Identification method of crowded passenger flow based on automatic fare collection data of Nanjing Metro

Lu Jia   Ren Gang   Xu Linghui

(Jiangsu Key Laboratory of Urban ITS, Southeast University, Nanjing 211189, China)
(Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Nanjing 211189, China)
(School of Transportation, Southeast University, Nanjing 211189, China)

Abstract: To relieve traffic congestion in urban rail transit stations, a new identification method of crowded passenger flow based on automatic fare collection data is proposed. First, passenger travel characteristics are analyzed by observing the temporal distribution of inflow passengers each hour and the spatial distribution concerning cross-section passenger flow. Secondly, the identification method of crowded passenger flow is proposed to calculate the threshold via the probability density function fitted by Matlab and classify the early-warning situation based on the threshold obtained. Finally, a case study of Xinjiekou station is conducted to prove the validity and practicability of the proposed method. Compared to the traditional methods, the proposed comprehensive method can remove defects such as efficiency and delay. Furthermore, the proposed method is suitable for other rail transit companies equipped with automatic fare collection systems.

Key words: travel characteristic; identification method; crowded passenger flow; automatic fare collection

DOI: 10.3969/j.issn. 1003－7985.2019.02.014

Traffic congestion problems in the metropolises of China have made it imperative to renovate and reconstruct the existing urban rail transit (URT) systems. It is common for many URT stations to suffer from very high pedestrian density, which will lead to the increased risk that URT stations will encounter crowded passenger flow. In addition, the safety and comfort of passengers will always be under threat, and the attraction of URT will also be reduced. The optimal method to solve these problems is to identify and relieve crowded passenger flow. Nevertheless, not all passenger flows should be regarded as crowded passenger flow. Thus, it is necessary to determine the threshold of passenger flow scientifically.

Automatic fare collection (AFC) has become increasingly popular in public transportation. It provides a vast

amount of continuous and dynamic data, such as the number of inflow passengers. It is also detailed and accurate compared to manual surveys. Early applications of AFC data appeared in developed countries such as America, England, and France. The existing literature conducted on AFC mainly focused on passenger trip origin-destination (OD). It was divided into two periods. Before 2010, only the data of inflow passengers can be supplied. Barry et al. [1-3] has been researching on the estimation of passenger trips OD. After 2010, the data of outflow passengers can be obtained. Rao et al. [4-6] tended to research the real-time OD matrices. In addition to the passenger trip OD, AFC data was also applied to study travel characteristics. Hasan et al. [7] used the AFC data and observed the spatial and temporal passenger travel patterns. They also simulated two important passenger decisions: where to go and how long to stay. Sun et al. [8] utilized AFC data visualization methods to reveal the characteristics of passenger trips and evaluate travel time reliability from the passengers' perspective.

Currently, the identification of crowded passenger flow is a hot topic along with the rapid increase of traffic demand during recent years. A vast amount of literature focused on the operation state of stations and the service level of facilities. Ma[9] evaluated the passenger flow state of transfer stations in terms of three aspects, stabilizing intensity, balanced intensity, and the smooth intensity based on the entropy theory. Huang[10] completed the quantitative analysis of the safety state of passenger flow, and utilized fuzzy comprehensive evaluation to identify the safety states of the passenger flow. Likewise, the existing literature also researched the crowded passenger flow caused by large-scale activities. Xu et al. [11-12] took many factors into account, such as safety capacity, safety density, safety speed, density gradient and speed gradient, to study the early-warning method of passenger flow. Xu et al. [13] established the analysis model of train delay to calculate the identification threshold for the influence imposed on the distribution of passenger flow. At the same time, some scholars quantified the capacity of station facilities to identify crowded passenger flow. Davidich et al. [14] evaluated the effects of waiting pedestrians, and proposed a new cellular automata model for ana-

lyzing and predicting the capacity of waiting areas in critical situations. Seriani et al.[15] analyzed the capacity of transfer space between the subway and bus, and made a planning guide for Santiago, Chile. Fernández et al.[16] proved the existence of pedestrian saturation flow in the platform gates of public transportation, and showed the capacities of the train doors under different conditions.

