Research Article

Wireless Sensor Network-Based Mathematical Modeling for Communication Intelligent Monitoring System

Ying Shen [ID]

School of Mathematics and Computer Application Technology, Jining University, Shandong, Jining 273155, China

Correspondence should be addressed to Ying Shen; shenying@jnxy.edu.cn

Knowing the behavioral patterns of city residents is of great value in formulating and adjusting urban planning strategies, such as urban road planning, urban commercial development, and urban pedestrian flow control. Based on the high penetration rate of cell phones, it is possible to indirectly understand the behavior of city residents based on the call records of users. However, the behavioral patterns of large-scale users over a long period of time can present characteristics such as large dispersion, difficult to discover patterns, and difficult to explain behavioral patterns. In this paper, we design and implement a human behavior pattern analysis system based on massive mobile communication data based on serial data modeling method and visual analysis technology. For the problem that it is difficult to capture the behavioral patterns of residents in cities in call records, this paper constructs base station trajectories based on users' cell phone call records and uses users' long-time base station trajectories to mine users' potential behavioral patterns. Since users with similar activity characteristics will exhibit similar base station trajectories, this paper focuses on the similarity between text sequences and base station trajectory sequences and combines the word embedding method in natural language processing to build a Cell2vec model to identify the semantics of base stations in cities. In order to obtain the group behavior patterns of users from the base station trajectories of group users, a user clustering method based on users' regional mobile preferences is proposed, and the results are projected using the Stochastic Neighbor Embedding (t-SNE) algorithm to expose the clustering features of large-scale cell phone users in the low-dimensional space. To address the problem that user behavior patterns are difficult to interpret, a visual analysis model with group as well as regional semantics is designed for the spatial and temporal characteristics of user behavior. Among them, the clustering model uses the distance between scatter points to map the similarity between users, which helps analysts to explore the behavioral characteristics of group users.

1. Introduction

With advances in science and technology and computer hardware, today's computer systems are capable of storing massive amounts of data. Researchers at Berkeley University estimate that the world generates about one million terabytes of data per year, mostly stored in digital form. In the social sector, the average monthly volume of data exchanged on mainstream social media platforms worldwide has reached 2.072 billion since September 2020, and in the communication sector, data exchange between humans has become easier and faster as modern communication standards continue to be upgraded, with more than 50 million

communication data in Beijing in August 2020, according to data from communication carriers. According to the International Data Corporation, by 2021, the amount of data produced and copied globally will increase to approximately 44 trillion GB, and the world has entered the era of "big data" [1]. In recent years, as urbanization accelerates worldwide, people around the world are constantly generating, disseminating, and exchanging massive amounts of high-dimensional heterogeneous data, from human social data involving microblogs, WeChat, and QQ to communication data generated by cell phone calls. Human behavior data has become one of the most common types of data used in urban visual analytics. The study of human behavior has

been an enduring topic in various fields. For business, the study of crowd flow patterns facilitates the development and adjustment of business marketing strategies, such as the accurate placement of advertisements and the precise location of businesses, thus promoting business profitability [2]. For urban management, understanding the gathering and movement patterns of crowds can help urban planners to formulate urban planning strategies and make timely road risk assessments and urban emergency predictions to adjust urban traffic strategies and urban emergency plans [3].

Based on different data sources, human behavior data can be divided into traffic data, commuting data, mobile communication data, geotagged data, and social media data. Among them, mobile communication data refers to the data records between cell phones and the base stations built by telecommunication operators. With the rapid development of communication technology, the way of human communication has been changing, from the initial transmission of letters to fixed-line communication to today's mobile communication devices. Cell phone has become an essential item in people's life and social life, and as a sensor of human activities, cell phone can reflect the activities of urban residents [4]. With the dense coverage of base stations, a large amount of spatiotemporal data related to user socialization and behavior is generated when people use mobile devices to make calls, such as the geographic location of nearby base stations when users talk, the length of user calls, and who users talk to. This provides researchers with an unprecedented information resource to study human mobility and granular solutions for urban populations [5]. However, how to find the valuable information hidden in the huge amount of heterogeneous data is also a pressing problem. Daily data is usually recorded automatically by sensors and monitoring systems, even for simple daily actions such as making phone calls and forwarding messages. Typically, computers record multiple parameters to ensure the accuracy of the data, which in turn generates a large amount of high-dimensional data. The traditional method of displaying data in text only keeps the amount of data within 100 items, but it is only a drop in the ocean for dealing with data containing millions of data items [6]. The more widely used data mining is more suitable for data sets with a priori models and certain regularity, but crowd behavior data are often dynamic, diverse, multidimensional, and uncertain, and the patterns and mechanisms are difficult to be discovered, so building a priori models is a great challenge for researchers. Traditional pattern recognition methods are effective in discovering patterns in small-scale data, but they are powerless for large-scale data, which on the one hand involves very large entities and on the other hand may span a long period of time. During this period, individual patterns may be found to change, while different entities have their own unique patterns. In order to effectively discover potential patterns from large-scale data, visual analytics has emerged, which can help automated methods to discover useful information more effectively with the help of human experience and knowledge, and on the other hand, visual analytics can also improve the interpretability of patterns. The current academic definition of visual analytics is "a method that combines automated analysis with interactive visual analytics techniques to enable effective information mining, reasoning, and decision making on extremely large and complex data sets [7]." In the years of rapid development of the visual analytics discipline, visual analytics has become an integrated approach that combines visual design, data science, human-computer interaction, and analytical reasoning [8–10].

