# Train delays prediction based on feature selection and random forest

Yuanyuan Ji, Wei Zheng, Hairong Dong, *Senior Member, IEEE* and Pengfei Gao

***Abstract*—Although trains are more efficient and convenient than other transportation, delays often occur. Accurately predicting the delay time of trains is of great significance to both dispatchers and passengers. The method for predicting the arrival delay time of trains is based on feature selection algorithm and machine learning. First, we collect train delay cases to sort out the delay factors. In addition to internal factors, external factors such as weather and signal failure are also considered. Then, an improved max-relevance and min-redundancy method (mRMR) is used for feature selection. Finally, we apply the method of weighted random forest (wRF) to predict the delay time. The results demonstrate that the feature selection algorithm has a prominent effect on improving the accuracy of the model, and the mean square error based on the weighted random forest has an improvement potential in forecast precision.**

## I. Introduction

With the development of high-speed railway technology, railway transport is faster and more convenient than it used to be, and passengers are demanding higher train punctuality. However, seen from the real-world operation records, trains often deviate from the trajectory of the planned diagram due to some interference factors which can cause delays. After the interference event, dispatchers need to adjust the train diagram to minimize the delay time. Forecasting the delay of trains is an important basis for revising train diagrams. At present, dispatchers predict the arrival time of trains using experience only, without adequate computer support, which not only increases the workload on dispatchers, but also reduces the reliability of the prediction results. Therefore, it is necessary to use scientific methods to predict the delay time of trains.

### *A. Related Works*

Various techniques have been used to analyze train delays. They can be roughly classified into two groups: One is the mathematical modeling method based on probability distributions. However, this method has a small generalization ability, and it must be complemented by Monte Carlo sampling method [1], [2] or maximum sum algebra simulation [3] in the application scenario. The rise of machine learning and artificial intelligence technology provides new methods and theories for train operation prediction which can make up for the low universality of mathematical model and the lack of automatic prediction.

The machine learning models use the presented data to learn a general rule that maps inputs to outputs, which typically provide better fit. Karlaftis analyzed the differences and similarities between machine learning methods and statistical methods [4]. Marković proposed a support vector regression prediction method. The author used the evaluation score of each line by experts as a feature input. The experimental results showed that the average prediction accuracy is 10 minutes [5]. Barbour et al. applied support vector regression to estimate the delay of freight traffic. They considered the fixed characteristics of departure station, train priority and the number of trains as influencing factors [6]. Huang used the operation data of Wuhan-Guangzhou high-speed railway to predict the recovery time of delays [7]. In 2019, Huang further studied the neural network model for real-time prediction of train delays. Based on the train operation performance data of Wuhan Guangzhou high-speed railway, a prediction model of train delays based on recurrent neural network (RNN) was established. The research results showed that the prediction accuracy of the proposed deep learning model is significantly higher than that of the existing models such as artificial neural network, support vector regression and Markov [8]. Zhou was interested in the prediction of train running time in sections under different speed limit conditions. They constructed the model based on random forest algorithm and combined with the influencing factors of running time. The results showed that the prediction model based on random forest is feasible and effective [9]. Most of the above researches are based on the train actual operation data. Most of the selected features are characteristics of trains and their operating laws without partial influencing factors caused by emergencies.

Based on train delay cases, this paper selects factors that affect train delays, including common factors and unexpected factors. Unexpected factors are the main information affecting the delay time related to equipment failure, bad weather and limit violation. The improved mRMR feature selection algorithm is used for feature processing. A weighted random forest algorithm is applied to establish the regression prediction model for realizing the prediction of train arrival delays. The prediction model can play a role in railway dispatching center. When an emergency occurs, dispatchers will start the

*This work is supported by the National Natural Science Foundation of China (No. 61790573)

Yuanyuan Ji is with the National Research Center of Railway Safety Assessment, Beijing Jiaotong University, Beijing, China. (e-mail: 18125107@bjtu.edu.cn).

