

Article

# A Visual Analysis Approach for Inferring Personal Job and Housing Locations Based on Public Bicycle Data

Xiaoying Shi <sup>1,2,\*</sup> , Zhenhai Yu <sup>1,2</sup>, Qiming Fang <sup>1,2</sup> and Quan Zhou <sup>1,2</sup>

<sup>1</sup> School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China; yuzh@fingard.com.cn (Z.Y.); fangqiming@hdu.edu.cn (Q.F.); d3zhouquan@163.com (Q.Z.)

<sup>2</sup> Key Laboratory of Complex Systems Modeling and Simulation, Ministry of Education, Hangzhou 310018, China

\* Correspondence: shixiaoying@hdu.edu.cn

Academic Editors: Maria Antonia Brovelli, Hussein M. Abdulmuttalib, Peter Mooney, Chris Pettit and Wolfgang Kainz

Received: 2 May 2017; Accepted: 29 June 2017; Published: 7 July 2017

**Abstract:** Information concerning the home and workplace of residents is the basis of analyzing the urban job-housing spatial relationship. Traditional methods conduct time-consuming user surveys to obtain personal job and housing location information. Some new methods define rules to detect personal places based on human mobility data. However, because the travel patterns of residents are variable, simple rule-based methods are unable to generalize highly changing and complex travel modes. In this paper, we propose a visual analysis approach to assist the analyzer in inferring personal job and housing locations interactively based on public bicycle data. All users are first clustered to find potential commuting users. Then, several visual views are designed to find the key candidate stations for a specific user, and the visited temporal pattern of stations and the user's hire behavior are analyzed, which helps with the inference of station semantic meanings. Finally, a number of users' job and housing locations are detected by the analyzer and visualized. Our approach can manage the complex and diverse cycling habits of users. The effectiveness of the approach is shown through case studies based on a real-world public bicycle dataset.

**Keywords:** visual analysis; job-housing places; public bicycle data; urban mobility; spatio-temporal visualization

## 1. Introduction

With the widespread availability of location-aware technologies, the acquisition of massive spatio-temporal trajectory data over a long time period has become possible. The adoption of a data analysis method to derive new insights from these data and solve geography-related problems has become a research hotspot [1–3].

Acquiring the home and work-place locations of residents is the basis of analyzing the urban job-housing spatial relationship. Traditional methods use questionnaire surveys [4,5] to obtain personal job and housing data; these surveys are not only time-consuming, but also inaccurate when people move to a new location or change jobs. Big data analysis methods provide a new approach to this problem. Data containing rich information on human daily travel behaviors can be used to study the microscopic mobility patterns of individuals and understand the home-work dynamics of a city. Based on various trajectory data, researchers have tried to find personal places through defining rules [3,6,7] or fixed movement patterns [8]. The problem of finding semantic meanings of personal visited places is ill-defined [9], and rule-based methods have failed to manage complex travel modes because of the variable travel patterns of residents.

The public bicycle system (PBS) [10,11] is a new and important component of the urban public transit system in big cities; it provides more flexible and sustainable mobility, and is environmentally friendly and cost-effective. The PBS is a locally customized provision of affordable short-term access to bicycles on an ‘as-needed’ basis that could extend the reach of public transit services to final destinations [12]. More specifically, the user can rent a bike from a station near the starting place, use it for a short journey, and drop it off at any station in the city. The presence of users in each station is traced digitally. Research results show that the PBS was widely used by citizens in work and life: approximately 30% of users incorporated PBS into daily commuting, and the most frequently used stations were closest to either home (40%) or work (40%) [13]. Thus, the PBS data is useful for analyzing individual job and housing locations. Efforts have been devoted to understanding the daily routines of citizens and city dynamics based on PBS data [14–17], but the underlying station-related semantic meanings remain undetermined.

The problem of finding job and housing locations based on PBS data is interesting and challenging. It cannot be solved algorithmically without human wisdom. First, users can borrow bikes from any station at any time in the city, which covers a broad geographical range. How do we identify relevant candidate stations from all visited stations? Second, users’ travel habits are diverse and are based on different travel purposes. How do we understand unique individual bicycle hiring habits? Finally, personal semantic places reflect high-level information and cannot be extracted directly from raw trajectory data. How do we design analysis procedures to assist the analyzer in inferring job and housing locations interactively? Visual analysis techniques [18,19] combine the advantages of machine and human wisdom to gain insights from the dataset and reveal hidden patterns. To our knowledge, few studies have reported detecting job and housing locations based on PBS data.

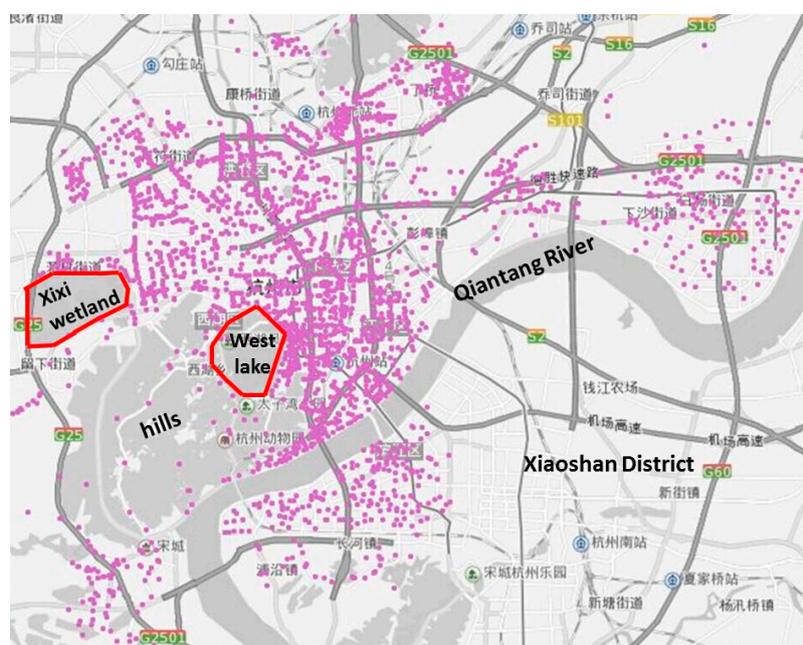


Figure 1. The distribution of stations in Hangzhou.

In this paper, we present a visual analysis approach to explore the personal job and housing locations based on PBS data. The PBS data for three consecutive months (April to June 2014) in Hangzhou, China, is used. The public bicycle system in Hangzhou was built by the government in 2008 and classed as a successful implementation [20]. Hangzhou is the capital city of the Zhejiang province, which has an estimated population of 9,018,000. By 2014, the number of public bikes exceeded 78,000, and there were more than 2000 bicycle service stations in Hangzhou. Users could find a service station every 300 m, and the average daily number of bicycle hires was about 260,000.

Figure 1 shows the distribution of stations in Hangzhou. The famous West Lake scenic spot was located in the city center, and the distribution was very dense in the downtown area. People could find stations in most parts of the city, except for the Xiaoshan District, hills and some water areas. Our approach begins by processing raw data. Then, users are clustered to find commuting users. Next, key candidate stations are recognized automatically, and several visual views are designed to help in analyzing the temporal pattern of visited stations and unique cycling habits of users. Based on the visual clues, the analyzer can recognize stations near home and work locations. Case studies are conducted to verify the effectiveness of our method.