Visual analytics places a higher priority on analyzing data and discovering information than visualization, integrating human cognitive and perceptual abilities into the data analysis process rather than just presenting data through recognizable graphic images. The value behind the data is uncovered by combining the strengths of humans and computers, as well as interaction analysis methods and interaction techniques. In this paper, user behavior is defined as a sequence of behavioral events that are linked together. For example, a sequence of user behaviors is "arrive at work from home at 9 am," "go out for lunch at 12 noon," and "return home from work at 6 pm," where "go to work from home at 9 am" is a behavioral event. If a person exhibits this behavior on weekdays, the behavior sequence is the behavior pattern of the user on weekdays. In order to explore the crowd behavior patterns and integrate visual analysis, clustering algorithm, and natural language processing, this paper analyzes the semantic recognition of urban areas, crowd trajectory mining, and crowd behavior display under specific conditions in both time and space dimensions; designs and implements a visual analysis system for urban crowd behavior patterns; and visualizes urban crowd behavior display through time series, heat map, cluster map, and other visual graphics. We design and implement a visual analysis system for urban crowd behavioral patterns and display urban crowd behavioral patterns through time series, heat map, cluster map, and other visual graphics, which solves the problems of highly dynamic, high-dimensional, and massive data, and helps users understand potential semantic information and discover more user behavioral patterns from the analysis results.

2. Related Work

From the early days of analyzing the association between patients and water sources by labeling cholera cases in the population to the present-day human behavior analysis based on spatiotemporal data to understand population flow patterns and explore human activity patterns, regardless of the time period, human behavior analysis has been a popular research topic in China and abroad. At present, most of the research is focused on human travel data, such as cab data and public transportation data to solve the urban traffic congestion problem, but there are relatively few analyses and studies on human behavior [11]. At the same time, with the improvement of spatiotemporal data accessibility and analyzability, visual analysis based on spatiotemporal data has taken shape, and the mobile communication data used in this paper is also a kind of spatiotemporal data. Human behavior analysis is an enduring topic that has been a hot topic for researchers at home and abroad, and with the rapid

development and continuous innovation of big data technology, the feasibility of crowd behavior analysis under large-scale data has gradually increased, and researchers have begun to focus their attention on it. With the continuous development of perception technology, the availability of human movement data has gradually increased, and the analysis of human movement patterns has become the focus of research [12].

The researchers investigated human activities from a new perspective by extracting individual movement trajectories from cell phone data and delineating city boundaries through the spatial distribution of base stations. The results showed that the distribution of human trajectories within cities showed an exponential pattern in general, with different exponents for different cities, and finally, the relationship between human mobility and urban morphology within cities was further validated through Monte Carlo simulation analysis. The researchers propose an innovative data mining framework that can identify macroscopic movement patterns of people from the initial call records of cell phone users, call detail records, i.e., call detail call data, and help planners obtain effective urban information from big data by developing a data mining pipeline to quantify the spatial distribution of residents' travel patterns in different areas of the city and facilitate more precise improvement of the city's infrastructure [13]. In addition, the application of human movement pattern analysis is becoming more and more widespread. Based on the GPS trajectory data of cabs and point of interest data, researchers have proposed a DRoF framework based on the trajectory data as a crowd movement, which makes a functional division of urban areas and contributes greatly to the planning of urban functional areas. Since late January 2020, a novel coronavirus outbreak has swept through more than 180 countries and regions worldwide. As a global emergency response, governments have taken various epidemic prevention measures to control the rapid spread of the epidemic [14]. Researchers have designed and implemented the EpiMob interactive visual analytics system based on a large amount of human movement data and urban POI data. The system is based on city restriction policies for epidemics and simulates human mobility and changes in the number of infected people in an interactive manner, allowing users to develop spatial and temporal scales based on different mobility restriction policies and dynamically displaying epidemic infections under different policies [15].

With the rise and growth of social media, location-based check-in services in various social media applications enable users to share activity-related information, thus providing a new source of human activity data. Based on large-scale geographic data in social media, researchers used a data-driven modeling approach based on topic modeling to classify user activity patterns to infer individual social patterns. Researchers modeled human activity patterns through relevant semantics based on Twitter data from Toronto and showed that the number of users with similar activity patterns decreased by 57.13% after incorporating regional semantic information into the corresponding spatiotemporal patterns, providing quantitative evidence that similar

movement patterns may have different motivations for activity [16]. At the application level, researchers used complex networks to model and analyze the call relationship network, given a graph $G(V, E)$, where users are considered as vertices V of G, relationships between users as edges E of G, and frequency of calls between users as relationship weights, and partition the user group based on the association partitioning algorithm in complex networks, using association members as the basic analysis objective to identify and analyze users' behavior patterns, and the users' behavioral patterns and social relationships among users are identified and explored by using association members as the basic analysis target [17, 18]. Based on a similar idea, they also serialize users' base station trajectories and determine the real relationship between users by calculating the similarity between the trajectory sequences of two users and the frequency of calls [19]. There are also many research works on user behavior pattern recognition based on sparse trajectory correlation; for example, researchers proposed a progressive visual analysis system for analyzing cell phone users' movement patterns based on public social media data with geolocation tags and combined spatial and textual information extracted from the data to analyze the semantic information of users' movement patterns [20].

With the continuous development of mobile communication technologies, existing wireless networks have been able to sense human activities and generate a large amount of data related to human activities, including mobile communication data [18, 21, 22]. Currently, this type of data is the main source of information for conducting human mobility and behavior studies and is mainly collected and stored by base stations built by telecommunication operators. Researchers [23] proposed a sequential model based on call data to capture correlations between geographic regions based on the presence of significant covariances and applied Venn diagrams and trajectory maps for visual presentation in maps, aiming to analyze the interconnections between different regions due to high temporal variations in local population density.

