

Article

Identification of Urban Jobs–Housing Sites Based on Online Car-Hailing Data

Shuoben Bi , Luye Wang , Shaoli Liu, Lili Zhang and Cong Yuan

School of Geographical Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China

* Correspondence: bishuoben@163.com

Abstract: With the development of cities, the organization of jobs–housing space is becoming more complex, and the rapid, effective identification of both residences and workplaces is crucial to sustainable urban development. The long time series of online car-hailing data conveys a large amount of activity trajectory information about urban populations, which can represent the social functions of urban areas, including workplaces and residences. This paper constructs a jobs–housing site identification model based on human activity characteristics. This model uses a time series dataset of online car hailing that characterizes the changes in regional passenger flow and implements the similarity measure and semi-supervised learning of time series to determine the classification of urban areas. Then, the jobs–housing factor method is introduced to extract the jobs–housing characteristics of different regions, which achieves the jobs–housing site identification. Finally, the empirical analysis of Chengdu city shows that the proposed model method can effectively mine the distribution of urban jobs–housing sites. The identification results are consistent with the actual situation, and the combination of the time series similarity and the jobs–housing feature variable improves the identification effect, providing a new way of thinking about urban jobs–housing space research.

Keywords: urban jobs–housing sites; online car-hailing data; time series mining; similarity measurement; jobs–housing factor method



Citation: Bi, S.; Wang, L.; Liu, S.; Zhang, L.; Yuan, C. Identification of Urban Jobs–Housing Sites Based on Online Car-Hailing Data. *Sustainability* **2023**, *15*, 1712. <https://doi.org/10.3390/su15021712>

Academic Editor: Isam Shahrour

Received: 8 November 2022

Revised: 13 January 2023

Accepted: 13 January 2023

Published: 16 January 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a complex system, cities host different kinds of social activities. Dwelling and employment are the most basic types of activities in a city. The spatial configuration relationship between them is important for interpreting the internal spatial structure of the city and sustainable urban development, and this area of research has received extensive attention. In the global context, the increasingly close ties between countries and regions and the deepening degree of economic exchanges have promoted the optimization and upgrading of urban industrial structures and attracted more people to cities, making the jobs–housing space more complex and placing higher demands on urban development and urban management. At the same time, limited urban land resources are being used for more and more residential and employment activities, urban ecology is being damaged, and healthy urban development is under threat. Healthy urban development requires optimizing the layout of jobs–housing space and establishing a harmonious and orderly social environment. The appropriate spatial pattern for jobs–housing sites is the foundation for healthy urban development. Therefore, many scholars at home and abroad have conducted research on the jobs–housing distribution in major cities around the world, looking for the direction that can guide healthy urban development, help urban development and urban management, and repair urban ecology. Research on the jobs–housing distribution in China began in the 1980s. Due to the introduction of the reform and opening-up policy, China relaxed its population mobility control policy, and there was a massive influx of the population into cities and a concentration of social and economic activities. Chinese cities entered a period of dramatic spatial reconstruction and scale expansion [1]. The spatial

pattern of the planned economy period with the unit compound as the basic territorial unit gradually disintegrated, and the distribution of urban jobs–housing sites experienced major changes. In the absence of adequate knowledge of the spatial characteristics of urban jobs–housing sites, urban management and construction were no longer considered appropriate for the rapid development of the city, and the urban ecological spatial structure has suffered damage. In addition, with the increasing improvement in urban transportation facilities, there were more possibilities for residents to choose as their jobs–housing sites. What followed was the increasingly prominent phenomenon of spatial separation and matching dislocation between urban residences and workplaces. The resulting problems include long-term, long-distance commuting, the deterioration of the urban ecological environment, reduction in livability, and other types of urban challenges to sustainable urban development. Therefore, we constructed an empirical method to obtain information about urban jobs–housing site distribution to provide a basis for relieving the pressure on urban jobs–housing sites, solving urban development problems, and protecting urban ecology.

Traditionally, jobs–housing site identification was conducted using the data of residents' travel survey reports or special questionnaire surveys [2,3]. The rapid development and wide application of information and communications technology (ICT) and the global positioning system (GPS) led to the large-scale collection of long-term mobile positioning data. These data broke the traditional information collection barriers and spatial scales, providing a more refined database for the research on identifying the distribution of jobs–housing sites and urban spatial structure, and promoting the formation of consensus theories and viewpoints. Research on the identification methods of jobs–housing sites for different data sources is still a primary concern of urban researchers. Jiang et al. [4] constructed a time–space bilayer clustering analysis method based on the information entropy method and combined it with the mobile phone signaling data in Tianjin to identify jobs–housing sites. Niu et al. [5] used mobile phone signaling data to calculate the recurrence rate of each user and identified the workplace and residence of Shanghai residents to obtain employment density and commuting data. Chen et al. [6] constructed a comprehensive decision matrix of residences and workplaces based on the characteristics of the stay time slot and stay time of mobile phone signaling trajectory points and identified the most likely semantic stay points as the residence and workplace. Based on smart card data, Sari et al. [7] constructed a location recognition model with the number of trips, visit frequency, and stay time as features, which detected the residences and workplaces in London with a high accuracy. Liu [8] formulated the identification rules for public transportation commuting in Wuhan city and used a rule-based decision tree approach to identify jobs–housing sites. Zhou et al. [9] estimated the residence and workplace of public transport workers in Beijing, which they used to evaluate the jobs–housing balance in Beijing and established a land use model for areas with a jobs–housing imbalance. Researchers also investigated the distribution of jobs–housing sites using Weibo check-in data [10,11], taxi GPS trajectory data [12–14], and point of interest (POI) data [15–17], obtaining rich research results.

The breakthrough of spatial data mining technology and the production of massive data sources containing residents' travel behavior have provided a new perspective on the distribution of jobs–housing sites, but current research does not focus on online car-hailing data. As typical demographic activity data, online car-hailing data contain the spatial activity characteristics of residents' travel and urban zoning characteristics, which can be used to reflect transportation demand, regional social functions, and urban dynamic characteristics [18,19]. Their use for jobs–housing site identification is conducive to expanding the theories and perspectives of jobs–housing space research. In addition, previous research on jobs–housing site identification tends to focus on the travel characteristics of the peak period and ignores the activity characteristics of residents during the rest of the time. As such, the representation of jobs–housing functions has not been well explored. It is necessary to carry out time series mining on the temporal characteristics of the complete time period to improve the effect of jobs–housing site identification.

This research is based on the traditional methods of jobs–housing site identification. The research objectives are as follows. How can we identify urban jobs–housing sites quickly and effectively to discover new opportunities for jobs–housing site identification and urban sustainable development supported by big data? How to provide real-time information of jobs–housing site distribution for urban development and management, and to improve the timeliness and effectiveness of urban management and construction decisions, thus alleviating the jobs–housing separation and maintaining urban ecology? This paper examines the travel characteristics of human activity using one month of Chengdu Didi Chuxing order data. It explores the time series similarity measurement method that adapts to the characteristics of online car-hailing data, and it combines the time series mining algorithm and jobs–housing factor method to make up for the shortcomings of the traditional method. Finally, this paper constructs a jobs–housing site identification model and explores the jobs–housing relationship in Chengdu city with examples.

The paper is organized as follows. First, a description of the study area and data is given. Second, the method and process of the jobs–housing site identification model is introduced. Third, the distribution of the jobs–housing sites in the central urban area of Chengdu is mined and the results are analyzed and discussed. Finally, we summarize the conclusions of our research.

2. Study Area and Data

2.1. Study Area

Chengdu is one of the two cores of the Chengdu–Chongqing urban agglomeration, and it is a typical monocentric city in western China. At the beginning of the 21st century, under the guidance of the “land replacement” policy, industrial land in the central urban area of Chengdu gradually migrated outward. The nature of this land was changed to residential, financial, or green land, while some government units and residential houses in the central urban area were transformed into commercial land with higher returns [20]. Figure 1 shows that during the spatial restructuring of land, Chengdu carried out the expansion from the center to the outside in the urban development pattern, which caused the separation of the jobs–housing sites in the central urban area. Therefore, it is important to identify the jobs–housing sites in the central urban area for the healthy development of western cities. Considering the current urbanization level in Chengdu and the integrity of the data, we selected the central urban area (i.e., the interior of the Fourth Ring Road) of Chengdu city, Sichuan Province as the study area, as shown in Figure 2.

