Evaluation Methodology of Transit Based on Big Data Mining

Kang ZHOU* and Xiao PENG
China Academy of Transportation Sciences, Beijing, 100029
*Corresponding author

Keywords: Urban traffic, Transit evaluation, Index, Big data, Bayesian network.

Abstract. Analysis of the operating conditions of public transit, potential problems and potential; evaluation of the existing public transport network scheme to meet the future urban public transport passenger traffic demand; feedback and verification of the implementation effect of public transport network planning. These are very important to traffic system management, especially bus operation management. Therefore, it is necessary to establish a scientific and practical evaluation index system and evaluation method for the public transport network, and make a scientific evaluation of the public transport network, and optimize the bus routes according to the evaluation results. Based on the big data mining and Bayesian network, a new method for evaluating public transport service is proposed. So, the service of transit can be evaluated reasonably from the point of quantitative analysis.

Introduction

Now, the research focus on bus line and bus operation evaluation has gradually shifted from network structure to service quality, and more attention has been paid to the theories and methods of passenger psychology, evaluation methods and model establishment. Botzow (1974) proposed a method for evaluating the quality of public transport trip service and applied it to the evaluation of San Francisco Autonomous rail Rapid Transit (ART) [1]. Alter (1976) believed that it is need to improve transit accessibility, reliability, direct ratio, frequency, density and reduce travel time in order to increase the competitiveness of transit, and this index is used to evaluate the level of service quality [2]. Dhingra (1987) mainly considered 9 indexes, compared and analyzed the operation situation of 3 bus lines in Delhi, India [3]. Ray (1994) mainly considered the influence of the type of bus on the service quality, and evaluated the 2 bus lines in India, Calcutta [4]. David (2003) studied the potential factors affecting the quality of public transport services and their relative importance, and established a quantitative model for calculating the quality of public transport services [5].

Because of the difficulty of data collection, the evaluation methods only evaluate the public transport network or operation from different angles, and it is difficult to make a comprehensive evaluation of the public transport network from many angles. With the development of Internet technology, the collection and processing of bus operation information based on big data provide objective and detailed data support for the evaluation of public transport network, and can give more comprehensive and objective evaluation.

Organization of the Text Public Traffic Network and Operation Evaluation Index System

Based on the analysis and research of the influencing factors of the evaluation object, and according to the primary and secondary factors, the evaluation index system expresses the results of the factors materialized. For the evaluation of public transport network, it is necessary to construct the evaluation index system from the angle of systematicness, scientificalness, comparability and operability. Each index should be independent, quantifiable and universal. The commonly used evaluation methods of bus network include analytic hierarchy process (AHP), expert evaluation method, Delphy, and so on.
Public Transport Network Evaluation Index System

From surface to line and then to point, the traditional evaluation index system of macro public transportation network is divided into 3 kinds (Figure 1): The first kind of index refers to the indexes that reflect the rationality of the overall layout of the network, which can reflect the scale of urban public transport construction, development base and potential respectively from the line density, accessibility of public traffic, the length of bus lines, the number of double track lines, and the uneven coefficient of passenger flow; The second kind of index refers to the indexes that reflect the level of public traffic service, from the aspects of bus line carrying capacity, coefficient of line non-linear, line repeatability, transfer coefficient and punctuality rate, the characteristics of convenient, rapid, comfortable, economical and efficient passenger transport network are reflected, this is the most direct embodiment of the structural rationality of the public traffic network. The third kind of index relates to a single bus stop, from the bus stop density, passenger travel time, vehicle access rate, and the proportion of bays, etc., to reflect the degree of residents approaching the bus.

Evaluation Index System of Bus Line Operation

The quality of public traffic service refers to the extent to which the public traffic enterprises meet the passengers travel demand. In bus operation evaluation, the weight of this index is higher. The evaluation of public traffic service quality usually involves safety, speed, accuracy, economy, convenience and comfort. Usually construct bus operation quality index system from the two major aspects of the service level and operational efficiency.