Our method makes the following contributions:

- (1) A new kind of trajectory data: PBS data is studied to infer personal job and housing locations, and a visual analysis approach is presented to process such data.
- (2) Specific to the characteristics of PBS data, different visual views are designed to present meaningful abstractions from the raw data, and assist the intuitional and interactive reasoning process.

This paper is structured as follows. Section 2 reviews the related work. The data description and system pipeline are presented in Section 3. The details of user clustering and visual analysis procedures are described in Sections 4 and 5. Case studies are presented in Section 6 to demonstrate the system's usability. Section 7 concludes the paper with suggestions for future work.

## 2. Related Work

### 2.1. Discovering Personal Job and Housing Locations from Movement Data

As extensive human movement data became available, they were used to discover individual home and work locations. Based on different types of movement data, most methods define rules to determine the semantics of a place. Long et al. [2] combined bus card data with a household travel survey to identify job and housing locations by supposing that the departure bus stop of the first trip in one day was the home location and the location where the user spent more than 6 h (other than home) was the work location. Hasan et al. [6] assumed that there existed a fixed probability for a person to visit his home and work locations. Ahas et al. [7] and Isaacman et al. [3] used cellular network data to identify home and work places based on the frequencies of personal calls from each place. Yan et al. [8] defined a fixed movement pattern: 'home-work-shop-home', and tried to extract the most likely places to satisfy this pattern.

Because the rule-based methods were unable to deal with the complex travel modes of various users, visual analysis techniques were incorporated. Yu et al. [1] designed a visualization tool called 'iVizTRANS' to detect the home-work dynamics of a city based on bus travel data. Andrienko et al. [9] defined a set of rules for home and work-place detection and visualized the visited features of places. The analyzer could adjust the weight of rules and check the updated visual results to decide the optimal weight combination.

The above methods were mainly based on mobile phone or smart card transaction data. PBS data is a new kind of mobility data with unique features. Beecham et al. [21] tried to find a user's workplace area based on PBS data, and adopted three density-estimation algorithms to label a scope around an individual workplace. Their method required a known personal home location in advance and only detected the work-place location. The visual representation was used to compare the effectiveness of three algorithms for creating workplace areas. The assignment of the work-place still used predefined rules, which were unable to cope with various situations. In contrast with [21], visual analysis is used for reasoning and decision support in our method.

### 2.2. Analysis of the Public Bicycle System

Overviews of the existing research on the public bicycle system [10,11] informed us of both the popularity of PBS and the numerous research publications focusing on it. Ricci et al. [10] considered

that the increasing availability and quality of PBS data, coupled with the development of data mining and visualization techniques, could dramatically enhance the ability to understand the operation of the PBS. China established the first public bicycle system in Hangzhou and other cities in 2008. There are more than 100 cities and regions with a public bicycle system, making China's public bicycle program the largest in the world [22]. Zhang et al. [20] analyzed the public bicycle systems in five Chinese cities and explored their characteristics and commonalities. The impacts of built environment factors on the trip demand and the ratio of demand to supply at bike stations were investigated in the paper [23]. The service quality of the PBS in Hangzhou was evaluated and user satisfaction was measured; positive feedback was received from the respondents [24].

Visual analysis techniques offer an opportunity to extract meaningful information from PBS data. A group of studies considered its spatio-temporal usage patterns, such as observing the daily routines of citizens and city dynamics [14,15,25], and identifying the space-time variance of travel behaviors [16]. A web-based visual analytics application was designed for comparing the usage patterns of bike sharing programs in different cities [26]. Other studies focused on visualizing the spatial communities of biking flows [27], or observing the change of user profiles [28] and the usage differences among different user types [17,29]. In our previous work [30], we designed an interactive visual analytic system for exploring flows generated by PBS, which helped to visually classify stations with different flow patterns and investigate abnormal behaviors. The above methods mainly focused on users' cycling patterns or the visited features of the stations, but ignored the underlying semantics of stations. In this paper, we try to infer useful semantic meanings of stations for different users.

### 3. System Overview

#### 3.1. Data Description

This research uses PBS data (April to June 2014) from Hangzhou, China, which was provided by the Hangzhou Bus Group. Two separate data sources are involved: a complete set of hire records and a full database of station information. The station database includes the station ID (*statID*), station name (*statName*), station address (*statAddr*), longitude (*lng*), and latitude (*lat*) for every station. The hire database stores rental and return records generated by users' hire behaviors with six fields: user ID (*uID*), bike ID (*bikeID*), lease station (*leaseStat*), lease time (*leaseTime*), return station (*returnStat*), and return time (*returnTime*). The hire records and the station records are related by station ID (Figure 2).

station records				
statID	statName	statAddr	lng	lat
6145	North LinJiang Garden	...	120.153289	30.193087
3122	The intersection of JiChang road and QiuTao road	...	120.203716	30.296576
...	...	...	...	...

Hire records					
uID	bikeID	leaseStat	leaseTime	returnStat	returnTime
***647	928694	6145	2014/4/23 18:27:14	6075	2014/4/23 18:52:41
***397	930315	1004	2014/3/23 17:01:47	3212	2014/3/23 18:15:25
...	...	...	...	...	...

Figure 2. The station records and hire records.

#### 3.2. System Pipeline

In this work, we aim to explore personal home and work locations based on PBS data, and develop a visual analytic system to support this aim. The system pipeline is shown in Figure 3. First, data is

preprocessed to improve the quality of the raw data. Second, because different users borrow bikes for different purposes, and some of them seldom use this system, this could result in information that is not useful. To identify potential commuting users from the large user population, a cluster algorithm is implemented to classify users with obvious commuting properties. Next, a visual analysis procedure is designed to help infer a specific user's job and housing locations. Several visual views, such as a geographical map, temporal polylines, and Gantt chart are integrated to show data characteristics from different perspectives. Different from traveling by bus or subway, the bicycle hire fee is related to time duration. Users may return bikes in immediate stations to avoid an extra fee, and multiple short trips would appear even in a real sense of travel trip. In addition, cycling behavior is always combined with other means of transport. Therefore, special consideration should be taken in designing visual views. Finally, the analyzer can determine stations near the home and work locations of multiple users by summarizing users' visiting characteristics and their travel habits, and by visualizing the results.

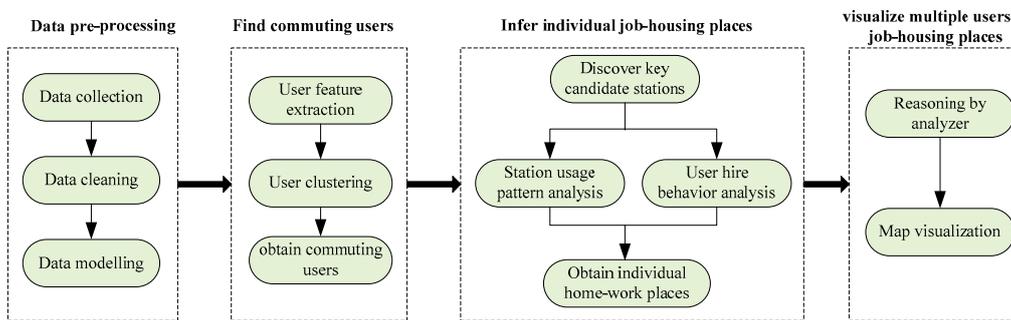


Figure 3. The system pipeline.

