Analysis of High-Speed Railway Passenger’s Travel Choice Behavior Based on Deep Learning Model

Shujie Wang and Zhenhuan He*
School of Traffic and Transportation, Beijing Jiao tong University, Beijing, 100044, China

*Corresponding author email: zhhe@bjtu.edu.cn

Abstract: The analysis of railway passengers’ travel choice behavior plays an important role in railway passenger product design and transport organization. Many researchers focused on it and lots of models and methods had been proposed. With the rapid development of information technology in recent years, deep learning models are more and more widely used in many research fields, and also applied in this research. Based on the OD train ticket data of Beijing-Shanghai high-speed railway line, it studied travel choice behavior of passengers in Beijing-Shanghai high-speed railway line during the weekday, and then took attributes of ticket data (arrive-departure station, arrival-departure time, etc.) as the mapping features to study passengers' choice of different trains, The result shows that the method given by the research has quite high fitting accuracy.

Keywords: Travel choice behavior analysis; Deep learning; High-speed railway ticket data.

1. Introduction
The analysis of railway passengers’ travel choice behavior can be helpful to achieve the travel choice preference of passengers to different factors, such as departure time, travel time, price, etc., which are essential for railway passenger product design and transport organization. There are mainly two types of models in related researches: disaggregate models based on survey (RP/SP) data and aggregate models based on large-scale data represented by passenger ticket data. Wang et al. constructed a MNL model for passenger’s travel plan using Beijing-Shanghai high-speed railway passenger travel RP survey data in 2013. The model assumes that the travelers are related to some explanatory variable, and achieves them based on a series of user travel choice models represented by Logit model. In fact, the travel choice behavior of passengers is very complex, and it involves a large number of influencing factors, many of which are difficult to clearly reflect the functional relationship in the model, which causes a certain impact on the accuracy and applicability of the model. With the development of big data processing technology, various machine learning models have been continuously developed and applied, which provides new ideas for solving the above problems. Machine learning models, which do not depend on the relationship between hypotheses or any prior knowledge of the potential relationship, and have excellent ability of iterative recognition patterns and extract rules from data, are increasingly applied to the analysis of passenger choice characteristics. They don’t relate travelers to any influencing factors, and takes various factors that influence travelers' travel choices as mapping features of actual travel choices, which has a wider applicability. Sun et al. predicted passengers' choice of different classes of trains by support vector machine (SVM) and artificial neural network (ANN) of machine learning based on online booking data, and proved that machine learning methods have better prediction results than discrete selection model through practical examples. Xie et
al. proposed a recurrent neural network based on the long short-term memory neural unit to predict the taxi travel behavior by using the GPS data of Beijing taxi, and verified that the prediction accuracy of this method is higher than that of the traditional travel prediction method.

After years of implementation of the railway ticket sales system, a large amount of railway ticket data has been accumulated which has the characteristics of large amount of data, comprehensive time and space coverage, and provides a solid foundation for the study of user choice behavior. In this study, the ticket data of Beijing-Shanghai Hongqiao high-speed railway during the usual period are used to study passengers' preference for different types of trains by taking attributes of ticket data as the mapping features of deep learning model.

2. Analysis of Influencing Factors of High-speed Railway Passenger Choice

The influencing factors of passenger travel choice can be mainly divided into two aspects: passenger socioeconomic characteristics and train characteristics. Shi et al. divided the factors into passenger main factors, train characteristics and random factors, and studied railway passengers' choice behavior. In general, it generally includes: arrival and departure station, arrival and departure station’s GDP (Trillion yuan/year), arrival and departure time, ticket fare, travel time, seat type.

In addition, train capacity constraints and railway ticket sales strategies can also have a certain impact on passengers' actual choice behavior. A lot of information can be reflected from ticket data, including:

1. OD mileage, total mileage of the train and distance coefficient (Dis_Coef). The distance coefficient refers to the ratio of travel distance to the total mileage of train. These factors are related to railway ticket sales strategies. In order to ensure the consistent of traffic flow, the short-distance tickets (with a lower distance coefficient) in long-distance trains are restricted in the process of ticket sales to ensure that the actual train passenger flow is consistent with the target passenger flow.

2. Seat-capacity. It related to the ticketing sale strategy, the larger the train capacity, the greater the possibility of selecting passengers.

3. Load factor. It reflects the train tension and train capacity, the higher the load factor of the train, the more strain the train capacity, the less freedom passengers can choose.

4. Origin station coefficient. The origin station has more tickets, tickets are easier to buy, and passengers also have more ample boarding time and better train environment.

