

Research Article

Visual Analytic Method for Metro Anomaly Detection and Diffusion

Yunhui Li,¹ Yong Zhang ,¹ He Shi,¹ Yun Wei,² and Baocai Yin¹

¹Beijing Key Laboratory of Multimedia and Intelligent Software Technology Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

²Beijing Urban Construction Design & Development Group Co. Limited, Beijing 100124, China

Correspondence should be addressed to Yong Zhang; zhangyong2010@bjut.edu.cn

Received 11 November 2019; Revised 12 July 2020; Accepted 3 August 2020; Published 1 September 2020

Academic Editor: Luigi Dell'Olio

Copyright © 2020 Yunhui Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of urbanization in recent years, thousands of people have flooded into the city, which has brought tremendous pressure on the supervision and operation of relevant traffic management departments. In particular, the unexpected events in the urban rail transit system have caused great troubles for city managers. Aiming at the problem of abnormal passenger flow in the metro, this paper proposes a visual analytic method to support the abnormal passenger flow detection, verification, and diffusion analysis in the metro system. The method provides an intuitive visual metaphor and allows users to perform simple interactive operations to verify abnormal passenger flow. In addition, the method reveals the diffusion law of abnormal passenger flow in time and space in a two-dimensional diffusion view. The Beijing Rail Transit AFC data are used to validate the developed system, and two reliable analysis cases are presented. The system can help users quickly grasp the abnormal propagation rules and help them to develop different scheduling strategies for different anomalous propagation paths.

1. Introduction

With the rapid development of urban rail transit, rail transit has become the preferred mode of transportation for more and more people to travel. The large number of passengers entering the rail transit system has brought a huge impact on it. It is of great significance for the normal operation of the rail transit system to master the changing rules of rail transit passenger flow in real time, issue passenger flow warning information in time, and implement contingency plans for passenger flow changes. With the expansion of the scale of rail transit networks in megacities such as Beijing, daily emergencies are increasing. Under this circumstance, once a sudden accident occurs in a local area, it not only disorders the operation of the metro in this region but also causes a serious impact on adjacent stations and lines and even triggers the embarrassment of ground transportation. Therefore, how to identify and control the impact of such events is an urgent practical requirement.

Anomaly detection is a crucial interdisciplinary research area in many fields such as equipment fault detection and

software security. In this paper, anomalies refer to a mutation (sudden increase or decrease) of the passenger flow due to unexpected events (such as train fault, bad weather, major event, etc.) that occur at the stations. The metro system is a net system with very close connections among different stations. Once an abnormal passenger flow propagates through the orbital system, it will have an extremely bad influence on a large number of passenger flows in the system.

At present, there are some research studies on the detection methods of abnormal passenger flow in rail transportation; however, there is little research on the spread of abnormal passenger flow on the orbit. Many visual analysis methods can be used to spread abnormal information, mainly concentrate in the fields of social science, device monitoring and software analysis, transportation, and so on. However, few articles can display anomalous information in combination with geographic space and describe the propagation of anomalous passenger flow in the rail transit networks. It is difficult for the operation management

department to quickly grasp the abnormal propagation rules of passenger flow according to the existing studies, to formulate different scheduling policies for different types of abnormal propagation routes.

To solve the aforementioned problems, a method of metro anomaly detection and diffusion is proposed in this paper and a visual analysis system is developed. The main contributions of this paper are summarized as follows:

- (i) Considering the intrinsic characteristics of metro flow data, an improved RPCA model is introduced to detect anomalies at stations with abnormal passenger flows
- (ii) An anomaly diffusion algorithm is also proposed to obtain the diffusion of abnormal passenger flow in the metro and to understand the diffusion relationship between abnormal stations
- (iii) A visual analytic system is developed, which provides a two-dimensional diffusion view to show the spread of abnormal passenger flow in time and space

2. Related Works

This section reviews the prior approaches to visualize traffic spatiotemporal data and the visual analysis methods of abnormal information diffusion.

2.1. Visualization of Traffic Spatiotemporal Data. Much research has been done on the visualization of spatiotemporal traffic data [1]. Andrienko and Andrienko [2] and Slingsby et al. [3] both proposed a visual analysis method. In their methods, the map is divided into regular rectangles by using multiple heatmaps, and each heatmap represents a time overview of traffic volume in designated areas. Wang et al. [4] also used pixel table view to visualize temporal changes in traffic speed of selected road sections. The TripVista system developed by Guo et al. [5] provided an interactive method to analyze the trajectory data of different types of vehicles and pedestrians at intersections. TrajectoryLenses [6] allow users to select specific regions and time to analyze the trajectory data from the origin to the destination. Zeng et al. [7] proposed a series of visualization techniques to analyze passenger routes and traffic efficiency in the public transport system. Chen et al. [8] proposed an interactive visualization to discover motion patterns. Tominski et al. [9] showed the usefulness of 3D trajectory bands to visualize trajectory data. In their visualization, attribute data of individual trajectories are visualized as color-coded bands, and sets of trajectories are visualized by stacking the bands. Most of the relevant research studies focus on the data of road networks and vehicle trajectories, but the data in this study are based on smart card data rather than trajectory data, which is different from the general passenger flow analysis. The use of big data visualization in passenger flow analysis from network, line, station, and section is only part of big data visualization technology. It is very common to show traffic flow through points [10], lines [11], and regions [12].

The existing visual analytic methods of traffic spatiotemporal data display data not only in two-dimensional space but also in three-dimensional space. However, there are not many studies on the combination of anomaly detection and geographic information. This article aims to visually display the source of the anomalies on the map and monitor the status of each station. In this way, the stations that the user interested in can be found directly, and the abnormal details of the stations can also be grasped intuitively.

2.2. Diffusion Visualization of Anomaly. In recent years, there has been a lack of research on the propagation of orbital anomalies, and there have been fewer studies on the visualization of anomalous diffusion. Shi et al. [13] proposed multiple designs to visualize and analyze anomalies in sensor networks to allow different aspects of data to be investigated. The temporal expansion model graph displays the network as a directed tree, the correlation graph visualizes the correlations among attributes, and the dimension projection graph maps the sensor nodes to a scatterplot. Whisper [14] used a sunflower metaphor to represent spatiotemporal information diffusion on Twitter in 2012. Thom et al. [15] detected and visualized spatiotemporal anomalies based on geo-located Twitter messages. A cluster analysis approach is used to distinguish between global and local messages. The aggregated messages are visualized as term clouds on a geographic map. Wang et al. [16] developed a visualization system for studying traffic congestion based on trajectory data, which includes a graph list view that combines a series of maps of traffic congestion. van den Elzen et al. [17] extended Massive Sequence Views to analyze the temporal and structural aspects of dynamic networks. But these views cannot reveal the topology of the network very well. In 2014, Itoh et al. [18] proposed a novel visual diffusion method to explore the passenger flow changes in the Tokyo Metro and their spread in time and space. And in 2016, it was further improved [19]. Their approaches enabled us to extract and visualize passenger flows on a complicated metro network from large-scale data of the smart card system and unusual phenomena and their propagation on a spatiotemporal space. However, 3D visualization will cause many issues like occlusion and perception errors, especially in a megacity. Zhao et al. [20] developed #FluxFlow, an interactive visual analysis system for revealing and analyzing anomalous information dissemination in social media. In 2016, Chen et al. [21] presented D-Map (Diffusion Map) that can summarize the historical information diffusion processes initiated by a central user in a social space context considering the communities of the influenced users. Tao et al. [22] proposed a visual analysis framework based on high-order correlation graphs to detect, analyze, and reason collective anomalies in 2018. The views provide better interactivity compared with the previous correlation graph. Users can use this framework to get multiple levels of relationships between multiple types of objects and events.

