Comparative Study of Short-term Forecasting Models of Online Car-hailing Demand

Shiyu Chang¹, Zhili Liu¹, Dongyu Kuang¹ and Hui Zhao²

¹Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing, China
²Beijing Transportation Institute
¹Email: cambiosyang@163.com

Abstract. Based on the time sequence of the online car-hailing data and the change trend characteristics of the online car-hailing data, the short-term prediction of the demand for online car-hailing is carried out. Established the short-term prediction model of the online car-hailing demand, which is based on ARIMA, single feature LSTM and multi feature LSTM respectively. The optimal parameters and structure of the model are determined by using the actual demand data of drip-drip taxi platform, and the comparative experiments are carried out. The experimental results show that the multi feature LSTM model is the best.

1. Introduction

Compared with the traditional touring taxi, the online car-hailing market improves the long-term information asymmetry in the traditional taxi market, and improves the supply-demand relationship. The new operation characteristics of online car-hailing also provide a new idea for improving the balance of transportation travel supply and demand. Therefore, it is of great significance to use the massive travel data of online car-hailing to study the matching of supply and demand in the new situation. The research on short-term traffic flow has achieved rich results, and the research direction is focused on road traffic flow prediction. Wang Jun and Guan Wei (2006) set up a short-term prediction model based on Kalman filter, verified the model with real data, and expounded the applicability of the short-term traffic flow model based on Kalman filter. Yang Gaofei et al. (2012) combine ARMA and Calman filter to build a combined model based on the characteristics of short-term traffic flow. The prediction results show that the combined model is better than the single model. Zheng Xuanxuan et al. (2012) conducted comparative experiments on different models, and found that the radial basis function neural network model is more suitable for short-term traffic flow prediction. With the informatization construction of taxi industry, there is still a lot of research space to apply the short-term traffic flow prediction method to the demand prediction of online car-hailing.

2. Building A Short-term Forecast Model of Demand for Online Car-hailing Based on ARIMA

Based on ARIMA model, the modeling process of the demand short-term prediction model for online car-hailing is divided into three parts: sequence stabilization, model identification and model inspection. The modeling process is shown in Figure 1. In modeling, it is necessary to judge whether the demand data of online car-hailing is stable data, and test the smoothness and white noise of the sequence data; then use autocorrelation function and partial autocorrelation function to determine the lag value p and moving window value Q of ARIMA model, and estimate the model parameters to get the final ARIMA model. Finally, we use the model to predict, test and diagnose the results of the
model, and judge whether ARIMA model is suitable for the prediction of demand data of online car-hailing.

3. Construction of Short-term Forecast Model of Network Car Demand Based on LSTM

The use of sections to divide the text of the paper is optional and left as a decision for the author. Where the author wishes to divide the paper into sections the formatting shown in table 2 should be used.

3.1. Single Feature LSTM Model

As shown in Figure 2, the overall process of the short-term prediction model of car hailing demand of single feature LSTM network is mainly divided into three parts. The sliding time window splits the data set, trains and tests the single feature LSTM model, and evaluates the results, which are in turn progressive relationship. In this paper, each piece of data is generated in a specified time step to complete the construction of feature engineering, and all the data of training single feature LSTM are also obtained. The data of online car-hailing demand is divided into training set and test set. The training set is used to determine the parameters of short-term prediction model of online car-hailing demand with single feature LSTM, and the test set is used to verify, evaluate and compare the short-term prediction model of online car-hailing demand with single feature LSTM.

![Figure 1. ARIMA model modeling process](image-url)
3.2. Multi Feature LSTM Model

The overall process of the short-term prediction model of single feature LSTM online car-hailing demand is shown in Figure 3. Firstly, the influencing factors of online car-hailing demand are extracted, the hidden features are extracted, and the input vector of multi feature LSTM prediction model is obtained. Then the input vector is normalized. The data set is divided into training set and test set. The training set is used to train the model, and the optimal parameters of multi-feature LSTM model are obtained. Finally, the multi-feature LSTM model is used to predict the demand data of online car-hailing, and the prediction results are evaluated and analyzed.

In the aspect of feature extraction, all variable features are counted, among which weather and date variables are classified variables, and the rest variables are continuous variables. The dummy variable processing is used to classify the variables, that is, m categories of the variables are transformed into m-1 variables, and the variable value is 0 or 1. The input features of multi feature LSTM model include weather features, road congestion features, similar time slice features and date features, and the feature extraction of short-term prediction model of network car demand based on multi feature LSTM is completed. In this paper, Adam (adaptive motion estimation) algorithm is used as the optimization function of loss function. In the Adam optimization algorithm used in this paper, the parameter values $\mu=0.9$, $\nu=0.999$, $\eta=10^{-8}$.