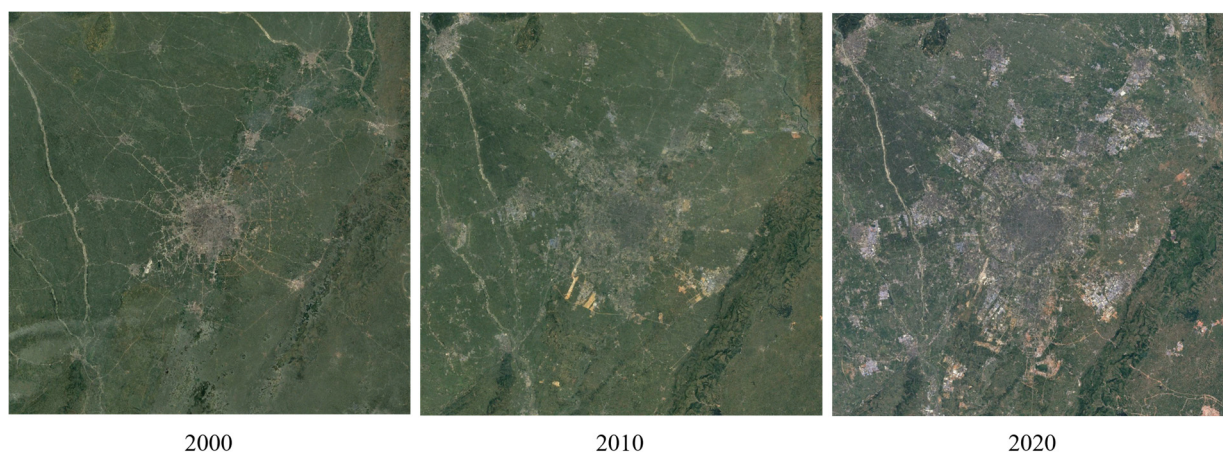


Figure 1. Maps showing the urban development of Chengdu at different periods (from Google Map historical images).

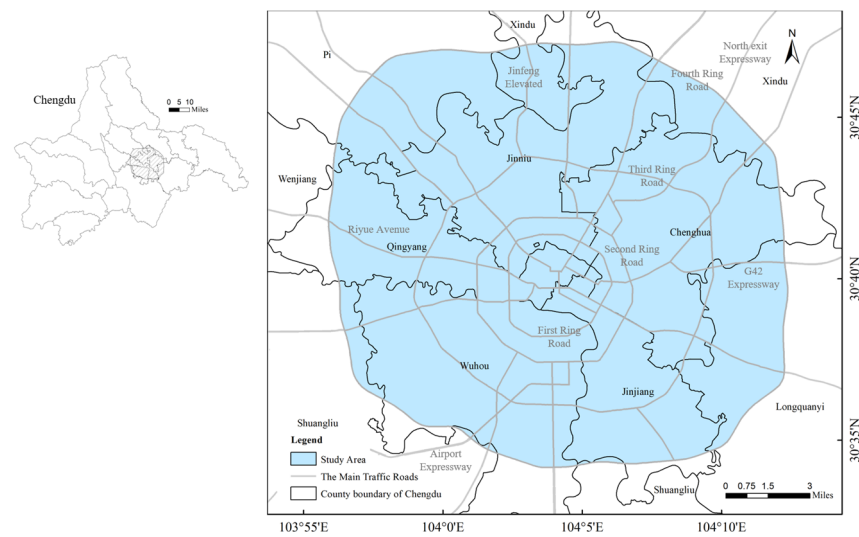


Figure 2. Schematic diagram of the study area, showing the main traffic roads and administrative division in the central urban area of Chengdu. The administrative division data are from the National Geomatics Center of China (<http://www.ngcc.cn/ngcc/> (accessed on 22 July 2021)).

2.2. Study Data and Preprocessing

2.2.1. Road Network Data

Before identifying the urban jobs–housing site distribution, the study area needed to be divided into several basic study units. The complex and closely connected road system of a city divides the entire city into different areas, and different functional buildings, such as residential quarters and industrial buildings, are distributed in their respective areas. Therefore, we used the actual road network as the boundaries to divide the study area. The road network data were obtained from the OpenStreetMap official website (<http://openstreetmap.org> (accessed on 5 July 2021)). After extracting the centerline of the double-lane roads in the original road network data, removing minor elements such as the interior of the community, overpasses, and broken roads, and performing topology inspection, the processed road network data were used to segment the study area, and the study area was divided into 4455 study units, as shown in Figure 3.



Figure 3. The study units, which were obtained by cutting the study area with the centerline of the roads, served as the basic spatial unit for this research.

2.2.2. Online Car-Hailing Data

The online car-hailing data in this study, which comprise the order data of Chengdu Didi Chuxing from November 2016, were acquired from the GAIA open data initiative [21] of Didi Chuxing Company and were collected or produced by Didi Chuxing Company using an irreversible de-identification process. Didi Chuxing is a one-stop travel platform covering taxi, limousine, carpool, designated driving, bus, and freight services, which has changed the traditional way of taxi hailing and established a modern way of travel for users in the era of the mobile Internet. According to a Smart Travel Report released by Didi Chuxing Company, in the third quarter of 2016, the total number of online car-hailing trips in Chengdu reached 130 million, and 6 out of every 10 people had used the service of Didi Chuxing Company. Chengdu has the highest penetration rate of online taxi trips in the western region. Online car hailing is an important part of commuting trips, with 415,000 trips in the morning peak and evening peak every weekday, equivalent to the capacity of 138 subway trains or 3715 buses, which is very convenient for residents' travel.

The dataset contains information for a single passenger trip, which includes the unique order ID, latitude, longitude, and time information of origin and destination. In the original data, the time information was presented in the Unix timestamp format, which is defined as the total number of seconds from 00:00:00 GMT on 1 January 1970 to the present. For statistical convenience, we converted the Unix timestamp format to the common time format. To encrypt the geographic location, the latitude and longitude information in the original data was adapted to the GCJ-02 coordinate system. We needed to uniformly convert the latitude and longitude information into the commonly used CGCS2000 coordinate system through the coordinate correction algorithm so as to facilitate the subsequent calculation and statistics. In addition, the dataset had some quality issues. Some origin and destination times did not fall on any day in November, and these instances with incorrect time information needed to be deleted. After preprocessing the original data, a total of 5,635,815 pieces of valid order data were obtained.

The temporal variation in picking-up and dropping-off volumes in an area contains rich information on social dynamics. It characterizes the travel patterns of urban residents and can reflect the urban functional structure, including the residences and workplaces. The traffic volume refers to the number of people participating in online car-hailing trips in a region, which represents the amount of population activity. The inflow volume refers to the difference between the number of people dropping off and the number of people picking up in a region. Both volumes are the integration features of the picking-up and dropping-off volumes and are effective in representing the travel patterns of residents [22,23]. By considering the uneven area of the study unit, the traffic volume density and inflow volume density were used to understand the activity characteristics of residents. The equation for this is as follows:

$$D_{traffic} = \frac{N_{pick-up} + N_{drop-off}}{area}, \quad (1)$$

$$D_{inflow} = \frac{N_{drop-off} - N_{pick-up}}{area}, \quad (2)$$

where $D_{traffic}$ is the traffic volume density, D_{inflow} is the inflow volume density, $N_{pick-up}$ is the pick-up volume, $N_{drop-off}$ is the drop-off volume, and $area$ is the area of the region. We extracted the hourly traffic volume density and hourly inflow volume density in a month, and then averaged the density for the same weekdays (Monday–Friday) and weekends (Saturday–Sunday) to obtain the human activity characteristics for a total of 168 h in 7 days. The bivariate time series dataset of 4455 study units was constructed, as shown in Figure 4. It characterizes the travel pattern of each plot and serves as the basic data for the division of urban areas.