For the metro network, the structure is naturally a graph. However, its characteristics are completely different from

road traffic and other graphs mentioned above. So the above visual analysis method is difficult to be applied to urban rail transit directly. A visualization method for the metro is needed, which not only shows the details of the anomaly in combination with geospatial space but also describes the propagation of the anomaly in the rail transit network topology.

3. Overview

The workflow of the visual analytic method introduced in this paper is shown in Figure 1. The entire method mainly consists of two modules: data preprocessing and visualization. The first module carries out data processing and analysis of card data, mainly including data preprocessing, passenger flow statistics, and diffusion analysis. The second module is the visualization part, which is mainly divided into detection visualization of abnormal passenger flow, verification visualization of abnormal passenger flow, and diffusion visualization of abnormal passenger flow.

In the first module, the Beijing Rail Transit is taken as an example to preprocess and preliminary analysis the Beijing smart card swiping data in the third quarter of 2015. The data come from the Beijing Traffic Information Center covering 11 travel routes in Beijing, 17 operating lines, and 270 stations for passenger travel records. Each record has a total of 37 attributes, including user card number, entry and exit time, entry and exit lines and stations, etc. The relevant records of passenger entry and exit station time and location information are very complete. The raw data have a lot of redundant information. After deleting the useless information and merging the pedestrian card travel record, the final pedestrian travel record data is obtained. The data mainly include five attributes: card number, inbound time, inbound station number, outbound time, and outbound station number. According to the travel records of passengers, the passenger flow of each station at a certain time interval is calculated to support for anomaly detection and analysis of diffusion relationship and to provide data support for the visualization of the metro passenger flow.

The visualization part in the second module mainly implements many functions, as follows:

- (i) The anomaly detection module provides the model's detection results and visually displays the detected anomalies in a hierarchical matrix view.
- (ii) A verification view is designed to verify the suspected abnormal stations. This view displays historical passenger flow information while presenting current passenger flow information.
- (iii) A two-dimensional dynamic diffusion view is proposed, through which the user can visually observe the propagation of anomalies in time and space.
- (iv) Multiple ways to interact with the system are provided. The system supports user-defined

monitoring areas and the interaction to deepen the understanding of abnormal passenger flow diffusion.

4. Design of Visualization System

The visual analytic system is shown in Figure 2. The system can be divided into five parts, which are diffusion view, interactive panel and overview view, matrix view, and verification view.

4.1. Interaction. In this visualization system, the interactive panel has three buttons with functions of selecting the detection area, detecting anomalies, and showing details of anomalies, respectively. When the user clicks the button of selecting the detection area, the system allows the user to create a rectangular area on the map. Then, the system will perform anomaly detection on the stations in the rectangular area. After clicking the button of detecting anomalies, the system will automatically detect the stations with abnormal passenger flow and display the result in the overview view. Users can select the station for anomaly validation in this view. The diffusion relationship of the abnormal passenger flow in all detected stations will be shown to the users by clicking the button. Besides, our system also provides the interactive method between multiple views. For example, users can zoom and drag the dynamic anomaly diffusion view. Time scaling is also supported in the matrix view for users to further observe. Generally, it is believed that interacting with the system can help users better understand anomalies and diffusion relationship.

4.2. Anomaly Detection Module

4.2.1. Abnormal Passenger Flow Detection. The traffic of the station shows obvious periodic changes with time. Therefore, the traffic data matrix essentially satisfies low-rank characteristics. The real traffic data are usually affected by abnormal events, which will cause traffic fluctuations and result in the data not matching the low-rank characteristics. This paper uses the improved RPCA model [23] to decompose the original traffic flow matrix into low-rank traffic flow and abnormal traffic flow, to study the detection and extraction of metro abnormal traffic. Traffic flow data are structured into a matrix $D \in R^{m \times n}$, which the horizontal axis is the date e and the vertical axis is the continuous time interval t , and therefore each element $d_{e,t}$ in the matrix is passenger flow value in a specific time interval of one day. $x_{e,t}$ is the normal passenger flow value in a time interval of one day, and $a_{e,t}$ is the abnormal passenger flow value. The value measured in a certain period of time interval is expressed as $d_{e,t} = x_{e,t} + a_{e,t}$, so traffic matrix can be decomposed as

$$D = X + A \quad (1)$$

The daily metro passenger flow shows a relatively fixed change rule, and the flow tends to be consistent with that of the same period of the week. Therefore, the time periodicity

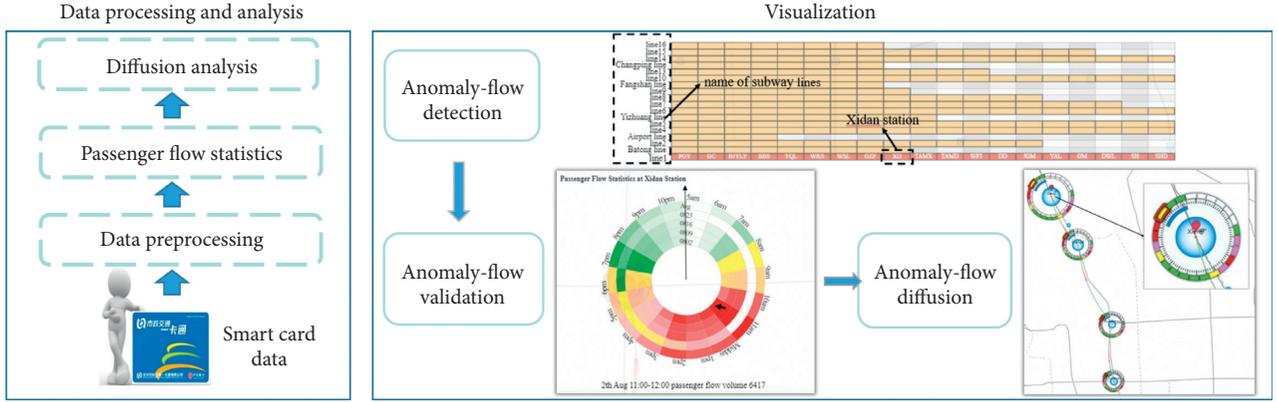


FIGURE 1: The workflow of our framework for analysis of anomaly diffusion.