3. Crowd Behavior Patterns Based on Wireless Sensing Communication Intelligence

3.1. Wireless Data Preprocessing Techniques. With the rapid development of perception technology, the size of data is getting bigger and the data sources are getting more complex, leading to an increase in the possibility of abnormal data in the data. High-quality data leads to better predictions and models; therefore, data preprocessing has become crucial and also a fundamental step in the process of data science, machine learning, and artificial intelligence. In this section, the various steps of data preprocessing and the methods will be discussed. When performing data collection, there are three factors that usually affect the quality of data. The more the feature variables of the data, the more difficult it is to visualize the training set. Most of the feature variables contained in real-world data are correlated with each other, making it very easy to create data redundancy. The purpose of dimensionality reduction algorithms is to compress the

dimensionality of the data with maximum protection of the data information.

3.1.1. Accuracy.
Erroneous values may appear in the data that deviate from expectations. The main causes of data inaccuracy include human errors during data entry and transmission, computer errors, and memory variations triggered by natural factors. For example, users submit incorrect data, incorrect formatting of input fields, duplicate data, etc.

3.1.2. Completeness.
The data is missing attribute values or functional values. The main reasons for this include data unavailability, removal of inconsistent data, and removal of data that was initially considered irrelevant.

3.1.3. Consistency.
There are inconsistent data formats, inconsistent data attributes, and inconsistent data aggregation. In addition, factors such as timeliness, trustworthiness, and interpretability also affect data quality, so data preprocessing is crucial to ensure high-quality data. Generally speaking, data preprocessing is divided into three stages: data cleaning, data integration, and data streamlining.

Data cleaning usually "cleans up" data by removing anomalous attributes, adding missing attributes, smoothing out noise in the data, and correcting inconsistencies between data and predefined attributes. After integrating or concatenating multiple data sets, the final data set is prone to dirty data, including data with missing values, data containing errors, and nonstandard data. Usually caused by data entry errors, data update errors, and data transfer errors, the data must be cleaned to remove or repair dirty data before formal data processing, as shown in Figure 1.

Missing values are usually caused during data collection or as a result of changes in some data validation rules. The current field of data analysis deals with missing values in two main ways: one is to directly eliminate the rows of missing data while the other is to fill in the missing values. The former is simpler and more efficient, but it does not have the advantage for handling large amount of data. The latter is relatively stable for missing values and is usually filled by interpolation or by using the mean, median, or plurality of the corresponding features. The most common problematic data in the data are outliers, which are usually referred to as outliers when the data range is outside the specified region, also known as outliers. They generally come in two forms, pseudoanomalies and true outliers. Pseudoanomalies are not true outliers but are generated by the normal operating state of the data under some mechanism; true anomalies are outliers caused by the data itself and are not related to the objective state. The error between the unprocessed data and the true data is usually defined as noise. There are three methods to deal with it: binning operation, equal frequency binning, and equal width binning. All three methods group the data, and different methods use different kinds of data instead of all the numbers in the group and thus achieve the role of smoothing the data. The binning method uses the mean, the equal frequency binning uses the median, and the equal width binning uses the boundary values.

3.2. Data Modeling and Cluster Analysis.
In machine learning classification, it involves classifying data based on attributes, which are often referred to as feature variables of the data. The more feature variables of the data, the more difficult it is to visualize the training set. Most of the feature variables contained in real-world data are correlated with each other, making it very easy to create data redundancy. The purpose of dimensionality reduction algorithms is to compress the dimensionality of the data with maximum protection of the data information, as shown in Figure 2. Feature selection refers to finding a smaller subset of the original set of variables or a subset of the feature set. This subset models the problem using filtering, wrapping, and embedding; feature extraction refers to the projection of data located in a high-dimensional space to a lower dimensional space, usually 2 or 3 dimensions. Principal component analysis (PCA) and t-distributed neighborhood embedding algorithm (t-SNE) are the most commonly used linear dimensionality reduction algorithms, which can reduce the dimensionality of data while maintaining the maximum amount of information. In this coordinate system, the dimension with the largest variance in the projection is called the first principal component, followed by the second principal component, and so on, until all coordinates contain all the original information of the data.

In order to help understand the principle of the dimensionality reduction method, here, we choose to explain the detailed execution of PCA. The following is a data X consisting of N data points, each of which is a D-dimensional vector, which will be dimensionally reduced using the PCA algorithm.

$$Q_{i*j} = \left\{ \begin{array}{l} q_{11}\, q_{12} \cdots q_{1j} \\ q_{21}\, q_{22} \cdots q_{2j} \\ \quad \cdots \\ q_{i1}\, q_{i2} \cdots q_{ij} \end{array} \right\}. \tag{1}$$

