Analysis on Train Stopping Accuracy based on Regression Algorithms

Lin Ma
National Engineering Research Center of Rail Transportation Operation and Control System, Beijing Jiaotong University, Beijing, 100044, China.
linma@bjtu.edu.cn

Xiangyu Zeng
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China.
11120319@bjtu.edu.cn

***Abstract*—Stopping accuracy is one of the most important indexes of efficiency of automatic train operation (ATO) systems. Traditional stopping control algorithms in ATO systems have some drawbacks, as many factors have not been taken into account. In the large amount of field-collected data about stopping accuracy there are many factors (e.g. system delays, stopping time, net pressure) which affecting stopping accuracy. In this paper, three popular data mining methods are proposed to analyze the train stopping accuracy. Firstly, we find fifteen factors which have impact on the stopping accuracy. Then, ridge regression, lasso regression and elastic net regression are employed to mine models to reflecting the relationship between the fifteen factors and the stopping accuracy. Then, the three models are compared by using Akaike information criterion (AIC), a model selection criterion which considering the trade-off between accuracy and complexity. The computational results show that elastic net regression model has a best performance on AIC value. Finally, we obtain the parameters which can make the train stop more accurately which can provide a reference to improve stopping accuracy for ATO systems.**

***Index Terms*—data mining, train stopping accuracy, ridge regression, lasso regression, elastic net regression**

I. INTRODUCTION

In recent years, urban railways have been developed rapidly due to high-speed, punctuality and safety in public transportation systems [1]. Most developed cities in China such as Beijing and Shanghai have to depend on metros to reduce the public traffic jam [2]. Automatic train operation (ATO) systems have been widely used in current urban rail transit systems to improve their performances. Stopping accuracy is one of the most important performance indexes in evaluating the efficiency of ATO systems. Some measures and solutions are conductive to increasing the efficiency of the metro operation, such as analyzing the relationship between the stopping accuracy and train operation parameters, finding out the reasons which lead to inaccurate stopping and putting forward some suggestions for improving stopping accuracy.

The stopping accuracy analysis has been mainly studied by some subway operating companies or research institute. H.-P Yu analyzed the stopping accuracy of urban rail train based on the braking system, signal system, traction control system and line conditions, then presented some suggestions for the improvement of train braking system [3]. X.-Y Li analyzed the failure of inaccurate stopping in ATO systems, then found out the reasons affecting the stopping accuracy in terms of route and speed, at last provided improvement methods on signal professional [4]. A simple model for train stopping is proposed by Richard Banach by investigating the development of train control systems [5]. D.-W. Chen presented least square estimation and an adaptive network based fuzzy inference system (ANFIS) to estimate the train station parking error in urban rail transit [6,7] and employed some online learning algorithms to dynamically change braking rate to reduce the stopping error [8]. X.-J. Jiang found the reasons why the train stopped inaccurately by analyzing the real time running data in the station, then presented a method to find the installation location of the platform screen door (PSD) to make the train stop more accurately [9].

The previous works have achieved some results about the stopping accuracy analysis. However, the previous works mainly laid emphasis on the locomotive performance in the ATO systems. And, only a few of factors, such as speed, time, location, were studied. Furthermore, the field-collected data were too small to mine enough useful rules for stopping accuracy. Some specific factors which could influence stopping accuracy still cannot be found and explained.

Different from previous research, we use large amount of field-data and regression methods in data mining to establish some models which reflect the relationship between stopping accuracy and many other factors. Moreover, we concretely analyze which factors influence the stopping accuracy and how to influence it in order to provide a basis for improving stopping accuracy.

doi:10.4304/jsw.9.5.1237-1244

II. DATA PREPROCESSING

A. Data Cleaning

Some abnormal points and noise will be inevitably added to the process of data acquisition. On one hand, the data inputted manually will be puzzled by input errors; on the other hand, noise data also exists in the process of automatic data collection caused by equipment malfunction. Furthermore, there may be some error data in the collection of precise stopping data, e.g., deviation in the recording of ATO operation data, or failure of data collecting device. Hence, processing for some abnormal points is required to deal with the above situations.

As the stopping accuracy is a continuous variable, the method used in this paper is regression, where the data is processed smoothly on the basis of data fitting. Multiple linear regression is the most popular data analysis method. The linear relationship between variables and the output is attained by using the sample data fitting, after distinguishing the difference between the normal and abnormal value input through the function of in-out relations. The data with big errors are considered as abnormal data, and they will be eliminated.

