Division Method of Urban Customized Bus Passenger Flow Distribution Area for Commuter-oriented Demand

Zhong Jianfeng
Hefei University of Technology, Department of Planning and Managing of Transportation
Tunxi Road 193, Hefei, China
1360700172@qq.com

Bai Haijian
Hefei University of Technology, Department of Planning and Managing of Transportation
Tunxi Road 193, Hefei, China
232949587@qq.com

Wu Zhenyu
Hefei University of Technology, Department of Planning and Managing of Transportation
Tunxi Road 193, Hefei, China

ABSTRACT
To optimize customized bus routes, the passenger flow distribution area which is suitable for opening customized bus routes should be determined firstly. Based on the characteristics of customized bus, this paper determines the appropriate division method of passenger flow distribution area. It establishes a suitable clustering method of passenger flow distribution area, considering the traffic volume of the site and the regional capacity allocation. According to traditional clustering method and k-means clustering algorithm, it can get an appropriate customized bus routes. An example analysis, in a 5000*5000 area randomly generated 500 points, the passenger flow of each point is randomly loaded between 5-20 and the maximum adjustment threshold of capacity allocation is 350, is used to verify the algorithm. Through it, based on determining the up and down area, the appropriate customized bus routes are planned, which represent the trend of customized bus routes between the upper and lower areas.

KEYWORDS: Urban customized bus; Commuter-oriented demand; Division method; Passenger flow distribution area

1 INTRODUCTION
Customized bus is a direct route to the region, which is a supplement and development of existing public transport and is able to meet multiple travel needs.
In recent years, there are many studies in customized bus. Kirby and Bhatt analyzed ten cases of public transport services in detail and provided some experience of customized bus in the organization, operation, management and other aspects. McCall analyzed a commuter bus service model in California, the United States, which travelled only a little longer than a private car, but it
costs far less than a private car. Potts put forward the theory of customized bus service model applying in different areas, which could be in combination with the actual situation to determine whether the need to open customized bus. Bao Jian analyzed the development trend of customized bus, explored the executive conditions of customized bus transportation, expounded the management characteristics of customized bus, and provided advice for operators in the operation and management of customized bus. Hu Liege determined the location and layout of customized bus stations by using k-means clustering method and range covering formula.

All researchers illustrated the concepts of the new mode of transportation — customized transportation and expounded the difference between customized bus and traditional bus in detail, but all of them did not analyze in depth and get a division method of passenger flow distribution area of customized bus. Therefore the studies could not guide the optimization of the customized bus network from the theoretical and technical level. In order to solve such problems, this paper analyzes the network of urban customized bus for the commuter-oriented demand and establishes a division method for passenger flow distribution area. K-means clustering algorithm is suitable to establish the model to achieve this goal, considering the traffic volume of the site and the regional capacity allocation,

2 COMMUTING INFORMATION DATA PROCESSING

The operation of the customized bus should meet the actual passenger demand, and the comprehensive and accurate data is the basis for the determination of the distribution area. Through online questionnaire survey and mobile phone app, morning and evening peak travel information of passengers is collected. Through quantitative analysis of commuting travel information, it can get the location information of all reservation requests and the total number of passengers required for the site.

Processing the demand data of the commuter customized bus and quantifying the destination location information of passengers and the time window requirements for arriving at the vehicle arrival, it can provide data support for the determination of operation area. Before clustering, the information of passengers should be turned into the information of stations.

![Figure 1: Commuting information transformation](image)

From Figure 1, the location and time window information of all passengers are arranged, from which the location of the passenger’ actual riding position \((x_i, y_i)\) are got. The passengers are allocated to the corresponding bus stations according to the shortest walking distance, and the total demand of each station \(q\) is calculated. The time window information of the station is composed of the time window information of the \(z\) passengers allocated. So that we can know

\[
q_i = \sum z
\]

\[
(e_i, l_i) = \sum(e_z, l_z)
\]
In the formula above, the element $q_i$ represents the demand of bus station $i$, $z$ represents the passengers assigned to the site $i$, $(e_i, l_i)$ is the time window information of the station and $(e_p, l_p)$ is the time window information of the passenger.

### 3. DIVISION METHOD OF PASSENGER FLOW DISTRIBUTION AREA

The purpose of customized bus is to meet passengers with similar travel demand. After quantifying the location and the demand information of the site, on the basis of determining the get-off area, the area which is suitable to open the customized bus is determined, considering the maximum capacity allocation and the distance between sites.