Wei Zheng is with the National Research Centre of Railway Safety Assessment and the Beijing Key Laboratory of Intelligent Traffic Data Security and Privacy Protection Technology, Beijing Jiaotong University, Beijing, 100044, China (corresponding author to provide phone: +86-010-51685292; e-mail: wzheng1@bjtu.edu.cn).

Hairong Dong is with the State Key Lab of Rail Traffic Control & Safety, Beijing Jiaotong University, Beijing, 100044, China (e-mail: hrdong@bjtu.edu.cn).

Pengfei Gao is with the National Research Centre of Railway Safety Assessment and the Beijing Key Laboratory of Intelligent Traffic Data Security and Privacy Protection Technology, Beijing Jiaotong University, Beijing, 100044, China (e-mail: pfgao@bjtu.edu.cn).

prediction of delay time according to the current situation to adjust the train diagram. The forecast results can also be transmitted to stations, providing timely information to passengers (the scenario diagram of delay prediction is shown in Fig. 1). In case of emergency, the first train involved is mainly predicted. The subsequent train delays can be calculated by the running time interval. Therefore, the research object of this paper is a single train, without considering the interaction between trains.

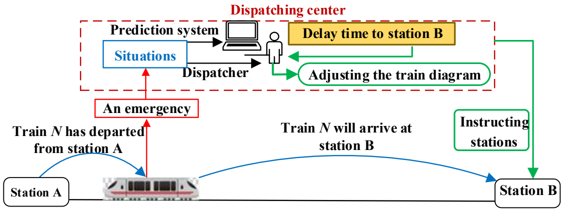


Figure 1. Scenario diagram of train delays prediction

## B. Paper Structure

The rest of this paper is arranged as follows. The second part introduces the selection of delay factors and the deformation of mRMR algorithm. The weighted random forest prediction model is described in Section III. Section IV shows the implementation process and prediction results. Finally, the conclusion and the future work are given in Section V.

# II. Delay Factors and Feature Selection Algorithm

This part mainly analyzes cases of train delays and finds out the factors that affect the delays. We use the improved mRMR algorithm to achieve feature selection, which makes the features finally input into the prediction model with high correlation and low redundancy. The research framework of this paper is shown in Fig. 2.

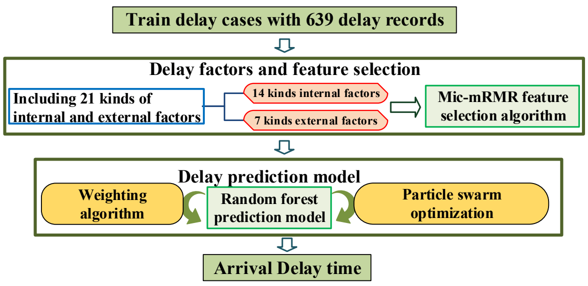


Figure 2. Train delays prediction system framework

## A. Train Delay Factors

The railway system consists of many subsystems, which are characterized by complexity, dynamics and openness. The factors of train delays are even more complicated. They are mainly divided into two aspects: one is external factors, the other is the common factors. The external factors mainly include equipment failure, bad weather, limit invasion and human factors. For the common factors, many researchers have conducted relevant analysis. Liu believes that the running time and the dwell time of the train are related to the delay time by analyzing the train actual operation data [10]. Meng selected four dispatching stations operation data of high-speed railway to explore the law in train operation delays. The results showed that the running time delay rate was positively correlated with the average running delay time and the average increase running delays [11].

The cases of train delays studied in this paper comes from 639 delay records of Wuhan-Guangzhou railway line. The main information covered in the delay records includes date, train number, delay description, delay time, etc. Some original records intercepted are shown in Table I. Based on the researches of the above scholars and combining the cases of train delays, we selected 21 kinds of influencing factors as pre-selected features. The specific features and descriptions are listed in Table II. The bold fonts are reference features, and the normal fonts mean delay cases features. In the last column of the table, static represents the fixed common factors, and dynamic shows the external factors that change with unexpected conditions. The specific types of incident are shown in Fig. 3 and Fig. 4, including 13 kinds of signal failures (Fig. 3), 3 kinds of weather causes and 2 kinds of limit invasion (Fig. 4). In Fig.4, the coded numbers 1-13 correspond to 13 signal faults in Fig.3.