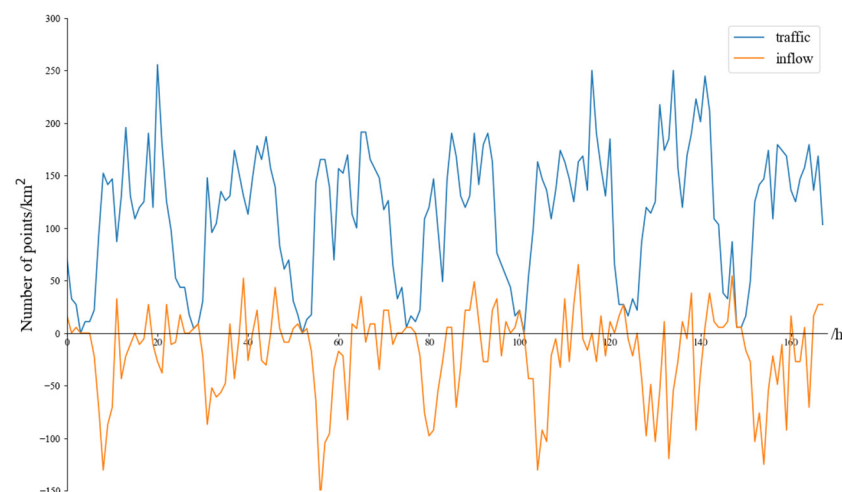


Figure 4. Time series of the regional sample: “traffic” is the traffic volume density time series and “inflow” is the inflow volume density time series.

3. Identification Model of Jobs–Housing Sites

The morning peak and evening peak are the most frequent time periods for commuting between workplaces and residences and contain rich jobs–housing characteristics. The jobs–housing factor method [24] describes the jobs–housing characteristics of an area through human activity characteristics in a particular time period. It is easy to understand, has a high operating efficiency, and is a common algorithm for identifying jobs–housing sites (and most of the other algorithms are extensions of it). However, it is greatly restricted by characteristics of the peak period and threshold selection, and the mining of temporal characteristics is not sufficient, so it is deficient in the representation of jobs–housing characteristics. Time series mining can adequately extract the features of each time period and classify urban areas based on the similarity of the time series features, but the mining results do not clearly identify the workplace and residential functions of each regional category. Therefore, we combined the time series mining method with the jobs–housing factor method and constructed an identification model of jobs–housing sites based on the time series similarity and the jobs–housing feature variable. The similarity of the time series features was used to make up for the shortcomings of traditional jobs–housing characteristics extraction. The jobs–housing factor method was used to determine the jobs–housing characteristics of the mining results. Thus, the influence of the particular time period characteristics was weakened and the accuracy of the jobs–housing site identification was improved.

The specific process of the model is as follows. First, we implement the similarity measure for the online car-hailing time series by using the LB_Lance distance and dissimilarity index of the time series to construct the similarity matrix. Second, based on the similarity matrix, K-medoids clustering is used on the online car-hailing time series to obtain the training samples. Third, based on the training samples and similarity matrix, K-nearest neighbor (KNN) classification is carried out on the online car-hailing time series to obtain the urban area classification results. Then, to clearly determine whether each classification area is a jobs–housing site or not, the jobs–housing factor is introduced as a feature variable to extract the jobs–housing characteristics of each classification area, and thus identify residences and workplaces. Finally, the model is validated and evaluated by the Kappa coefficient. The technical route of the model is shown in Figure 5.

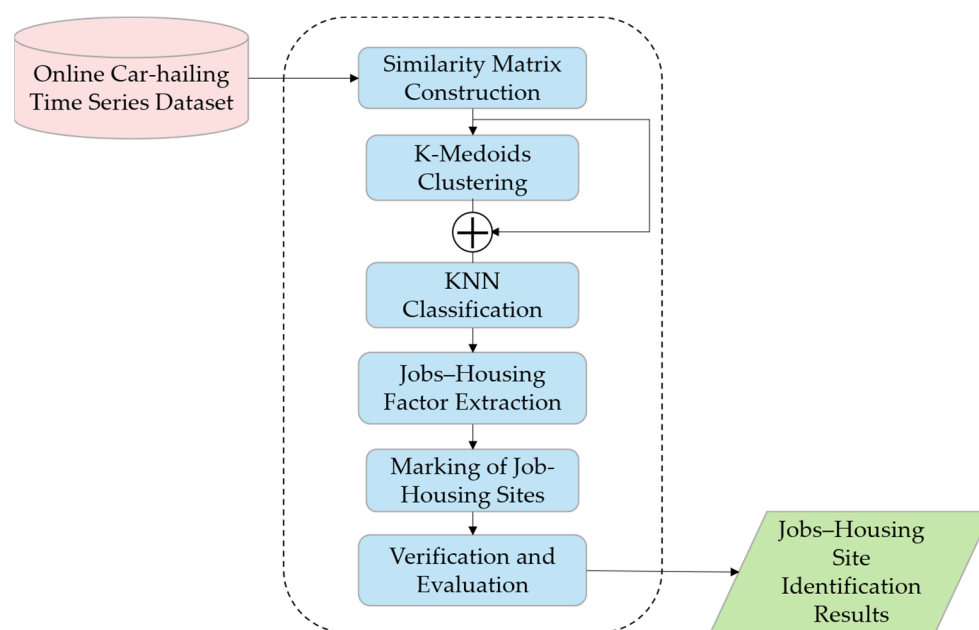


Figure 5. Technical process of the jobs-housing site identification model: “ $\oplus$ ” indicates that the similarity matrix is to be added to the KNN classification.

3.1. Method for Similarity Measure of Time Series

The similarity measure is often quantified and intuitively implemented through the distance function. The representation and characteristics of the data are directly related to the effectiveness of the distance function on the similarity measure. Most of the existing research on time series mining of online car hailing does not consider the characteristics of online car-hailing data. In addition, the Euclidean (Euc) distance and dynamic time warping (DTW) distance [23,25] commonly used in research are not applicable to the characteristics of online car-hailing time series. We needed to establish a new measurement method to better portray the similarity between the time series and obtain a better time series mining effect.

3.1.1. LB_Lance Distance Function

The first obstacle that needs to be overcome in the distance calculation based on the raw time series data is the operational efficiency, because the long period of the time series will increase the time complexity. Therefore, researchers have proposed using the lower bound function to filter out the parts of the time series that do not satisfy the similarity requirement, improve the performance of time series similarity search, and improve the operation efficiency of the algorithm. Yi [26], Kim [27], Keogh [28], and others proposed lower bound functions supporting a distance metric. The LB_Keogh function (LB means “lower bound”) is the lower bound function proposed by Keogh. It is based on the global time warp constraint and constructs the upper and lower boundaries of the dynamic time series through a sliding window as a way to filter the time series. The distance calculation is more compact and efficient than LB_Kim and LB_Yi designed by Kim and Yi. In addition, the randomness of online car-hailing travel behavior will cause some abrupt changes in the time series, resulting in some singular values in the time series, and the Euclidean distance will make the singular values have a larger weight, reducing the accuracy of clustering. Lance and Williams first proposed a dimensionless distance function, which was named the Lance distance. It can overcome the shortcoming that the algorithm is sensitive to singular values, and it weakens the negative impact of the absolute value difference of the samples on the measurement results.

Based on the above analysis, we propose a LB_Lance distance function based on the LB_Keogh lower bound function concept, which is to process the time series within a

sliding window of $2w + 1$, take the maximum and minimum eigenvalues of the time series within each window to construct the upper and lower boundaries, and then calculate the sum of the Lance distances between the boundary lines of the feature points that do not fall within the upper and lower boundaries. There are two time series of equal length: $P = (p_1, p_2, \dots, p_i, \dots, p_m)$ and $Q = (q_1, q_2, \dots, q_i, \dots, q_m)$. The LB_Lance distance metric between the two time series is shown in Equation (3):

$$D_{LB_Lance} = \frac{1}{n} \sum_{i=1}^n \begin{cases} \frac{|L_p(i) - U_q(i)|}{|L_p(i)| + |U_q(i)|} & L_p(i) > U_q(i) \\ \frac{|L_q(i) - U_p(i)|}{|L_q(i)| + |U_p(i)|} & L_q(i) > U_p(i) \\ 0 & \text{others} \end{cases} \quad (3)$$

where D_{LB_Lance} is the LB_Lance distance between the time series, n is the number of time series point pairs that satisfy the conditions and participate in the calculation, and U_p , L_p , U_q , and L_q are the upper and lower boundaries of time series P and Q . U is the upper boundary of the time series, and L is the lower boundary of the time series.