5. Service frequency. It is the number of trains that can be served for the same OD in a day, and the higher the service frequency, the higher the degree of selective freedom of passengers.

6. OD total passenger. This reflects the degree of competition of passengers for tickets. The larger the total number of passengers, the more intense the competition, and the smaller the degree of freedom for passengers to choose.

3. Passenger Selected Behavioral Analysis Based on Deep Learning Model

3.1. Model Overview

Deep learning is a new method to learn representations from data and it emphasizes learning from continuous layers, which correspond to increasingly meaningful representations. The foundation of the deep learning model is the perceptron (neuron). The perceptron receives input signals from n other perceptrons x_n. These input signals are transmitted through weighted connections w_n. The input value received by the perceptron is compared with the bias b_n of the perceptron, and then via ‘activation function’ (σ(z)) output the perceptron result. It generally has three or more layers of neural networks, including one layer of input, one layer of output and several intermediate layers. The output of each layer of the neural network is:

\[ \sigma(w_n \cdot x_n + b_n) \]

Currently, deep learning network model can be mainly divided into dense feedforward network model, convolution network model and recurrent network model.

Since the data in this study do not take the time into account and there was no sequence relationship between the data, the feedforward network model is adopted. The feedforward network model of deep learning can be mapped to the target by a series of input features through a series of simple data transformations (layers).
The deep learning feedforward network adopted in this study is fully connected with the upper and lower layers, and the generated signals are transmitted forward while the feedback error signals are transmitted backward. The output value of the loss function is calculated as the feedback signal, and the weight and bias are fine-tuned to reduce the loss value, once the minimum loss is achieved, thus obtaining the trained network. This adjustment is done by the optimizer.

3.2. Model Construction

The attributes of Beijing-Shanghai high-speed railway ticket data are used as input features of deep learning fully connected network after data processing. Through a series of layer transformations, they were finally mapped to the number of people who chose different trains, and finally obtain the selection probability of passengers for different trains.

The steps for building a deep learning network model are as follows:

1) Data processing. Numerical and textual data are processed into tensor types suitable for model input. The time of train arrive-departure time is processed as the number of minutes from 0 to 1440, and the distance coefficient is expanded 100 times for the convenience of processing. Text data is processed by one-hot coding. In this study, the arrival and departure station, seat type and whether it is the originating station are processed by one-hot coding. Finally, we standardized each feature, that is, for each feature of the input data (the columns of the input matrix), we subtracted the average value of the feature, and then divided it by the standard deviation, and so that the average value of the feature is 0 and the standard deviation is 1.

2) Loss function

In this study, the loss function is MAE (mean absolute error), and the calculation formula is as follows:

$$\text{MAE} = \frac{\sum |y_i - x_i|}{n}$$  

(1)

Where $y_i$ is the number of predicted selected train passenger number, $x_i$ is the number of actual selected train passenger number, and $n$ is the size of sample data.

3) Model hyper-parameters adjustment

The model hyper-parameters include: the number of network layers, the number of nodes per intermediate layer, and the optimizer.

Deep learning network models include an input layer, an output layer and several intermediate layers. In order to fully capture the network layers most suitable for model data, relatively few layers (3 layers) is selected. In this study, the data has a 64-dimensional data format after one-hot processing. Therefore, the choice of the intermediate point must be larger than 64 to better represent the three-dimensional space formed by the data, and gradually increase the number of intermediate nodes in the step of 16, and according to the average loss gradually adjust the network layers with different number of nodes in the intermediate layer until the average loss has no obvious change, finally achieve the optimal combination of intermediate nodes number is 256-128-128-128-64-64-1.

The average minimum prediction loss value of different network layers is shown in Fig. 1:
As can be seen from Fig. 1, with the gradual increase of the number of intermediate network layers, the average minimum loss value predicted by the model gradually decreases, while the number of intermediate network layers of the model is greater than seven, the prediction result does not improve significantly, so the model finally chooses the combination of the seven intermediate layers.

Currently, Adam optimizer, SGD (stochastic gradient descent) and RMSprop algorithm are commonly used in deep learning models. Select the optimizer that best fits your model data according to your loss under different optimizers. The loss curves obtained by different optimizers are shown in Fig. 2.

As can be seen from Fig. 2, RMSprop has a better optimization effect than SGD, while Adam gets a smaller loss value, so Adam is selected as the optimizer of this model.