The data cleaning and regression analysis are performed by Matlab. To do regression analysis, we first need to collect input-output data pairs. For the train stopping accuracy, we obtained 1600 samples with 15 inputs and one output. Of course, the output variable is the stopping accuracy. As to the other 15 input variables, more details can be found in Section 3. After processing, the goodness of fit and residual of stopping accuracy are shown in Fig.1 and Fig.2.

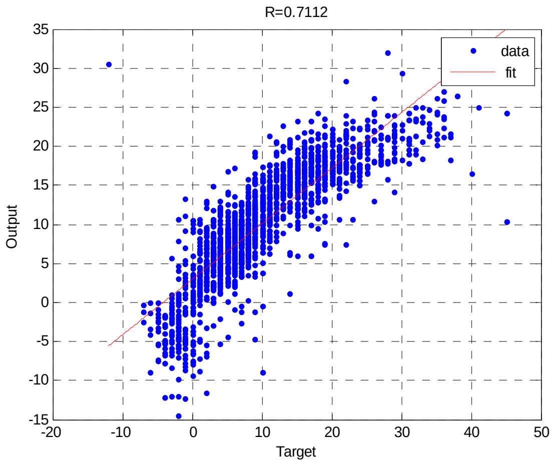

Fig.1 Fitting of stopping accuracy

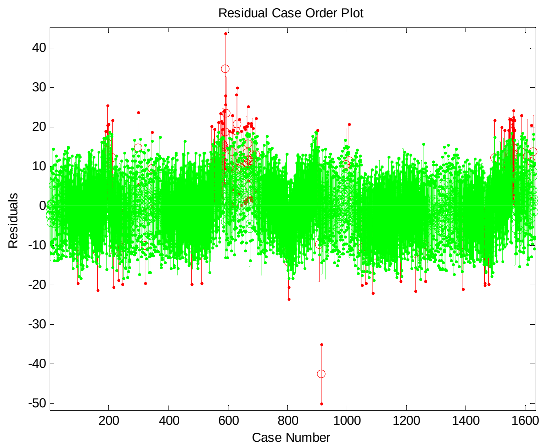

Fig.2 Data cleaning residual

The model fitting value is 0.7112. As it is shown in Fig.2, the data with big residual errors are marked by circular mark. The area through the vertical line indicates the 95% confidence interval of the residual validation. The red represents a departure from the regression of observed value, and these data will be eliminated.

After several rounds of data cleaning, the data with big errors have been eliminated. The goodness of fit of stopping accuracy after cleaning is shown in Fig.3. The model fitting value has increased to 0.8509.

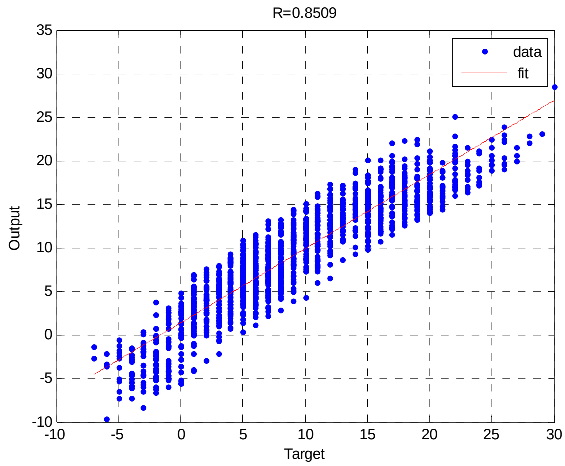

Fig.3 Fitting of stopping accuracy after cleaning

B. Data Normalization

In some cases, the data will be standardized to make them change in a certain range. For some algorithms, we need to compare and operate among the variables. If there is no normalization processing, the final results will be distorted. The method employed in this paper is range normalization, which is commonly used in regression analysis.

The method linearly changes the data between 0 and 1. Find a variable in the maximum (a_{max}) and minimum (a_{min}), and then utilize the formula below:

$$a' = \frac{a - a_{\min}}{a_{\max} - a_{\min}} \quad (1)$$

Since this method has linear features, it makes the distribution of the original variables unchanging.