3.1.2. Dissimilarity Index of Time Series

The LB_Lance distance is based on the proximity of the values. However, the eigenvalues at each time node of the time series are correlated, and the growth pattern of the time series likewise characterizes the degree of similarity of the time series. Chouakria et al. [29] introduced an adaptive dissimilarity index based on the temporal neighboring relationship coefficient of the first order, which measures the proximity of the dynamic growth behavior of the time series through the temporal neighboring relationship coefficient of the first order so as to correct the conventional distance function. The coefficient is shown in Equation (4):

$$CORT(P, Q) = \frac{\sum_{i=1}^{T-1} (p_{i+1} - p_i)(q_{i+1} - q_i)}{\sqrt{\sum_{i=1}^{T-1} (p_{i+1} - p_i)^2} \sqrt{\sum_{i=1}^{T-1} (q_{i+1} - q_i)^2}}. \quad (4)$$

The value range of $CORT$ is $[1, -1]$. When the $CORT$ value tends to 1 from 0, the two time series increase or decrease simultaneously under the same magnitude of change; that is, the behavior is similar, so the time series distance obtained from the conventional distance function needs to be reduced. When the $CORT$ value tends to -1 , the conventional time series distance increases. When the $CORT$ value is equal to 0, the two time series do not have monotonicity and are stochastically linearly independent, and so the conventional measure is unchanged. Therefore, an adjustment function is needed to adjust the original $CORT$ value to be used as a coefficient of the conventional distance metric. The exponential adaptive tuning function is usually used, as shown in Equation (5):

$$F(x) = \frac{2}{1 + e^{kx}} (k \geq 0). \quad (5)$$

3.1.3. Similarity Metric Function for Time Series of Online Car-Hailing Data

This study combines the LB_Lance distance and dissimilarity index to measure the similarity of the time series of online car-hailing data, as shown in Equation (6):

$$Dis(P, Q) = F[CORT(P, Q), k] * D_{LB_Lance}(P, Q, w), \quad (6)$$

where $Dis(P, Q)$ is the distance measure, $CORT(P, Q)$ is the correlation coefficient of the time series growth pattern, $F(x)$ is the exponential adaptive tuning function of the $CORT$ function, and $D_{LB_Lance}(P, Q)$ is the LB_Lance distance function based on the lower bound function. This measurement method not only has better adaptability to the characteristics of the time series of online car hailing but also has symmetry, which avoids the influence of the determination of the initial clustering centers and the order of data clustering on the

final clustering results. It reduces the number of inter-sample distance calculations and the complexity of the distance determination.

3.2. Semi-Supervised Learning Method

The time series mining algorithm was used to mine the travel characteristics of residents for complete time periods. Clustering is a commonly used time series data mining algorithm among the time series pattern mining algorithms. The K-medoids clustering algorithm avoids the influence of singular data points on the clustering results because it selects the actual object as the clustering center instead of the mean value. In terms of the distance measurement, the K-medoids clustering algorithm can flexibly define the proximity status of the metric values. However, the clustering algorithm does not have data labels prepared for training samples, resulting in clustering results that may have more inaccurate categorization. To make the results more credible, the KNN classification algorithm was added on the basis of K-medoids clustering to form a semi-supervised learning method. In time series data mining tasks, the KNN classification algorithm is considered to be the most appropriate approach, of which the 1-NN classifier is the most widely used [30]. The KNN algorithm only determines the class of the sample to be classified according to the class of the nearest one or several samples in the classification decision. Therefore, for the traffic volume time series sample set with many intersections or an overlap in the class domain, the KNN method is more suitable than other methods [31].

We performed K-medoids clustering on the time series dataset based on the similarity metric function for the time series of online car-hailing data to generate preliminary results of the urban area division, selected the data with a better clustering effect as training samples according to the time series of each clustering center in the clustering result, and then performed a KNN supervised classification on the remaining samples based on the training samples and the similarity matrix to obtain the classification results of the time series dataset, with the urban regions divided into several categories.

3.3. Jobs–Housing Factor

The results of semi-supervised learning cannot determine whether each cluster is a jobs–housing site or not. To obtain the distribution results of the jobs–housing sites, we needed to extract the jobs–housing characteristic of each category from the urban area classification results. In this study, the jobs–housing factor JHF was used as a feature variable to identify jobs–housing sites, quantify the migration characteristics between regions, and explore the jobs–housing characteristics in the regions. Residents migrate from their residences to other areas by online car hailing during the morning peak and migrate from their workplaces to other functional areas by online car hailing during the evening peak. Therefore, the online car-hailing data generated during the morning and evening peak hours contain obvious characteristics of residences and workplaces. The jobs–housing factor method based on the characteristics of human activity can express the characteristics of workplaces and residences by analyzing the dynamic change and migration characteristics of urban population, so as to identify the distribution of jobs–housing sites. The jobs–housing factor of an area is calculated as shown in Equation (7):

$$JHF = \frac{inflow_m \times outflow_e - inflow_e \times outflow_m}{totalflow_m \times totalflow_e}, \quad (7)$$

where $inflow_m$ and $inflow_e$ denote the inflow volume of online car hailing in the morning peak (7:00–10:00) and in the evening peak (17:00–20:00), respectively; $outflow_m$ and $outflow_e$ represent the outflow volume of online car hailing in the morning peak and in the evening peak, respectively; and $totalflow_m$ and $totalflow_e$ are the traffic volume of online car hailing during the morning peak and evening peak, respectively. When a certain area is dominated by the inflow of online car hailing during the morning peak and the outflow during the evening peak, JHF takes values within the range of (0,1], indicating that the area tends to have workplace characteristics. The more JHF tends to 1, the stronger the

area is workplace-related. Conversely, a certain area associated with a large number of morning outflows as well as evening inflows will render JHF within the range of $[-1, 0)$, implying that the area is associated with residences. The closer JHF approaches -1 , the stronger residential characteristics the area exhibits. In this paper, based on the origin and destination information of the online car-hailing data in each urban cluster, we counted the average values of inflow volumes and outflow volumes in the peak period for each classification area within a month, and then calculated the jobs–housing factor of each classification area to describe the jobs–housing characteristics.

3.4. Model Validation and Evaluation

For the identification results of jobs–housing sites, the accuracy can be evaluated by the Kappa coefficient based on the confusion matrix, whose numerical magnitude reflects the level of classification accuracy, with a value range of $[0, 1]$. There is a lack of true values because there is no official urban jobs–housing site distribution map of Chengdu as a reference for verification. Therefore, relevant information, such as that from Google Earth images, AutoNavi maps, and real-life photos of landmark areas, was used as the true value of the confusion matrix for the purpose of verifying the effectiveness of the method used in this study for identifying jobs–housing sites.

4. Experimental Results and Analysis

4.1. Parameter Setting

In the time series similarity metric function, the time window w determines the size of the constructed time series boundary and the number of feature points participating in the time series similarity measurement. Figure 6 shows the construction results of the time series boundaries under different w values. We found that with the increase in w , the range between the lower and upper boundaries gradually becomes larger, making the final nodes involved in the similarity calculation decrease and gradually deviate from the original time series morphology, resulting in the decrease in the accuracy of the time series similarity measurement. The time series interval constructed by the online car-hailing data is 1 h. To retain the original morphological features of the time series in the process of similarity measurement, we took $w = 1$, that is, when constructing the time series boundary, we took the maximum and minimum feature points within 1 h before and after the node as the boundary point. The weight k in the distance function regulates the contribution of the behavior and values to the time series distance measurement, and when $k = 0$ the contribution of conventional distance is 100%. Figure 7 shows the effect of the adaptive adjustment for several values of k greater than 0 in the exponential adaptive tuning function. $k = 2$, which has the middle degree of change in the function curve, was selected as the weight to obtain the similarity matrix of the time series dataset for online car hailing, and the similarity measurement was implemented.

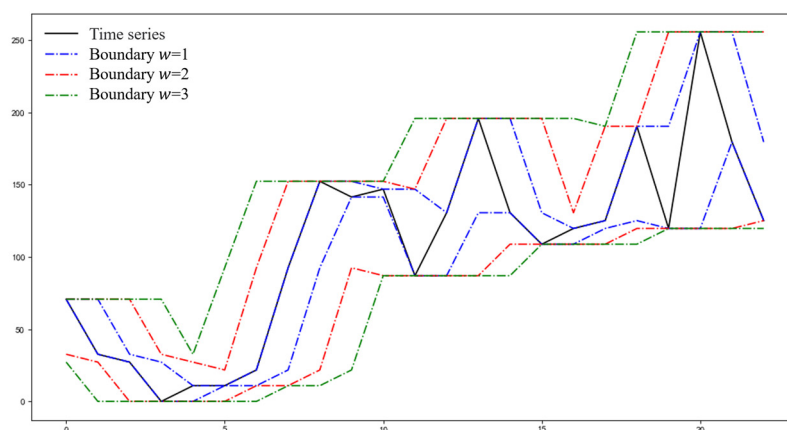


Figure 6. Time series boundaries constructed by different w values.