There are a total of 5617 training samples in this study. In order to get the best weight and bias as soon as possible and speed up the training efficiency, the batch-size is taken to be 128.
The hyper-parameters structure of the deep learning model selected after debugging is as follows: one layer of input, seven intermediate layers and one layer of output; The nodes of each layer are: input layer -256-128-128-128-64-64-1 (output layer); The optimizer is Adam; The activation function is the ‘Relu’ function, which is commonly used in machine learning.

The deep learning development environment is: TensorFlow 2.0 + PyCharm

3. Case analysis
OD passenger ticket data of the Beijing-Shanghai high-speed railway line on Tuesday, April 2018 are used as training data. The reason is that this period is usual, with abundant train capacity and fewer restrictions for passengers, so they can choose the trains they are satisfied with as much as possible. The sample number is 5,617. Using the Tuesday in May 2018 as the forecast data, the sample total was 5,373. In order to reduce the error caused by the accidental nature of personal subjective choices, the three weeks of Tuesday in April (20180410, 20180417, 20180424), the average number of passengers selected by different trains is taken as the goal of training data. And the next three Tuesdays after the May-Day (20180508, 20180515, 20180522), the average number of people is selected as the prediction target, and the OD with fewer people (generally <10) is eliminated to reduce random errors. Part of original ticket data is shown in Table 2.

Table 2. Part of ticket data of Beijing-Shanghai high-speed rail line.

| Date       | Train | Departure Station | Departure time | Arrival Station | Arrival time | Mile | Seat-type | Passengers | Fare |
|------------|-------|-------------------|----------------|-----------------|--------------|------|-----------|------------|------|
| 20180514   | G270  | Cangzhou west     | 20:15          | Beijing south   | 21:08        | 210  | Second class | 45         | 94.5 |
| 20180514   | G270  | Xuzhou east       | 18:02          | Zaozhuang       | 18:20        | 66   | Second class | 25.3       | 29.5 |
| 20180517   | G270  | Zaozhuang         | 18:22          | Beijing south   | 21:08        | 627  | Second class | 81.7       | 284  |
| 20180517   | G1914 | Nanjing south     | 12:01          | Changzhou north | 12:33        | 130  | Second class | 21         | 59.5 |
| 20180424   | G1940 | Nanjing south     | 18:21          | Xuzhou east     | 19:35        | 331  | First class  | 13.3       | 254.5 |

The processed data is shown in table 3.

Table 3. Processed data.

| Dep_time | Arr_time | Mile | Time (min) | Capacity | Fare | Service | Frequency | OD-Passenger | Dep. GDP | Arr. GDP | Total Mile | Load Factor | Dist Coef |
|----------|----------|------|------------|----------|------|---------|-----------|--------------|----------|----------|------------|-------------|-----------|
| 1215     | 1268     | 210  | 53         | 871      | 94.5 | 39      | 2960.2    | 3944        | 30320    | 1306     | 0.78       | 16.1        |           |
| 1082     | 1100     | 65   | 18         | 871      | 29.5 | 19      | 493.5     | 6755        | 2402     | 1306     | 0.78       | 5           |           |
| 1102     | 1268     | 627  | 166        | 871      | 28.4 | 9       | 732.8     | 2402        | 30320    | 1306     | 0.78       | 48          |           |
| 721      | 753      | 130  | 32         | 1036     | 59.5 | 44      | 1374.2    | 12820       | 7050     | 1509     | 0.59       | 8.6         |           |
| 1101     | 1175     | 331  | 74         | 118      | 254.5 | 103    | 4660.3    | 12820       | 6755     | 1509     | 0.59       | 21.9        |           |

Because passengers in the same regions have similar selection characteristics, in order to better classify statistics, it uses the fuzzy clustering analysis method to classify the station ranks of the Beijing-Shanghai high-speed railway line into four levels based on arrival and departure traffic flow, the location and economic and population level of station nodes. The classification results are shown in table 4.

Table 4. Node class of Beijing-Shanghai high-speed rail line.

| Node level          | Station name                        |
|---------------------|-------------------------------------|
| The First-class nodes | Beijing south, Jinan west, Nanjing south, Shanghai hongqiao, Tianjin south, Xuzhou east |
| The second-class nodes | Bengbu south, Cangzhou west, Changzhou north, Danyang north, Dezhou east, Kunshan south, Suzhou north, Wuxi east, Zhenjiang south |
| The third-class nodes | Chuzhou, Langfang, Qufu east, Tai’an, Tengzhou east, Suzhou east, Zaozhuang |
| The forth-class nodes | Dingyuan |

According to the passenger ticket data, the actual selection probability of different trains for different ODs of the Beijing-Shanghai high-speed railway line can be obtained, comparing with the fitting results.
obtained from the deep learning network model. It can be concluded that the overall fitting result is good. Taking Beijing south-shanghai hongqiao as an example, the comparison between actual passenger selection and fitting results is shown in Figure 3.