III. Appling Data Mining Methods to Analyze Stopping Accuracy

Data mining [10,11] is a kind of intelligent data analysis techniques, rising at the end of twentieth century. Due to its advantage in automatically extracting and finding useful information from the mass data, data mining has been applied in many fields. Data mining methods includes regression analysis [12], classification analysis [13], clustering analysis [14], association analysis [15] and so on. In this section, three classic regression methods will be used and compared to find the best one for analyzing the train stopping accuracy.

By analyzing the data from ATO systems, we find that there are 15 variables which can affect the stopping accuracy. We define the 15 input variables as x1, x2, … , x15, the meaning of each input variables are shown in Table 1. Obviously, there is only one output variable: the stopping accuracy or stopping error.

TABLE I.

Value of Parameters in Ridge Regression Model

Variable	Name
x_1	Distance from the target point at the time of air braking (cm)
x_2	Speed at the time of electro-pneumatic switching (cm/s)
x_3	Distance from the target point at the time of electro-pneumatic switching (cm)
x_4	Air braking time (s)
x_5	Mean acceleration of controller's output in air braking (cm/s^2)
x_6	Time (s)
x_7	Air braking rate
x_8	Mean between goal and actual values in electric braking stage (cm)
x_9	Variance between goal and actual values in electric braking stage
x_{10}	Mean between goal and actual values in air braking stage (cm)
x_{11}	Variance between goal and actual values in air braking stage braking stage
x_{12}	Weight of train braking stage (t)
x_{13}	Mean of network voltage in braking stage (V)
x_{14}	Variance of network voltage in braking stage
x_{15}	Mean of gradient in braking stage (‰)

A. Ridge Regression

The estimation of coefficients in the multiple linear regression model relies on the independence of variables in the model. When the variables are linked to each other and the variables approximate linear correlation, the matrix $(x^T * x)^{-1}$ is closed to a singular matrix. Then, least squares estimation is as follow:

$$b = (x^T * x)^{-1} * x^T * y \tag{2}$$

The least squares estimation makes it extremely sensitive to the large random error of the observed value y and produces a larger error.

Ridge regression solves this problem by the following formula:

$$b = (x^T * x + kI)^{-1} * x^T * y \tag{3}$$

where k is ridge parameter, and I is unit matrix. This is the ridge regression [16-18] proposed by Hoerl and Kannard in 1970. And, they proved that there was a parameter k (k>0) which satisfied Eq. (4)

$$MSE(b(k)) < MSE(b) \tag{4}$$

Hence, the main task of the ridge estimated and analytical method turns out to find the minimum value k.

Attaining the k value, we can calculate the model parameters of ridge regression. We select 60% of 1366 samples after cleaning randomly as the training data set, and the remaining 40% are employed as the testing data set. The fitting diagram of the training data set is shown in Fig.4.

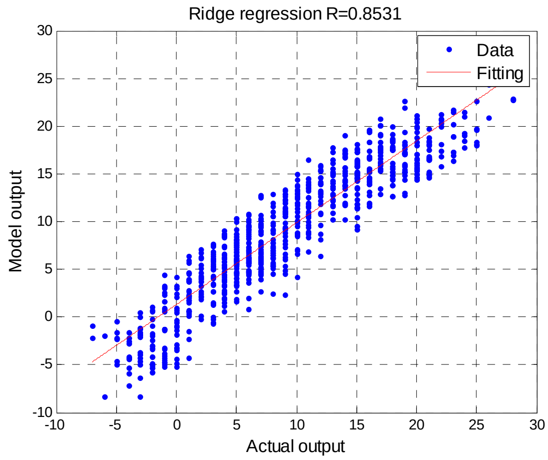

Fig.4 Outputs of ridge regression model for training data

For the training data set, the R (correlation coefficient) of ridge regression model is 0.8531, and the RMSE (root mean square error) is 2.7301. Model testing is performed on the remaining 40% testing data, and the fitting diagram of model validation is shown in Fig.5.

