Subway Energy Consumption Prediction based on XGBoost Model

Jinbing Ha and Ziyi Zhou *

Department of Information Management, Nanjing University of Science and technology, Nanjing, China

* Corresponding author: 996480118@qq.com

Abstract. In the process of urban rail transit operation and management, accurate prediction of subway energy consumption is beneficial for establishing a reasonable operational organization mode and evaluating energy efficiency. Due to the multitude of factors affecting train energy consumption, traditional mathematical regression methods struggle to guarantee predictive accuracy. Thus, a energy consumption prediction method based on XGBoost is proposed. To enhance model training efficiency and accuracy, the Lasso model is utilized for feature selection of subway energy consumption influencing factors. Additionally, the K-means++ algorithm is employed for clustering subway energy consumption. Using the operational energy consumption data of Qingdao Subway Line 3 as an example for validation, XGBoost algorithm is employed to predict subway energy consumption. The results are then compared with those of the SVR and LSTM algorithms using three evaluation metrics. It is found that the XGBoost algorithm provides predictions of subway energy consumption that are closer to the experimental values.

Keywords: Urban rail transit, consumption prediction, XGBoost.

1. Introduction

As one of the crucial public transportation systems in urban areas, the subway has witnessed rapid development in major cities due to its advantages such as high speed, large capacity, and minimal land occupancy. At the current stage, urban subway networks are expanding rapidly, leading to an increase in energy consumption. Therefore, research on subway energy consumption prediction has garnered significant attention. Accurate and rapid prediction of subway energy consumption is beneficial for setting reasonable line conditions and devising transportation organization models.

Numerous scholars have conducted extensive research on energy consumption prediction using various methods to propose better energy-saving strategies. Hickman et al. [1] proposed a formula for estimating subway energy consumption based on the analysis of average speed and station spacing. Lukaszewicz P [2] selected traction force, speed, acceleration, and train physical parameters as inputs to establish an energy consumption estimation model based on train driver's operational behavior. This model was applied to different types of trains and tracks. Philip C.M. Leung et al. [3] used a Multi-Layer Perceptron neural network (MLP) to establish a model for predicting the average weekly power consumption of stations in urban rail transit. This provided a reference for building a subway energy consumption prediction model. In order to address the non-linear relationship issues arising from multiple factors affecting subway energy consumption, L(U) [4] employed two machine learning methods: Support Vector Regression (SVR) and Random Forest Regression (RFR) to predict subway energy consumption.

The aforementioned literature mainly focuses on subway energy consumption prediction through the establishment of energy consumption analytical models, or the analysis of historical data to build regression models. However, the subway energy consumption process is complex, and there are numerous factors influencing energy consumption. Using the aforementioned methods may lead to less precise predictions or overly limited perspectives. In order to fully consider the factors affecting train traction energy consumption and predict subway energy consumption more accurately, this paper employs the XGBoost algorithm for subway energy consumption prediction. This method can effectively address issues related to high dimensionality and non-linearity, making it widely

applicable to datasets, demonstrating good performance, and minimizing the risk of overfitting, ultimately resulting in high accuracy.

2. XGBoost Model

The XGBoost algorithm belongs to the Boosting branch of ensemble learning. Its goal in each iteration is to reduce the residual error from the previous iteration. This process aims to establish a new tree model in the direction of the negative gradient. In other words, the training and prediction results of the previous decision tree will influence the sample input for the next tree model. The specific steps of the XGBoost algorithm are as follows[5]:

Suppose we have a training set (X, y), where X is the input feature matrix, and y is the target variable. The objective is to learn a powerful predictive model F(x) such that the predicted value F(xi) is as close to the true value yi as possible. The model can be described as:

$$F(x) = \sum_{m=1}^{M} f_m(x) \tag{1}$$

Where F(x) is the final predictive model, $f_m(x)$ is the prediction output of the m-th base learner (which is a decision tree), and M is the number of base learners.

Each base learner $f_m(x)$ is a decision tree, and its output depends on the input features x and the tree's parameters. To train the LightGBM model, the following objective function is minimized:

$$Obj(w, f) = \sum_{i=1}^{n} L(y_i, F(x_i)) + \sum_{m=1}^{M} \Omega(f_m) \tag{2}$$

Where $L(y_i, F(x_i))$ is the loss function, used to measure the error between the predicted value and the true value y_i. $\Omega(f_m)$ is the regularization term, used to control the complexity of the model and prevent overfitting.

3. Subway Energy Consumption Prediction

3.1. Data Preprocessing

First, normalization is applied to the data using min-max scaling. Then, the Lasso algorithm is used for feature selection of the energy consumption of rail transit trains. The dataset includes 14 independent variables: train weight, traction force, braking force, maximum speed, average speed, interval time, number of trains, transfer stations, starting stations, normal station A, normal station B, average stop time, average passenger flow, and peak hour maximum passenger flow. The variable selection path based on the Lasso energy consumption evaluation index model is shown in Fig.1.