(1) First, the data needs to be centralized

$$q_{ij} = x_i - \alpha_j,$$
$$\alpha_j = \frac{\sum_{j=1}^{n} x_j}{n} \tag{2}$$

(2) Then, the covariance matrix S of the centered data is computed and the eigenvectors b of this matrix and the corresponding eigenvalues are obtained, and the centered data is denoted as X
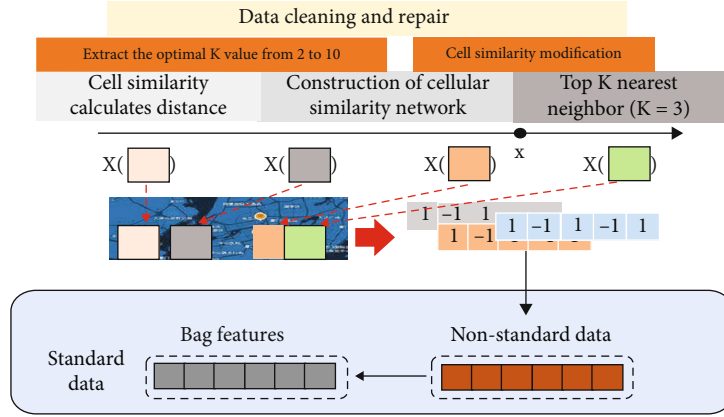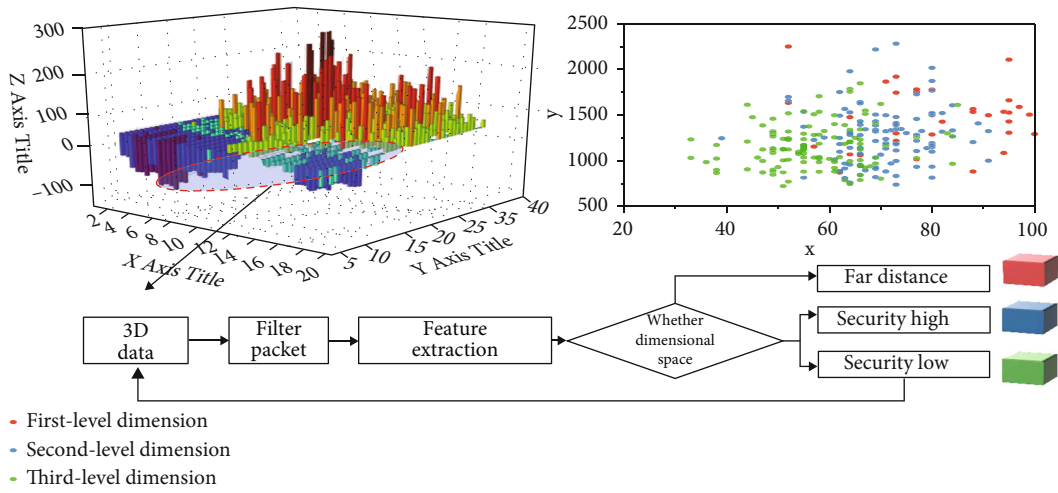
FIGURE 1: Data cleaning and repair.

- First-level dimension
- Second-level dimension
- Third-level dimension

FIGURE 2: Data dimensionality reduction technology.

$$Q_{ij} = \frac{\sum_{i=1,j=1}^{n,n} (x_i + x_j)^2}{n},$$
$$Q_{ij} = \beta q_{i,j}$$

(3)

(3) The obtained feature vectors are sorted in a descending order according to the corresponding numerical magnitude of the feature values

(4) Construct matrix B using eigenvectors and perform dimensionality reduction

$$Q_{M*N} = W^T{}_{M*N} \cdot W_{M*N}$$

(4)

Compared with PCA, t-SNE is a more advanced and effective method, which first converts the distance between points into a conditional probability to express the similarity between these points, and the distance metric between points is obtained by Euclidean operator, where $E(x_i, x_j)$ denotes the Euclidean distance between x_i and x_j. Next,

the inner similarity matrix of the original high-dimensional data X and the transformed low-dimensional data Z is obtained.

$$X(x_i|x_j) = \frac{\sum_{j=1}^{n} x_j}{\sum_{i=1}^{n} x_i},$$
$$Y(y_i|y_j) = \frac{\sum_{j=1}^{n} y_j}{\sum_{i=1}^{n} y_i}.$$

(5)

In order to compare the performance of PCA algorithm and t-SNE algorithm in dimensionality reduction and their potential capabilities for visualization, a comparison of the results of t-SNE and PCA on handwritten digital bodies is shown next.

$$L_X = \sum_{i}^{n} X(x_i|x_j) \times Q^T{}_{M*N},$$
$$L_Y = \sum_{i}^{n} Y(y_i|y_j) \times Q_{M*N}.$$
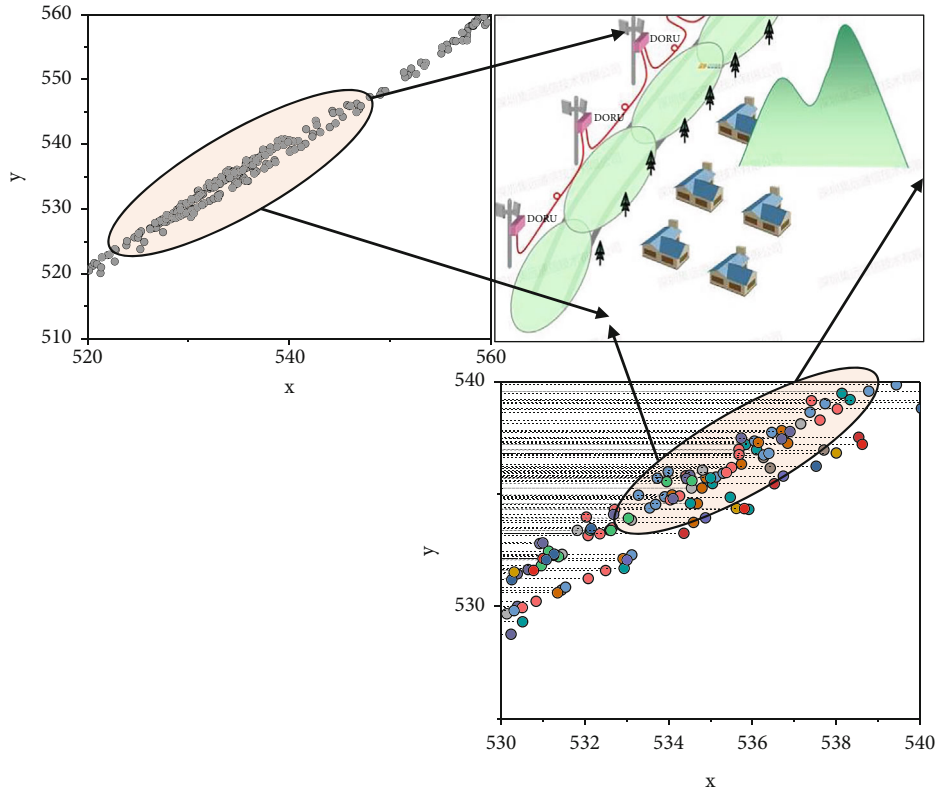
(6)

FIGURE 3: Data set for downscaling projection to 2D scatter plot.