### 3.3. Data Preprocessing

Before analysis, data is cleaned by deleting three types of incorrect hire records (a) *leaseStat* or *returnStat* is null, which suggests it is not a complete record; (b)  $leaseTime \geq returnTime$ ; and (c) the interval between *leaseTime* and *returnTime* in a record is less than 3 min, which indicates that a bike is not really being borrowed for some reason, such as a bike malfunction.

After cleaning, the data is modeled to meet the input needs of different visual views. User-level and station-level information are extracted. Station-level information includes geographic information and the temporal aggregation of data by hourly intervals under rental and return conditions. User-level information includes a user's cycling trajectories aggregated by day. A record is defined as follows:

$$userTraj_r = \{uID, t\_Date, \{s\_Ori_i, s\_Dest_i\}_{1 \leq i \leq n}\}$$

which includes the whole hire records of the user with *uID* in a specific day *t\_Date*.  $\{s\_Ori_i, s\_Dest_i\}_{1 \leq i \leq n}$  denotes the origin and destination information of the user's *i*th hire route in that day, and *n* is the total number of rentals in that day.

## 4. User Clustering

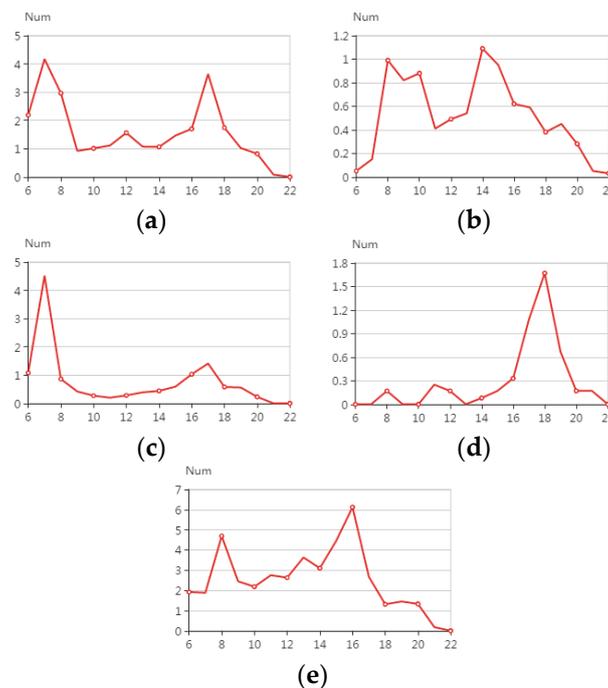
Obvious commuting properties can help in determining the home and work locations of users. Thus, we develop a clustering algorithm to classify users. The algorithm first extracts bicycle rental features from cycling trajectories, and then aggregates users according to these features.

The feature vector  $f_{uID}$  for each user has 21 dimensions. It includes several attributes: the rental ratio during workdays (*weekdayRatio*), the rental ratio during weekends (*weekendRatio*), the rental ratio during holidays (*holidayRatio*), the *rentalMode* (if *weekdayRatio* is the largest, then *rentalMode* = 1, else if *weekendRatio* is the largest, *rentalMode* = 2, else *rentalMode* = 3), and the rental frequency ( $rentalFreq = \text{total rental numbers} / \text{number of days}$ ). Because the service time of most bicycle stations in Hangzhou is from 6:00 a.m. to 10:00 p.m., we query the total hourly number of bicycle rentals during

workdays of the given period for each user, which contains 16 values in our case, corresponding to the length of the rental from 6:00–7:00 a.m. to 9:00–10:00 p.m. Then, by dividing these values by the total rental number during the workdays of the given period, the hourly rental ratio ( $hourRatio_6, hourRatio_7, \dots, hourRatio_{21}$ ) is calculated. Therefore, the feature vector  $f_{uID}$  could be represented by:

$$f_{uID} = \{workdayRatio, weekendRatio, holidayRatio, rentalMode, rentalFreq, hourRatio_i (6 \leq i \leq 21)\}$$

The feature vectors of each user are combined to generate the feature matrix  $FM$ :  $FM = [f_{uID_1}; f_{uID_2}; \dots; f_{uID_n}]$ , in which  $n$  is the number of users. Each element in  $FM$  is normalized by row. Finally,  $FM$  is processed by a  $k$ -means clustering algorithm to classify users. We analyze the clustering results and choose the cluster number as five, which represents a reasonable classification. The cluster results are shown in Figure 4. Each sub-figure represents a certain kind of user rental characteristic. The  $x$ -coordinate represents the hour, and the  $y$ -coordinate represents the hourly rental number during workdays of the statistical period. Cluster 1 and cluster 5 tend to be commuters, who have high rental numbers during the morning and evening peaks and have slight differences in peak hours, so the users in these two clusters are chosen for further analysis.



**Figure 4.** User cluster results. (a) cluster 1; (b) cluster 2; (c) cluster 3; (d) cluster 4; (e) cluster 5.

## 5. Visual Analysis Procedure

To obtain individual job and housing locations, a visual analysis procedure is designed. We focus on analyzing the active stations visited by users and users' bicycle hire patterns. To be more specific, a bicycle rental network is first constructed according to a specific user's cycling trajectories, and key nodes in the network are detected to form a candidate set of key stations. Then, several visual views are designed to show the spatio-temporal attributes when visiting these key stations and the user's travel pattern. Finally, the analyzer infers the user's home and work locations with the help of these visual results.

### 5.1. The Discovery of Key Candidate Stations

Based on a specific user's cycling trajectories, a weighted network  $G = (V, E)$  is constructed. The node set  $V$  represents the stations that a user has visited. If the user borrows or returns bikes

between nodes  $v_i$  and  $v_j$ , then there exists an edge  $e_{ij}$ . All edges form the edge set  $E$ .  $w_{ij}$  (weight of edge  $e_{ij}$ ) represents the sum of the user's rental and return number during the given period.

To automatically find key candidate stations, we adopt node degree  $k_i$  to measure the importance of node  $i$ , because users often have a relatively higher rental/return number near work and home locations.  $k_i$  is defined as the weighted sum of adjacent edges for node  $i$ :

$$k_i = \sum_j w_{ij}$$

a larger  $k_i$  value indicates a higher usage frequency of the station, which also implies the importance of that station. We calculate the top five stations with the largest  $k_i$  values to form the candidate set of key stations.