The existing literature mainly applied station operation state (such as density) and facility capacity to identify crowded passenger flow and calculate the threshold by video recognition technology. However, these methods lack accuracy and flexibility, and do not consider the travel characteristics of AFC data completely. Therefore, to solve these problems, this paper proposes a new method to obtain the threshold rapidly by analyzing the travel characteristics and five-minute passenger flow. This paper presents the passenger travel characteristics including spatial distribution and temporal distribution, and proposes a new identification method for crowded passenger flow based on a five-minute passenger flow.

1 AFC Data

All AFC data involved is provided by the Nanjing Metro Operation Co., Ltd. Nanjing Metro only accepts smart cards (either single-trip card or the other 49 categories of smart cards). That is to say, AFC data can cover all the inflow passengers and outflow passengers. Therefore, this AFC data includes the information as follows: date of travel, tap-in time at origin station, tap-out time at destination station, line IDs and station IDs.

The dataset contains the information of all the passengers involved in Nanjing Metro across a year's duration of service.

2 Passenger Travel Characteristics

2.1 Spatial distribution

Cross-section passenger flow is the key indicator for analyzing the characteristic of spatial distribution and estimating whether new lines are worthwhile constructing. It can also help to compile train operating plan, train formation plan and the train timetable. The definition of cross-section passenger flow is the number of passengers actually on a section (geographic link between two adjacent stations in a given direction) of an URT line at a specific point in time.

Cross-section passenger flow is computed by several factors including the tap-in time at the origin station, tap-out time at the destination station, the path in the URT system, train timetable and relative capacity, and the distances between stations. The data source of calculating cross-section passenger flow at morning and evening peaks is the AFC data and basic information of the rail transit network provided by Nanjing Metro Operation Co., Ltd. Considering the service time and

steady passenger flow, only Line 1 is selected. The morning peaks selected are from 8:00 to 9:00, and the evening peaks are from 18:00 to 19:00. Fig. 1 shows cross-section passenger flow in the up and down directions at morning and evening peaks. The up direction is from Maigaoqiao to China Pharmaceutical University. The down direction is from China Pharmaceutical University to Maigaoqiao.

From Fig. 1, some interesting phenomena can be found. There are some similarities at morning and evening peaks. The distributions of cross-section passenger flow in the down direction at morning peaks and in the up direction at evening peaks. The maximum cross-section


(a)
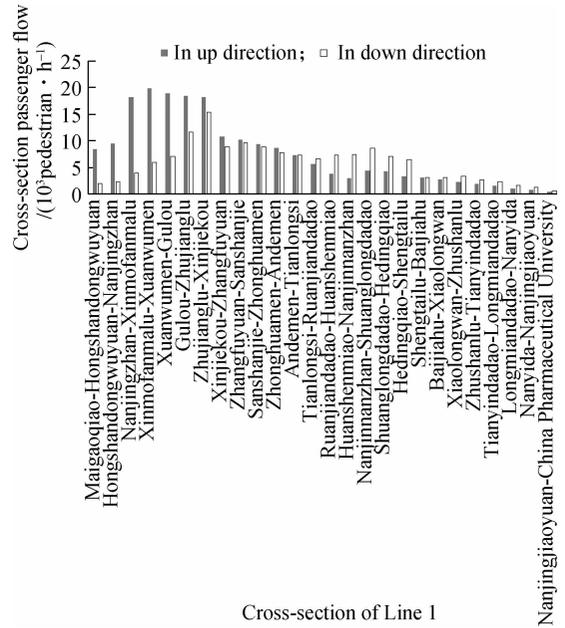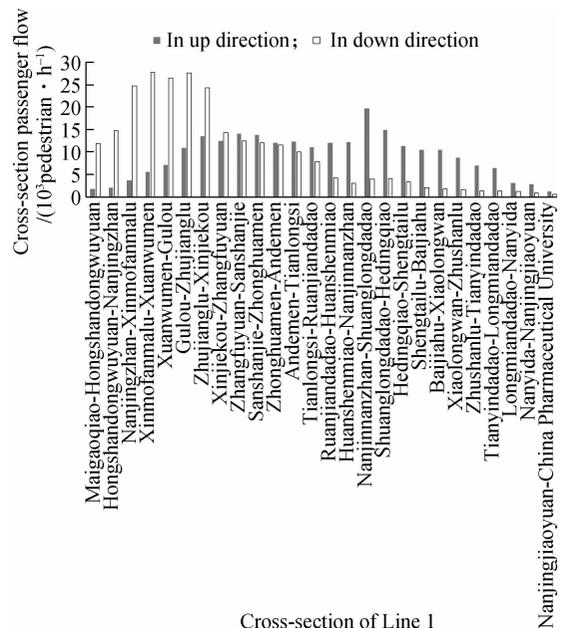
(b)