of the passenger flow makes the matrix X exhibit a low-rank characteristic. Anomalous events are less likely to occur and usually do not last long. In this way, only a small number of data elements in the obtained matrix traffic data D are abnormal, so the decomposed matrix A has sparse characteristics, which is considered as abnormal in this paper. In the traffic flow data matrix, the traffic data at the same time interval on the same day of the weeks are generally approximately equal, so an improved RPCA model is used to constrain the differences among rows in matrix X . Considering that the ι_1 norm is more robustness for mutation detection, the optimal matrices X and A are obtained by the following improved RPCA model and the model is as follows:

$$\begin{aligned} \min_{X,A} \quad & \|X\|_* + \lambda_1 \|A\|_1 + \lambda_2 \|HX\|_1 \\ \text{s.t.} \quad & D = X + A, \end{aligned} \quad (2)$$

where λ_1 and λ_2 control the weight. The above temporal differential matrix $H = \text{Toeplitz}(0, 1, -1)$, in which the central diagonal is defined as ones, and the first upper diagonal is defined as negative ones. This time constraint matrix $\|HX\|_1$ captures the difference value by using the characteristics of the flow values of the same period on the same day of adjacent weeks.

4.2.2. Performance Tests of the Model. To test the performance of the improved RPCA, Wang [23] compared it with the RPCA, the wavelet transform method, and the threshold method. These methods are also used to perform anomaly detection on real passenger flow data and tested performance to validate the model. The data set chosen to verify the robustness of the model is synthetic data constructed by adding noise to real metro traffic data. Extract the inbound or outbound traffic at each station at 1-hour intervals, arrange the weekly passenger traffic into a row vector, and superimpose the 12-week passenger traffic row vector to form a data matrix. A four-layer wavelet filter is used to filter the noise of the original data matrix, and the average value is taken as the ideal data set. An anomaly matrix with a corrupted rate of 10% to 50% is then added to the data set to

form the test data set. Three data sets are selected from the city center to the suburbs, including 20, 18, and 12 stations, respectively, with sizes of 240×126 , 216×126 , and 144×126 . Precision rate pr [24] is used to evaluate the recognition accuracy of anomalies, which is defined as follows:

$$pr = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}, \quad (3)$$

where $\text{precision} = a_{\text{true}}/a_{\text{all}}$, $\text{recall} = a_{\text{true}}/a$, a_{all} and a_{true} denote the number of anomalies recognized by our model and the number of true anomalies among them, respectively, and a represents the actual number of anomalies. Take the average of the accuracy rate of these stations as the experimental result, as shown in Figure 3. It can be seen from the figure that the improved RPCA method is significantly better than the other three methods as the data corrupted rate increases. And the improved RPCA is more robust than the RPCA at the same corrupted rate. So, the improved RPCA method is chosen as the anomaly detection model.

4.2.3. Abnormal Passenger Flow Visualization. The improved RPCA model can be used for anomaly detection to obtain the information of abnormal stations in the whole orbital system, including the time and intensity of abnormal occurrence. This paper designs a diffusion view that combines abnormal stations with geographic information. Abnormal stations detected by the model will be marked by red jumping points, as shown in Figure 2, which coincide with the geographical location of stations on the map. The jumping red points also help users distinguish the abnormal station and affected stations. This view also provides a flexible drag and map zooming.

In order to master the abnormal passenger flow of all stations on the map as a whole, an overview view is also designed in this paper, as shown in Figure 4. In this view, the horizontal coordinate represents the stations, the vertical coordinate represents the Beijing rail transit lines, the yellow grid represents the normal station, and the red grid represents the abnormal station. When users click on the red grid in this view, it interacts with the diffusion view and

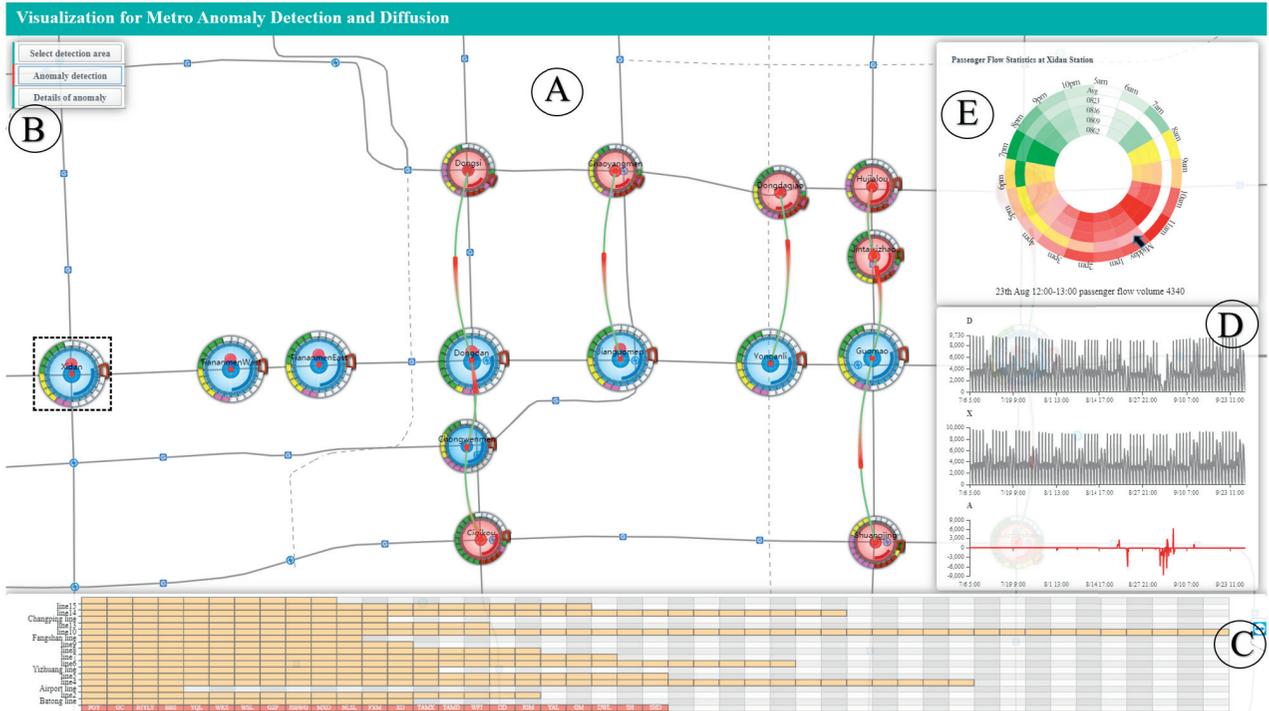


FIGURE 2: System interface. (a) Anomaly diffusion view: display anomalies and demonstrate the spatiotemporal law of anomalies. (b) Interaction panel: provide some interactive functionality. (c) Overview view: grasp the running status of all stations. (d) Matrix view: show the result of anomaly detection. (e) Verification view: show passenger flow changes on the day compared with historical passenger flow.