As shown in Figure 3, the MNIST data set is chosen to compare the PCA algorithm and the t-SNE algorithm, which is a large-scale handwritten digital image database collected and created by this paper. The data set contains a training set of more than 60,000 samples and a test set of more than 10,000 samples, and the two algorithms are used here to try to reduce the dimensionality of the test set in MINST, respectively. From the results of applying the PCA algorithm to compress the data to two dimensions, we can find obvious visual confusion, and it is difficult to distinguish different categories as an observer, while the effect of t-SNE is relatively better. Therefore, in this paper, the t-SNE algorithm will also be used to project the clustering results of base stations and users in order to discover the features of both in two-dimensional space and facilitate the effective extraction of similar base stations and similar users.

3.3. Trajectory Monitoring Modeling of User Communication Base Stations. The similarity between users' regional mobile preference vectors is used to cluster the city residents to help the analysts discover the corresponding group behavior patterns. The algorithm first calculates the similarity probability of points in the high-dimensional space and then calculates the similarity probability of corresponding points in the low-dimensional space. In order to extract the behavioral characteristics of users from their call records, the base station trajectories of all users are first extracted and grouped from the data set, and the base station trajectories of users with low activity levels are screened out here because it is difficult to show valuable information over a long period of time. Then, after one-hot coding the base station trajectories of high active users, this chapter draws on the method of word embedding in natural language processing, analogizes base stations to words in language, analogizes trajectories to sentences, and designs and proposes a Cell2vec model to extract and cluster the semantics of base stations. The model in this paper takes the assumption that the neighboring base stations before and after the sequence of base stations have similar semantic information as the premise. Each dimension of the base station embedding vector can characterize the probability or intensity of the base stations belonging to different semantics. Let the base station used by user u at the ith call be b_i; then, the sequence of base stations generated by user u in a finite time period T is $u_c = \{c_1, c_2, c_3, \cdots, c_n\}$, where n is the length of the user's base station trajectory. For example, if a user makes a call with other users around base station a at 8 a.m., 2 calls are made around base station b at 11 a.m. and 13 a.m., and the last call comes from a call received around base station c at 11 p.m., then this person's base station sequence for today's day can be written as $\{c_a, c_b, c_b, c_c\}$.

It is observed that when using the effective base station sequence of cell phone users to identify the semantics of base stations, different cell phone users may show different activity in their calling activities due to their different types of work and different social preferences. Here, the call frequency of users in the experimental data set is counted, and the results are shown in Figure 4. Among these users,
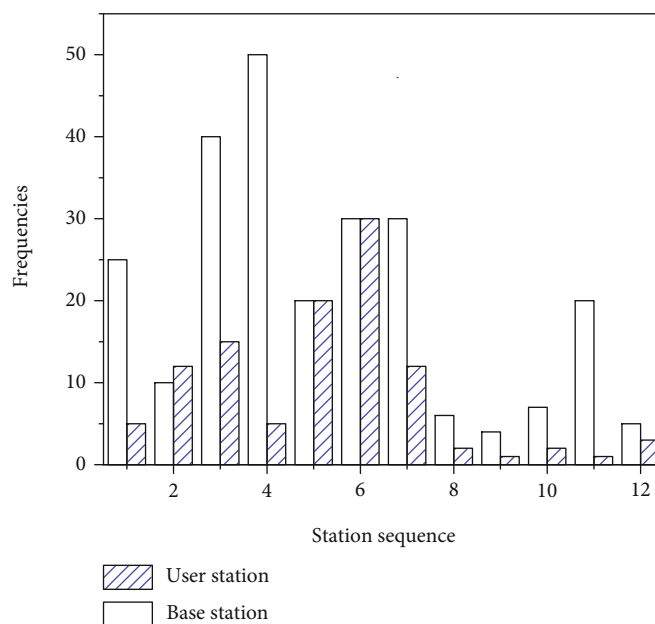
Figure 4: Histogram of user base station access frequency.

the minimum value of the number of base station visits is 1 and the maximum value is 152; the average value of the number of visits is 40, and the median value is 33. Theoretically, users with lower call frequency will generate uncertainty information for the analysis of user behavior patterns brings errors. In other words, it is difficult for low active users to show fixed or long-term behavioral patterns in the call records, and using them as the basis for semantic recognition by the base station will lead to errors in the recognition results, so users who talk more than 20 times a week can be selected as valid users, i.e., representatives of urban residents' mobility patterns, based on the call frequency distribution of all users. This filtering operation not only maintains a large base of users but also improves the correct rate of semantic recognition of the base station.

4. Results and Analysis

The final output of this function is the user area movement preference vector, as shown in Figure 5. The user area movement preference vector can well reflect the movement state of users in the city, and the similarity of movement patterns among users can be obtained indirectly by calculating the difference between user area movement preference vectors. In order to present the similarity between users' behavioral patterns and facilitate the selection of user groups of interest, this paper uses the t-SNE algorithm to cluster users based on their regional movement preference vectors. t-SNE algorithm's projection results can effectively present the similarity between data points, and this method is mainly used for dimensionality reduction and exploration of high-dimensional data, which is an effective method for visualizing and presenting high-dimensional data. It is an effective method for visualizing high-dimensional data. The neighboring points in the high-dimensional space correspond to

the nearby embedded low-dimensional points, and the far points in the high-dimensional space correspond to the distant embedded low-dimensional points. The algorithm first calculates the similarity probability of points in the high-dimensional space and then calculates the similarity probability of corresponding points in the low-dimensional space. The similarity between points is measured by the conditional probability; that is, the conditional probability that B belongs to these neighboring points is the similarity between point A and point B if the surrounding neighboring points are selected under a Gaussian distribution centered on A in a manner proportional to their probability density. The algorithm is designed to optimize these two similarity measures using a cost function to achieve the correct representation of data points in a low-dimensional space.