![Selection Comparison](image)

**Figure 3.** Comparison of actual passenger selection and fitting results from Shanghai hongqiao to Beijing south.

From the figure above, it can be seen that deep learning model has a good fitting accuracy in predicting the passengers’ selection probability of Shanghai hongqiao to Beijing south. The selection accuracy data of different OD of different nodes are shown in Table 5.

| OD level                        | Average accuracy of different car number selection |
|---------------------------------|--------------------------------------------------|
| Between the first class nodes   | 72.2%                                            |
| Between the first and second class nodes | 74.9%                                            |
| Between second class nodes      | 81.7%                                            |
| Second class nodes and other nodes | 83.2%                                            |

Due to the first class nodes are usually relatively large comprehensive hub nodes, with more passenger flow and more complex passenger type, the passenger’s travel choice behavior is very complicated, and involves a large number of influencing factors. Therefore, the fitting effect between the OD of the first-level node and other nodes is not very good. For this type of OD, the mapping feature dimension will be further increased in subsequent studies and the amount of ticket data will be increased for more accurate fitting.

4. Summarize

Due to its unique pattern recognition ability and ability to extract rules from data, deep learning models do not require domain knowledge with the support of a large amount of data, more and more such non-parametric machines are used in emerging behavior choice study. While the deep learning model still has its drawbacks: 1 Lack of theoretical support. 2 Can not explain the weight of mapping features to results. In this study, a deep learning network model is used to analyze the Beijing-Shanghai high-speed railway market and the attributes of ticket data are used as the feature vector of deep learning model, which can map the selection probability of high-speed railway passengers for different type of trains. The results show that machine learning model can predict design indicators well.

This article only analyzes the ticket data on Tuesday, and the data volume is limited. In subsequent studies, we can continue to increase the prediction accuracy by increasing the feature vector, and increases the data training amount. Several aspects of the high-speed railway operations and
management can benefit from the developed methods in this study, one of which is high-speed railway passenger flow distribution research.

Acknowledgement
This research is financially supported by the "Belt and Road" Land Corridor International Transport Research and Exchange Center (2016YFE0201700); Technology for Improving Railway Passenger and Freight Benefits and Service Levels(2018YFB1201402); Research on Key Technologies for Improving Transportation Capacity and Efficiency of High-speed Railway Network (P2018X001).

Reference
[1] Cao yu, wang cheng, wang xin, gao yueler. Short-term traffic flow prediction of urban roads based on spatio-temporal node selection and deep learning [J/OL]. Computer application: 1-10[2019-12-23].
[2] Wu Fanyi, wang, zhao Di Xu Kaiming, wu jianjun. Reinforcement learning based on the depth of cellular unmanned aerial vehicle (uav) in a network of trajectory design [J/OL]. Radio communication technology: 1-6 [2019-12-23].
[3] Zhang chao. Research on type identification of railway freight cars based on deep learning [D]. Beijing Jiaotong University, 2019.
[4] Zhao bing, dai mingrui, li ping, ma xiaoning, wu yanhua. Research on defect detection of key railway components based on deep learning [J]. Journal of railway science, 2019, 41(08):67-73.
[5] Wu haiping, huang shicun. Experimental study on information extraction of newly added construction land based on deep learning -- innovation exploration of national remote sensing monitoring project of land use [J/OL]. Remote sensing of land resources, 2019(04):159-166[2019-12-23].
[6] Huang xin, zhang zhiqiang, Dan xinghua, zhang tao, li shida. Research on intelligent railway passenger travel based on e-ticket [J]. China railway, 2019(11):1-6.
[7] Qiang lixia. Research on high-speed railway passenger travel choice behavior based on passenger ticket data [J]. Railway transportation and economy, 2008, 40(04):52-57.
[8] Liu gaoyuan. Analysis and research on passenger travel behavior of China's high-speed railway based on passenger ticket characteristic data [D]. Beijing Jiaotong University, 2012.
[9] Zhang hang, zhao peng, qiao ke, et al. Logit model and analysis of high-speed railway passenger travel time selection [J], railway transportation and economy, 2017,39 (1) : 55-60.
[10] The Machine Learning; New Machine Learning Study Results Reported from Tongji University (Analyzing High Speed Rail Passengers' Train Choices Based On New Online Booking Data In China) [J]. Computers, Networks & Communications, 2019.