Fig. 1 All cross-section passenger flow of Line 1. (a) At morning peaks; (b) At evening peaks

passenger flow in the up direction at morning peaks is different from that at evening peaks. The former is at the cross-section from Nanjingnanzhan to Shuanglongdadao and the latter is from Xinmofanmalu to Xuanwumen. Moreover, the distribution of cross-section passenger flow in the up direction at morning peaks is uniform. So is that in the down direction at evening peaks.

2.2 Temporal distribution

In order to consider the characteristic of passenger flow and to better predict the crowded passenger flow, this paper mainly employs the number of inflow passengers each hour to analyze temporal distribution. In Fig. 2, the number of inflow passengers for all lines of Nanjing Metro in each hour is shown. For each colored area, the height represents the number of inflow passengers of the corresponding metro line each hour.
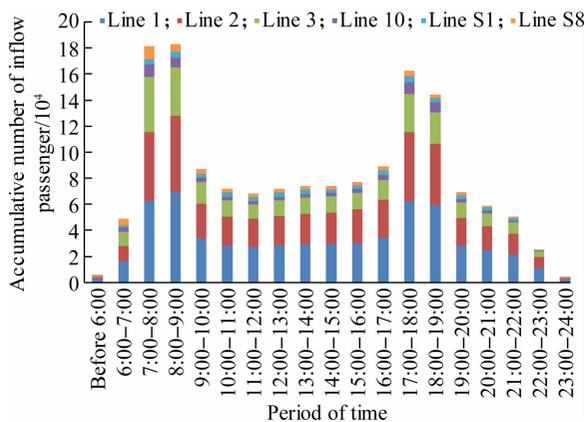

(a)
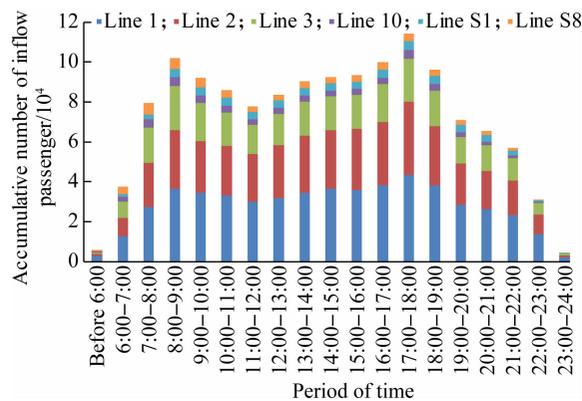
(b)

Fig. 2   Temporal distribution of the inflow passenger flow. (a) On workdays; (b) On weekends

Through observation, we find that on workdays, due to commuting traffic, the number of inflow passengers clearly has two peaks. One is from 7:00 to 9:00 and the other is from 17:00 to 19:00. The passenger flow decreases slightly at noon and gradually lessens at night from 19:00 to 24:00. On weekends, the number of inflow passengers in all periods is particularly well-balanced.