Clustered views are usually suitable for grouping crowds, while for information related to the spatial movement range of crowds, such maps are difficult to display them, and map-based display of the spatial movement characteristics of crowds is undoubtedly the best choice. From a geographical point of view, heat mapping is a geographic clustering method for displaying phenomena, which mainly shows the locations with high density of geographic entities, and is also a method for geographic visualization of geographic locations, as shown in Figure 6. It is commonly used in maps to show the crime activities in different areas of the city, the probability of traffic accidents. Users who are often in the university area at night are screened on the map, and their behaviors are observed. User behaviors are first decomposed according to school days and rest days, and then, their behaviors are explored in the two time periods separately. The situation after the decomposition of user behavior is presented. Heat maps usually take point data as input, and when coloring points, in order to make the chromaticity smoother, it is usually necessary to create smooth density
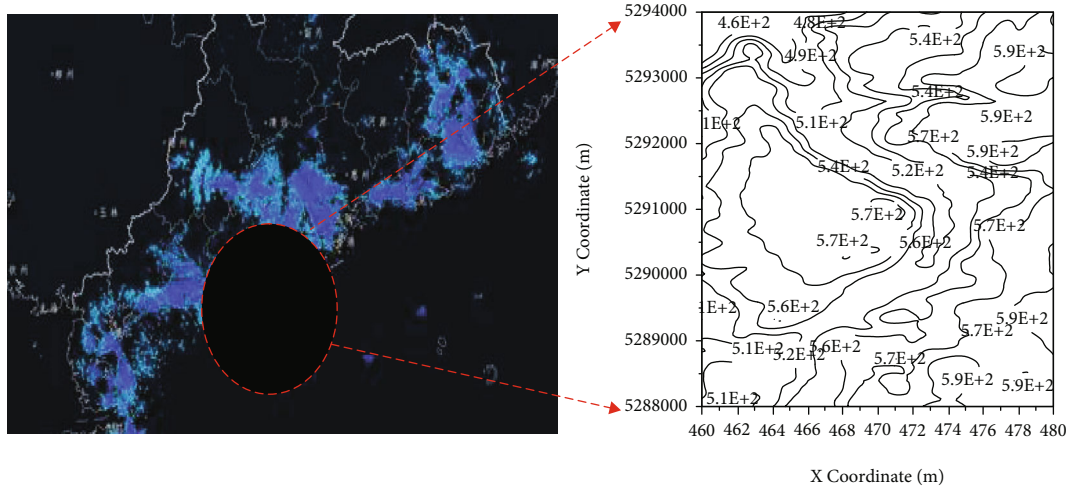
Figure 5: Multigranularity group activity area visualization.
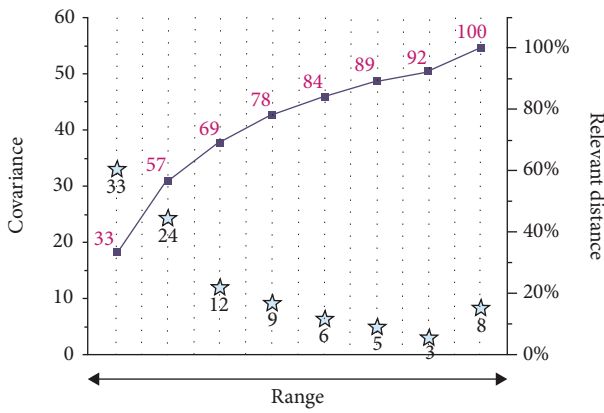

Figure 6: Location map.

surfaces using interpolation methods to derive the data of unmeasured points based on known point coordinates. One of the most classical methods is kriging interpolation, which is a statistical inference method and is often used for processing large matrices, analyzing historical data for noise estimation, etc. Its main advantage is the ability to fill the data with the best solution using linear unbiased estimation (BLUE).

Rectangular plots are a visualization method for exploring association patterns of two-dimensional relationships in data, especially for the exploration of temporal correlation patterns. The system in this paper is aimed at displaying the temporal characteristics of users, including active time and call time distribution, using matrix plots, as shown in Figure 7, where the horizontal coordinates of the rectangle plots are used to represent a day of the week and the vertical coordinates are represented as an hour of the day. In other words, each small rectangular block will be used to show the frequency of users using a particular base station at a particular hour on a particular day of the week, and the frequency of the base station being accessed is presented by the color of the block, with a darker color for a higher frequency and a lighter color for the opposite. In addition, the system is

also equipped with interactive tools embedded in the map, users can use circles or rectangles to circle or box the area of interest in the map, and the relevant information of the selected area will be filtered and presented in the background of the system to generate the corresponding bead map.