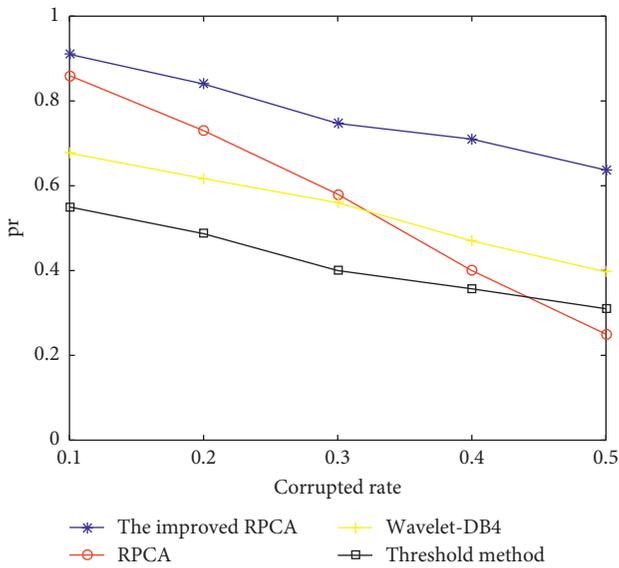


FIGURE 3: Performance comparison of anomaly detection.

locates the abnormal station in this view and focuses the viewer’s perspective on the geographic location of the abnormal station on the map. The purpose of the anomaly overview view is to help the users always monitor the anomalous stations in the orbital system, especially when zooming in or dragging the map with a limited field of view. At the same time, the right side of the system interface shows the matrix view and verification view of the abnormal traffic.

To further demonstrate the results of anomaly detection, the matrix view and the verification view are designed. The matrix view is shown in Figure 2(d), which shows the user the result of the anomaly detection. In this figure, three broken line graphs are used to display the normal passenger flow D , the passenger flow matrix X with low-rank characteristics, and the anomaly matrix A with sparse characteristics. For example, anomaly detection on the passenger flow of Xidan station for 12 consecutive weeks is conducted. The detection results are shown in Figure 5. The outbound traffic D is counted at intervals of every one hour and is decomposed by the improved RPCA method to obtain traffic matrices X and A . In the figure, three broken line graphs are used to represent the three flow matrix changes over time. From top to bottom, it represents the measured data D , the flow matrix X , and the abnormal flow matrix A . The horizontal axis represents a time series of one-hour intervals for twelve consecutive weeks, and the vertical axis represents changes in passenger flow. Observing the measured data D , it can be seen that the outbound traffic of the Xidan station presents a more obvious periodic traffic change on weekdays and weekends. It is expected that the flow matrix X will better recover the desired flow value and the sparse matrix A better captured the anomalies. It is clear to see the four anomalies circled by the ellipse from the figure. Then, users can zoom the view by turning the mouse to see the results in a more accurate time range. As shown in Figure 6, the results of the anomaly detection on August 23 are represented, and the passenger flow detection result is shown between the two dotted lines. Hovering the mouse to a location will show the passenger flow and abnormal flow value at the moment.

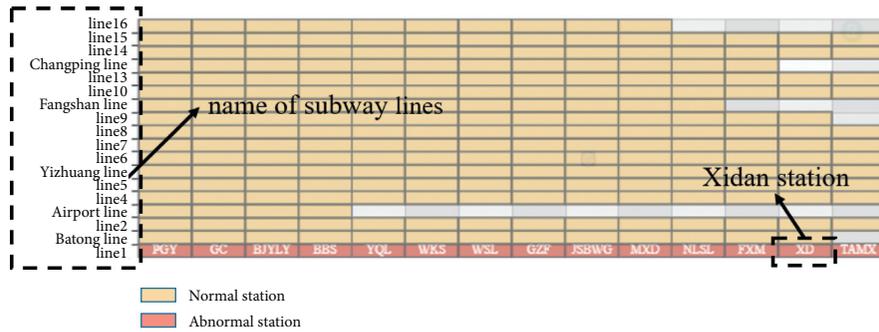


FIGURE 4: The overview of anomalies: this view shows the overview of abnormal stations on the whole metro system at a certain time. The longitudinal coordinates represent the line name, and each grid represents the corresponding station of the line. The normal stations are expressed by yellow, the abnormal stations are expressed by red, and the abbreviations of the station names are displayed.

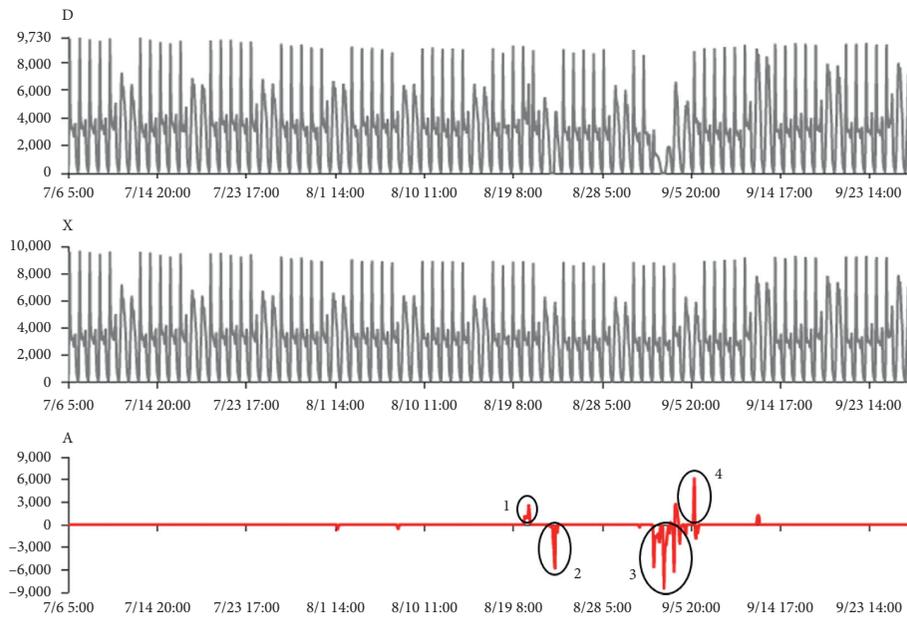


FIGURE 5: Component decomposition of exit flow of Xidan station by the improved RPCA. The horizontal axis is time interval covering 12 weeks, and the vertical axis is traffic value.