Since the correlation between the feature factors we selected and the predicted values is unknown. There may be redundancy between the features. In order to ensure that each feature input into the machine learning model is as unique as possible, we first analyze the correlation and redundancy of the delay data.

The Pearson correlation coefficient measures the degree of correlation between two things (variables). The closer the correlation coefficient is to 1 or -1, the higher two variables are correlated. On the contrary, the closer the correlation coefficient is to 0, the weaker the correlation is.

TABLE I. Train Delay Records

| Num. | Date | Train number | Vehicle number | Time of receiving fault information | Delay description |
|---|---|---|---|---|---|
| 45 | 01-12 | G6031 | CRH3-76 | 18:40 | At 18:37, G6031 stopped due to an ATP fault at the K2052 + 351 section between Lechang East and Shaoguan. It was driven at 18:48 after the system was restarted. The train should arrive at Shaoguan Station at 19:23, the actual time is 19:28, and 5 minutes later. |
| 46 | 01-20 | G6006 | CRH3-65 | 8:30 | At 8:32, G6006 (CRH3 / 65) timed out at K2219 + 000 between Guangzhou South and Guangzhou North, which caused C3 to C2, and then returned to C3 after the departure of Guangzhou North. The train should arrive at Guangzhou South Station at 8:56 Minutes, real time is 8:59 minutes. |

TABLE II. TRAIN DELAY FACTORS

| Class | Factors and number | Description |
|---|---|---|
| 1 | Train type-0, Types of locomotive-1, **Train delay history-2, The current station-3, Arrival station-4, Plan running time-5, Plan dwell time-6,** Departure delay-7, **Line attributes-8, Station size-9, Station history delays-10, Section length-11, Section atttrubutes-12, Festival-20** | Static |
| 2 | Fault location-13, Time of occurrence-14, Incident type-15, Speed limit level-16, Speed limit length-17, Stop-18, Failure duration-19 | Dynamic |

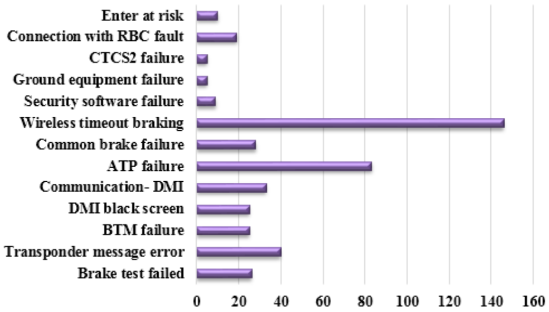


Figure 3. Signal failure statistics

| Data encoding | Types of Incident | Number of samples |
|---|---|---|
| 14 | Strong wind | 79 |
| 15 | Heavy rain | 47 |
| 16 | Heavy snow | 16 |
| 17 | True invasion limit | 19 |
| 18 | False invasion limit | 31 |

Figure 4. Bad weather and violation limits types and sample sizes

We assume that there are two variables X and Y. The Pearson correlation coefficient between the two variables can be calculated by formula (1):

$$\rho_{X,Y}=\frac{cov(X,Y)}{\sigma_X\sigma_Y}=\frac{E(XY)-E(X)E(Y)}{\sqrt{E(X)^2-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \quad (1)$$

where $E$ is the mathematical expectation and $cov$ is the covariance. The thermodynamic diagram of Pearson correlation coefficient of factors is shown in Fig. 5.

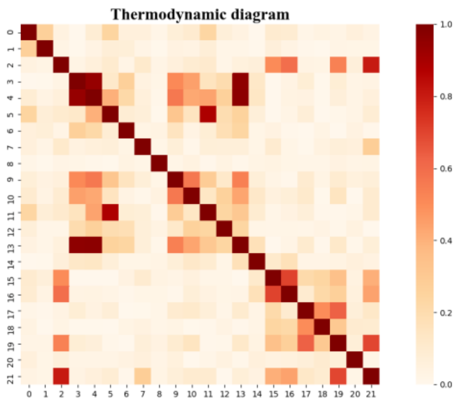


Figure 5. Thermodynamic diagram of Pearson correlation coefficient among variables

In the thermodynamic diagram, if the color is dark, it represents a large correlation between two variables. Conversely, the lighter the color, the weaker the correlation between them. It can be seen from Fig. 5 that among the 21 feature variables selected, there are redundant features and uncorrelated features. How to select the most relevant feature among the feature variables and ensure the least correlation between the selected variables is the focus of next part in this paper.