This case explores the changes in the movement patterns of people in the city through the presence of real events in the data, thus verifying the credibility of this paper's approach in terms of its ability to capture the behavioral patterns of people. It is common knowledge that most people's behavior patterns will be found to change significantly during longer national holidays, such as shifting from the inherent pattern of going to work to resting at home for an extended period of time or arriving at another place far away for a trip. In order to find such changes, this paper investigates user behavior patterns during the May Day holiday period by constructing user trajectory semantic vectors using users' call records in the week before and during May Day, respectively, calculating the difference between the vectors and finding the group of users with the largest difference. Find the users adjacent to this user from the projection view, select them all, and view their temporal features. From the temporal characteristic view, we can see that the frequency of calls of this group of users is significantly higher on the eve of holidays and more frequent during holidays than usual. Next, looking at the geospatial activity characteristics of the users in Figure 8, we can see that the activity range and frequency of these users have increased significantly compared to the preholiday period. Finally, in order to determine the destinations of these users, using the time filter to focus only on the users' movement trajectories during the holiday period, it can be found that the destinations reached by these users are mostly scenic spots. According to the actual situation, it can be seen that during national holidays, the crowd traffic in scenic spots will show a significant increase compared to the usual ones, and the frequency of base station visits will increase even more as a result, while the opposite is true for working areas. This case effectively verifies that the method in this paper can effectively capture and explain the changes of crowd behavior patterns.
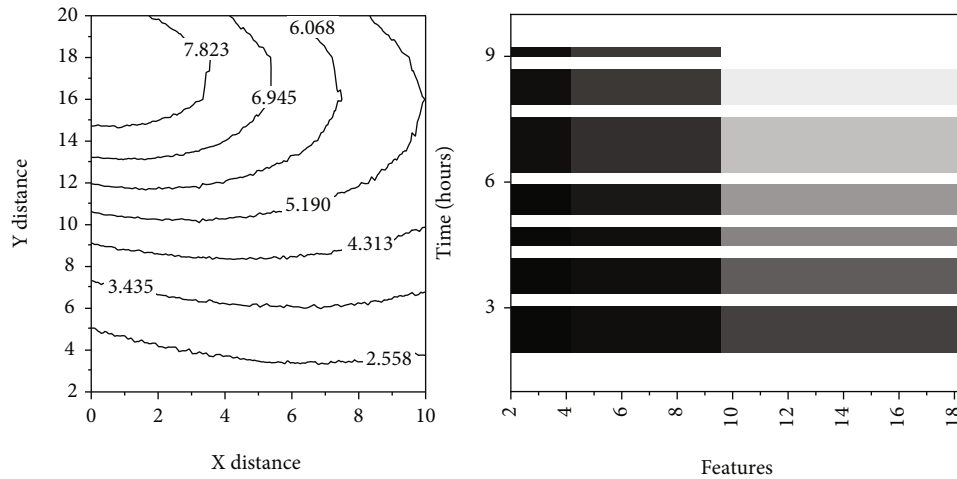
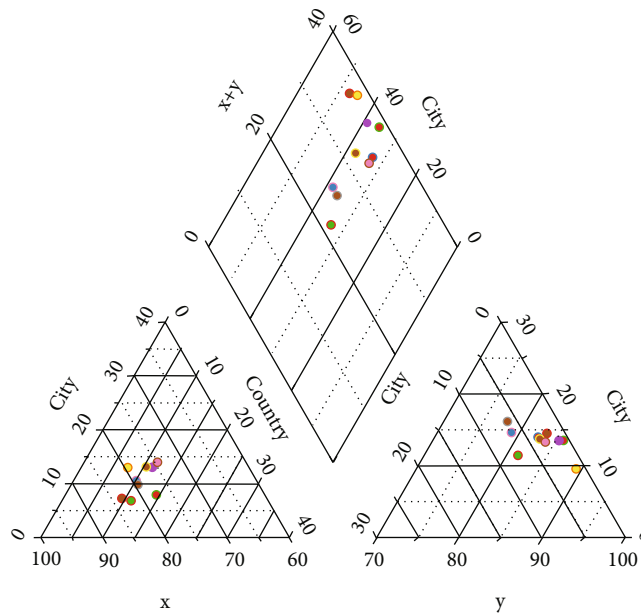Figure 7: Visualization model related to user timing features.

Figure 8: Intelligent monitoring of urban crowds.

This case evaluates the system's ability to discover users' specific patterns through a priori knowledge, thus verifying the credibility of this paper's method for exploring specific behavioral patterns of regional populations. University students, a group with dominant behavioral patterns, are chosen here for analysis. Due to the time control measures on university campuses, the majority of groups entering and leaving the campus at night are mainly students. Therefore, users who are often in the university area at night are screened on the map, and their behaviors are observed. User behaviors are first decomposed according to school days and rest days, and then, their behaviors are explored in the two time periods separately. The situation after the decomposition of user behavior is presented. From the time-series fea-

ture matrix map, we can find that the frequency of calls on school days is very low for this group of users, and the change of call locations is also small. At the same time, we can find from the map that this group of users visits bus stations, bookstores, shopping plazas, etc. on their rest days. From the analysis results, it is very consistent with the behavior pattern of college students, and this case effectively verifies the ability of this paper to identify and explain the behavior pattern of people in urban areas.

There are three methods to deal with it: binning operation, equal frequency binning, and equal width binning. All three methods group the data, and different methods use different kinds of data instead of all the numbers in the group and thus achieve the role of smoothing the data. The binning

method uses the mean, the equal frequency binning uses the median, and the equal width binning uses the boundary values.

5. Conclusion

Users move in the city for a long time showing significant regularity, and when the number of user groups is large enough, their behavioral patterns can well reflect the population mobility of city residents. In this paper, we use user call records and the proposed user trajectory modeling method to effectively mine the behavior patterns of group users in the city, and design a visual analysis model with good interaction style to interpret the extracted behavior patterns. In this paper, we summarize the spatiotemporal data analysis, human behavior patterns, and visual analysis methods based on spatiotemporal data and find that the mobile behavior patterns of group users are difficult to extract and explain after researching related literature. The Cell2vec algorithm is proposed to model users' base station trajectories by combining the word embedding method in natural language analysis technology and to identify the base station semantics and regional mobile preferences based on the correlation knowledge between text sequences and user trajectory sequences. In order to help analysts identify and interpret user behavior patterns, we use the user area movement preference calculation method to obtain the movement characteristics of the population in different areas and different time periods, and use the t-SNE algorithm to obtain the user area movement preference vector and cluster the users to help analysts discover the population clustering characteristics in low-dimensional space. To improve the interpretability of user behavior patterns, visual analysis models of group as well as regional semantics are designed to present the spatial and temporal characteristics of user behavior. User clustering and projection of the results using a method based on user area movement preferences are used to help analysts explore the characteristics of group user behavior. A heat map is used to show the movement characteristics of the population in space and the dispersion of users in regions. In order to present user area distribution and area-related base station information in detail, a map-based multigranularity visual analysis method of group area distribution is designed, with each level progressively presenting relevant information.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work.