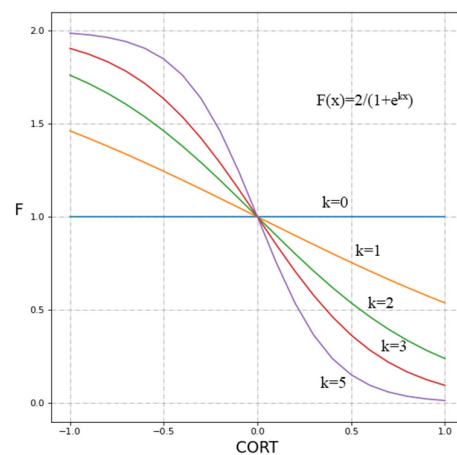


Figure 7. Exponential adaptive tuning function curve graph.

The determination of the number of clusters needs to comprehensively consider the size of the dataset, the purpose of classification, and the validity of the clustering effect. In this paper, the silhouette coefficient [32] and the Davies–Bouldin Index (BDI) [33] were mainly used to select the optimal number of clusters. These two indicators comprehensively consider the degree of separation of different categories and the degree of cohesion of the same category in the cluster structure. Based on the similarity matrix constructed above, the changes in the silhouette coefficient and DBI under different cluster numbers were calculated by repeated clustering operations for the time series, as shown in Figure 8. The larger the silhouette coefficient and the smaller the DBI, the better the clustering effect [32,33]. After a comprehensive consideration of the results of the two clustering indicators and the size of the sample, we selected seven as the cluster number.

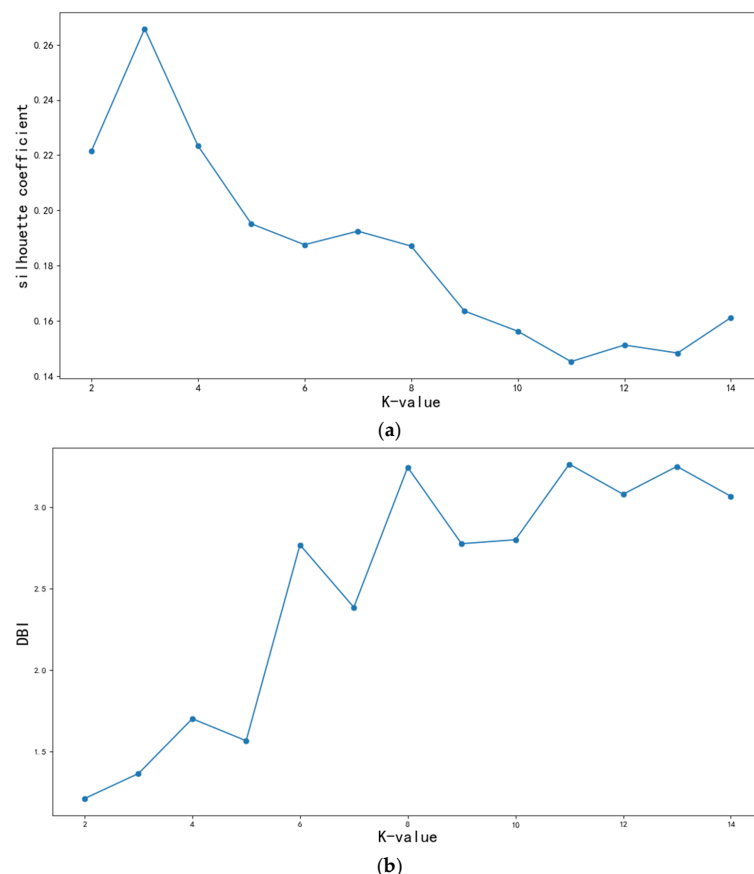


Figure 8. (a) Silhouette coefficients and (b) Davies–Bouldin Index under different numbers of clusters.

4.2. Identification Results

4.2.1. Results of Semi-Supervised Learning

Based on the above methods and parameters, time series mining was carried out on the online car-hailing time series dataset, and the final result of the urban area classification was obtained (Figure 9). Using the time series of the clustering centers of each cluster, the traffic volume density and inflow volume density of the seven clusters (C0–C6) on each day were counted (Figure 10) to find the residents' travel characteristics of each cluster and make a preliminary judgment on the jobs–housing sites. Observing the spatial distribution and residents' travel characteristics of each regional category, we found that the urban area of C0 is concentrated in the central urban area, and the population activity is mainly intensive from 7:00 to 20:00, with the peaks at 9:00 and 18:00. The population outflow is large from 7:00 to 10:00, and the inflow volume density reaches a positive value in the evening, which is consistent with the commuting characteristics of residents who go to work by car in the morning and return to their residences after work or recreation in the evening. It also has obvious residential characteristics, such as the Ma'an Community. C6, which has similar travel characteristics to C0, also has the residential characteristics of outflow in the morning and inflow in the evening. Its traffic volume density is higher than that of C0, which may be related to the fact that its spatial distribution is far away from the urban center and it requires a longer commuting distance, so more residents choose online car-hailing travel, including the Wangjiang Community. The urban area of C5 is close to the urban center and has a greater resource advantage for workplaces, such as the Raffles Office Building. The population activity is mainly concentrated from 8:00 to 17:00, and the maximum inflow volume density occurs around 9:00. The traffic volume density and inflow volume density on weekends are significantly lower than those on weekdays, with the travel characteristics of a workplace. The urban area of C1 is distributed at the edge of the central urban area, which is typical of the spatial distribution characteristics of industrial-type dominated workplaces, which can reduce the impact on the downtown area, such as the Tianfu Business Park. The traffic volume density peaks at 9:00 and 18:00, and residents inflow in the morning and outflow in the evening. The population activity is low, which may be related to the fact that industrial areas are generally equipped with staff dormitories. The time series of the remaining categories do not reflect the regular jobs–housing characteristics, so the remaining categories are attributed to other regions.

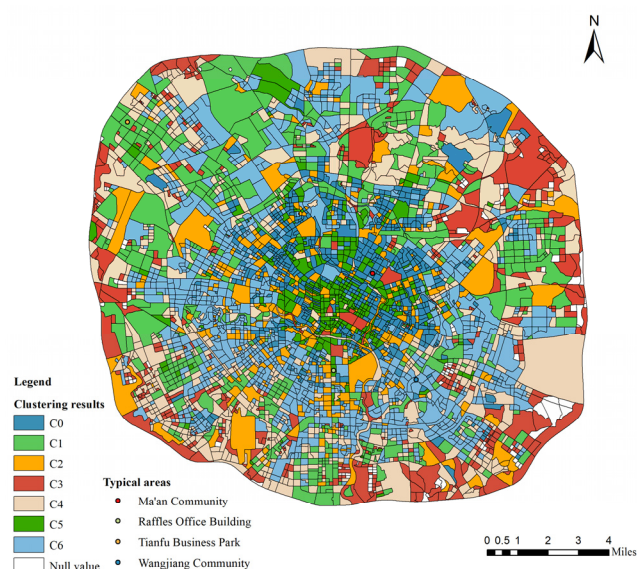


Figure 9. The results of the urban area classification.