## *B. Feature Selection Algorithm*

An ideal feature selection algorithm should remove irrelevant, weak and redundant features and retain some relevant non-redundant features [12] [13]. In this paper, an improved mRMR algorithm is used to realize feature selection. The original mRMR algorithm [14] is a feature selection method based on a search strategy, which uses the mutual information between variables to select the features that are most correlated with the prediction variable. At the same time, it also removes the features with great redundancy. The mutual information value in mRMR represents the information quantity of the common part between two variables, which is an important measurement method in information theory. Despite the theoretical value of mutual information, it is often hard to get an ideal feature set, because mutual information is sensitive to discrete values and can't deal with continuous features very well.

In the process of feature selection, the evaluation criteria of features are employed to appraise the quality of features. Different methods have a great influence on the selected feature subset. Evaluation criteria can not only describe the correlation between features and predicted values, but also measure the redundancy between features. The existing evaluation criteria can be divided into distance measure (such as Euclidean distance, Mahalanobis distance, etc.), consistency measure [15], dependence measure (such as Pearson coefficient), information measure (e.g. mutual information [16], information gain, maximum information number [17], etc.) and classification accuracy measure. We do not consider consistency measure and distance measure, since the consistency measure only works in the classification models, and the classification performance of the distance measure is low. Although classification accuracy measurement has quite classification accuracy, its time complexity is very high. The maximum information coefficient (MIC) can overcome the problems of mutual information coefficient which has good generalization performance. It can not only deal with discrete features but also continuous features, and can be applied for classification and regression problems. The MIC of features $f_1$ and $f_2$ can be expressed as follows:

$$MIC(D)=\max_{XY<B(n)} M(D)_{X,Y}=\max_{XY<B(n)}\frac{I^*(D,X,Y)}{In(min(X,Y))} \quad (2)$$

where, $D=\{(f_{1i}, f_{2i}, i=1, 2, ...), n)\}$ is an ordered set. $X$ and $Y$ divide $f_1$ and $f_2$ into $X$ and $Y$ segments respectively. $B\ (n)$ represents the number of grids. The denominator normalizes the mutual information values under different partitions.

The flow of MIC-mRMR algorithm: Let $K$ be the number of features to be selected. First, we take all feature sets and number each feature. Second, we calculate the MIC values of all features to form the mRMR pool, including the selected features and the remaining features. Then features with the maximum MIC value are selected to add to the selected feature set, and the remaining features with the method of traversal are extracted. The final subset of mRMR features meets the following equations:

$$max\ D(S,c),\ D=\frac{1}{|S|}\sum_{x_i\in S} I(x_i,c) \tag{3}$$

$$max\ R(S),\ R=\frac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i,x_j) \tag{4}$$

$$max\ \Phi(D,R),\ \Phi=D/R \tag{5}$$

where $S$ represents a subset of mRMR features, $c$ means a predictor variable, and $x_i$, $x_j$ $(i, j = 1, 2, ..., K,\ i \neq j)$ are any two features in $S$. The mRMR algorithm framework based on MIC is shown in Fig. 6.