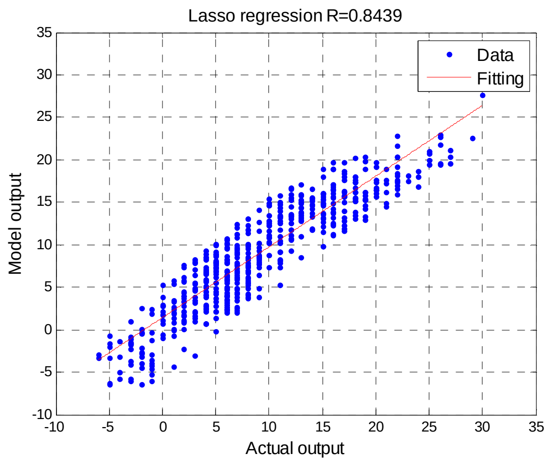

Fig.5 Outputs of ridge regression model for testing data

For the testing data set, the R of ridge regression model is 0.8439, and the RMSE is 2.8913, which are very close to those in training data set.

From the above results, we can find that ridge regression model achieves good results with high R and less RMSE for the train stopping error regression. The parameters in the model are shown in Table 2.

TABLE II.

VALUE OF PARAMETERS IN RIDGE REGRESSION MODEL

Variable	Coefficient
x_1	6.418904
x_2	-3.95443
x_3	-0.29837
x_4	-6.9465
x_5	1.767661
x_6	0.738984
x_7	-0.03553
x_8	1.350614
x_9	-0.33472
x_{10}	1.521577
x_{11}	-2.83665
x_{12}	0
x_{13}	-0.39322
x_{14}	-0.15784
x_{15}	-0.52106
Constant term	1.41254

As can be seen in Table 2, the correlation coefficients are high for the variables x1, x2, x4. Hence, these three variable are x1(distance from the target point at the time of air braking) which is proportional to the stopping accuracy, x2 (speed at the time of electro-pneumatic switching) and x4 (air braking time) which are inversely proportional to stopping accuracy. We find the variation of the three parameters are [359 521] (cm), [78 127] (cm/s) and [0.8 4] (s), respectively according to the analysis of the training data. Hence, to achieve high stopping accuracy, the variables x1, x2 and x4 should be 359cm, 127cm/s and 4s respectively. When adopting the above values, the stopping accuracy will be better or the stopping error will be less.

B. Lasso Regression

Lasso (Least Absolute Shrinkage and selection operator) regression is a shrinkage estimation method which is firstly proposed by Tibshirani in 1996 [19].

Lasso is a regularization technique for performing linear regression. Lasso includes a penalty term that constrains the size of the estimated coefficients. Lasso is a shrinkage estimator: it generates coefficient estimation which biases to be small. Nevertheless, a lasso estimator can have smaller MSE than an ordinary least squares estimator for new data. Unlike ridge regression, as the penalty term increases, lasso sets more coefficients to zero. This means that the lasso estimator is a simpler model with fewer explanatory variables. Hence, lasso regression is an alternative to stepwise regression and other dimensionality reduction techniques.

For a given value of λ, a nonnegative parameter, lasso regression try to solve this regularization problem.

$$\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - x_i^T\beta) + \lambda\sum_{j=1}^{p}\beta_j\right\} \tag{5}$$

where N is the number of observations, y_i is the response at observation i and x_i is data, a vector of p values at observation i. λ is a positive regularization parameter corresponding to one value of λ. With the increase of λ, the number of nonzero components of β decreases.

The model of lasso regression is adopted to analyze the stopping accuracy data, and the error estimation method of lasso regression model is mainly used K-fold cross-validation. The relationship between λ and MSE (mean square error) after cross-validation of lasso regression is shown in Fig.6.

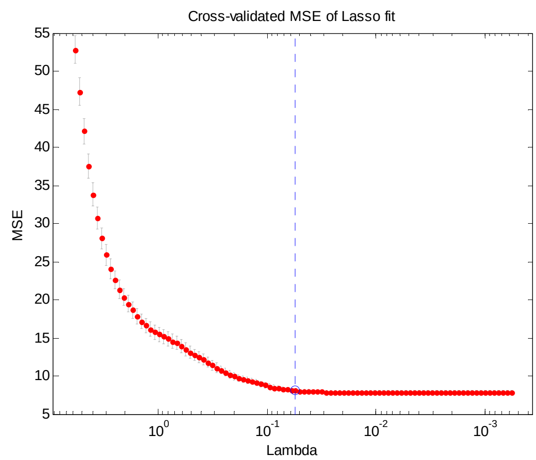

Fig.6 MSE of lasso regression

We use the MSE to evaluate the lasso regression model. In Fig.6, the MSE increase quickly with the increase of λ and the nonzero parameters are too few to fit the model. When the value of λ is very small, the

MSE is almost constant with the decrease of λ. Hence, overfitting is possible with very small λ. For this stopping accuracy problem, we select $\lambda = 0.0586$ by considering the trade-off of MSE and the number of nonzero parameters. Then, the lasso regression for training data set is shown in Fig.7.