Determination Method of Evaluation Index Based on Big Data Mining

With the rapid development of computer hardware technology, the continuous improvement of enterprise information technology and the improvement of database technology, the amount of data accumulated by human beings is increasing exponentially. The use of data mining can discard the false data and retain the true data from a large number of data, extract useful information and convert it into knowledge. In determining data of public traffic network evaluation index, by using IC card as the main source of data can get a lot of information, so as to reduce the data amount of the traditional investigation, can not only improve data acquisition efficiency and can obtain more accurate and reliable information resources. Big data information processing cases of major indicators are as follows.

1) Waiting flow of bus stop
   By the video acquisition, recognition and other technologies, the number of bus stop waiting can be got. In addition, we can count the number of IC card users at different sites and express them by
MarkPoint images. The image consists of the following parts, as shown in Figure 2: (1) Taking Beijing as base map, marked with bus network and sites; (2) A range roaming block was created at the lower left of the image, and range roaming was realized according to different colors so as to distinguish the waiting range; (3) In the upper left of the interface, through the establishment of manual input function box, we can focus on monitoring the flow of top ranked TOP5, TOP10, TOP15 and other regions.

![Figure 2](image)

Figure 2. Waiting volume monitoring of important bus stops in the city.

2) Passenger flow of bus line and section

Through monitoring the representative section passenger flow in real time, the change of different lines can be observed, combined with the congestion degree of the road network, intelligent regulation of bus dispatching can be realized. Section passenger flow monitoring and indication are shown in Figure 3.

The passenger flow of real-time section on the key lines in the city is shown in Figure 3 (a): (1) Taking the map of Beijing as the base to establish the X and Y axes, bus lines and bus stops are distributed in the map, Z axis reflects the size of the section passenger flow. (2) The line function box, which was set in the upper left corner of the interface, can search the lines you want to see, and set the button to view the key lines. After the button is clicked, three-dimensional dynamic diagram of several key lines’ sectional passenger flow can be displayed simultaneously. (3) A range roaming block was set up at the lower left of the interface, the range roaming was realized according to different colors, in order to distinguish the range of section passenger flow. When the cursor moving to the specific section, it will show the specific passenger flow value of this section.

Figure 3 (b) is the passenger flow dynamics of a section in a short period of time, the horizontal axis represents the time, the vertical axis indicates the line section, and the color indicates the size of the passenger flow, through the picture, we can see the trend of passenger flow in short time on the line.

![Figure 3](image)

Figure 3. Real-time section passenger flow monitoring on the line.

3) The fulfil load rate of a bus on the line

Figure 4 (a) describes a dynamic 3D graph of real-time passenger flow information. Taking Beijing city as the base map. The fulfil load rate is represented by the Z axis and the contour lines are drawn in different colors. There is a roaming scroll bar in the lower left, the passenger flow information can be
viewed in real time. By clicking on the different lines, the results will be shown in Figure 4 (b), that is, the variation trend of fulfil load rate of a bus in short time.

![Real-time change of fulfil load rate of a bus on the key lines](image1)

(a) Real-time change of fulfil load rate of a bus on the key lines

![Dynamic fulfil load rate of a bus in a certain direction of a line](image2)

(b) Dynamic fulfil load rate of a bus in a certain direction of a line

Figure 4. The real-time fulfil load rate of a bus on the line.

4) Day and night variation rhythm of section passenger flow on the line

(1) Exhibition of day and night variation rhythm about historical section passenger flow on important lines

Day and night variation rhythm of section passenger flow can be obtained by the superposition of real-time section data and displayed by 3D images. As shown in Figure 5, the background is the map of Beijing, Z axis expresses passenger flow of section on the line. The function box of the line was set on the upper left corner of the interface, it can search the lines that you want to see and set up the buttons for showing key lines, after clicking the button, it can simultaneously display three-dimensional dynamic map of section passenger flow on several key lines; The time axis was set at the lower part of the interface. After it was clicked, the section passenger flow at different times of the day can be shown, the timeline can be pushed forward or backward after clicking the left or right button, the timeline also can be pushed forward automatically after clicking the middle button. When the cursor is moved to a specific section, the specific number of section passenger flow on the road will be shown.
(2) Day and night variation rhythm of historical section passenger flow on important lines

After clicking the particular line in Figure 5, the day and night variation of section passenger flow on the line is described in Figure 6.