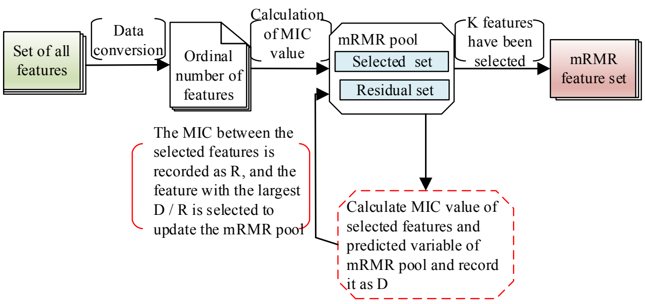


Figure 6. MIC-mRMR algorithm framework

## III. RANDOM FOREST PREDICTION MODEL

### A. Random Forest

Random forest is an integrated decision tree model belonging to the bagging framework [18]. Randomness is the core of the random forest. By randomly selecting samples and features, the correlation between decision trees can be reduced. There are two meanings of randomness in random forest. One is to select the same amount of data as the training sample from the original training data randomly. The other is to randomly select features to establish a decision tree. These two kinds of randomness can reduce the correlation among the decision trees and further improve the accuracy of the model. Finally, these decision trees are generated to form a random forest, which is a classifier performs classification and regression prediction on new data. For classification problems, the final result is determined by voting on multiple tree classifiers. The average predicted value of multiple trees determines the final result in the regression problems. The randomness idea is shown in Fig. 7.

Despite the fact that random forest algorithm has the advantages of high accuracy and is not prone to overfitting, the accuracy of this algorithm largely depends on the parameters, and parameters adjustment generally requires manual selection. There is no optimal parameter based on the optimization theory at present. In addition, the final prediction value of the regression model is to take the average value of each decision tree, without accuracy distinction. Therefore, we have improved the random forest model for the above problems.

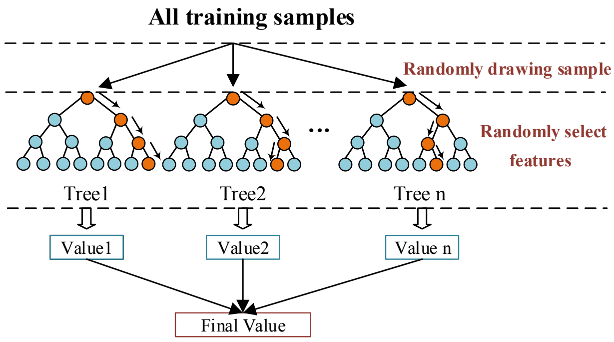


Figure 7. Random ideas in a random forest model

### B. Particle Swarm Optimization for Hyper-parameters

There are three main hyper-parameters in the random forest algorithm: the decision tree size *n_tree*, the maximum depth of the decision tree *max_depth*, and the number of feature subspaces $K$. The model accuracy obtained by training with different combinations of hyper-parameters can vary widely. The particle swarm optimization (PSO) algorithm is adopted to optimize the hyper-parameters to improve the accuracy of random forest.

PSO is a heuristic optimization method [19] [20], which simulates the way of birds foraging. It has been widely used in various combinatorial optimization problems in recent years. In this algorithm, the entire bird swarm is compared to a particle swarm. A particle represents a feasible solution. Each particle continuously updates its position and speed through its own experience and group experience, and finally reaches a globally optimal position. It is assumed that the total number of particles is $N$, and the position of the *ith* particle in $M$-dimensional space is expressed as $x_i = (x_{i1}, x_{i2}, ..., x_{iM})$, and the velocity is expressed as $v_i = (v_{i1}, v_{i2}, ..., v_{iM})$. The particles are updated according to the following equations:

$$v_i=\omega\cdot v_i+c_1\cdot rand()\cdot(pbest_i-x_i)+c_2\cdot rand()\cdot(gbest_i-x_i) \tag{6}$$

$$x_i=x_i+v_i \tag{7}$$

where $w$ is the inertia factor, *rand* () is a random number between (0, 1), $c_1$ and $c_2$ are learning factors, *pbest* and *gbest* are their own optimal values and group optimal values respectively.

### C. Weighted Optimization Algorithm

The final prediction value of original random forest is the average of each decision tree. Since the prediction accuracy of each tree in the random forest model is different, some of which are high and some of which are low, taking the mean directly will increase the error. This paper gives a certain weight to each tree and finally calculates the weighted average value of all decision trees to reduce the prediction error.