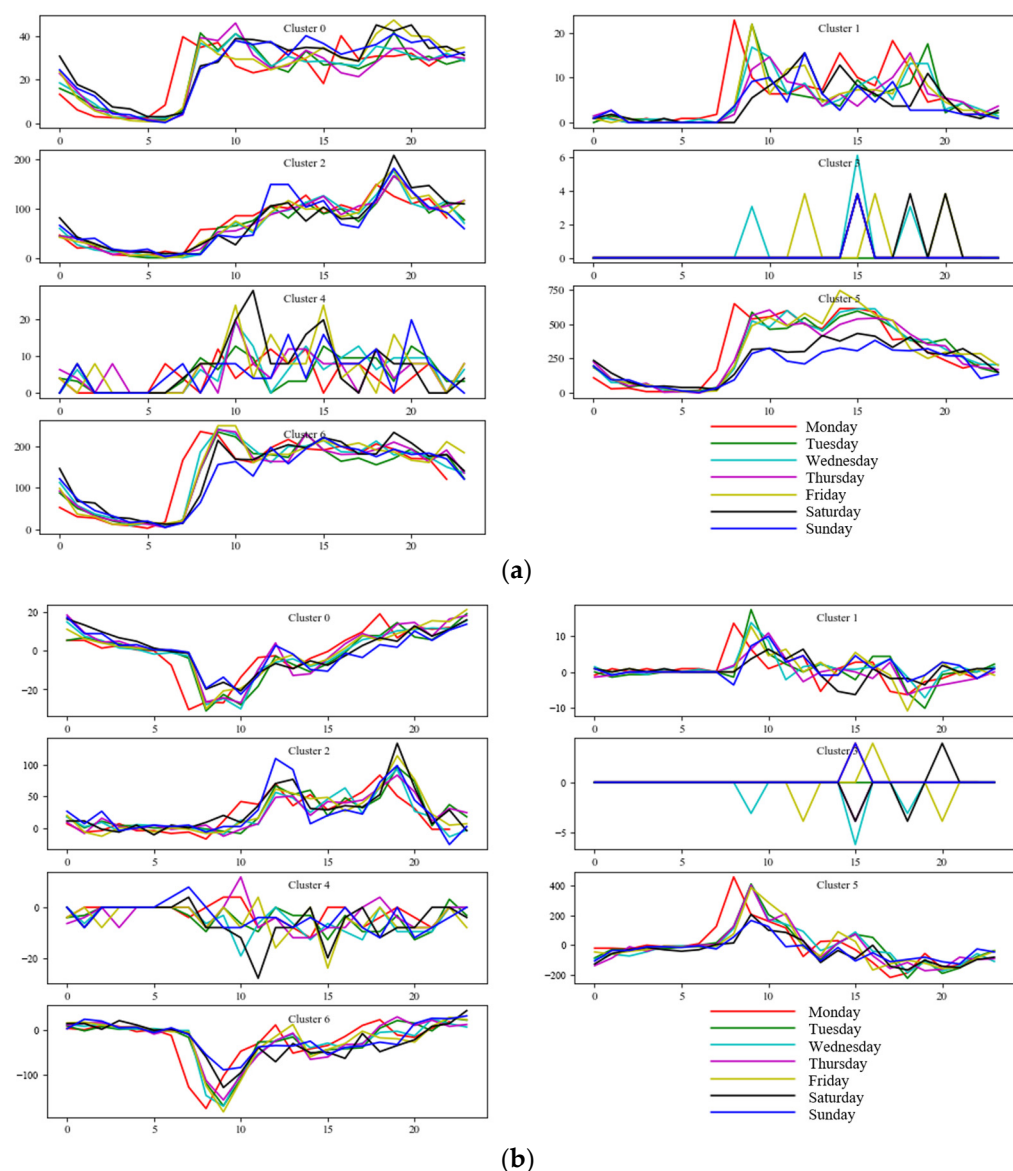


Figure 10. Curves of the (a) traffic volume density and (b) inflow volume density for each cluster-center.

4.2.2. Results of Jobs–Housing Site Identification

To determine the urban jobs–housing sites more clearly, we further defined the jobs–housing characteristics of each cluster. The jobs–housing factor distribution of each category is shown in Figure 11.

We determined the jobs–housing sites by comparing the *JHF* of each regional category. The curve shows that the *JHF* of C3–C4 is near 0, and the jobs–housing characteristics are not obvious. The possibility of belonging to the jobs–housing sites is low. The largest *JHF* is C5, followed by C1. C5 and C1 have a small gap and their values are much higher than 0. Both have strong workplace characteristics. The *JHF* values of C0 and C6 are far less than 0 and show strong residential characteristics. Although the *JHF* of C3 is less than 0, the residential characteristics are not as strong as those of C0 and C6. Therefore, C1 and C5 were identified as workplace areas and C0 and C6 as residential areas, which is consistent with the above analysis of residents' travel characteristics. The final identification results are shown in Figure 12. There are a total of 1730 plots of residence and 882 plots of workplace identified.

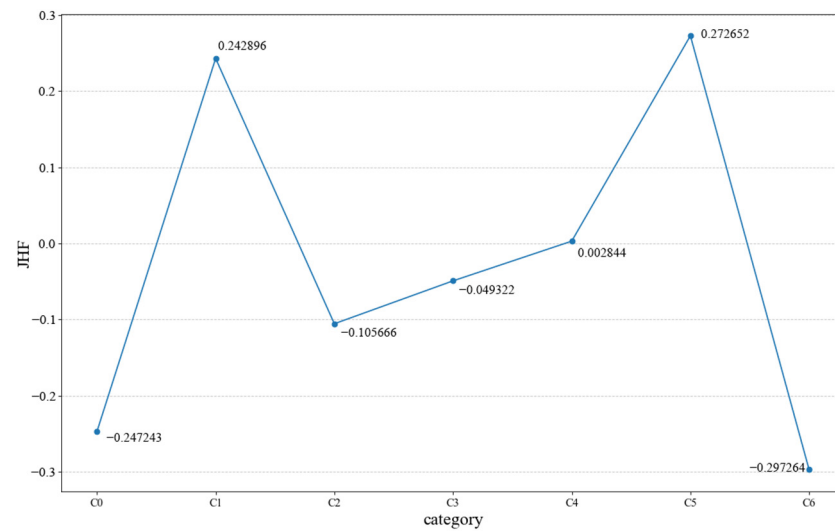


Figure 11. Jobs-housing factor curve diagram for each regional category.



Figure 12. Distribution map of the jobs-housing sites (verification plots are the true values selected in the confusion matrix).

4.3. Accuracy Verification

Due to the challenges of visual interpretation and the complexity of urban areas, it was difficult to obtain the true value of all plots in the study area with the available information. Therefore, we randomly selected 100 typical plots from the identification results; these plots were compared and a confusion matrix was constructed, as shown in Table 1.

Table 1. Confusion matrix of jobs–housing site identification results.

Actual Class	Predicted Class		
	Residence	Workplace	Other Regions
Residence	39	1	5
Workplace	0	17	1
Other regions	5	3	29

The Kappa coefficient calculated based on the confusion matrix is 0.7637, and the accuracy of jobs–housing site identification is 88.89%, which proves that this study has a high accuracy in identifying jobs–housing sites. To reduce the effect of randomness on the accuracy of the result validation, the plots were randomly selected several times for the Kappa coefficient calculation. In the extraction results, the highest Kappa coefficient reaches 0.8101 and the average value is 0.7493, and the highest accuracy reaches 93.33% and the average value is 86.79%, so the overall reliability of this study is relatively high. The identification results have a high consistency with the actual jobs–housing sites in Chengdu. In previous research on jobs–housing site identification, Song [34] obtained 88.9% accuracy based on cell phone signaling data, and Jian [35] obtained 90.4% accuracy based on social media data, indicating that the identification accuracy of the jobs–housing site identification model proposed in this study is excellent. The online car-hailing data are reliable and accurate for identifying jobs–housing sites.

4.4. Distribution Characteristics of Jobs–Housing in Chengdu City

We analyzed the results of the identified distribution of the jobs–housing sites. Affected by the natural topography and urban development model, the overall distribution of the urban jobs–housing sites in the central urban area of Chengdu is circular in shape, which is a typical plain city layout. The residence and workplace areas show different distribution patterns.

4.4.1. Residences with the Coexistence of Continuous Piecewise Aggregation and Small-Scale Aggregation

Tianfu Square is the center of Chengdu city, around which the Sichuan Provincial Government and other units have developed an aggregation area. It is the political center and economic center of Chengdu city. Its high-speed growth in regional economic development drives the continuous spread of residences around it, showing the characteristics of continuous piecewise aggregation. There is a clear tendency of a distribution close to the side of the West Third Ring Road and in the city center around Shuangnan Street, Hongpailou Street, and other high-density residence-concentrated contiguous distributions. However, the agglomeration of residences in the fringe area is significantly lower than that around the city center, and the distribution of residences decreases towards the outer ring area. In the Fourth Ring Road area—located far away from the city center due to the establishment of ecological environment elements such as the giant panda breeding base, wetland parks, and forest parks—the residence areas are isolated from each other, showing the characteristics of multi-point and small-scale aggregation distribution. This area converges mainly along important traffic roads, ensuring a good traffic environment to connect with the city center. Residence aggregation areas like this are represented by the Kaiyuan International and Huamei Community on Dafeng Street and the aggregation areas represented by Longjingyuan and Zhonghai Jincheng on Cuqiao Street. In addition, because the Fourth Ring area is far from the city center and has a beautiful environment, residences such as Lafei Manor which cover a large area of greenery have been built to meet the modern concept of living. Therefore, the layout of these residences forms a continuous piecewise aggregation around the center and a small aggregation at multiple points away from the center.