To verify the existence of abnormal passenger flow, the abnormal passenger flow of the station is compared with the passenger flow of the last three weeks. The passenger flow is shown in a radial view as shown in Figure 7. The top-left corner of the view shows the name of station which will be verified. Five rings are drawn in the middle of the view. The inner four rings represent one day, and the date represented by the ring gradually increases from the inside to the outside. Date labels are drawn on the circle. The fourth loop visualizes the passenger flow to be verified. The outermost ring represents the average passenger flow value for the past three weeks. Besides, the rings are divided into 18 parts at an hourly interval, representing 5 am to 11 pm. In the figure, the traffic of the station is rendered in five colors of white, green, yellow, pink, and red. By observing the view, users can find that the passenger flow of the same day of every week has the

same change law. If there is an anomaly in the passenger flow to be verified, it is not difficult to find it and intuitively observe it. This view supports mouse hover events. It shows the traffic of the site at the current moment when the mouse is placed on each arc. Figure 7 shows the passenger flow of Xidan station on August 23. As can be seen from the figure, the fourth ring is markedly different from the three inside, indicating an anomaly at the station. In contrast to the other four rings, it can be found that there is no passenger flow from 6 am and until noon. Although the passenger flow in the afternoon has a certain rebound, there are still some differences compared with the historical passenger flow. Hovering over the various locations, we can observe changes in passenger flow over time to verify the anomalies. The position pointed by the arrow in the figure represents the passenger flow from 11:00 to 12:00 on August 2, which is 6417.

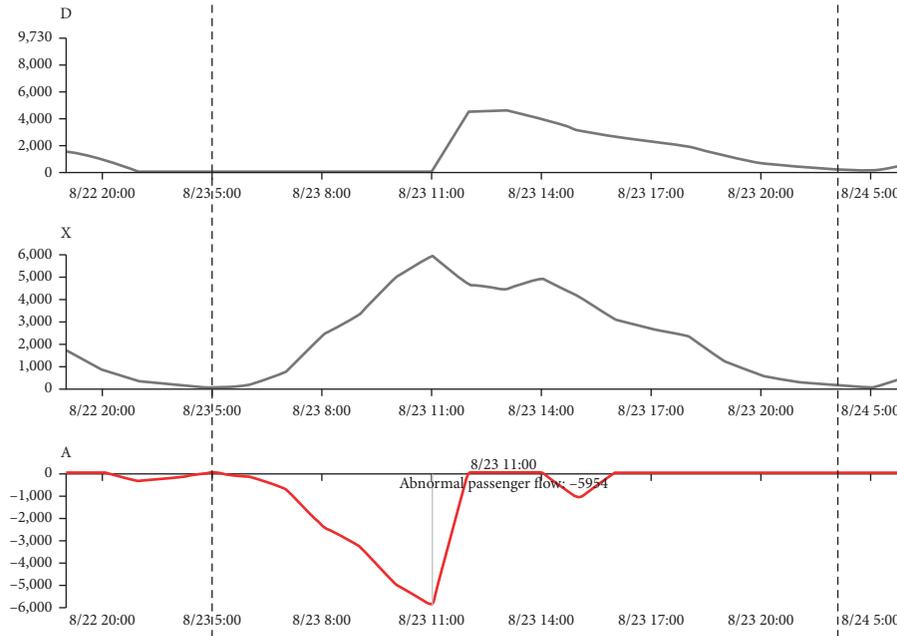
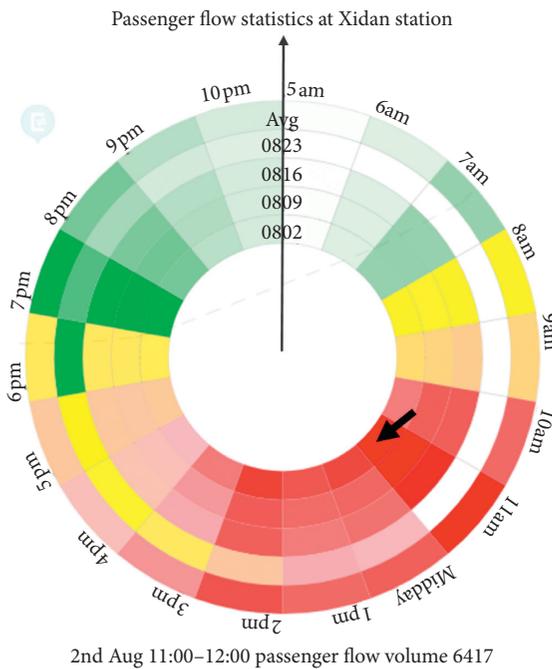


FIGURE 6: The detection results of passenger flow for Xidan station on August 23.



2nd Aug 11:00–12:00 passenger flow volume 6417

FIGURE 7: The verification view: verification of passenger flow anomalies at Xidan station on August 23.

4.3. Anomaly Diffusion Module. This module mainly consists of two parts. The first part explores the spatiotemporal diffusion characteristics of abnormal passenger flow, especially for the diffusion process of abnormal passenger flow in some time and space inside the orbit. The second part visualizes the entire diffusion process.

4.3.1. Spatiotemporal Anomaly Diffusion Analysis. In our method, the time period in which the abnormal passenger flow occurs at the station is defined as the abnormal point. Assuming that the abnormal point of the site s_i is set $(ANOM_{i,t})$, where i is the station number and t is the time point. That is, at time t , the traffic of s_i is detected as anomaly. The definition of diffusion includes the diffusion of space and time. The rules of diffusion are as follows: the point adjacent to s_i in space is $s_m, s_{m'}, s_{m''}$. For s_i , it can be considered that s_i can diffuse to the adjacent points in space at time t and can be spread to $s_m, s_{m'}, s_{m''}$ at time $t + 1$. Similarly, s_i at time $t + 1$ can be simultaneously affected. $s_m, s_{m'}, s_{m''}$ can also be spread to $s_i, s_m, s_{m'}, s_{m''}$ at time $t + 2$, and so on. The spatiotemporal anomaly diffusion graph is shown in Figure 8.

In this paper, the spatiotemporal diffusion algorithm is used to detect the diffusion behavior of anomalous points. The basic idea of the algorithm is as follows: First, assign a flag variable to each abnormal point $ANOM_{i,t}$, which is initially 0, indicating that it is not accessed and traverses the abnormal point when the abnormal period is n . For each spatiotemporal anomaly point, the algorithm finds other spatiotemporal anomalies that are directly or indirectly connected to the node in space and then detects whether there are abnormal points connected to these abnormal points in the $n + 1$ period. If there are, continue the spatiotemporal diffusion analysis according to the above rules; if there are not, output the set of abnormal points. The algorithm steps are shown in Algorithm 1. The spatiotemporal diffusion algorithm is applied to all the anomalous points obtained by the improved RPCA model to study the diffusion pattern inside the abnormal region.

