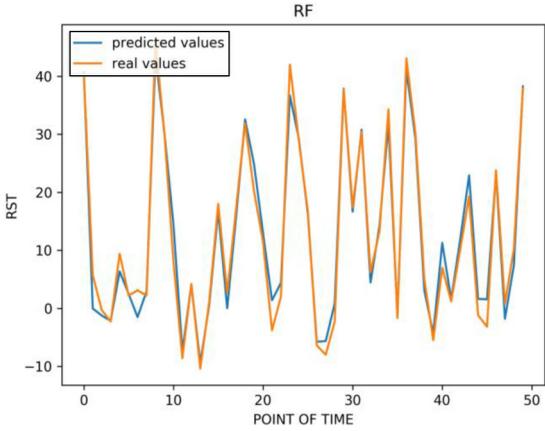


Figure 12. Line chart of real value and predicted value with RF

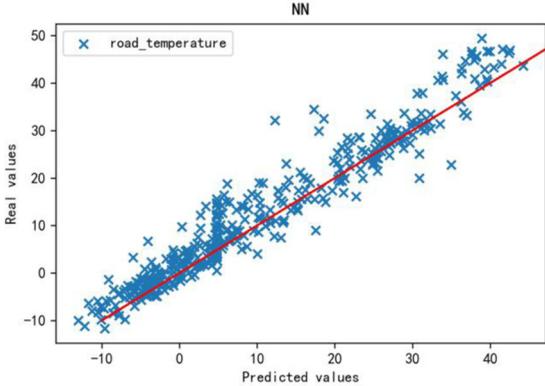
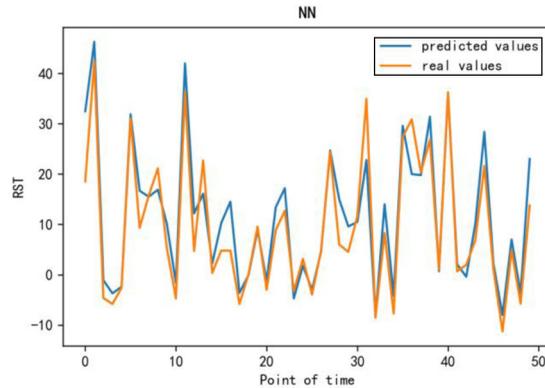


Figure 13. Regression plot for data samples with NN

**Figure 14.**  
Line chart of real  
value and predicted  
value with NN



Figures 5 and 6 show the regressive results of MLR model. It can be seen from Figure 5, the prediction result of using MLR model to predict the RST is instable, and the predicted value has a large deviation from the true value. The reason is that the model is a simple linear model and cannot capture the nonlinear relationship among the features; therefore, the overall prediction effect is general. As can be seen from Figure 6, the model does not predict well for the inflection point of RST.

Figures 7 and 8 show the regression results of LASSO model. Maybe it is because we have done feature selection, the results of LASSO and MLR are similar. Figure 7 shows the prediction result of LASSO. Although LASSO has a regularization compared with MLR, LASSO does not have a positive improvement in the MLR prediction results. The reason is that the input vector has only seven dimensions, after regularized, the dimension may be reduced, then the features will provide less information to the model, resulting in the final prediction results getting worse. It can also be seen from Figure 8, the predicted value differs greatly from the true value.

Figures 9 and 10 show the regression results of GBRT model. Figure 9 shows the best results of this paper, we can see that the prediction effect of GBRT relative to MLR is better, the difference between predicted value and real value of GBRT is smaller, and the stability of the model is also better. Figure 10 shows that the prediction of RST is very accurate, and the subtle changes in temperature can be learned. The reason why GBRT works well is that the model can constantly adjust the weight of features according to the results during the training process, and this Boosting strategy can reduce the instability of prediction. Therefore, when the dimension of our data set is not high and the amount of data is not large, ensemble algorithm is a good choice for us.