4.4.2. Workplaces with Spatial Differentiation and Polycentricity Clustered Significantly

Workplaces are mainly distributed in the Fourth Ring area, but the distribution is not as concentrated as that of the residences. They are scattered in different regions, and the distribution is relatively scattered between aggregation sites, which are spatially independent of each other. The spatial scales are all relatively close, such as Fester Enterprise Park, Chengdu Intelligent Information Industrial Park, and Western Intelligence Valley. These areas show different distribution forms, such as clusters and contiguous sheets, and are mainly dominated by industrial parks, manufacturing, and other company enterprises. In addition to these dense workplace areas, there are many workplace areas scattered within the Third Ring Road. These regions are concentrated towards the city center, mainly in the service industry, such as government units and office buildings, where they occupy a small area and are mixed with other functions. Therefore, the spatial distribution of workplaces has the characteristics of obvious differentiation and significant polycentric aggregation.

4.5. Comparative Analysis

To compare the jobs–housing site identification model proposed in this study with the traditional jobs–housing site identification method and establish that the extension of the jobs–housing factor method in this study is effective, we used the jobs–housing factor method for jobs–housing site identification on the online car-hailing data. We calculated the *JHF* of each study unit and plotted a histogram of the calculated results (Figure 13). It can be seen that *JHF* follows a skew-normal distribution in the research area, with a mean value of -0.1 , a standard deviation of 0.3 , a first quartile of -0.3 , and a third quartile of 0.1 . This indicates that the number of residences in Chengdu is greater than the number of workplaces, which is consistent with the identification results above. Therefore, to expand the probability of residences being identified and reduce the probability of workplaces being identified, we determined the threshold based on the quartiles and moved the quartiles to the right by one mean value unit (that is, we set the areas with *JHF* values greater than 0.2 as a workplace area and less than -0.2 as a residential area). We then carried out confusion matrix construction and Kappa coefficient calculation based on the same 100 plots mentioned in Section 4.3, and the results are shown in Table 2.

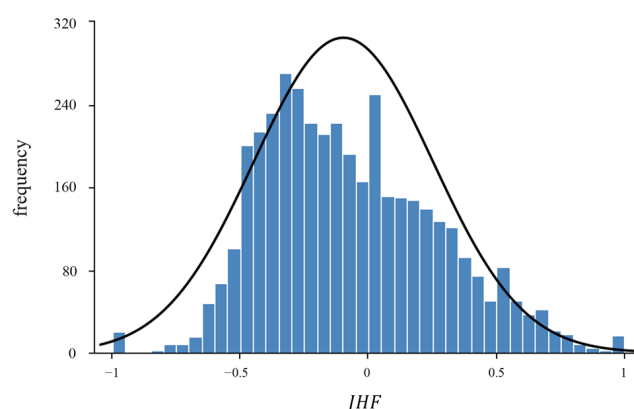


Figure 13. The histogram of *JHF*.

Table 2. Confusion matrix of the division results of the jobs–housing factor method.

Actual Class	Predicted Class		
	Residence	Workplace	Other Regions
Residence	36	3	6
Workplace	0	13	5
Other regions	10	8	19

Table 2 shows that the identification results of the jobs–housing sites obtained directly by the jobs–housing factor method were more over-identified, with 48.65% of the other regions being incorrectly identified as jobs–housing sites. We found that 27.78% and 20% of the plots in the jobs–housing sites were incorrectly identified. There were areas with a weaker jobs–housing characteristic in the peak period that were not identified, and the Kappa coefficient was 0.4990. To reduce randomness, the average Kappa coefficient obtained after multiple plot sampling was 0.5074, which was lower than the Kappa coefficient obtained by the method in this study, indicating that the identification effect of the method in this study is better than that of the jobs–housing factor method.

5. Discussion

In the context of globalization and urbanization, the research on jobs–housing distribution plays an important role in urban development and urban management. The concept of jobs–housing balance can be traced back to Howard’s “Garden Cities” theory in the late 19th century [36], which advocated the idea of mutual proximity and a balanced layout of jobs–housing sites. Mumford [37] further elaborated this and proposed the concept of “balance”, and jobs–housing balance has become an important guiding ideology for alleviating urban problems. As the basic research work of the jobs–housing balance planning problem, the proposed jobs–housing site identification method can provide jobs–housing site distribution data to support the mitigation of urban problems and play a positive role in facing the complex global jobs–housing space problems and sustainable development challenges, which is conducive to the promotion of jobs–housing balance in major cities around the world and the expansion and enrichment of jobs–housing balance related theories. In particular, with the promotion of big data, it is convenient to obtain the spatio-temporal behavior characteristics of urban residents’ travel. There is a great possibility that the model of this paper can be used in other cities, and its research ideas have a certain potential for understanding the current situation of the jobs–housing space and carrying out jobs–housing balance planning.

This study identified urban jobs–housing sites based on online car-hailing data. The results demonstrate that the online car-hailing data have a high reliability for the identification of a large range of residents’ jobs–housing sites, which is an ideal data source for quickly obtaining the jobs–housing sites of urban residents. To make a more accurate identification of jobs–housing sites, this paper explored a similarity measure method adapted to the characteristics of the online car-hailing data, fully considering the operational efficiency, mutagenicity, and growth patterns of the time series. The results demonstrate that the time series mining algorithm based on the LB_Lance distance and dissimilarity index of the time series has great potential in identifying jobs–housing sites.

The jobs–housing factor method belongs to the self-defined time period models, and its empirical division of the time periods cannot tap into the complete temporal pattern of regions. It is easily restricted by the peak period characteristics and the threshold selection, and it is not conducive to the accurate reflection of jobs–housing functions. For example, in a scenic area, tourists flood into the scenic area during the morning peak and migrate to other areas by online car hailing after sightseeing during the evening peak. If we only consider the inflow and outflow of the area during the peak periods, it is likely we would identify this area as a workplace area, resulting in over-identification. Therefore, it is not sufficient to consider only peak period characteristics. The proposed identification model of jobs–housing sites in this study divides urban areas based on the similarity of the online car-hailing time series. Time series mining is performed on the time series characteristics of residents throughout the day, which can allow the general areas and jobs–housing sites to be divided into different categories, avoiding the above-mentioned accidental situation and making up for the shortcomings of the traditional method. Then, we introduced the jobs–housing factor as a feature variable to distinguish jobs–housing sites from other regions in the urban classification results, so that the semi-supervised learning results can reflect the urban jobs–housing sites characteristics and improve the effectiveness of identifying

jobs–housing sites. Therefore, it is important to make full use of the similarity of the time series and the feature variable of jobs–housing sites to more clearly describe the impact of human activities on jobs–housing sites, which is meaningful for the identification of urban jobs–housing sites.

These research results can be used as a basis for urban planning and as basic data for researchers. Through jobs–housing site identification to obtain accurate information on the distribution of jobs–housing sites, urban managers can make scientific decisions, determine the appropriate construction land and functional types, and reasonably plan traffic on roads to improve urban ecology and optimize resource allocation. This will improve the quality of life for urban residents. It can also provide basic data for researchers to evaluate the urban structure, facilitate deeper excavation of urban jobs–housing space relationships, explore the related factors affecting jobs–housing separation, and contribute to the realization of jobs–housing balance. This research provides a new idea for the construction of a jobs–housing site identification method, which can be extended to the research on the jobs–housing space in other cities of different scales. The method is based on the characteristics of human activity to determine the distribution characteristics of urban jobs–housing sites, and it can provide support and guidance for healthy urban development and refined urban management and promote the digitalization of urban management.

6. Conclusions

The identification of jobs–housing sites is necessary to evaluate urban spatial structures and maintain sustainable urban development, but most previous studies have focused on the peak periods. Prior studies seldom deeply explored time series characteristics. In this work, the study area was divided based on the OSM road network data, and online car-hailing order data were converted into a time series dataset. We performed semi-supervised learning on the online car-hailing time series by combining K-medoids clustering and KNN classification. The similarity measure was implemented by using the LB_Lance distance and dissimilarity index of the time series to obtain the urban area classification results. On this basis, we combined the jobs–housing factor to identify residences and workplaces and conducted a comprehensive verification with real ground object information. Our results indicate that each process of the method is reasonable and effective, which makes up for the shortcomings of the traditional method. The identification results can be used as the evaluation basis for a further analysis of the jobs–housing structure, which is of great significance to the management and planning of Chengdu and other western cities. A total of 1730 plots of residence and 882 plots of workplace were finally identified, and the average kappa coefficient was 0.7493, which is better than the traditional method. The overall distribution of the urban jobs–housing sites in the central urban area of Chengdu is circular in shape. Residences exist in both piecewise aggregations and small-scale aggregations, while workplaces with spatial differentiation and polycentricity are clustered significantly.