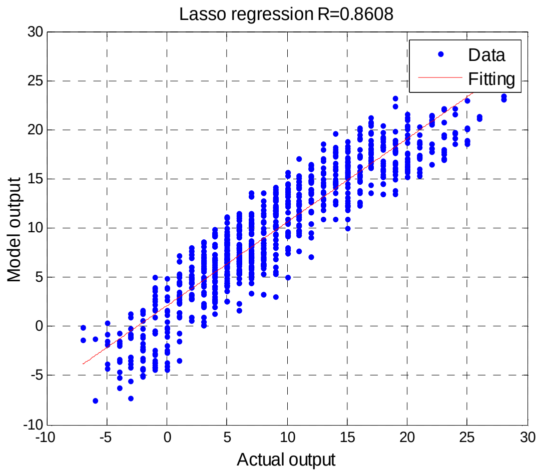

Fig.7 Outputs of lasso regression model for training data

Finally, we find the optimal lasso model with the following performance indices: the R of lasso regression is 0.8608 and the RMSE is 2.7123. Model testing is performed on the remaining 40% testing data, and the fitting diagram of model validation is shown in Fig.8.

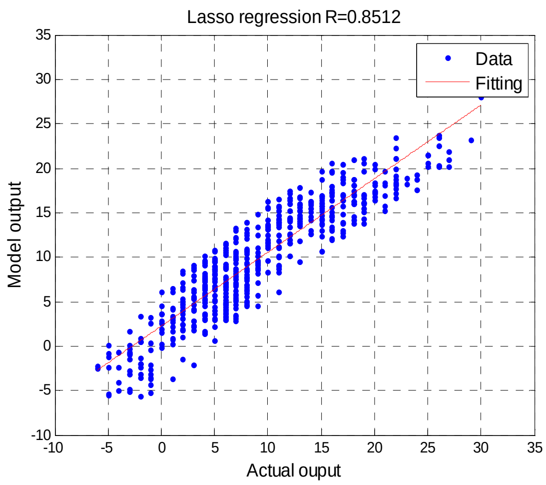

Fig.8 Outputs of lasso regression model for testing data

For the testing data set, the R of lasso regression model is 0.8512, and the RMSE is 2.8779. The parameters of the lasso regression model are shown in Table 3.

TABLE III.
VALUE OF PARAMETERS IN LASSO REGRESSION MODEL

Variable	Coefficient
x_1	0.18814
x_2	-0.44023
x_3	0
x_4	-9.1019
x_5	34.6974
x_6	0.16401
x_7	0
x_8	0.40185
x_9	-0.0851
x_{10}	0.11510
x_{11}	-0.4563
x_{12}	0
x_{13}	-0.01246
x_{14}	0
x_{15}	-0.41085
Constant term	2.37614

In the obtained optimal model by lasso regression, the number of model parameter is 11. The variables x3 (distance from the target point at the time of electro-pneumatic switching), x7 (air braking rate), x12 (weight of train) and x14 (variance of network voltage in braking stage) have nothing to do with train stopping accuracy. The stopping accuracy is highly correlated with variables x4, x5. The variable x4 (air braking time) is inversely proportional to the stopping accuracy, and variable x5 (mean acceleration of controller's output in air braking) is proportional to the stopping accuracy. The variation of the these two parameters are [0.8 4](s) and [0.0959 0.3491](cm/s^2), that is to say, when the variable x4 is closed to 4s and variable x5 is closed to 0.0959cm/s^2, the stopping accuracy will be better.

C. Elastic Net Regression

Elastic net regression [20,21] is a related technique which proposed by Hui Zou. Elastic net is a hybrid of ridge regression and lasso regression. Like lasso regression, elastic net regression can generate reduced models by generating zero-valued coefficients. Empirical studies have suggested that the elastic net regression is better than lasso regression on data with highly correlated predictors.