### 5.2. The Visual Analysis of Station Usage Pattern

After determining the key candidate stations, visual views are designed to show the spatio-temporal attributes when users visit these stations. To visualize spatial features, a trajectory aggregation map is created (Figure 5b). Filtering functions are provided. The analyzer can choose the time period (*leaseDate*), the minimum aggregated *bikeNum*, and the calendar properties of *leaseDate*. The calendar properties are divided into three categories: workday, weekend, and public holiday. Then, the trajectory aggregation map displays the user's cycling trips that satisfy the filtering conditions. Stations are represented by blue dots on the map according to their longitude and latitude. Two kinds of user trips exist: OD trips (with different lease and return stations) and circle trips (with the same lease and return station). For OD trips, arcs are used to connect the origins and destinations, and the center arrows denote the cycling directions. The arcs are double encoded. The thicker and more opaque arc indicates a larger number of rentals. If a station has circle trips, an outer orange ring is added on the blue dot. A thicker orange ring indicates a larger number of circle trips. By using the trajectory aggregation map, the analyzer can observe the station locations, and also understand individual movements among stations in the geographical context.

To visualize the temporal pattern, the station visited temporal view (Figure 5a) is created to present the temporal pattern of users' presence in the key candidate stations under different calendar properties. Each row corresponds to one of the key candidate stations, and contains three subfigures showing the user's hourly rental/return number under different calendar properties. The  $x$ -coordinate of the subfigure represents the hour, while the  $y$ -coordinate represents the hourly rental/return number. By utilizing this view, the analyzer can compare station usage functions.

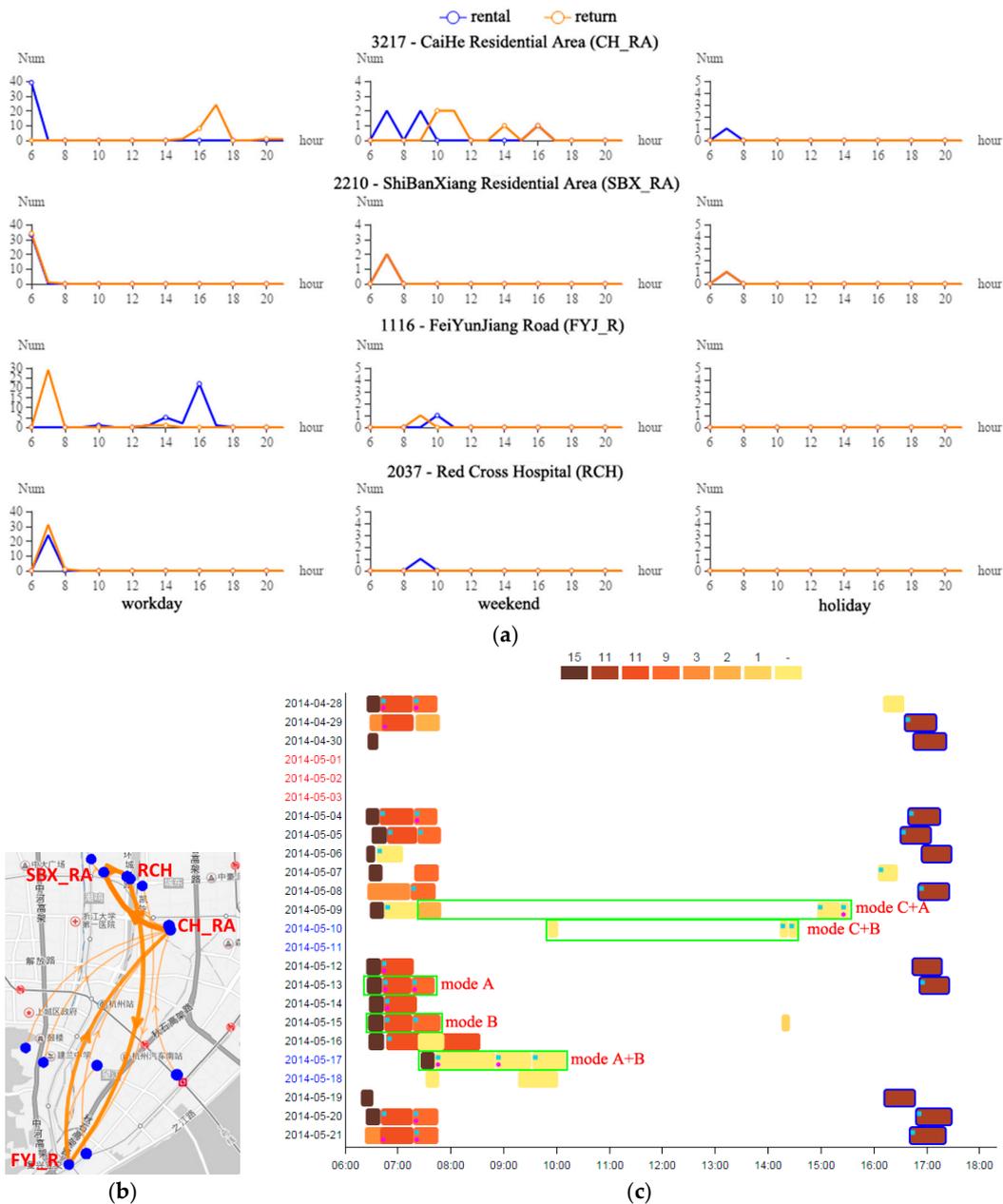
POI (place of interest) information is used for place meaning validation in other works [9,31], because it partly reflects the function of a location. In our experiment, we also visualize the POI information around a station. However, we find that it has little effect on the analysis results for the PBS data, because public bicycle stations are densely distributed in the city and there are various POIs around them. For example, except for the category 'residential area', the 'food' and 'shopping' categories are also common for a station near home. The surrounding POI information of a station has a lower frequency of the category 'residential area' that could not deny its semantic meaning of home, and vice versa.

### 5.3. The Visual Analysis of User Hire Behavior

In order to better understand user hire behavior in a detailed way and obtain more visual clues, we design a user hire sequence view to visualize daily cycling trajectories and explore individual unique cycling patterns. The main part of this view is a revised Gantt chart (Figure 5c) with two components: statistical charts and a map.

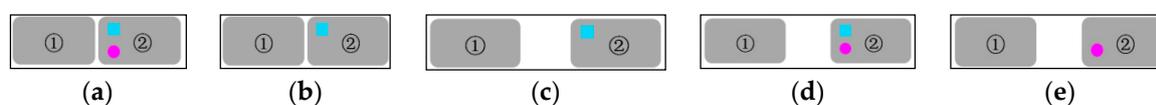
The revised Gantt chart provides an informative detailed visualization of daily hire trips. The  $x$ -axis and  $y$ -axis represent the hour and date, respectively. The color of the date is used to distinguish its calendar property: red for a public holiday, blue for a weekend, and black for a workday.

Each horizontal bar indicates a hire trip, whose length is related to its time duration. When hovering on the bar, detailed information is shown, including the borrow /return station and time. The bar color is assigned according to the trip occurrence frequency. To be more specific, the trips are first aggregated with the same origin and destination stations. Then, the aggregated trips are ordered based on their occurrence frequency. Finally, the color legend is drawn on the top center of the chart. The higher the frequency of the trip is, the deeper red color the bar is. The top seven trips with the highest frequencies correspond to the first seven color segments, and the numbers above the color segments denote the frequency values. All other trips with small probabilities are drawn using light yellow (last color segment). When clicking one color segment, the borders of the corresponding horizontal bars are highlighted with different colors.