4. Experimental Results and Analysis

4.1. Analysis of Prediction Results

From figures 4, 5 and 6, it can be seen that the predicted values of the three models basically match the changes in the demand for cars in the actual value network. In the off-peak period, the prediction error is relatively stable, and the error in the peak period is significantly greater than that in the off-peak period, and the error in the late peak period is the largest. The multi feature LSTM model has the best prediction result and the least prediction error.
Figure 4. ARIMA model prediction results and error variation chart

Figure 5. Prediction results and error variation of single feature LSTM model

Figure 6. Prediction results and error variation of multi feature LSTM model
4.2. Comparative Experimental Analysis

Figure 7 shows that the single feature LSTM model and ARIMA model have similar prediction accuracy in the late peak period, while the multi feature LSTM model is better, which shows that in the late peak period, there are more factors affecting the demand of online car-hailing, only relying on the change trend of online car-hailing can not predict the demand of online car-hailing in the peak period very well. It is necessary to add feature variables to accurately predict the demand of online car-hailing in the peak period.

In this paper, RMS error (RMSE), mean absolute error (MAE) and determination coefficient (R-square) are selected as the criteria to evaluate the short-term prediction experiment of the demand for car Hailing. Table 1 below is the evaluation coefficient table of the three models.

|                | Multi feature LSTM | Single feature LSTM | ARIMA  |
|----------------|--------------------|---------------------|--------|
| RMSE           | 93.544             | 95.738              | 98.512 |
| MAE            | 56.221             | 61.017              | 68.543 |
| R-square       | 0.932              | 0.923               | 0.921  |

Observe the determination coefficient R-square from the table, the closer it is to 1, the higher the model fitting degree is, so the multi feature LSTM model has the highest fitting degree. Comparing RMSE and MSE, the prediction error of multi feature LSTM model is the smallest, and that of single feature LSTM model is lower than that of ARIMA model.

To sum up, the short-term prediction model based on multi-feature LSTM has the best prediction result, which can better fit the trend of the demand for online car-hailing, and the prediction error is the smallest.

5. Conclusion

In this paper, through the processing of the original data to get the data of the demand for online car-hailing, using the data to make short-term prediction of the demand for online car-hailing in a certain region. Based on the time sequence and trend characteristics of online car-hailing data, a short-term prediction model of online car-hailing demand based on ARIMA, single feature LSTM and multi feature LSTM model is established. The correlation analysis of online car-hailing demand data is carried out, the features are extracted, and the model structure is determined. The parameters of the model are determined by the actual demand data of the online car-hailing. It is verified that in the
short-term prediction of the online car-hailing demand, the multi-feature LSTM model has the best prediction result and the least prediction error. However, this paper only forecasts the short-term demand of online car-hailing in the region with the largest demand, and does not make a deep research on the change of online car-hailing demand between different regions. According to the characteristics of different regions, different models can be established to predict the demand of regional network for car Hailing.

6. References
[1] Wang Jun and Wang Wei 2006 *J. Computer and Communications*. 05 16-19
[2] Yang Gaofei, Xu Rui, Qin Ming, Zheng Kaili, Zhang Bin 2017 *J. Journal of Zhengzhou University(Engineering Science)*. 38(02) 36-40
[3] Zheng Xuanchuan, Han Baoming, Li Dewei 2012 *J. Shandong Sci.* 25(03) 23-28
[4] Xu Na and Qian Chao 2015 *J. Technology of Highway and Transport*. 04 141-145
[5] Hu Jihua, Xie Haiying 2011 *J. Communications Standardization*. 18 50-53
[6] Liu Lina, Chen Yanyan, Zhang Wenge 2010 *J. Communications Standardization*. 13 89-92
[7] Wang Guanghong and Jiang Ping 2004 *J. Journal of Tongji University*. 02 246-252
[8] Lu Zenan, Shang Yulin 2017 *J. Science & Technology Vision*. 13 116-117
[9] Zhang Honghong 2017 *J. Journal of Guizhou University of Finance and Economics*. 13(05) 312-315
[10] Wang Shuqi 2018 *J. Journal of Guizhou University of Finance and Economics*. 14(12) 326-328