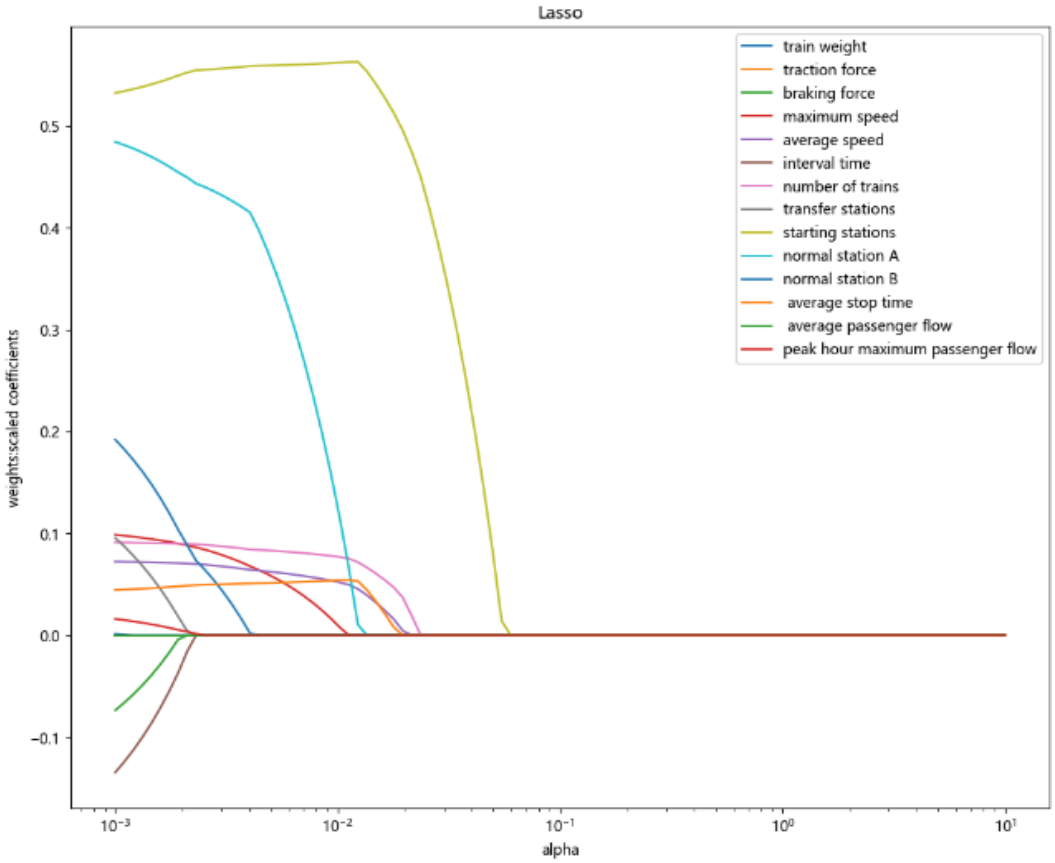

Figure 1. Lasso Algorithm Variable Selection Chart

The horizontal axis can be seen as the degree of compression coefficient. When the compression coefficient alpha is 0.002, the variables selected by the Lasso algorithm are: starting stations,maximum speed, average speed, interval time, number of trains, transfer stations, normal station A, normal station B, average stop time, and peak hour maximum passenger flow. The Lasso algorithm reduces the independent variable set from 14 to 10.

3.2. Energy Consumption Pattern Clustering

The K-Means++ algorithm is used to cluster the preprocessed energy consumption of rail transit trains. The clustering number is determined based on the SSE and SC under different clustering numbers. According to Fig.2 and Fig.3, the optimal clustering number is finally determined to be 5.