**Figure 5.** The process of inferring home and work locations in example A. (a) The station visited temporal view; (b) The trajectory aggregation map; (c) The user hire sequence view.

Two optional glyphs are added on the horizontal bar if necessary. The blue square indicates that the return station of the previous trip (bar) is the same as the rental station of the next trip (bar), and the purple circle indicates that the bike used in contiguous trips has not changed. By using these glyphs, several hire modes can be determined (Figure 6). In mode A, the trips are contiguous and two glyphs exist simultaneously. In this situation, the user borrows a bike from station *SA*, and returns it to station *SB*. After that, he borrows the same bike from station *SB* immediately. Because the charge for using the public bicycle is related to time duration, a trip of more than one hour would charge fees. Some users often visit an intermediate station between the origin and destination, and return and borrow the same bike to save money. In mode B, only the blue square exists. Suppose the user borrows a bike *B1* from *SA*, and returns *B1* to *SB*. Then, he immediately borrows a new bike *B2* from station *SB*. In this situation, the user could not only avoid an extra charge, but also change an uncomfortable bike. In mode C, only the blue square exists and the two bars have a certain distance, which indicates that the user rides to a station *SA*, returns bike *B1*, handles affairs near that station, and finally returns to station *SA* and borrows a new bike *B2* for a new trip. The above three modes are commonly seen. In mode D, the two glyphs exist simultaneously, but the two bars have a certain distance. Suppose the user borrows a bike *B1* from station *SA*, and returns *B1* to station *SB*. After a time period *T1*, one borrows the same bike *B1* from station *SB*. This situation is uncommon, and occurs only when bike *B1* is not being borrowed during the period *T1*, and the user chooses the same bike *B1*. In mode E, only the purple circle exists and the two bars have a certain distance. Suppose the user borrows a bike *B1* from station *SA*, and returns *B1* to station *SB*. After a period of time, bike *B1* appears in station *SC*, and is borrowed by the same user. Because the probability of a user borrowing the same bike in two non-adjacent trips with a different intermediate station is very low, this mode may only be valid for staff who transit bikes among stations or may be caused by PBS recording errors. These modes can be integrated to present trips with different characteristics.



**Figure 6.** Explanation of different hire modes. The grey block stands for a horizontal bar in the revised Gantt chart, and two optional glyphs (blue square and purple circle) are added on the bar if necessary to indicate different user hire patterns. (a) Mode A; (b) Mode B; (c) Mode C; (d) Mode D; (e) Mode E.

The additional statistical charts show the ranking results of the first/last visited stations and trips, which help the analyzer to better determine regular features. To understand the trip trajectories in the geographical context, a map is linked with the revised Gantt chart. When a horizontal bar is clicked, the related trajectory is shown synchronously on the map by an arc. The arc has an initial transparency. When clicking on multiple bars, if they correspond to the same trip, then the arc is drawn again with a reduced transparency. Otherwise, a new arc representing the trip is rendered. When clicking the date label on the *y*-axis of the Gantt chart, the user's trip route for that day is rendered on the map by gradient colors.

Using the above visual views, the analyzer can infer and obtain job and housing locations for multiple users. The final results are demonstrated in the next section.

## 6. Case Studies

This system is implemented based on B/S architecture. The dataset is stored in Oracle 11g. The J2EE technical architecture is used at the server side to process the data, and JDBC is employed to connect to the database server. The digital map is developed based on a Baidu map (<http://lbsyun.baidu.com/>), and echarts (<http://echarts.baidu.com>) and d3.js (<https://github.com/d3/d3/wiki/Gallery>) are used to develop the visualization views.

By using our system, the job and housing locations of commuting users with simple patterns can be found quickly. In the following section, we present the inferring processes of two more complex examples for job and housing location discovery. The case studies demonstrate that our system can help find personal locations from cycling trajectories, which is a new approach, in contrast with the traditional survey method.

### 6.1. Example A

Different from traveling by bus or taxi, the origin and destination of a trip generated by riding a public bicycle may not include the true starting and ending places. In example A, we present how to use our system to analyze a user who does not ride directly from home to work and visits several intermediate stations while commuting. First, we select the trips between 1 April and 31 May, and the minimum rental amount is set to be 2. The station visited temporal view is shown in Figure 5a and the top four key candidate stations are the CaiHe Residential Area (CH\_RA), the ShiBanXiang Residential Area (SBX\_RA), the FeiYunJiang Road (FYJ\_R), and the Red Cross Hospital (RCH). Analysis of the temporal hire pattern indicates that the station CH\_RA seems to be near home, because it has a higher number of rentals in the morning peak and higher return number in the afternoon peak of workdays; the FYJ\_R station seems to be near the workplace, and has the reverse pattern.