For an α strictly between 0 and 1, and a nonnegative λ, elastic net regression solves the problem

$$\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - x_i^T\beta) + \lambda P_\alpha(\beta)\right\} \tag{6}$$

Here,

$$P_\alpha(\beta) = \frac{1-\alpha}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1 = \sum_{j=1}^{p}\left(\frac{1-\alpha}{2}\beta_j^2 + \alpha|\beta_j|\right) \tag{7}$$

When $\alpha = 1$, elastic net regression is the same as lasso regression. As α shrinks toward 0, elastic net regression approaches to ridge regression. For other values of α, the penalty term $P_\alpha(\beta)$ interpolates between the L^1 norm of β and the squared L^2 norm of β. Hence, the elastic net regression is similar to ridge regression. The relationship between λ and MSE after cross-validation for elastic net regression is shown in Fig.9.

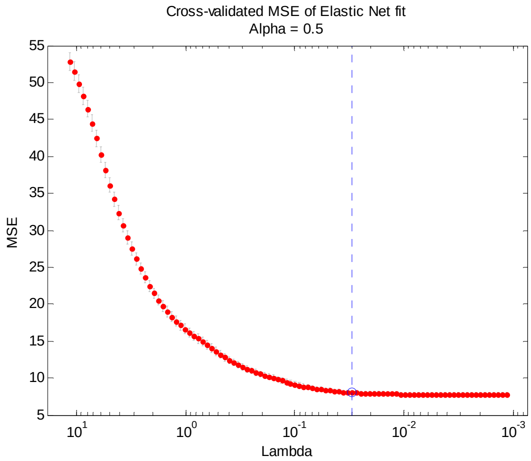

Fig.9 MSE of elastic net regression

We set $\alpha = 0.5$ and also use the MSE to evaluate the elastic net regression model. There are different results in elastic net regression and lasso regression according to the coefficient changing trend. It reflects a qualitative analysis of the variables. When the value of λ increases, there are multiple variables tend to be zero. Furthermore, it reflects the generality of parameters variation in elastic net regression. Along with the variation of λ, elastic net regression often remains or removes the coefficient in group which is highly correlated with. For this stopping accuracy problem, we select $\lambda = 0.0219$ to achieve the trade-off of accuracy and complexity, as shown in Fig.9. Then, the results of elastic net regression for training data set are shown in Fig.10.

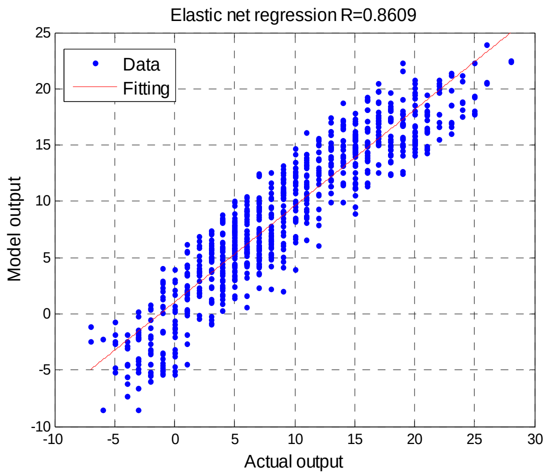

Fig.10 Outputs of elastic net regression model for training data

We find the optimal elastic net model with the following performance indices: the R of elastic net regression is 0.8609 and the RMSE is 2.6842. Model testing is performed on the remaining 40% testing data, and the fitting diagram of the model validation is shown in Fig.11.

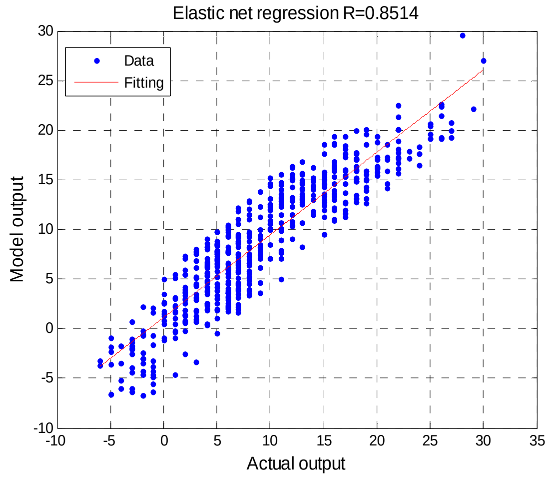

Fig.11 Outputs of elastic net regression model for testing data

For the testing data set, the R of elastic net regression model is 0.8514, and the RMSE is 2.8443. The parameters of the elastic net model are shown in Table 4.