3   Threshold of Crowded Passenger Flow

3.1   Method

Currently, the existing identification methods of crowded passenger flow mainly focus on the station operation state (such as density) and station facility capacity. These methods are qualitative rather than quantitative, and also lack accuracy and flexibility. Meanwhile, the crowded passenger flow of URT stations is closely related to people's life. The threshold of crowded passenger flow is determined by many factors such as AFC, the waiting area of the platform, and staircase area. However, these factors are difficult to evaluate in practice. Therefore, to address the shortcomings of the existing methods, a new identification method of crowded passenger flow is proposed based on the five-minute passenger flow. Compared to previous methods, this method can identify the crowded passenger flow accurately and flexibly by calculating threshold quantitatively. All data involved must be cleaned so that abnormal data records can be removed. The main steps are as follows:

1) Analyze one station's travel characteristics including temporal distribution and spatial distribution, and determine the morning and evening peaks of this station.

2) Calculate the number of inflow and outflow passengers every five minutes at the morning and evening peaks on each day over a period.

3) Divide the maximum value of five-minute passenger flow into n sections. Then count the frequency of five-minute passenger flows located at each section, and generate a frequency histogram.

4) Find all the possible function curves by observing the frequency histogram. Then conduct the correlation analysis, and finally determine the suitable function.

5) Use Matlab to fit the probability density curve of passenger flow, and acquire the function formula of probability density.

6) Determine the appropriate threshold based on the capacity of the automatic ticket checker (ATC), as shown in Tab. 1.

Tab. 1   Standard of capacity of an automatic ticket checker

Automatic ticket checker types	Card types	Maximum design capacity/ (pedestrian · h^{-1})
Three-wing revolving door	Magnetic card	1 500
	RF card	1 800
Door leaf style	Magnetic card	1 800
	RF card	2 100

3.2   Case study

Xinjiekou (XJK) station is the key station located in the CBD of Nanjing. It supports many links of passengers, which leads to a crowded passenger flow. Therefore, XJK station is selected for conducting a case study.

The AFC dataset of XJK station covers a period starting

from April 1, 2015 to March 31, 2016. Combined with the analysis of temporal distribution and spatial distribution in Section 2, the morning and evening peaks of XJK station are defined as 7:00—9:00 and 17:00—19:00, respectively.

Fig. 3 shows the frequency histogram of the five-minute passenger flow during peaks. The morning and evening peaks are divided into 48 five-minute time periods, such as 7:00—7:05. Then, the passenger flow for 48 five-minute time periods per day during one year is calculated. Namely, 17 520 (48 × 365) five-minute passenger flow values are obtained. The maximum value of five-minute passenger flow approaches 2 100. Then, 2 100 is divided into 70 sections (30 pedestrians in each section) from 0 to 2 100. This is the abscissa. Then, 17 520 values are analyzed to obtain the number of values from different sections. This is the ordinate. In summary, there are 70 rectangles and the width of each rectangle is 30, and the height of each rectangle represents the number of values located at each section. Due to the limitation of space, every 10 rectangles is separated by a light white vertical line, i. e. 300 or 600. The value of 300 corresponds to a point, that is, the value of the abscissa corresponding to the light white vertical line in the figure. The width of each rectangle corresponds to the section of passenger flow. For example, the abscissa of the 10th rectangle ranges from 270 to 300. By analogy, the abscissa of the 20th rectangle ranges from 570 to 600.
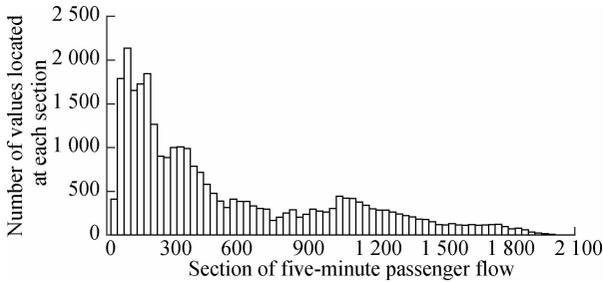
**Fig. 3**   Frequency histogram of five-minute passenger flow

After observation, we find that the tendency of the obtained histogram is consistent with log-normal distribution. Then, the P-P plot test is applied to analyze its fitting degree. The P-P plot is drawn based on the relationships between the cumulative proportions of the variable and the designated distribution. The equation for the probability density of log-normal distribution is

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad (1)$$

where μ is the expectation of $\ln x$, and σ^2 is the variance of $\ln x$.