![Figure 5: Day and night variation of section passenger flow.](image1)

In Figure 6 (a), the horizontal axis represents the time, and the vertical axis represents a different section of a line, each cell represents the passenger flow at one time at a section, different colors represent different passenger flow number. Figure 6 (b) is a helical graph, similar to the function of Figure 6 (a), taking a day as a cycle, the variation rhythm of passenger flow at different sections on the line in day time and night is analyzed. Color represents the magnitude of passenger flow, and can be used to compare the generality and dissimilarity of passenger flow at the same time. Through Figure 6 (a) or Figure 6 (b), we can find the largest section and the corresponding time of a line passenger flow, for the larger section of passenger flow, the passenger flow of specific section can be further selected, as shown in Figure 6 (c). In Figure 6 (c), the passenger flow of each section on the line during a peak period on weekdays and non-work days are calculated.
Model of Evaluation

Model Construction and Determination of Variables

After determining the index system of bus network and operation, it is necessary to determine the advantages and disadvantages of network through the analysis of the index value obtained. The commonly used methods are fuzzy comprehensive evaluation method, expert scoring method, etc. These methods are subjective and difficult to reflect the problems of the real network objectively. Bayesian network and Support Vector Machine can quantify the qualitative problems, and can analyze the evaluation indexes more objectively. Bayesian network is a graphical network based on probabilistic reasoning, and it can be used for decision evaluation of multiple control factors [6].

Combining the evaluation index system of public traffic network and operation, the evaluation model was constructed using the Bayesian network as shown in Figure 7. The evaluation model of public transit network can find and solve problems in time by observing the three factors of pressure, state and response. The first level determines whether the index is normal according to the “pressure”, “condition”, “response” and other factors, and the second level includes the specific indicators under the three categories.

<table>
<thead>
<tr>
<th>pressure</th>
<th>condition</th>
<th>response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passenger Flow demand of rush hour $X_1$</td>
<td>Running speed $X_7$</td>
<td>Improve the service level of driver and conductor $X_{n}$</td>
</tr>
<tr>
<td>Passenger Flow of other transport modes $X_6$</td>
<td>Complaints and suggestions times $X_{18}$</td>
<td>Increase the departure frequency of rush hour $X_{19}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>whether or not to adjust $S$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Figure 7, each indicator node $X_i$ and $S$ are variables, and the indexes $X_1$-$X_6$ analyze the pressure of the bus operation mainly from the external environment (passenger flow demand of peak hours, congestion status of peak hours, passenger boarding efficiency of peak hours, the unbalanced coefficient of passenger flow, average waiting time of passengers, the passenger volume of other transportation modes, etc.). Indexes $X_7$-$X_{18}$ are the status of public traffic service. Indexes $X_{19}$-$X_n$ are some improvement measures to the external pressure under the condition of not adjusting the line. The directed arcs between nodes represent the causal relationship between variables, and no arc links indicate conditional independence; according to the threshold of each evaluation index variable, the variable set and the corresponding range are determined. In combination with the established evaluation level, the $X_1$-$X_6$ range of each index variable is divided into five grades, they are grade one (1), grade two (2), grade three (3), grade four (4), and grade five (5), the $S$ is defined as “non-adjustment” and “adjustment”, and the two states are denoted as 1 and 2 respectively.

Calculation of Local Probability Distribution

According to historical statistics and expert experience, the prior probability and conditional probability are calculated, so that the probability of any combination can be calculated. First, the probability of each index variable (a priori probability) is determined, the expression is shown in Table 1.