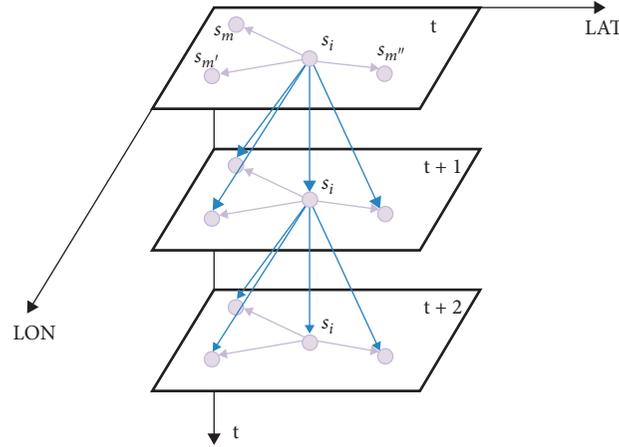


FIGURE 8: Spatiotemporal anomaly diffusion graph.

```

Input: the abnormal points  $ANOM_{i,t}$  for all stations.
Output: the diffusion relationship for all stations.
(1) for each  $i \in [1, N]$  do
(2)   for each  $a_i \in ANOM_{i,t}$  do
(3)     if  $a_i.flag = 1$  then?
(4)        $a_i$  is inserted into templist
(5)        $a_i.flag = 1$ 
(6)     while templist != null do
(7)       delete first element  $ANOM_{m,n}$  in the templist
(8)       insert  $ANOM_{m,n}$  into anomalyest
(9)       for  $ANOM_{R_m, s_j, n}$  do
(10)        if  $ANOM_{R_m, s_j, n} \neq 0$  then
(11)           $ANOM_{R_m, s_j, n}$  is inserted into templist
(12)           $ANOM_{R_m, s_j, n}.flag = 1$ 
(13)        end
(14)        for set ( $ANOM_{R_m, s_j, n+1}$ ) do
(15)          if  $ANOM_{R_m, s_j, n+1} \neq 0$  then
(16)             $ANOM_{R_m, s_j, n+1}$  is inserted into templist
(17)             $ANOM_{R_m, s_j, n+1}.flag = 1$ 
(18)          end
(19)        end
(20)       if  $len(anomalyest) \geq 2$  then
(21)         return anomalyest
(22)     end
(23)   end

```

ALGORITHM 1: Spatiotemporal anomaly diffusion algorithm for our system.

4.3.2. Abnormal Passenger Flow Diffusion Visualization.

In Section 4.2.1, the improved RPCA model is used to obtain information about the abnormal passenger flow, and the anomaly diffusion model in Section 4.3.1 is used to obtain the diffusion relationship of the abnormal passenger flow in the orbit. After obtaining the diffusion relationship between the abnormal stations, the diffusion relationship is visually displayed in the two-dimensional space. According to the characteristics of metro data, a dynamic anomaly diffusion view is designed as shown in Figure 9. This view shows detailed anomalous diffusion information including the direction of the anomalous propagation, the time when the anomalous passenger flow arrives at each site, the duration of anomalous

passenger flow at each station, the strength of the anomaly, and the change of the anomalous passenger flow at each station.

This view relies on the geographic information of the Beijing map and can well demonstrate the diffusion relationship among the anomalous stations in the orbital system. In our visual coding, the source of the anomaly is represented by a red jumping point. Centered on the source of anomalies, there will be a dynamic diffusion circle, which will continue to shrink and expand. When the diffusion circle shrinks to a minimum, it becomes a point. When it expands to the maximum, it spreads to the outer circle. The frequency of contraction and expansion is determined by the intensity of the anomaly. The greater the abnormal intensity

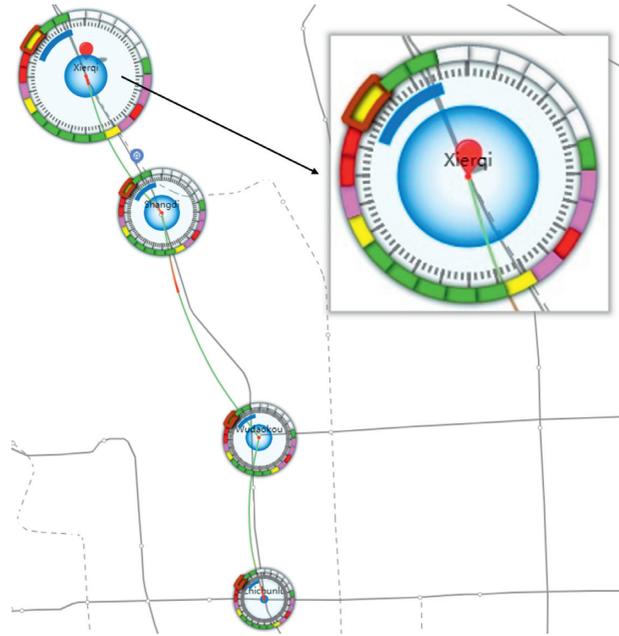


FIGURE 9: The diffusion view: August 13th, 2015, the anomaly spread of Xierqi station.

is, the greater the frequency is. If the diffusion circle is red, the abnormal situation is the sudden increase of the flow, and blue indicates the sudden decrease of the flow.

The biggest difficulty in the diffusion visualization of anomalous passenger flow is the reasonable display of the time of anomalous propagation. The time scale is drawn on the circle, which represents 24 hours a day. It is divided into four blocks at an interval of fifteen minutes per hour. At the same time, coloring the rings divided into 24 blocks to indicate different passenger flow levels, which is red, pink, yellow, green, and white, which in turn means that the traffic is gradually decreasing, and white means that the traffic is zero. The closer the value of the passenger flow to 0 is, the closer the color is to white. The block that represents the start time of abnormal propagation on the ring will be highlighted. The passenger flow has a different performance and varies within a certain range due to different stations. So when coloring passenger flow, this view only targets the station that is currently being colored. In the way, the same red at different stations in the diffusion view represents a different passenger flow value, but the running status of each station can be observed. Besides, the arc corresponding to the time scale on the ring is drawn inside. The color of the arc is consistent with the increase or decrease of passenger flow. By observing this arc, the duration of the anomaly can be obtained. In this way, the stations affected by the abnormal passenger flow can be obtained and the influence relationship among each abnormal station can be understood in the anomaly diffusion view. The intensity of anomalies, the time of abnormal propagation, and the influence of abnormal propagation to the station on its passenger flow can be grasped.

5. Case Study

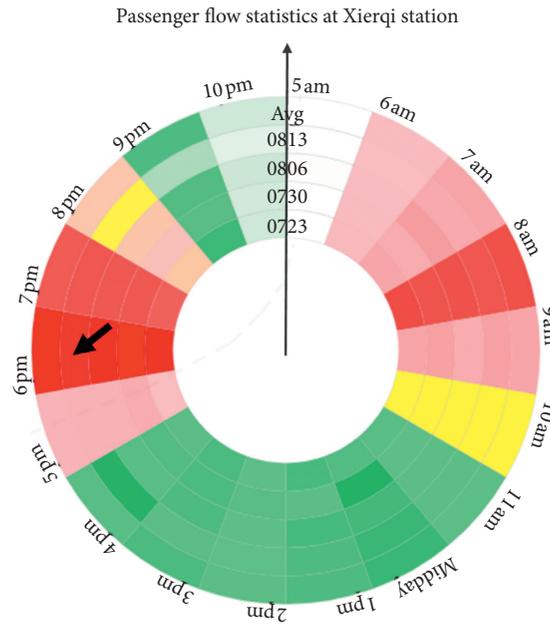
The system presented in this paper provides some interactive views that show the users the detection and diffusion of abnormal passenger flows that occur in the

metro network. Here, two cases are provided to study the abnormal events (emergency, traffic control), which visually analyze the anomalous diffusion of the Xierqi station and the impact of the Beijing Metro Line 1 sealing station event.