There are four main evaluation criteria for the accuracy of regression tree:

- Mean square error (MSE): The statistical parameter is the mean of the sum of squares of the corresponding point errors of the predicted data and the original data. The smaller the value, the higher the accuracy of the prediction result. The calculation equation is:

$$\mathrm{MSE}=\frac{1}{m}\sum_{i=1}^{m}\left(f_i-y_i\right)^2 \tag{8}$$

In equation (8), $m$ is the number of predicted samples, $f_i$ is the predicted value, and $y_i$ is the true value of the sample.

- Root mean square error (RMSE): Standard error is the arithmetic square root of mean square error. It is used to measure the deviation between the observed value and the true value. The significance of RMSE is that after the square root, the error result is at the same level with the data, which can better describe the data. The calculation equation is:

$$\mathrm{RMSE}=\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(f_i-y_i\right)^2} \tag{9}$$

- R-squared: The calculation formula of the determination coefficient is (10), where $R$ is the correlation coefficient, and the square of it is the determination coefficient $R^2$. The value of $R^2$ is between 0 ~ 1. If the value is close to 1, it means that the model fits well. It is generally considered that the model with a good fit exceeds 0.8.

$$R^2=1-\frac{\sum_{i=1}^{m}(f_i-y_i)^2}{\sum_{i=1}^{m}(\bar{f}_i-y_i)^2} \tag{10}$$

The evaluation criterion value of the determination coefficient is proportional to the prediction accuracy, so we choose r-squared as the weight coefficient. The calculation equation of weighted random forest prediction result is as follows:

$$\mathrm{value}=\frac{1}{n}\sum_{i=1}^{n}(r_i^2\cdot f_i) \tag{11}$$

The improved random forest algorithm block diagram is shown in Fig. 8.

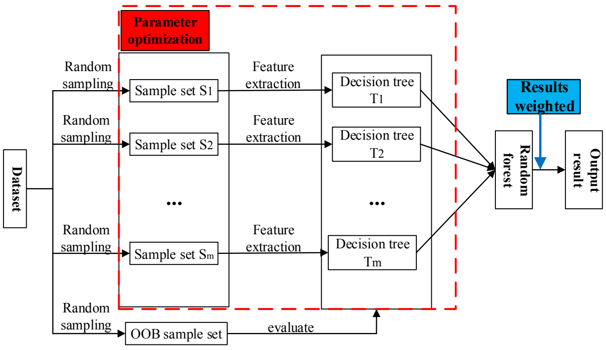


Figure 8. Improved random forest block diagram

## IV. EXPERIMENTS AND RESULTS

### A. Feature Selection Results

First of all, the particle swarm optimization algorithm was applied to the delay data with size 22 * 639. The optimal parameter random forest hyper-parameters combination was obtained after 200 iterations. All the algorithms and models in this article are built with the support of Python. The iteration results are: *n_tree* = 112, *max_depth* = 9 and $K$ = 12.

We tested three different methods to select the features: original mRMR algorithm, mRMR algorithm based on the MIC coefficient, and mRMR algorithm based on the MIC and Pearson coefficient (MIC and Pearson coefficients are used as evaluation criteria for $D$ and $R$ respectively). Since the best feature number of PSO algorithm is 12, we finally selected 12 features and applied three feature selection algorithms to the delay data. The results are shown in Fig.9.

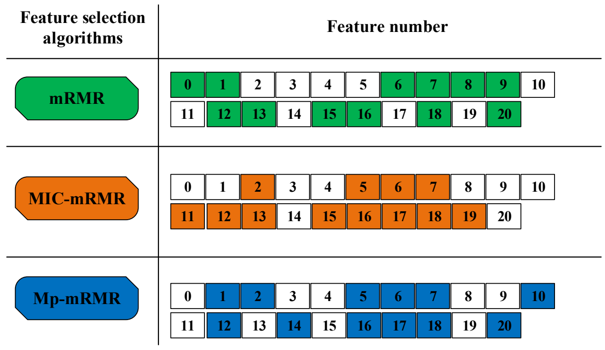


Figure 9. Improved random forest block diagram

In Fig. 9, each colored feature number represents that it was selected by the algorithm. Obviously, in the result graph, the features selected by the three algorithms have certain differences. We applied the feature data selected to the random forest model. The prediction accuracy is shown in Table III.