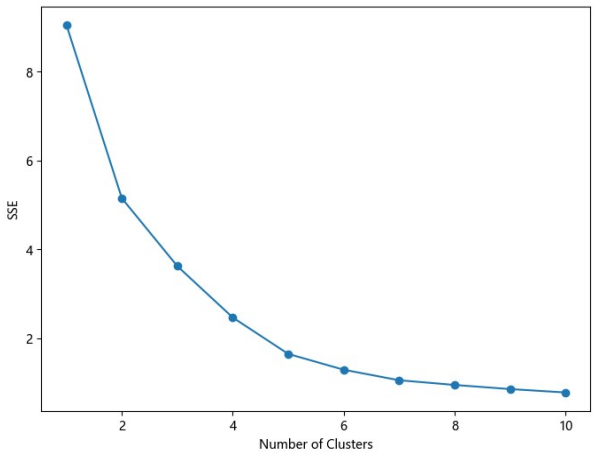

Figure 2. SSE under Different Clustering Numbers

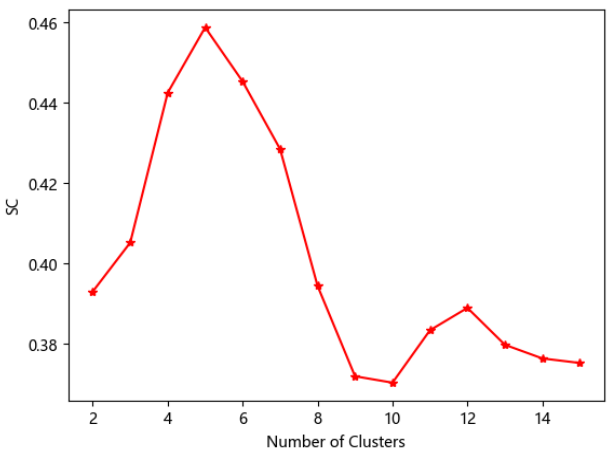

Figure 3. SC under Different Clustering Numbers

3.3. Subway Energy Consumption Prediction

In this experiment, the XGBoost algorithm is used to predict subway energy consumption. The results are compared with those obtained from the SVR and LSTM algorithms using evaluation metrics such as MSE, RMSE, and MAE.

Since both the LSTM and XGBoost algorithms require hyperparameter tuning, this experiment uses grid search to determine the parameters. In this study, the Keras deep learning framework is used to build the LSTM model, and the LSTM parameters are shown in Table 1.

Table 1. LSTM Algorithm Parameters for Each Energy Consumption Pattern

Category	Epoch	Batch Size	Layers	Optimization Algorithm	Activation Function	Learning Rate
1	300	10	5	Adam	Relu	0.1
2	200	5	4	Adam	Relu	0.05
3	200	5	5	Adam	Relu	0.05
4	150	5	3	Adam	Relu	0.1
5	400	15	7	Adam	Relu	0.1

The XGBoost algorithm parameters are shown in Table 2.

Table 2. XGBoost Algorithm Parameters for Each Energy Consumption Pattern

Category	Number of Weak Learners	Gamma Learning Rate	Max Depth	Min Child Weigh
1	40	0.3	0.1	5
2	50	0.5	0.1	4
3	40	0.3	0.1	4
4	40	0.1	0.1	2
5	50	0.5	0.1	6

Using cross-validation, the performance of the three algorithms in terms of MAE, MSE, and RMSE is shown in Table 3.

Table 3. Evaluation Results for Each Algorithm

Category	Metric	SVR	LSTM	XGBoost
1	MSE	0.05	0.26	0.05
	MAE	0.23	0.46	0.18
	RMSE	0.23	0.51	0.22
2	MSE	0.5	0.32	0.04
	MAE	0.23	0.51	0.16
	RMSE	0.23	0.56	0.20
3	MSE	0.12	0.18	0.10
	MAE	0.34	0.35	0.28
	RMSE	0.34	0.43	0.32
4	MSE	0.15	0.32	0.11
	MAE	0.39	0.49	0.28
	RMSE	0.39	0.56	0.33
5	MSE	0.06	0.21	0.06
	MAE	0.25	0.20	0.20
	RMSE	0.25	0.24	0.24

From Table 3, it can be observed that for the five categories of rail transit train energy consumption patterns, the XGBoost algorithm provides the best predictive results. This further demonstrates that the XGBoost algorithm is capable of more accurate predictions of subway energy consumption.

4. Summary

(1) The XGBoost algorithm demonstrates highly accurate prediction results in subway energy consumption forecasting, with MSE values for all five categories around 0.09. This high level of prediction accuracy affirms the feasibility of using the XGBoost algorithm for subway energy consumption prediction.

(2) The predictions generated by the XGBoost algorithm are closer to the experimental values. Through a comparison of the three-evaluation metrics, it is evident that the XGBoost algorithm provides more accurate predictions of subway energy consumption compared to the LSTM and SVR algorithms.

References

[1] Hickman J, Hassel D, Joumard R, Samaras Z, Sorenson S. Methodology for calculating transport emissions and energy consumption. Office for Official Publications of the European Communities, Luxembourg, Belgium. 1999.

[2] Lukaszewicz. P. Energy consumption and running time for tains: Modelling of running resistance and driver behavior based on full scale testing. Stockholm: Royal Institute of Technology, 2001.

[3] Leung Philip C. M., Lee Eric W. M. Estimation of electrical power consumption in subway station design by intelligent approach. Applied Energy, 2013, 101: 634 - 643.

[4] L (U) Huanhuan; ZHANG Yuzhao. Research on the prediction of traction energy-consumption of subway train based on machine learning. Journal of Railway Science and Engineering, 2019, 16 (7): 1833 - 1841.

[5] LIU Jiang; XU Kangzhi; CAI Baigen; GUO Zhongbin; WANG Jian. XGBoost-based fault prediction method for on-board train control equipment. Journal of Beijing Jiaotong University, 2021, 45 (4): 95 - 106.