If all the points in P-P plot appear as a straight line, the data will obey the designated distribution. The results of the P-P plot test are revealed in Fig. 4. Each point re-

presents a piece of data (see Fig. 3), and all the points are closely centered on the 45-degree line of the P-P plot. In other words, it means that the data satisfies the log-normal distribution.
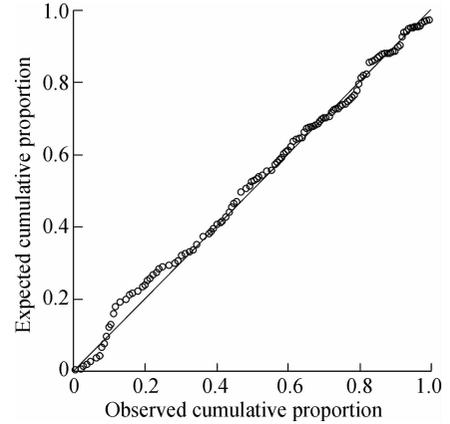
Fig. 4   P-P plot test of tendency of histogram

It is proven that the data is consistent with the log-normal distribution. Next, we use Matlab to conduct the curve fitting. The fitting result is shown in Fig. 5. The abscissa in this figure is similar to that in Fig. 3. There are 70 rectangles and the width of each rectangle is 30. Then, 17 520 values are analyzed to obtain the number of the values from different sections. The ordinate represents the ratio of the number of values up to 30. Namely, for each white rectangle, the area of each rectangle represents the probability that the five-minute passenger flow locates at the corresponding section. Due to the limitation of space, every 10 rectangles is separated by a light white vertical line, i. e., 300 or 600. The value of 300 corresponds to a point, that is, the value of the abscissa corresponding to the light white vertical line in the figure. The red curve is the probability density function curve of the five-minute passenger flow located at each section.
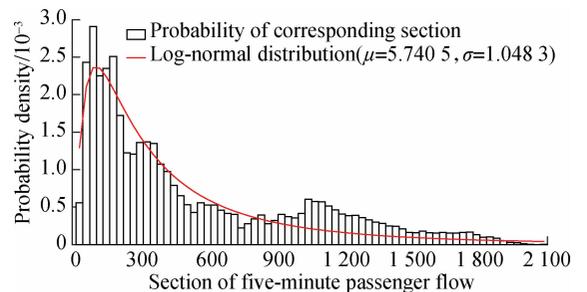
Fig. 5   Probability density curve fitting of the section of five-minute passenger flow

The probability density function of the passenger flow is

$$f(x) = \frac{0.953\ 1}{x\sqrt{2\pi}}\exp\left(-\frac{(\ln x - 5.739\ 1)^2}{2.201\ 6}\right) \quad (2)$$

where x is the passenger flow, and $f(x)$ is the probability

density function.

The probability density function can derive the cumulative distribution function by calculating the probability of passenger flow over the arbitrary sections.

In the field investigation of XJK station, all the ATCs in XJK station are door leaf style using a mix of magnetic card and RF card. The number of entry checkers and edit checkers are equal to 15. Based on Tab. 1 and the above situation, the maximum design capacity for each ATC is 1 800 pedestrian/h. So, the capacity for XJK station is 2 250 pedestrian/5min for all entry and edit checkers. The maximum five-minute passenger flow surveyed is 2 088 which is close to 2 250. It means that the data surveyed can be taken to be the threshold. However, it is necessary to point out that if the maximum five-minute passenger flow surveyed is less than 70% of the design capacity, the design capacity of each ATC should be regarded as the threshold.

The early-warning situations are classified into three classes: red level, yellow level, and green level. Among them, the yellow and red levels indicate that the station has a crowded passenger flow. In general, the probabilities of red and yellow levels are both 10% (see Fig. 6).