#### 3.1 Site Attributes and Distance

Clustering is categorized by individual or sample characteristics and mainly used to study the relationship between data. Site attributes are represented by a data matrix, the data matrix usually uses $m$ variables to represent $n$ objects and this data structure is the form of relational tables. The distance between stations can be represented by different distance metrics, which is used to reflect the distance between different cluster regions. The commonly used distance measurement methods are as follows

1. Euclidean distance

\[
d(i, j) = \left( \sum_{k=1}^{n} (x_{ik} - x_{jk})^2 \right)^{1/2}
\]

2. Manhattan distance

\[
d(i, j) = \sum_{k=1}^{n} |x_{ik} - x_{jk}|
\]

3. Minkowski distance

\[
d(i, j) = \left( \sum_{k=1}^{n} (x_{ik} - x_{jk})^p \right)^{1/p}
\]

The location of passengers' travel demand on the bus station and the requirement of each station are sorted out respectively. All sites constitute a collection of sites $X \{X_1, X_2, \ldots, X_n\}$, each sample in the site set has 2 attributes $X \{x_1, x_2\}$, so the sites constitute a $n \times 2$ data matrix. As follows
Use $S_{ij}$ to represent the distance between two stations, for bus stations, the distance is reflected by Euclidean distance and Manhattan distance. The distance between classes after clustering is directly represented by the distance between the barycenters of the two classes.

$$S_{ij} = \left\{ \alpha \sqrt{(x_{ij} - x_{ij})^2 + (y_{ij} - y_{ij})^2} + \beta \left[ |x_{ij} - x_{ij}| + |y_{ij} - y_{ij}| \right] \right\}$$

In the formula above, the elements $\alpha$ and $\beta$ are related to the structure of urban road network, $x_{ij}$ and $y_{ij}$ are the horizontal and vertical coordinates of $i$ point, $x_{ij}$ and $y_{ij}$ are the horizontal and vertical coordinates of $j$ point.

### 3.2 Clustering Algorithm Considering Transport Capacity Allocation

Reasonable regional division is the basis of customized bus network optimization. The location and demand of each station are different, which causes that it is difficult to meet the reservation requirements of all passengers for the coverage of customized bus stations and routes. If there are too many areas to be divided, the occupancy of opening customized bus lines will not meet the requirements, and it may reduce the efficiency of the use of road resources and be unable to ease traffic congestion in the morning and evening peak. If the partition area is too little, there are too many sites in the route of microcirculation, and frequent stopping of vehicles will waste passengers' travel time.

From Figure 2, before clustering, it is too difficult to solve the model of traffic station scale because of scattered stations, but after clustering, the buses only deliver passengers in the same area, and the number of driving stations is greatly reduced, which can greatly shorten the travel time of customized buses and improve their operation efficiency.
3.2.1 K-means algorithm and alternative sites

K-means algorithm is a clustering algorithm divided based on distance similarity measure, the core idea of which is to constantly adjust the cluster center until the sum of intra class squared error is minimum and unchanged.

![Diagram of K-means algorithm process]

Figure 3: The process of k-means algorithm

From Figure 3, the specific steps of k-means algorithm are as follows:

Step1. Initialize clustering centers, viz., randomly assign $k$ clustering centers $\{j_1, j_2, ... j_k\}$.

Step2. Distribute sites, for each site $X_i$, find out the nearest site to it and assign it to the class where the $j_v$ is the clustering center.

Step3. Modify cluster center, after the first site allocation, update the clustering centers to the position of barycenter of each class.

Step4. Calculate square error, $D = \sum_{i=1}^{n} \left[ \min_{v=1,2,...k} d(x_i, j_v)^2 \right]$.

Step5. Closing rule, if $D$ convergence, end the algorithm, otherwise, turn to step2.

$X_i$ represents a station, containing two data $x_i$ and $y_i$ of the plane coordinate of the station. The center after clustering $j_v$ represents the barycenter of each class, which also contains two data $x_j$ and $y_j$. The distance between sites $i$ and $j$, and the distance between site $i$ and center point $j_v$ can be measured by Equ.4. The distance between class centers $w$ and $v$ can be measured by Euclidean distance, as Equ.9.

$$\text{dis}(j_w, j_v) = \sqrt{(x_w' - x'_v)^2 + (y_w' - y'_v)^2}$$  \hspace{1cm} (9)

The sum of intra class squared error is shown as Equ.10.

$$D = \sum_{i=1}^{n} \left[ \min_{v=1,2,...k} d(x_i, j_v)^2 \right]$$  \hspace{1cm} (10)
The number of class centers is first set, and then the k-means clustering algorithm is used to calculate the central coordinates of k classes and the sites assigned to each class of k.

3.2.2 Cluster adjustment considering maximum capacity allocation

The clustering of bus stations needs to consider the distance between the stations and the passenger flow of each station. Some classes after clustering by k-means with higher passenger flow may not meet the transport capacity requirements of the public transport enterprises in this area. Based on this, passenger flow constraint should be added to adjust the class clusters. The algorithm flow of adjustment is shown in Fig.4, and its specific algorithm steps are as follows:

Step1. Select the class with highest passenger flow. Calculate the total passenger flow of the stations of each class, select the class with highest passenger flow $j_{\text{max}}$ and define passenger flow is $q_{\text{max}}$. 