TABLE IV.
VALUE OF PARAMETERS IN ELASTIC NET REGRESSION MODEL

Variable	Coefficient
x_1	0.195724
x_2	-0.45043
x_3	0
x_4	-9.11686
x_5	34.7798
x_6	0.16491
x_7	0
x_8	0.40448
x_9	-0.09522
x_{10}	0.11628
x_{11}	-0.45885
x_{12}	0
x_{13}	-0.0127
x_{14}	0
x_{15}	-0.41677
Constant term	1.24589

The relationship between parameters and stopping accuracy is mentioned above, and the number of model parameter is 11. As same as lasso regression, also variables x3, x7, x12 and x14 have nothing to do with train stopping accuracy, and the stopping accuracy is highly correlated with variables x4, x5.

IV. COMPARISONS OF METHODS

Ridge regression, lasso regression and elastic net regression are employed to analyze the stopping accuracy in the previous sections. We utilize, number of parameters, R, RMSE and AIC (Akaike information criterion) [22] one model selection method, to evaluate the performance of each regression method. All comparisons for training data and testing data are listed in Table 5 and Table 6 respectively.

TABLE V.
THE TRAINING DATA RESULTS OF METHODS

Method	Number of parameters	R	RMSE	AIC
Ridge	14	0.8531	2.7301	2.8660
Lasso	11	0.8608	2.7123	2.7857
Elastic net	11	0.8609	2.6842	2.7412

TABLE VI.
THE TESTING DATA RESULTS OF METHODS

Method	Number of parameters	R	RMSE	AIC
Ridge	14	0.8439	2.8913	2.9583
Lasso	11	0.8512	2.8779	2.9261
Elastic net	11	0.8514	2.8443	2.8872

We get the results from Table 5 and Table 6, the number of model parameters for ridge regression, lasso regression and elastic net regression are 14, 11 and 11 respectively. The number of model parameter for lasso regression and elastic net regression are less, and it can explain the model better. The parameters of lasso regression and elastic net regression are very similar, but the RMSE and AIC of elastic net regression are better than lasso regression, and the RMSE and AIC of elastic net regression are also the least in the three models. As mentioned above, the elastic net regression model makes the best effect. We analyze the elastic net regression model, and obtain the factors most related to the stopping accuracy which are variables x4 and x5. Considering various factors, the coefficient values which make the stopping accuracy tend to the best are shown in Table 7.

TABLE VII.
THE VALUES OF MAIN FACTORS

Variable	Value
x_4	2.048s
x_5	0.184 cm/s^2

V. CONCLUSION AND PROSPECT

This paper firstly explains the significance of stopping accuracy and introduces some of the previous research results. Previous research only studied a few of factors and field-data about stopping accuracy is limited. By analysis, we found there are fifteen factors which could affect the stopping accuracy. Based on the large amount of field-data, we put forward a new idea to study stopping accuracy using data mining methods.

The ridge regression, lasso regression and elastic net regression are employed to analyze the stopping accuracy respectively and the results of various methods are compared. By using AIC, we found that the elastic net regression model is the best one for this problem. And two most influential factors are found from 15 ones, which are helpful in increasing stopping accuracy for ATO systems.

Nevertheless, there are some future works to be worth exploring, such as choosing better methods of data mining, giving specific suggestions for improving stopping accuracy, and testing the research results in the field.

ACKNOWLEDGMENT

This work is partially supported by Beijing Laboratory of Urban Rail Transit, Beijing Higher Institution Engineering Research Center of Urban Rail Transit CBTC System, the National High Technology Research and Development Program (“863” Program) of China under grant 2012AA112800, by New Scientific Star Program of Beijing under grant 2010B015, by the Fundamental Research Funds for the Central Universities under grant 2012JBM016, by the independent research project from the State Key Laboratory of Rail Traffic Control and Safety under grant RSC2011ZT001.

REFERENCES

[1] M.Janic, “Multicriteria evaluation of high-speed rail, transrapid maglev and air passenger transport in Europe,” Transportation Planning and Technology, vol.26, no.6, 2003.

[2] “Large span of 5-years, the Beijing subway comes into a golden period of development,” Municipal Engineering Technology, vol.30, no.5, 2012.