5.1. Visual Analysis of Abnormal Diffusion in Xierqi Station.

At about 8:00 pm on August 13, 2015, the system detects that there may be a passenger flow anomaly in Xierqi station. The cell representing the Xierqi station in the overview view is selected, and the anomaly is verified in the verification view. As shown in Figure 10, the traffic volume of the Xierqi station around 8:00 pm is significantly lower than the historical traffic on the same day in the past three weeks. The same trend of passenger flow on the same day in the past three weeks is consistent, and at this moment, passenger flow has almost reached its peak. This moment belongs to the evening peak period. For such a sudden drop in passenger flow, it can be considered to be an anomaly. By checking the records of the relevant departments of the metro, it is found that the Xierqi station has a signal failure on the same day. The train travel time interval has become larger from Xierqi station to Shangdi station (direction to Xizhimen), resulting in a decrease in passenger flow at this station. Based on this, the impact of anomalies on surrounding stations is analyzed in a dynamic anomaly diffusion view.

Here, the abnormal outbound passenger flow at the Xierqi station (from Xierqi to Xizhimen) is analyzed for diffusion. As can be seen from the anomaly diffusion view shown in Figure 9, the diffusion circle of the Xierqi station is blue from 8:00 to 9:00 pm, indicating that the passenger traffic at the station has plummeted. The passenger flow of the Xierqi station changed from red to yellow during this period and continues to go down. From 8:00 to 9:00 pm, the abnormal traffic is transmitted to Shangdi, Wudaokou, and



13th Aug 18:00–19:00 passenger flow volume 15817

FIGURE 10: The changes of passenger flow at Xierqi station on August 13, 2015.

Zhichun Road in turn, causing the traffic of these three stations to drop. The duration of anomaly diffusion at each station also decreases successively. The abnormal propagation to Zhichun Road stopped, and the impact on the subsequent stations is almost negligible.

5.2. Visual Analysis of the Influence of the Closure of Beijing Metro Line 1. Case 2 analyzes the anomalies in the Beijing Rail System on August 23, 2015. First, the anomaly detection module in the visualization system detected an abnormal passenger flow event at all stations on the entire line 1 at around 5 am as shown in Figure 11. For such a large-scale anomaly, the rationality of the detection results is verified, first. For the abnormal stations which users are interested in, anomaly verification is performed in the matrix view and verification view. As shown in Figure 5, the Xidan station on Line 1 is taken as an example. The figure shows the detection result of passenger flow at the station over time. With the mouse zoom event, the time of the matrix view is frozen on August 23. In the row of the anomaly matrix, it can be found that there is an obvious abnormal passenger flow scene at Xidan station on August 23. The detailed results are shown in Figure 6. From five to noon, the traffic volume is almost 0. It can be also found from Figure 7 that the passenger traffic on August 23 is significantly different from the historical period. There is no doubt that the abnormal passenger flow event does exist and the same result can be seen by selecting other stations. For this special case, only the sealing station or the train failure may be caused. After consulting the records of the relevant departments of the metro, it is proved that the stations of Line 1 and some stations of Line 2 cease operation in the morning. The closure of Line 1 lasted until

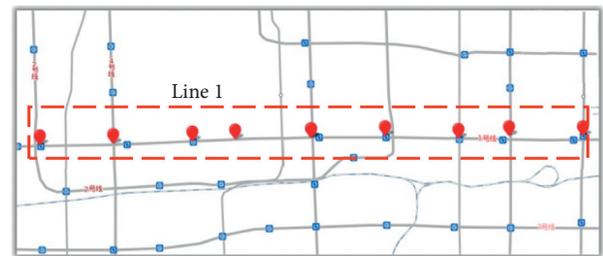


FIGURE 11: Anomalies on Line 1 at 5:00 on August 23, 2015.

noon. Although the abnormal cause of the abnormal flow of the station has been known, to understand the impact of the sealing station event on the entire metro system, the dynamic diffusion view can be observed for further analysis.

From the dynamic anomaly diffusion view shown in Figure 12, it can be seen that the blocking event of Line 1 causes a large number of traffic spikes at the stations with which it is connected. The blocking time occurs around 5 am, and the anomaly begins to spread to some nearby stations around five o'clock. The blue diffusion circle is the diffusion source, and the red diffusion circle is the affected stations. To allow the users to view the details of the diffusion map in more detail, only a part of the entire diffusion view is intercepted. In the view of the red dotted circle, the anomaly gradually spreads from the Guomao to Jintaixizhao and Hujialou, and from the Dawang Road to the Jintai Road, and caused an increase in the traffic of these stations. The passenger flow directly connected with the abnormal station immediately responds. From Figure 12, it can be seen that the passenger flow of Hujialou, Jintai Xizhao, and Jintai

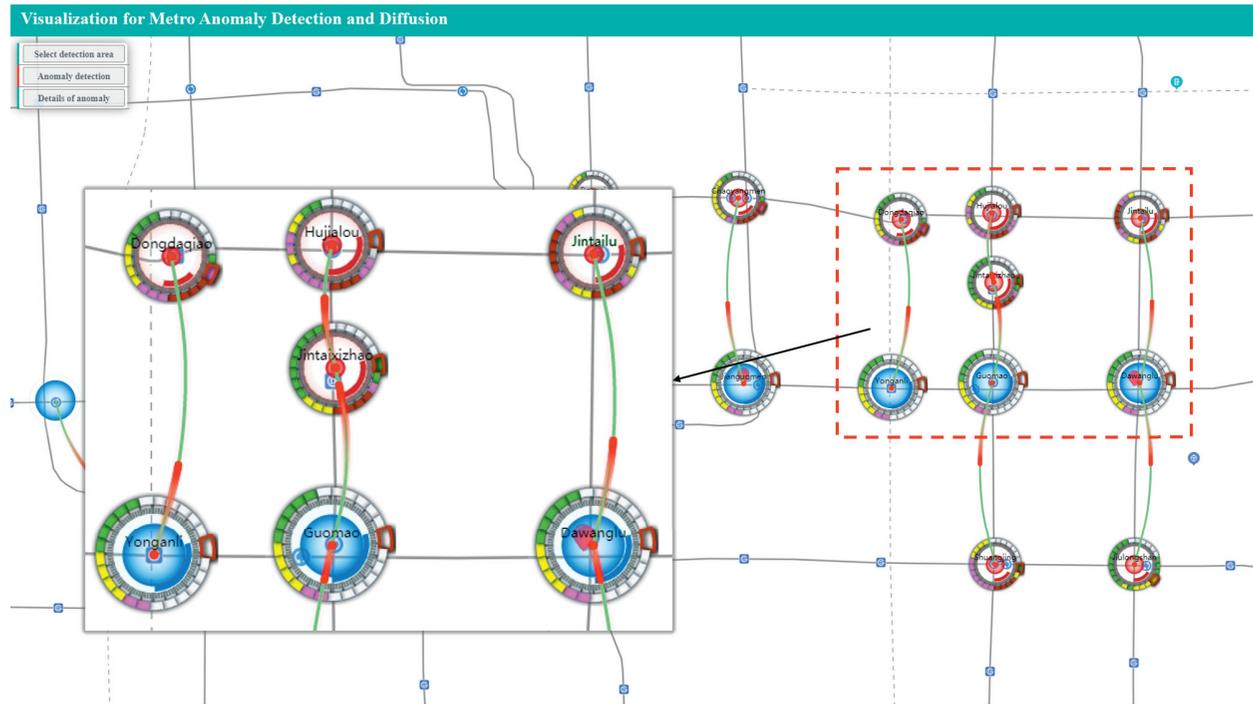


FIGURE 12: Anomalies of Line 1 caused anomalous diffusion of passenger flow at 5:00 on August 23, 2015.