Figure 7. Structure of evaluation model of public traffic network.
Table 1. Priori probability of parent node Xₙ.

<table>
<thead>
<tr>
<th>Xₙ</th>
<th>P(Xₙ=1)</th>
<th>P(Xₙ=2)</th>
<th>P(Xₙ=3)</th>
<th>P(Xₙ=4)</th>
<th>P(Xₙ=5)</th>
<th>∑</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>a₁</td>
<td>a₂</td>
<td>a₃</td>
<td>a₄</td>
<td>a₅</td>
<td>1</td>
</tr>
</tbody>
</table>

As we can see from Figure 7, the nodes X₁-Xₙ are the parents of the node S, or S is the child node, and the conditional probabilities between the nodes X₁-Xₙ and S can be represented in Table 2 [7].

Table 2. The conditional probability table of node S for node X₁-Xₙ.

<table>
<thead>
<tr>
<th>(X₁, X₂, …, Xₙ)</th>
<th>S</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>∑</td>
</tr>
<tr>
<td>(5, 5, …, 5)</td>
<td>a₁₁</td>
<td>a₁₂</td>
<td>1</td>
</tr>
<tr>
<td>(5, 5, …, 4)</td>
<td>a₁₂</td>
<td>a₂₂</td>
<td>1</td>
</tr>
<tr>
<td>……</td>
<td>……</td>
<td>……</td>
<td>1</td>
</tr>
<tr>
<td>(1, 1, …, 1)</td>
<td>aₙ₁</td>
<td>aₙ₂</td>
<td>1</td>
</tr>
</tbody>
</table>

We can see that through Table 2. Nodes X₁-Xₙ have 5 kinds of value states, the node S has 2 kinds of value states, and each combination of the parent nodes corresponds to a set of probabilities and both are 1.

Application of Model

After the construction of Bayesian network model and the calculation of local probability, the prediction can be carried out based on the model. The following steps can be used to predict whether or not the bus lines need to be adjusted.

The problem can be translated into calculating the probability that S is equal to 1 when the nodes X₁-Xₙ are given, and then compare the probabilities when S is equal to 2 to determine the adjustment. Particular process is like that:

1) The probability of the sub node S is calculated according to the prior probability of each parent node Xᵢ.

\[ P(S=1) = P(X₁, X₂, \ldots, Xₙ, S=1) = P(S=1 \mid X₁, X₂, \ldots, Xₙ) \cdot P(X₁, X₂, \ldots, Xₙ) \]

\[ = P(S=1 \mid X₁, X₂, \ldots, Xₙ) \cdot P(X₁) \cdot P(X₂) \cdots P(Xₙ) \]

(1)

\[ P(S=2) = P(X₁, X₂, \ldots, Xₙ, S=2) = P(S=2 \mid X₁, X₂, \ldots, Xₙ) \cdot P(X₁, X₂, \ldots, Xₙ) \]

\[ = P(S=2 \mid X₁, X₂, \ldots, Xₙ) \cdot P(X₁) \cdot P(X₂) \cdots P(Xₙ) \]

(2) Finally, the conditional probabilities are compared to determine whether the line needs to be adjusted. Compare the size of the \( P(S=1) \) and the \( P(S=2) \). If the probability of \( S=2 \) is larger, it indicates that it needs to be adjusted. Otherwise, there is no need to adjust.

Summary

The transit network evaluation method based on the analysis of big data technology can access to the site, passenger information and vehicle capacity using a big data mining, and can obtain the variation law of passenger flow and the uneven coefficient of passenger flow through the processing of real-time passenger flow and other information. Combining the public traffic network and operation evaluation index system, the evaluation model was constructed by Bayesian network, which can objectively describe the situation of public traffic network and operation with big data, and give the adjustment strategy.
Acknowledgement
This research was financially supported by the National Natural Science Foundation of China (No. 41471459).

References