[3] H.P. Yu, “Analysis on the Stop Inaccuracy of Urban Rail Train,” Mass Transit, vol.11, no.9, pp. 28–30, 2008.

[4] X.Y. Li, “Analysis of Inaccurate Train Locating Stopping in ATO Mode,” Railway Signalling & Communication Engineering, vol.7, no.4, pp.68-69, 2010.

[5] R. Banach, H. Zhu, W. Su, “Continuous KAOS, ASM, and formal control system design across the continuous /discrete modeling interface: a simple train stopping application,” Formal Aspects of Computing, pp.1-48, 2012.

[6] D.W. Chen, T. Tang, C.H. Gao, R.Q. Mu, “Research on the error estimation models and online learning algorithms for train station parking in urban rail transit,” Chin. Railway Sci., vol.31, no.6, pp.122-127, 2010.

[7] D.W. Chen, C.H. Gao, “Soft computing methods applied to train station parking in urban rail transit,” Applied Soft Computing, vol.12, no.2, pp.759-767, 2012.

[8] D.W. Chen, R. Chen, Y.D. Li, T. Tang, “Online Learning Algorithms for Train Automatic Stop Control Using Precise Location Data of Balises,” IEEE Transactions on Intelligent Transportation System, vol.13, no.6, pp. 1-10, 2013.

[9] X.J. Jiang, Y.Y. Xing, “Discussion on Relation between Installation Location of Screen Door and ATO Station Stop,” Railway Signalling & Communicati.on Engineering, vol.4, no.6, pp.38-40, 2007.

[10] J. Han, M. Kamber, J. Pei, “Data mining: concepts and techniques,” Morgan kaufmann, 2006.

[11] I.H. Witten, E. Frank, “Data Mining: Practical machine learning tools and techniques,” Morgan Kaufmann, 2005.

[12] Y.Q. Ding, D. Peng, “Quantitative analysis of near-infrared spectra by wavelet-based interferences removal and least

squares support vector regression," Journal of Computers, vol.7, no.4, pp.880-889, 2012.
[13] Q. Zhu, Y.Y. Zhang, "Construction of customer classification model based on Bayesian network," Journal of Computers, vol.8, no.5, pp.1200-1206, 2013.
[14] J.C. Xu, Y.P. Gao, S.Q. Li, L. Sun, "A greedy correlation measure based attribute clustering algorithm for gene selection," Journal of Computers, vol.8, no.4, pp.951-959, 2013.
[15] Y.R. Liu, M.L. Hou, Y.L. Fu, "Uncertainty time series' multi-scale fractional order association model," Journal of Computers, vol.7, no.11, pp.2617-2622, 2012.
[16] A.E. Hoerl, R.W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol.12, no.1, pp.55-67, 1970.
[17] A.E. Hoerl, R.W. Kennard, "Ridge regression: applications to nonorthogonal problem," Technometrics, vol.12, no.1, pp.69-82, 1970.
[18] A.E. Hoerl, R.W. Kannard, K.F. Baldwin, "Ridge regression: some simulations," Communications in Statistics-Theory and Methods, vol.4, no.2, pp.105-123, 1975.
[19] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), vol.58, no.1, pp.267-288, 1996.
[20] H. Zou, T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol.67, no.2, pp.301-320, 2005.
[21] H. Zou, H.H. Zhang, "On the adaptive elastic-net with a diverging number of parameters," Annals of statistics, vol.37, no.4, pp.1733-1751, 2009.
[22] K. Yamaoka, T. Nakagawa, T. Uno, "Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations," Journal of pharmacokinetics and biopharmaceutics, vol.6, no.2, pp.165-175, 1978.

Lin Ma was born in 1978, received the M.S. degree from Electronics and Information Engineering, Beijing Jiaotong University, China, in 2005. Now, she is a Senior Engineer and Associate Director with National Engineering Research Center of Railway Transportation Operation and Control System, Beijing Jiaotong University. Her research interests include carrying capacity of line and design of signal system in urban rail transit, automatic train operation.

Xiangyu Zeng was born in 1989. He received the B.S. in Electronics and Information Engineering, Beijing Jiaotong University, China in 2011. He is working toward the M.S. degree with the Electronics and Information Engineering, Beijing Jiaotong University. His research interests include machine learning, data mining.