TABLE III. TRAIN DELAY FACTORS

| Algorithm / evaluation criteria | MSE | RMSE | R2 |
|---|---|---|---|
| -- | 7.269 | 2.701 | 0.894 |
| mRMR | 5.236 | 2.288 | 0.913 |
| MIC-mRMR | **1.056** | **1.027** | **0.979** |
| Mp-mRMR | 2.369 | 1.539 | 0.957 |

We also calculated the errors of the data without feature selection for comparison. In Table III, the prediction accuracy of the data extracted by the feature selection algorithm is generally higher than that of the original data. Among the three feature selection algorithms, the MSE of MIC-mRMR algorithm is 1.056, which is significantly lower than the other two algorithms, and $R^2$ is higher than other algorithms, with a value of 0.979. Finally, we chose MIC-mRMR algorithm for feature selection, and then applied it to the later prediction model.

### B. PSO-wRF Prediction Model

We established the weighted random forest model, which used the feature data selected by the MIC-mRMR algorithm. The particle swarm optimization algorithm was used to optimize the hyper-parameters of the random forest. The forest size *n_tree*=12, the maximum depth of the decision tree *max_depth*=9, and the number of characteristic variables $K$=12.

In this paper, 70% of the delay data were used to train the model, and the remaining 30% were used to evaluate the training effect of the model. The result of the weighted random forest prediction model is shown in fig. 10. The horizontal axis in the figure is the sample of the test set, and the vertical axis is the delay time (minutes). The prediction accuracy of the model is MSE = 0.541, which represents that the accuracy of weighted random forest prediction is significantly improved compared with the original random forest algorithm.

In order to evaluate the results of the weighted random forest, we also considered the depth neural network (DNN) model, decision tree (DT) model and support vector regression (SVR) as the comparison. Three hidden layers are selected by depth neural network, the maximum depth of decision tree is 9 and the number of characteristic variables is 12. The kernel function of support vector regression is linear kernel function. The training and testing data set are the same as the weighted random forest model.

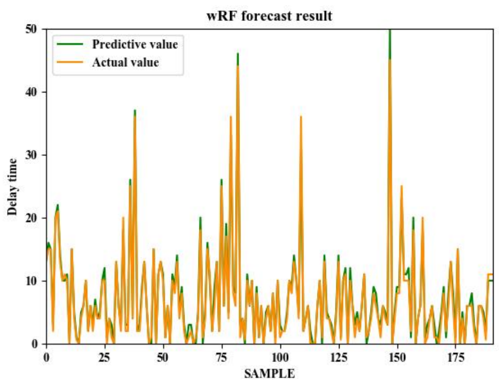


Figure 10. Weighted random forest prediction results

The prediction results of each model are shown in Fig. 11. The mean square error of the weighted random forest is 0.541, which is obviously smaller than DNN, SVR and DT, and the $R^2$ value of the weighted random forest is also higher. Indeed, our results show that wRF outperforms other models on the selected delay data.

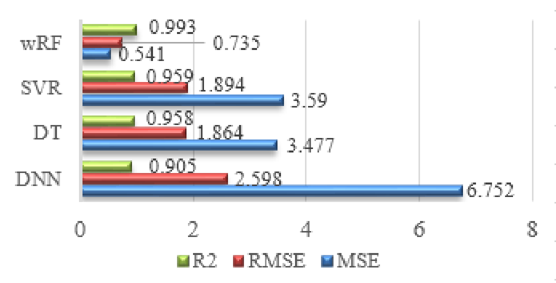


Figure 11. Prediction results of different models

## V. Conclusion and Future Work

By analyzing the records of train delays, we can comprehend the factors related to the delay time. For our train delay records, although signal failure, bad weather, and limit violations are the main factors affecting delay time, human factors and other equipment failures (e.g. catenary failure) can also be considered. On the practical side, pre-processing steps were necessary before the models were trained. The results of feature selection encourage further applications of the MIC-mRMR. The accuracy of the weighted random forest model optimized by PSO is better than that of decision tree, support vector regression and depth neural network model, with an average absolute error of 0.541. Several enhancements can be considered for future work. While our initial goal was to create a general prediction model for train arrival delays, the idea of departure delays could be explored. The model performance is sensitive to hyper-parameters, such as the decision tree size and the number of features. While we did tuning these, an improved PSO approach to get the globally optimal solution is likely to be beneficial.