Acknowledgments

References

[1] M. A. Al-Jarrah, M. A. Yaseen, A. Al-Dweik, O. A. Dobre, and E. Alsusa, "Decision fusion for IoT-based wireless sensor networks," *ieee internet of things journal*, vol. 7, no. 2, pp. 1313–1326, 2020.

[2] Z.-G. Du, J.-S. Pan, S.-C. Chu, H.-J. Luo, and P. Hu, "Quasi-affine transformation evolutionary algorithm with communication schemes for application of RSSI in wireless sensor network," *ieee access*, vol. 8, pp. 8583–8594, 2020.

[3] B. Duo, Q. Wu, X. Yuan, and R. Zhang, "Anti-jamming 3d trajectory design for UAV-enabled wireless sensor networks under probabilistic LoS channel," *ieee transactions on vehicular technology*, vol. 69, no. 12, pp. 16288–16293, 2020.

[4] W. Fang, W. Zhang, W. Chen, Y. Liu, and C. Tang, "TMSRS: trust management-based secure routing scheme in industrial wireless sensor network with fog computing," *wireless networks*, vol. 26, no. 5, pp. 3169–3182, 2020.

[5] G. Han, X. Yang, L. Liu, W. Zhang, and M. Guizani, "A disaster management-oriented path planning for mobile anchor node-based localization in wireless sensor networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 8, no. 1, pp. 115–125, 2017.

[6] E. H. Houssein, M. R. Saad, K. Hussain, W. Zhu, H. Shaban, and M. Hassaballah, "Optimal sink node placement in large scale wireless sensor networks based on Harris' hawk optimization algorithm," *IEEE Access*, vol. 8, pp. 19381–19397, 2020.

[7] S. M. Jenness, S. M. Goodreau, and M. Morris, "EpiModel: an R package for mathematical modeling of infectious disease over networks," *Journal of statistical software*, vol. 84, 2018.

[8] S. Karimi-Bidhendi, J. Guo, and H. Jafarkhani, "Energy-efficient node deployment in heterogeneous two-tier wireless sensor networks with limited communication range," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 40–55, 2020.

[9] A. Kavousi-Fard, W. Su, and T. Jin, "A machine-learning-based cyber attack detection model for wireless sensor networks in microgrids," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 650–658, 2020.

[10] K. H. Khan, D. Cunningham, B. Werner et al., "Longitudinal liquid biopsy and mathematical modeling of clonal evolution forecast time to treatment failure in the PROSPECT-C phase II colorectal cancer clinical trial," *Cancer discovery*, vol. 8, no. 10, pp. 1270–1285, 2018.

[11] H. Landaluce, L. Arjona, A. Perallos, F. Falcone, I. Angulo, and F. Muralter, "A review of IoT sensing applications and challenges using RFID and wireless sensor networks," *Sensors*, vol. 20, no. 9, p. 2495, 2020.

[12] L. Li, G. Xu, L. Jiao et al., "A secure random key distribution scheme against node replication attacks in industrial wireless sensor systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2091–2101, 2019.

[13] S. Misra, S. K. Roy, A. Roy, M. S. Obaidat, and A. Jha, "MEGAN: multipurpose energy-efficient, adaptable, and low-cost wireless sensor node for the internet of things," *IEEE Systems Journal*, vol. 14, no. 1, pp. 144–151, 2019.

[14] S. N. Mohanty, E. L. Lydia, M. Elhoseny, M. M. Al Otaibi, and K. Shankar, "Deep learning with LSTM based distributed data mining model for energy efficient wireless sensor networks," *Physical Communication*, vol. 40, p. 101097, 2020.

[15] M. Numan, F. Subhan, W. Z. Khan et al., "A systematic review on clone node detection in static wireless sensor networks," *IEEE Access*, vol. 8, pp. 65450–65461, 2020.

[16] S. S. Preeth, R. Dhanalakshmi, and P. M. Shakeel, "An intelligent approach for energy efficient trajectory design for mobile sink based IoT supported wireless sensor networks," *Peer-to-Peer networking and applications*, vol. 13, no. 6, pp. 2011–2022, 2020.

[17] T. Stylianopoulos, L. L. Munn, and R. K. Jain, "Reengineering the physical microenvironment of tumors to improve drug delivery and efficacy: from mathematical modeling to bench to bedside," *Trends in cancer*, vol. 4, no. 4, pp. 292–319, 2018.

[18] M. C. Chen, S. Q. Lu, and Q. L. Liu, "Uniqueness of weak solutions to a Keller-Segel-Navier-Stokes system," *Applied Mathematics Letters*, vol. 121, p. 107417, 2021.

[19] G. Sun, Y. Liu, Z. Chen et al., "Energy efficient collaborative beamforming for reducing sidelobe in wireless sensor networks," *ieee transactions on mobile computing*, vol. 20, no. 3, pp. 965–982, 2021.

[20] S. I. Xakimovich and U. D. Maxamadjonovna, "Mathematical modeling of transient processes of the automatic control system of water level in the steam generator," *Universal journal of mechanical engineering*, vol. 7, no. 4, pp. 139–146, 2019.

[21] M. C. Chen, S. Q. Lu, and Q. L. Liu, "Uniqueness of weak solutions to a Keller-Segel-Navier-Stokes model with a logistic source," *Applications of Mathematics*, pp. 1–9, 2021.

[22] H. ZainEldin, "M M Badawy; M A El-Hosseini; H Arafat; A Abraham, An improved dynamic deployment technique based-on- genetic algorithm (IDDT-GA) for maximizing coverage in wireless sensor networks [J]," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 10, pp. 4177–4194, 2020.

[23] C. Zhan and Y. Zeng, "Aerial-ground cost tradeoff for multi-UAV-enabled data collection in wireless sensor networks," *ieee transactions on communications*, vol. 68, no. 3, pp. 1937–1950, 2020.