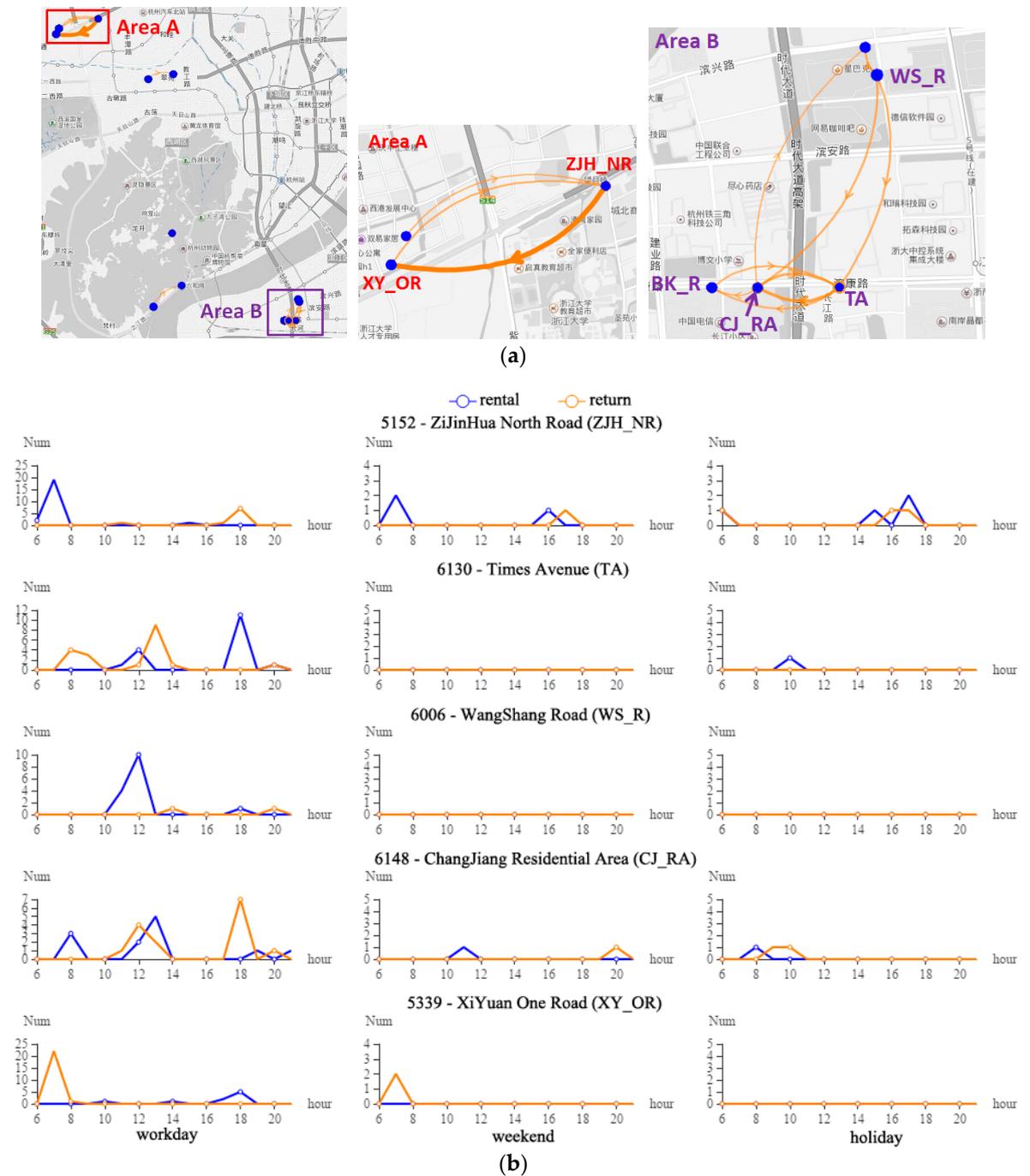
However, the stations SBX\_RA and RCH also have very high rental and return numbers in the morning peak, and we want to know their actual functions. We observe this by combining the trajectory aggregation map (Figure 5b) and the user hire sequence view (Figure 5c). The visited routes of these four stations are connected in the map. The top four color segments in Figure 5c are from CH\_RA→SBX\_RA, SBX\_RA→RCH, RCH→FYJ\_R, and FYJ\_R→CH\_RA respectively, from which we conclude that this user has a fixed cycling route during the morning peak. That route is CH\_RA→SBX\_RA→RCH→FYJ\_R. The last trip of the day is mostly from FYJ\_R to CH\_RA (blue border in Figure 5c) without visiting intermediate stations, and this pattern appears mostly on workdays. The starting and ending stations of the consecutive trips verify that CH\_RA is near home and FYJ\_R is near the workplace. The morning route is interesting, and includes two intermediate stations: SBX\_RA and RCH. We can also find the hire habit of the user from Figure 5c. It contains several hire modes, which can be found among most users. Mode A and mode B appear to be designed to avoid extra charges and change an uncomfortable bike. Mode C indicates that the user first rides to a station and returns the bike. After a period of time, the user goes back to the same station and borrows another bike. Combinations of modes are common, and we speculate that the user is frugal because mode A often appears.

### 6.2. Example B

In many situations, residents often use public bikes combined with other means of transport. The travel patterns of these users are more complicated, and are more difficult to infer without human wisdom. In example B, we explain the inferring process of a more complex situation. The trajectory aggregation map (Figure 7a) is checked first, and we find that the user frequently rents bikes in two distant regions. The enlarged maps of the two regions are shown.

Five key candidate stations are recommended: ZiJinHua North Road (ZJH\_NR), Times Avenue (TA), WangShang Road (WS\_R), ChangJiang Residential Area (CJ\_RA), and XiYuan One Road (XY\_OR). Seen from the station-visited temporal view (Figure 7b), the stations ZJH\_NR and CJ\_RA seem to be near home. Most bikes are borrowed in the morning and returned in the afternoon during workdays and weekends. The difference is that the CJ\_RA station has obvious noon trips. Station XY\_OR seems to be near the user's workplace; its morning return number is relatively high and it is not active during holiday periods. The station WS\_R is active during workdays at noon, which may be related to lunch travel. The station TA is active during workdays, and the return number in the noon period and the rental number in the afternoon are high. The last three stations seem to be near the office. From the

map (Figure 7a), we find that stations XY\_OR and ZJH\_NR are located far away from the other three stations. Therefore, using only the above results, we could not develop an accurate judgment.



**Figure 7.** The trajectory aggregation map and station-visited temporal view in example B. (a) The trajectory aggregation map; (b) The station-visited temporal view.

Next, we refer to the user hire sequence view (Figure 8). The most common trip is from ZJH\_NR to XY\_OR (green border), and often becomes the first journey in the morning. The histogram shows that ZJH\_NR is the station that is most often visited first. The above results indicate that ZJH\_NR is near home. By hovering the mouse on the bar, the starting stations of the noon trips are most often CJ\_RA, TA, and WS\_R. These stations are close to each other, but far away from the previous return

station XY\_OR. Combined with the map and temporal view, we infer that the station XY\_OR is a transition station to connect the home and workplace. The user may first reach this station and then take another traffic vehicle to a place near the office. The frequently used trips at noon are shown in area B (Figure 7a), which are related to lunch. By checking the map, we find that these stations are all near the Alibaba Group, which is a large company. So the user may choose any one of the stations as the starting station. The last trips in the evening are most often from TA to CJ\_RA (blue border) and TA to BK\_R (red border), as shown in Figure 8. We infer that the stations CJ\_RA and BK\_R also act as transition stations. In the evening, the user borrows a bike to first reach the transition station, and then takes another vehicle to reach home, because the starting station the next morning is ZJH\_NR. However, seen from the histogram in Figure 8, ZJH\_NR (home) is not involved in the top four of the last visited stations. While seen from the map in Figure 7a, the evening reversed trips from XY\_OR or a nearby station to ZJH\_NR are not obvious, and it is possible that another means of transport is used to arrive home. In this complex case, the last visited stations at the end of day are not near the home but are instead near transition stations. In general, we can infer the home location, the transition stations near home to go to work, and several frequently used stations near the company by adopting our visual analysis procedure.