The abscissa in Fig. 6 is similar to that in Fig. 3. There are 70 rectangles and the width of each rectangle is 30. Next, 17 520 values are analyzed to obtain the number of values in different sections. The ordinate represents the ratio of the number of the values to 30. Namely, for each white rectangle, the area of each rectangle represents the probability that five-minute passenger flow locates at the corresponding section. The blue curve is the probability density function curve of five-minute passenger flow located at each section. The red line represents the red level. When the five-minute passenger flow is greater than the value of the red line, the early-warning situation will occur at the red level. Also, the area of the right part of the red line is determined as 10%. The analysis of green and yellow levels are similar to that of the red level. Due to the limitation of space, every 10 rectangles is separated by a light white vertical line, i. e. 300 or 600. The value of 300 corresponds to a point, that is, the value of the abscissa corresponding to the light white vertical line in the figure.
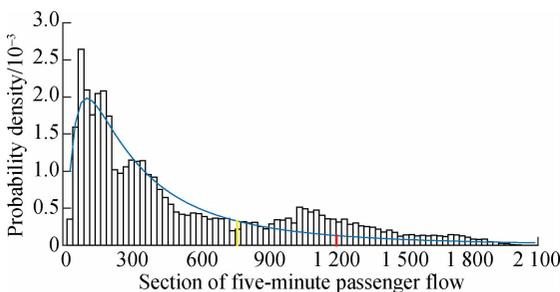

Fig. 6   Diagram of different classes of early-warning situation

The judgment criteria for the early-warning situation of crowded passenger flow in XJK station are shown in Tab. 2, where PV is five-minute passenger flow and the range of PV is calculated by Eq. (2).

Tab. 2   Judgment criteria of early-warning situation

Early-warning degree	Early-warning situation	Probability/%
$1\ 190 \leqslant PV$	Red level	10
$758 \leqslant PV < 1\ 190$	Yellow level	10
$PV < 758$	Green level	80

4   Conclusion

This paper firstly analyzes the passenger travel characteristics including temporal distribution and spatial distribution. For rail transit managers, the temporal distribution can be used to assign the transportation capacity to all lines in each period of the day. The spatial distribution can easily help us to identify and shut down the stations experiencing crowded passenger flow timely via analyzing the cross-section passenger flow of all lines. This paper also innovatively proposes a comprehensive method, acquires a probability density function by analyzing and matching the frequency of five-minute passenger flow to identify a crowded passenger flow. Through the analyses above, the judgement criteria of a early-warning situation are provided, which is convenient for managers to take measures for various situations. The case study of Xinjiekou station presented at the end proves the validity and practicability of the proposed method, and demonstrates how appropriate the analysis of AFC data is.

For future study, the transfer passenger flow at peaks should be taken into consideration. More factors must be included when analyzing the threshold, such as the station's capacity and the differences between workdays and weekends. Of course, the OD distribution and cross-section passenger flow should be combined together to analyze the spatial distribution, if available.

References

[1] Barry J J, Newhouser R, Rahbee A, et al. Origin and destination estimation in New York city with automated fare system data [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2002, **1817**: 183 −187. DOI:10.3141/1817-24.

[2] Zhao J H, Rahbee A, Wilson N H M. Estimating a rail passenger trip origin-destination matrix using automatic data collection systems[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2007, **22**(5): 376 − 387. DOI:10.1111/j.1467-8667.2007.00494.x.

[3] Chan J. Rail transit OD matrix estimation and journey time reliability metrics using automated fare data [D]. Cambridge, MA, USA: Massachusetts Institute of Technology, 2007.

[4] Rao H. Real-time estimation and prediction of OD matrix for public passenger flow based on AFC data [D]. Nanjing: Southeast University, 2014. (in Chinese)

[5] Yao X M, Zhao P, Yu D D. Real-time origin-destination matrices estimation for urban rail transit network based on structural state-space model[J]. *Journal of Central South University*, 2015, **22**(11): 4498 – 4506. DOI:10.1007/s11771-015-2998-4.