Road has increased to varying degrees from 5 am to 12 pm. The passenger flow at Dongdaqiao station has changed slowly. Besides, the period when the passenger flow is abnormal is 7:00 to 12:00. Differences in the duration of anomalies can be observed from the arc. The impact of the entire Line 1 station closure and the flow changes at the affected stations can be understood in the diffusion view.

6. Conclusion and Future Work

In this paper, an improved RPCA model is introduced for the detection of abnormal passenger flow. In addition, a visual analysis method of abnormal passenger flow is proposed for the visualization of abnormal passenger flow detection, verification, and diffusion. The AFC card data from the Beijing metro system is used to verify our method. The results show that we provide an intuitive and informative visualization tool for the rail transit department to monitor and manage passenger flow.

However, there is still room for further improvement. Further research can be conducted in the following two aspects: Our detection model only performs anomaly detection of a single station at a time. Tensor-based detection model can be used to detect multiple stations at the same time. On the other hand, the spatiotemporal diffusion algorithm can be further studied to make it more suitable for rail transit.

Data Availability

Raw data used for the research are available from the first author upon request by e-mail (s201739060@emails.bjut.edu.cn).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant nos. U1811463, 61602486, and 61876012 and in part by the Beijing Municipal Science and Technology Project under Grant no. Z171100004417023.

References

- [1] G. Andrienko, N. Andrienko, W. Chen, R. Maciejewski, and Y. Zhao, "Visual analytics of mobility and transportation: state of the art and further research directions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 2232–2249, 2017.
- [2] G. Andrienko and N. Andrienko, "Spatio-temporal aggregation for visual analysis of movements," in *Proceedings of the 2008 IEEE Symposium on Visual Analytics Science and Technology*, pp. 51–58, IEEE, Columbus, OH, USA, October 2008.
- [3] A. Slingsby, J. Wood, and J. Dykes, "Treemap cartography for showing spatial and temporal traffic patterns," *Journal of Maps*, vol. 6, no. 1, pp. 135–146, 2010.
- [4] Z. Wang, T. Ye, M. Lu et al., "Visual exploration of sparse traffic trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1813–1822, 2014.
- [5] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan, "Tripvista: triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection," in

- Proceedings of the 2011 IEEE Pacific Visualization Symposium*, pp. 163–170, IEEE, Hong Kong, China, March 2011.
- [6] R. Krüger, D. Thom, M. Wörner, H. Bosch, and T. Ertl, “Trajectorylenses—a set-based filtering and exploration technique for long-term trajectory data,” *Computer Graphics Forum*, vol. 32, no. 3pt4, pp. 451–460, 2013.
- [7] W. Zeng, C.-W. Fu, S. M. Arisona, A. Erath, and H. Qu, “Visualizing mobility of public transportation system,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1833–1842, 2014.
- [8] S. Chen, X. Yuan, Z. Wang et al., “Interactive visual discovering of movement patterns from sparsely sampled geotagged social media data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 270–279, 2016.
- [9] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, “Stacking-based visualization of trajectory attribute data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2565–2574, 2012.
- [10] G. Andrienko, N. Andrienko, G. Fuchs, and J. Wood, “Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 9, pp. 2120–2136, 2017.
- [11] B. Alper, N. Riche, G. Ramos, and M. Czerwinski, “Design study of linesets, a novel set visualization technique,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2259–2267, 2011.
- [12] C. Collins, G. Penn, and S. Carpendale, “Bubble sets: revealing set relations with isocontours over existing visualizations,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1009–1016, 2009.
- [13] L. Shi, Q. Liao, Y. He, R. Li, A. Striegel, and Z. Su, “Save: sensor anomaly visualization engine,” in *Proceedings of the 2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 201–210, IEEE, October 2011.
- [14] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu, “Whisper: Tracing the spatiotemporal process of information diffusion in real time,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2649–2658, 2012.
- [15] D. Thom, H. Bosch, S. Koch, M. Wörner, and T. Ertl, “Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages,” in *Proceedings of the 2012 IEEE Pacific Visualization Symposium*, pp. 41–48, IEEE, Songdo, South Korea, March 2012.
- [16] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, “Visual traffic jam analysis based on trajectory data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2159–2168, 2013.
- [17] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk, “Reordering massive sequence views: enabling temporal and structural analysis of dynamic networks,” in *Proceedings of the 2013 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 33–40, IEEE, Sydney, Australia, March 2013.
- [18] M. Itoh, D. Yokoyama, M. Toyoda, Y. Tomita, S. Kawamura, and M. Kitsuregawa, “Visual fusion of mega-city big data: an application to traffic and tweets data analysis of metro passengers,” in *Proceedings of the 2014 IEEE International Conference on Big Data (Big Data)*, pp. 431–440, IEEE, Washington, DC, USA, October 2014.
- [19] M. Itoh, D. Yokoyama, M. Toyoda, Y. Tomita, S. Kawamura, and M. Kitsuregawa, “Visual exploration of changes in passenger flows and tweets on mega-city metro network,” *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 85–99, 2016.
- [20] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins, “#Fluxflow: visual analysis of anomalous information spreading on social media,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1773–1782, 2014.
- [21] S. Chen, S. Chen, Z. Wang et al., “D-map: visual analysis of ego-centric information diffusion patterns in social media,” in *Proceedings of the 2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 41–50, IEEE, Baltimore, MD, USA, October 2016.
- [22] J. Tao, L. Shi, Z. Zhuang et al., “Visual analysis of collective anomalies through high-order correlation graph,” in *Proceedings of the 2018 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 150–159, IEEE, Kobe, Japan, April 2018.
- [23] X. Wang, Y. Zhang, H. Liu, Y. Wang, L. Wang, and B. Yin, “An improved robust principal component analysis model for anomalies detection of subway passenger flow,” *Journal of Advanced Transportation*, vol. 2018, Article ID 7191549, 12 pages, 2018.
- [24] W. Ye, L. Chen, G. Yang, H. Dai, and F. Xiao, “Anomaly-tolerant traffic matrix estimation via prior information guided matrix completion,” *IEEE Access*, vol. 5, no. 99, pp. 3172–3182, 2017.