Figure4: The process of cluster adjustment based on capacity allocation
Step2. Judge the threshold. Set an adjustment threshold $W$, if $q_{\text{max}} > W$, turn to step3, otherwise, finish the algorithm.

Step3. Adjust the classes. Select the farthest point from the cluster center in the group, calculate the distance between the rest cluster centers and the point and add the point into the shortest clustering center from it.

Step4. Update class center. Adjust the location of the cluster centers with culling points and join points.

Step5. Update passenger flow. Update the passenger flow $q_{\text{max}}$ of $j_{\text{max}}$, if $q_{\text{max}} > W$, turn to step3, otherwise, finish this adjustment and turn to step1.

4 CASE STUDY

In a 5000*5000 region, 500 points are randomly generated, and the passenger flow at each point is randomly loaded at 5-20, and the maximum adjustment threshold of the capacity allocation is 350. According to the plane coordinate of the data point, the position of the initial point is marked on the picture, as shown as Finger 5, each point in the graph represents the location of an appointment place for travel demand.

All demands of boarding stations are clustered by customized bus site clustering algorithm. First of all, the k-means algorithm is used to make the first cluster of all boarding stations and draw up 20 cluster centers. According to the distance criterion, the initial cluster centers and the boarding stations they contain are got and shown in Finger 6. Based on this, according to the principle of capacity adjustment, the new cluster centers and its included objects are obtained by adjusting the bus stations included in the cluster centers. If the threshold value of the passenger flow meets the requirements, then the results will be output, as shown in Figure 7.
Through the clustering results of k-means algorithm and final clustering, the number of sites and passenger flow corresponding to different clustering schemes can be obtained and shown in Table 1 and Table 2. The difference before and after clustering of the number of sites and passenger flow are shown in Figure 8 and Figure 9.

<table>
<thead>
<tr>
<th>Class number</th>
<th>Station number</th>
<th>Passenger flow</th>
<th>Class number</th>
<th>Station number</th>
<th>Passenger flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>393</td>
<td>11</td>
<td>18</td>
<td>222</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>435</td>
<td>12</td>
<td>23</td>
<td>255</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>353</td>
<td>13</td>
<td>15</td>
<td>213</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>279</td>
<td>14</td>
<td>22</td>
<td>252</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>452</td>
<td>15</td>
<td>31</td>
<td>399</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>249</td>
<td>16</td>
<td>17</td>
<td>220</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>427</td>
<td>17</td>
<td>16</td>
<td>159</td>
</tr>
<tr>
<td>8</td>
<td>27</td>
<td>338</td>
<td>18</td>
<td>16</td>
<td>177</td>
</tr>
<tr>
<td>9</td>
<td>26</td>
<td>325</td>
<td>19</td>
<td>26</td>
<td>315</td>
</tr>
<tr>
<td>10</td>
<td>37</td>
<td>423</td>
<td>20</td>
<td>21</td>
<td>273</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class number</th>
<th>Station number</th>
<th>Passenger flow</th>
<th>Class number</th>
<th>Station number</th>
<th>Passenger flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27</td>
<td>348</td>
<td>11</td>
<td>18</td>
<td>222</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>341</td>
<td>12</td>
<td>32</td>
<td>342</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>346</td>
<td>13</td>
<td>15</td>
<td>213</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>339</td>
<td>14</td>
<td>29</td>
<td>345</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>342</td>
<td>15</td>
<td>28</td>
<td>344</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>249</td>
<td>16</td>
<td>21</td>
<td>295</td>
</tr>
<tr>
<td>7</td>
<td>29</td>
<td>346</td>
<td>17</td>
<td>18</td>
<td>181</td>
</tr>
<tr>
<td>8</td>
<td>29</td>
<td>348</td>
<td>18</td>
<td>16</td>
<td>177</td>
</tr>
<tr>
<td>9</td>
<td>28</td>
<td>347</td>
<td>19</td>
<td>28</td>
<td>342</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>347</td>
<td>20</td>
<td>25</td>
<td>345</td>
</tr>
</tbody>
</table>
From Fig.8 and Fig.9, after adjustment, the number of stations contained in each class was decreasing overall, and the imbalance of passenger flow of all classes was significantly reduced. The capacity of the region will not be too large, and the overall situation is stable.

5 CONCLUSIONS

Based on the concepts and characteristics of customized bus, this paper introduces a division method of passenger flow distribution area considering the traffic volume of the site and the regional capacity allocation. The index of commuting analysis was obtained through quantization, viz., the distance indicator based on get-on and get-off sites and the time window indicator based on the different travel time of residents. Through k-means algorithm, morning and evening commuting customized bus distribution area clustering algorithm based on distance and passenger flow index is obtained.

REFERENCES

Li Yubo.(2012). The k-means algorithm is used to select the relevant problems[D]. Lanzhou Jiaotong University, 2012.