Figures 11 and 12 show the regressive results of RF model. Figures 11 and 12 show the prediction results of another ensemble algorithm RF, which also has a much-improved accuracy and stability relative to MLR and LASSO. The reason is that this ensemble strategy can combine the results of multiple sub-models, and finally give a more robust result, but the most obvious drawback of RF compared with GBRT is the long training time.

Figures 13 and 14 show the regressive results of NN model. Finally, this paper gives the prediction results of single-layer NN, and it is obvious that the results are not very good.

A very important reason for the poor prediction results is that our data are not enough, the feature dimensions are not complete, so we cannot predict RST from all aspects. We should work hard to look for more data and higher dimensions to get a more comprehensive prediction of RST through nonlinear learning.

## 5. Conclusions and future work

This paper compares the predictive accuracy of five algorithms on RST prediction. From the experimental results, it can be concluded that the generalization ability of the ensemble algorithm is stronger than that of the linear regression algorithm. At the same time, we can see that adjusting the parameters of integration strategy will have a great impact on prediction results and modeling time. In addition, because of the small amount of data, the performance of NN is not very good. In the future, we will try to use different basic learners to find the model with short modeling time and strong generalization capability, and use more data for deep learning.

## References

- Breiman, L. (1996), "Bagging predictors", *Machine Learning*, Vol. 24 No. 2, pp. 123-140.
- Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5-32.
- Dai, Y., Zeng, X., Dickinson, R., Baker, I., Bonan, G.B., Bosilovich, M.G. and Oleson, K., W. (2003), "The common land model", *Bulletin of the American Meteorological Society*, Vol. 84 No. 8, pp. 1013-1023.
- Edward, B. (1957), "Calculation of maximum pavement temperatures from weather reports", *Highway Research Board Bulletin*, Vol. 168, pp. 1-8.
- Feng, T. and Feng, S. (2012), "A numerical model for predicting road surface temperature in the highway", *Procedia Engineering*, Vol. 37, pp. 137-142.
- Friedman, J.H. (2001), "Greedy function approximation", *The Annals of Statistics*, Vol. 29 No. 5, pp. 1189-1232.
- Hansen, L.K. and Salamon, P. (1990), "Neural network ensembles", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12 No. 10, pp. 993-1001.
- Hastie, T. (2009), *The Elements of Statistical Learning*, Springer Science, USA.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015), "Deep learning", *Nature*, Vol. 521 No. 7553, pp. 436-521.
- Li, H. (2012), *Statistical Learning Methods*, Tsinghua University Press, Beijing.
- Liu, Y. (2005), "A mathematical model of multiple linear regression", *Journal of Shenyang Institute of Engineering*, Vol. 1 No. 2, pp. 128-129.
- Liu, B., Yan, S., You, H., Dong, Y., Li, J., Li, Y. and Gu, R. (2017), "An ensemble RBF extreme learning machine to forecast road surface temperature", *16th ICMLA, IEEE International Conference, IEEE*, pp. 977-980.
- Meng, C. and Liu, C. (2009), "Summer road temperature disaster forecast of expressway in Beijing", *J.of Institute of Disaster-Prevention Science and Technology*, Vol. 03 No. 7, pp. 26-29.
- Robert, T. (1996), "Regression shrinkage and selection via the LASSO", *Journal of the Royal Statistical Society*, Vol. 58 No. 1, pp. 267-288.
- Sass, B. (1997), "A numerical forecasting system for the prediction of slippery roads", *Journal of Applied Meteorology*, Vol. 36 No. 6, pp. 801-817.
- Tibshirani, R. (2011), "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 73 No. 3, pp. 273-282.

Wei, D., You, F., Fan, S., Yang, B. and Chen, M. (2010), "Evaluation and analysis of sounding quality of Beijing rapid renewal cycle prediction system (BJ-RUC) model", *Journal of Meteorological*, Vol. 36 No. 8, pp. 72-80.

Zhou, Z.H. (2009), "Ensemble learning", pp. 13-14.

Zhou, Z.H. (2016), *Machine Learning*, Tsinghua University Press, Beijing.

**Corresponding author**

Bo Liu can be contacted at: [boliu@bjut.edu.cn](mailto:boliu@bjut.edu.cn)