[6] Nagy V. Theoretical method for building OD matrix from AFC data[J]. *Transportation Research Procedia*, 2016, **14**: 1802 – 1808. DOI:10.1016/j.trpro.2016.05.146.

[7] Hasan S, Schneider C M, Ukkusuri S V, et al. Spatio-temporal patterns of urban human mobility[J]. *Journal of Statistical Physics*, 2013, **151**(1/2): 304 – 318. DOI:10.1007/s10955-012-0645-0.

[8] Sun Y S, Shi J G, Schonfeld P M. Identifying passenger flow characteristics and evaluating travel time reliability by visualizing AFC data: A case study of Shanghai Metro[J]. *Public Transport*, 2016, **8**(3): 341 – 363. DOI:10.1007/s12469-016-0137-8.

[9] Ma L. Analysis and evaluation of passenger flow operation state in urban railway transit hub[D]. Beijing: Beijing Jiaotong University, 2009. (in Chinese)

[10] Huang H C. Research on safety evaluation of passenger flow of railway passenger integrated transport hub[D]. Nanjing: Southeast University, 2011. (in Chinese)

[11] Xu X, Ma Y N, Li T, et al. Risk early-warning study of passenger flow in business district[C]//2010 *IEEE International Conference on Emergency Management and Management Sciences* (ICEMMS). Beijing, China, 2010: 310 – 313.

[12] Li T, Jin L Z, Ma Y N, et al. Study on method for monitoring and early-warning of passenger flow in large-scale activities[J]. *Journal of Safety Science and Technology*, 2012, **8**(4): 75 – 80. DOI:10.3969/j.issn.1673-193X.2012.04.014. (in Chinese)

[13] Xu R H, Ye J M, Pan H C, et al. Method for early warning of heavy passenger flow at transfer station of urban rail transit network under train delay[J]. *China Railway Science*, 2014, **35**(5): 127 – 133. DOI:10.3969/j.issn.1001-4632.2014.05.18. (in Chinese)

[14] Davidich M, Geiss F, Mayer H G, et al. Waiting zones for realistic modelling of pedestrian dynamics: A case study using two major German railway stations as examples[J]. *Transportation Research Part C: Emerging Technologies*, 2013, **37**: 210 – 222. DOI:10.1016/j.trc.2013.02.016.

[15] Seriani S, Fernández R. Planning guidelines for metro-bus interchanges by means of a pedestrian microsimulation model[J]. *Transportation Planning and Technology*, 2015, **38**(5): 569 – 583. DOI:10.1080/03081060.2015.1039235.

[16] Fernández R, Valencia A, Seriani S. On passenger saturation flow in public transport doors[J]. *Transportation Research Part A: Policy and Practice*, 2015, **78**: 102 – 112. DOI:10.1016/j.tra.2015.05.001.

基于南京地铁 AFC 数据的大客流识别方法

卢　佳　　任　刚　　徐凌慧

（东南大学江苏省城市智能交通重点实验室,南京 211189）
（东南大学现代城市交通技术江苏高校协同创新中心,南京 211189）
（东南大学交通学院, 南京 211189）

摘要:为了缓解城市轨道交通车站的交通拥堵问题,提出了一种基于自动售检票设备刷卡数据的大客流识别方法.首先,通过观察每小时进站乘客的时间分布和断面客流量的空间分布,分析乘客出行特征.其次,提出了大客流的识别方法.根据高峰时段的五分钟客流量进行频率直方图分析,通过 Matlab 拟合的概率密度函数计算阈值,根据获得的阈值划分预警等级.最后,以新街口站为例,证明了该识别方法的有效性和实用性.结论表明,与传统方法相比,所提出的综合方法消除了效率和延迟等缺陷.此外,所提方法也可适用于其他配备了自动售检票系统的地铁公司.

关键词:出行特征;识别方法;大客流;自动售检票设备

中图分类号:U239.3