## References


[1] J. Yuan, R. M. P. Goverde and I. A. Hansen, "Propagation of train delays in stations," *WIT Transactions on The Built Environment*, vol. 61, no. 1, pp. 975-984, 2002.

[2] Middelkoop, Dick, and M.Bouwman, "Simone: Large scale train network simulations," *Simulation Conference, 2001. Proceedings of the Winter IEEE*, 2001.

[3] W. Jiang and X. Li, "Research on the Delay Model of High Speed Train Operation Based on Maximum Algebra," *Railway Transport and Economy*, vol. 37, no. 1, pp. 66-71, 2015.

[4] Karlaftis, M.G., E.I.Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transportation Research Part C Emerging Technologies*, vol. 19, no. 3, pp. 387-399, 2011.

[5] N. Marković, S. Milinković, K.S.Tikhonov and P. Schonfeld, "Analyzing passenger train arrival delays with support vector regression," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 251-262, 2015.

[6] W. Barbour, J. C. M. Mori, S. Kuppa, et al., "Prediction of arrival times of freight traffic on US railroads using support vector regression," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 211-227, 2018.

[7] P. Huang, Q. Peng, C. Wen and Y. Yang, "Stochastic forest model for predicting the delay recovery time of trains on the Wuhan-Guangzhou high speed railway," *Journal of the China Railway Society*, vol. 40,no. 7, pp. 1-9, 2018.

[8] P. Huang, C. Wen, Z. Li, Y. Yang and Q. Pen, "Neural network model for real-time prediction of late time of high-speed railway trains," *Chinese Journal of Safety Science*, vol. 29, no. S1, pp. 20-26, 2019.

[9] X. Zhou, Q. Zhang and W. Xu, "Study on time-division prediction of train interval operation based on random forest under different speed limits," *Railway Transport and Economy*, vol. 40, no. 2, pp. 18-23+54, 2018.

[10] Y. Liu, J. Guo, C. Luo, et al., "Big data analysis of train operation performance and application prospects," *China Railway*, no. 6, pp. 76-79, 2015.

[11] Z. Meng, Z. Zuo, P. Yin, et al., "Analysis of characteristics of high-speed railway trains running late," *Chinese Journal of Safety Science*, vol. 28, no. S2, pp. 97-102, 2018

[12] J. Wang, L. Ci and K. Yao, "Overview of Feature Selection Methods," *Computer Engineering and Science*, vol. 27, no. 12, pp. 68-71, 2005.

[13] Y. Li, J. Du, B. Nie, W. Xiong, C. Huang and H. Li, "A Summary of Feature Selection Methods," *Computer Engineering and Applications*, vol. 55, no. 24, pp. 10-19, 2019.

[14] H. Peng, F. Long and C. Ding, "Feature Selection Based On Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, no. 08, pp. 1226-1238, 2005.

[15] A. Arauzo-Azofra, J. M. Benitez and J. L. Castro, "Consistency measures for feature selection," *Journal of Intelligent Information Systems*, vol. 30, no. 3, pp. 273-292, 2008.

[16] A. K. Jain, R. P. W. Duin and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37, 2000.

[17] D. N. Reshef, Y. A. Reshef, H. K. Finucane, et al., "Detecting novel associations in large data sets, " *Science*, vol. 334, no. 6062, pp. 1518-1524, 2011.

[18] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996.

[19] J. Kennedy R. Eberhart , "Particle swarm optimization," *Proceedings of ICNN'95 - International Conference on Neural Networks IEEE*, 1995.

[20] J. W. Wang, H. N. Li. "Overview of Particle Swarm Optimization Algorithms," *Modern Computer (Professional Edition)*, no. 2, pp. 24-29, 2009.