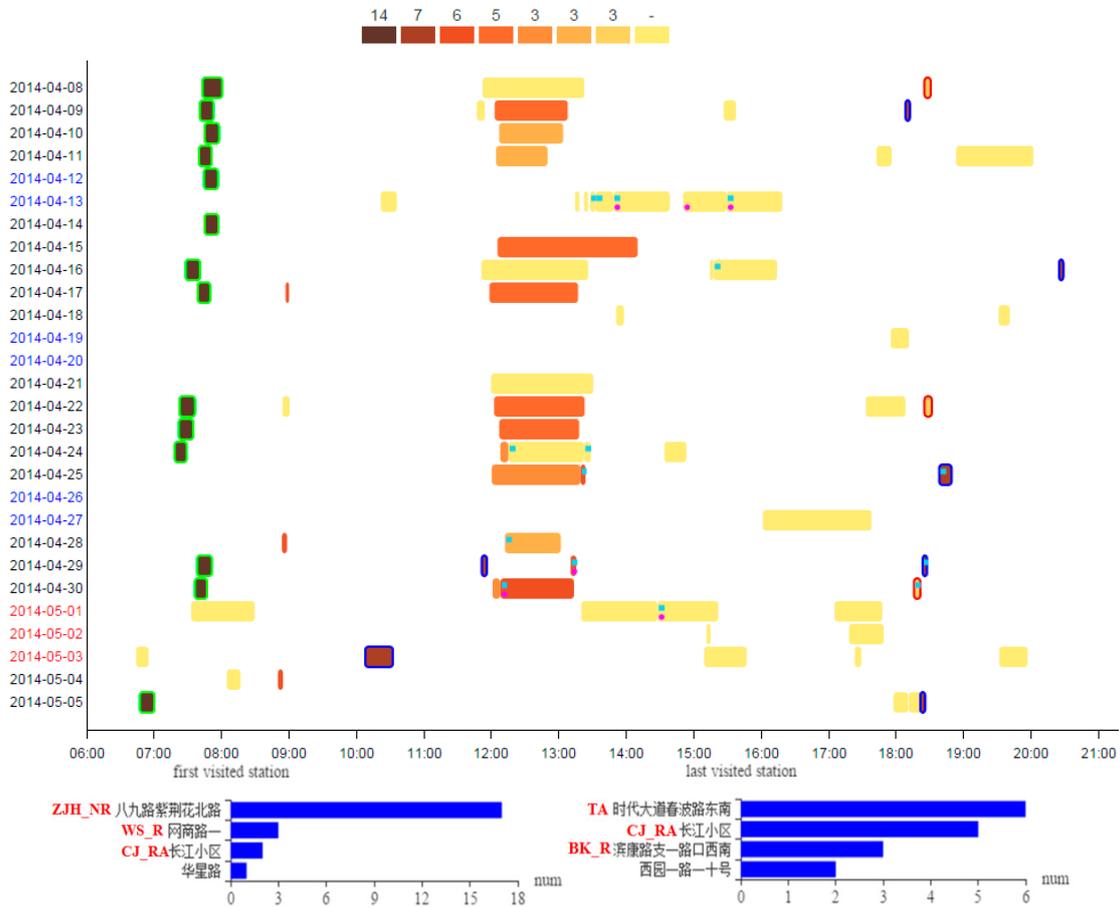


Figure 8. The user hire sequence view in example B.

### 6.3. Visualization of Multiple Users' Home and Workplaces

By using the visual analysis system, the analyzers find fifty users' home and workplaces. The visualization result is shown in Figure 9. In the map, the blue circle represents home, and the purple circle represents the workplace. A user's home and workplace are connected by a line. The bar chart shows the statistical result of the straight-line distance between the home and workplace

of all users. When clicking a bar in the chart, only the lines in that bar are shown on the map for better analysis. Analysis can be summarized as follows: (1) The home-work distances of cycling commuters are mostly less than 6.5 km; (2) For distances between 6.5 km and 7.5 km, users ride a longer distance, which is usually accompanied by several borrowing and returning behaviors, such as in example A; (3) For distances longer than 8 km, the commuting travels are accomplished by using multiple means of transport, such as in example B.

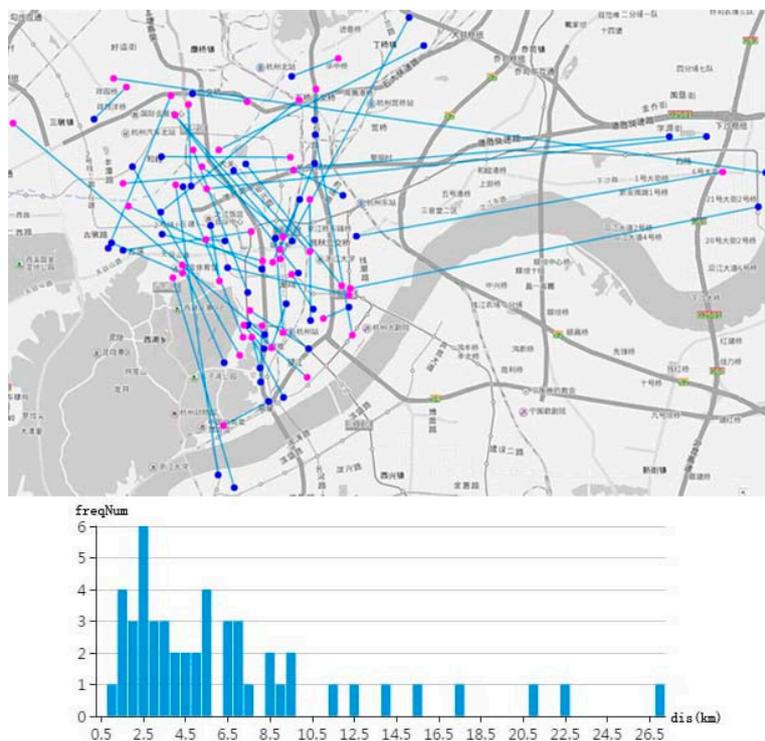


Figure 9. Visualization of multiple users' home and workplaces.

## 7. Conclusions

In this paper, we address the problem of adopting PBS data to identify personal home and work places. For this purpose, users are clustered and different visual views are designed to help the inference of job and housing locations. Through interactive visual mining, analysts can infer the job and housing locations of multiple users. Compared to the traditional way of obtaining personal locations through questionnaires, our method saves much time and effort. It provides an opportunity to extract important personal-related places from PBS trajectories, and can manage complex situations that cannot be managed by using simple rule-based methods. The results can be used to build a training set with personal home and workplaces labeled according to the cycling data, which is the basis of automated machine learning methods. Urban planners could also use the results to analyze the job-housing spatial mismatch in cities. In general, our method is a new attempt to solve the traditional geography problem by including emerging big data.

There are still some problems to be studied. The initial values of user clustering by  $k$ -means are randomly generated, so the clustering results depend on effective initial values, and we need to perform clustering several times to choose a better result. In the future, we will attempt to find a better user clustering algorithm with a higher stability. In addition, although our method helps in interactively discovering users' personal places, it still requires much time when coping with a large number of users. In the future, we will study the method of combining visual analysis and machine learning techniques to help in analyzing the spatial relationships between home and employment locations on a city scale.

**Acknowledgments:** This work is supported by the National Natural Science Foundation of China (No. 61603119, No. 61602141) and the Public Projects of Zhejiang Province (2016C33110, 2015C33067).

**Author Contributions:** Xiaoying Shi and Qiming Fang designed the visual analysis procedure; Zhenhai Yu implemented most of the system functions; Quan Zhou realized the user clustering algorithm; Xiaoying Shi wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yu, L.; Wu, W.; Li, X.; Li, G.; Ng, W.; Ng, S.; Huang, Z.; Arunan, A.; Watt, H. iVizTRANS: Interactive visual learning for home and work place detection from massive public transportation data. In Proceedings of the 2015 IEEE Conference on Visual Analytics Science and Technology (VAST), Chicago, IL, USA, 25–30 October 2015; pp. 49–56.
2. Long, Y.; Thill, J.C. Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing. *Comput. Environ. Urban Syst.* **2013**, *53*, 19–35. [[CrossRef](#)]
3. Isaacman, S.; Becker, R.; Cáceres, R.; Kobourov, S.; Martonosi, M.; Rowland, J.; Varshavsky, A. Identifying important places in people’s lives from cellular network data. In Proceedings of the 9th International Conference on Pervasive Computing (Pervasive), San Francisco, CA, USA, 12–15 June 2011; pp. 133–151.
4. Wang, D.; Chai, Y. The jobs-housing relationship and commuting in Beijing, China: The legacy of Danwei. *J. Transp. Geogr.* **2009**, *17*, 30–38. [[CrossRef](#)]
5. Zhao, P.; Lü, B.; De Roo, G. Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. *J. Transp. Geogr.* **2011**, *19*, 59–69. [[CrossRef](#)]
6. Hasan, S.; Schneider, C.M.; Ukkusuri, S.V.; González, M.C. Spatiotemporal patterns of urban human mobility. *J. Statist. Phys.* **2013**, *151*, 304–318. [[CrossRef](#)]
7. Ahas, R.; Silm, S.; Järv, O.; Saluveer, E.; Tiru, M. Using mobile positioning data to model locations meaningful to users of mobile phones. *J. Urban Technol.* **2010**, *17*, 3–27. [[CrossRef](#)]
8. Yan, Z.; Parent, C.; Spaccapietra, S.; Chakraborty, D. A hybrid model and computing platform for spatio-semantic trajectories. In *The Semantic Web: Research and Applications*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 60–75.
9. Andrienko, N.; Andrienko, G.; Fuchs, G.; Jankowski, P. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. *Inf. Vis.* **2016**, *15*, 117–153. [[CrossRef](#)]
10. Ricci, M. Bike sharing: A review of evidence on impacts and processes of implementation and operation. *Res. Trans. Bus. Manag.* **2015**, *15*, 28–38. [[CrossRef](#)]
11. Fishman, E. Bikeshare: A review of recent literature. *Trans. Rev.* **2016**, *36*, 92–113. [[CrossRef](#)]
12. Nikitas, A.; Wallgren, P.; Rexfelt, O. The paradox of public acceptance of bike sharing in Gothenburg. In *Proceedings of the Institution of Civil Engineers—Engineering Sustainability*; ICE Publishing: London, UK, 2016; Volume 169, pp. 101–113.
13. Shaheen, S.; Zhang, H.; Martin, E.; Guzman, S. Hangzhou Public Bicycle: Understanding Early Adoption and Behavioral Response to Bikesharing in Hangzhou, China. *Trans. Res. Rec.* **2011**, *2247*, 34–41. [[CrossRef](#)]
14. Froehlich, J.; Neumann, J.; Oliver, N. Measuring the pulse of the city through shared bicycle programs. In Proceedings of the UrbanSense 08, Raleigh, NC, USA, 4 November 2008; pp. 16–20.
15. Mooney, P.; Corcoran, P.; Winstanley, A.C. Preliminary results of a spatial analysis of Dublin’s bike rental scheme. In Proceedings of the 18th the GIS Research UK Annual Conference, London, UK, 14–16 June 2010; pp. 325–330.
16. Wood, J.; Slingsby, A.; Dykes, J. Visualizing the dynamics of London’s bicycle-hire scheme. *Cartogr. Int. J. Geogr. Inf. Geovis.* **2011**, *46*, 239–251. [[CrossRef](#)]
17. Beecham, R.; Wood, J. Exploring gendered cycling behaviours within a large-scale behavioural data-set. *Transp. Plan. Technol.* **2014**, *37*, 83–97. [[CrossRef](#)]
18. Chen, W.; Guo, F.; Wang, F.Y. A survey of traffic data visualization. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2970–2984. [[CrossRef](#)]
19. Andrienko, N.; Andrienko, G. Visual analytics of movement: An overview of methods, tools and procedures. *Inf. Vis.* **2013**, *12*, 3–24. [[CrossRef](#)]

20. Zhang, L.; Zhang, J.; Duan, Z.; Bryde, D. Sustainable bike-sharing systems: Characteristics and commonalities across cases in urban China. *J. Clean. Prod.* **2015**, *97*, 124–133. [[CrossRef](#)]
21. Beecham, R.; Wood, J.; Bowerman, A. Studying commuting behaviours using collaborative visual analytics. *Comput. Environ. Urban Syst.* **2014**, *47*, 5–15. [[CrossRef](#)]
22. Yang, R.; Long, R. Analysis of the Influencing Factors of the Public Willingness to Participate in Public Bicycle Projects and Intervention Strategies—A Case Study of Jiangsu Province, China. *Sustainability* **2016**, *8*, 349. [[CrossRef](#)]
23. Zhang, Y.; Thomas, T.; Brussel, M.; Van Maarseveen, M. Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China. *J. Transp. Geogr.* **2017**, *58*, 59–70. [[CrossRef](#)]
24. Zhang, D.; Xu, X.; Yang, X. User satisfaction and its impacts on the use of a public bicycle system: Empirical studies from Hangzhou, China. *Transp. Res. Rec. J. Transp. Res. Board* **2015**, *2512*, 56–65. [[CrossRef](#)]
25. Oliveira, G.N.; Sotomayor, J.L.; Torchelsen, R.P.; Silva, C.T.; Comba, J.L. Visual analysis of bike-sharing systems. *Comput. Graph.* **2016**, *60*, 119–129. [[CrossRef](#)]
26. Bargar, A.; Gupta, A.; Gupta, S.; Ma, D. Interactive visual analytics for multi-city bikeshare data analysis. In Proceedings of the 3rd Urbcomp, New York, NY, USA, 24 August 2014.
27. Zhou, X. Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago. *PLoS ONE* **2015**, *10*, e0137922. [[CrossRef](#)] [[PubMed](#)]
28. Goodman, A.; Cheshire, J. Inequalities in the London bicycle sharing system revisited: Impacts of extending the scheme to poorer areas but then doubling prices. *J. Transp. Geogr.* **2014**, *41*, 272–279. [[CrossRef](#)]
29. Vogel, M.; Hamon, R.; Lozenguez, G.; Merchez, L.; Abry, P.; Barnier, J.; Borgnat, P.; Flandrin, P.; Mallon, I.; Robardet, C. From bicycle sharing system movements to users: A typology of Vélo'v cyclists in Lyon based on large-scale behavioural dataset. *J. Transp. Geogr.* **2014**, *41*, 280–291. [[CrossRef](#)]
30. Shi, X.; Yu, Z.; Chen, J.; Xu, H.; Lin, F. The visual analysis of flow pattern for public bicycle system. *J. Vis. Lang. Comput.* **2017**. [[CrossRef](#)]
31. Krueger, R.; Thom, D.; Ertl, T. Visual analysis of movement behavior using web data for context enrichment. In Proceedings of the 2014 IEEE Pacific Visualization Symposium (PacificVis), Yokohama, Japan, 4–7 March 2014; pp. 193–200.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).