This study utilized big data in the quantitative research of jobs–housing sites. The fine-grained urban residents' spatial activity characteristics described by the online car-hailing data provide the possibility for a large-scale, large-sample-size, and real-time jobs–housing space evaluation. It can play a supporting role for promoting sustainable urban development. Our research explores the feasibility of using online car-hailing data for jobs–housing site identification and extends the traditional identification method. However, there were still some limitations to this study, and future work should address the following issues. Online car-hailing data only represent a part of the residents' travel pattern. In the future, we will combine other transportation data and surveys, consider the integration of multiple data types, and carry out in-depth research on the identification method of jobs–housing sites using commuting behavior and transportation mode. We will establish more objective and accurate evaluation indicators to verify the identification results to create a more accurate evaluation system.

Author Contributions: L.W., S.B. and S.L. conceived and designed the experiments; L.W. and L.Z. performed the experiments; L.W., S.B. and C.Y. wrote the Chinese paper; L.W., S.B. and S.L. translated the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the National Natural Science Foundation of China (No. 41971340 and No. 41271410).

Institutional Review Board Statement: Approval for the study was not required in accordance with local/national legislation.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, Shuoben Bi, upon reasonable request.

Acknowledgments: We thank LetPub (www.letpub.com (accessed on 4 May 2021)) for its linguistic assistance during the preparation of this manuscript. The authors would like to thank the handling editor and anonymous reviewers for their careful reading and helpful remarks.

Conflicts of Interest: The authors declare that they do not have any competing interests.

References

1. Ma, L.J.C. Urban transformation in China, 1949–2000: A review and research agenda. *Environ. Plan. A Econ. Space* **2002**, *34*, 1545–1569. [\[CrossRef\]](#)
2. Sultana, S. Job/Housing Imbalance and Commuting Time in the Atlanta Metropolitan Area: Exploration of Causes of Longer Commuting Time. *Urban Geogr.* **2002**, *23*, 728–749. [\[CrossRef\]](#)
3. Niu, Y.; Gao, Q. Study on the spatial relationship between work and residence in Zhengzhou based on commuting characteristics. *Shanxi Archit.* **2021**, *47*, 20–21+55. [\[CrossRef\]](#)
4. Jiang, Y.; Zheng, H.; Yu, S.; Tang, X. Relationship Between Job-Housing Spatial Distribution and Rail Transit Network in Tianjin: An Analysis Based on Cellular Data. *Urban Transp. China* **2018**, *16*, 26–35. [\[CrossRef\]](#)
5. Niu, X.; Ding, L.; Song, X.; Li, M. Development of Suburban New Towns in Shanghai: Jobs-Housing Spatial Relationship Analysis. *China City Plan. Rev.* **2018**, *27*, 15–23.
6. Chen, L.; Xiong, C.; Cai, M. A comprehensive decision-making algorithm of residence and workplace based on the identification of cellular signaling track points. *Acta Sci. Nat. Univ. Sunyatseni* **2022**, *61*, 106–116. [\[CrossRef\]](#)
7. Sari Aslam, N.; Cheng, T.; Cheshire, J. A high-precision heuristic model to detect home and work locations from smart card data. *Geo Spat. Inf. Sci.* **2019**, *22*, 1–11. [\[CrossRef\]](#)
8. Liu, Z. Research on Resident' Commuting and Station Classification in Wuhan Using SCD and POI. Master's Thesis, Wuhan University, Wuhan, China, 2018.
9. Zhou, J.; Long, Y. Jobs-housing balance of bus commuters in Beijing: Exploration with large-scale synthesized smart card data. *Transp. Res. Rec.* **2014**, *2418*, 1–10. [\[CrossRef\]](#)
10. Shi, G. Using Weibo Check-in Data Analysis Jobs-housing Balance and Commute Characteristics, A Case Study in Shenzhen. Master's Thesis, Wuhan University, Wuhan, China, 2017.
11. Zhao, X. Research on Spatial Distribution of Jobs-Housing and Commuting Behaviours Based on Urban Big Data Analyzing. Master's Thesis, Tsinghua University, Beijing, China, 2019. [\[CrossRef\]](#)
12. Wei, Y. Research on the Spatial-temporal Characteristics of Residents' Travel and the Relationship between Job and Housing Based on Multi-source Data Mining. Master's Thesis, Chang'an University, Xi'an, China, 2020. [\[CrossRef\]](#)
13. Mao, F.; Ji, M.; Liu, T. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. *Front. Earth Sci.* **2016**, *10*, 205–221. [\[CrossRef\]](#)
14. Fu, X.; Sun, M.; Sun, H. Taxi Commute Recognition and Temporal-spatial Characteristics Analysis Based on GPS Data. *China J. Highw. Transp.* **2017**, *30*, 134–143. [\[CrossRef\]](#)
15. Zhang, Y.; Liu, J.; Wang, Y.; Cao, Y.; Bai, Y. Research on the Method of Urban Jobs-Housing Space Recognition Combining Trajectory and POI Data. *Int. J. Geo-Inf.* **2021**, *10*, 71. [\[CrossRef\]](#)
16. Zhang, Y. Research on Method of Urban Jobs-Housing Space Identification Based on Open Travel and POI Data. Master's Thesis, Lanzhou Jiaotong University, Lanzhou, China, 2020. [\[CrossRef\]](#)
17. Zheng, Z. Spatial Distribution Characteristics and Influencing Factors of Occupation and Residence in Zhuhai: Investigation and Analysis Based on Interest Points of Gaode Map and Traffic Sensing Data. *Planners* **2019**, *35*, 51–56.
18. Qi, G.; Li, X.; Li, S.; Pan, G.; Wang, Z.; Zhang, D. Measuring social functions of city regions from large-scale taxi behaviors. In Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops, Seattle, WA, USA, 21–25 March 2011; pp. 384–388. [\[CrossRef\]](#)
19. Liu, Y.; Wang, F.; Xiao, Y.; Gao, S. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landsc. Urban Plan.* **2012**, *106*, 73–87. [\[CrossRef\]](#)
20. Yue, D. Urban Restructure and Renewal in Chengdu City. Master's Thesis, Southwest Jiaotong University, Chengdu, China, 2007.

21. The GAIA Open Data Initiative. Available online: <http://outreach.didichuxing.com:8080/app-vue/dataList> (accessed on 21 February 2021).
22. Liu, X.; Tian, Y.; Zhang, X.; Wan, Z. Identification of Urban Functional Regions in Chengdu Based on Taxi Trajectory Time Series Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 158. [\[CrossRef\]](#)
23. Li, Y.; Tu, Z.; Liu, Y.; Tang, M.; Wang, N. Urban Functional Area Identification Based on Similarity of Time Series. *Geospat. Inf.* **2021**, *19*, 4+22–29+47.
24. Mao, F. Mining Commuting Pattern and Urban Jobs-housing Balance from Multi-Source Mobile Trajectory Data. Ph.D. Thesis, East China Normal University, Shanghai, China, 2015.
25. Liu, X. Identification and Analysis of Urban Functional Areas in Chengdu Based on Multisource Data. Master's Thesis, Southwest University, Chongqing, China, 2020. [\[CrossRef\]](#)
26. Yi, B.K.; Jagadish, H.V.; Faloutsos, C. Efficient Retrieval of Similar Time Sequences Under Time Warping. In Proceedings of the 14th International Conference on Data Engineering, Orlando, FL, USA, 23–27 February 1998; pp. 201–208. [\[CrossRef\]](#)
27. Kim, S.W.; Park, S.; Chu, W.W. An index-based approach for similarity search supporting time warping in large sequence databases. In Proceedings of the 17th International Conference on Data Engineering, Heidelberg, Germany, 2–6 April 2001; pp. 607–614. [\[CrossRef\]](#)
28. Keogh, E.; Ratanamahatana, C.A. Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* **2005**, *7*, 358–386. [\[CrossRef\]](#)
29. Chouakria, A.D.; Nagabhushan, P.N. Adaptive dissimilarity index for measuring time series proximity. *Adv. Data Anal. Classif.* **2007**, *1*, 5–21. [\[CrossRef\]](#)
30. Chen, Y.; Hu, B.; Keogh, E.; Batista, G.E. DTW-D: Time series semi-supervised learning from a single example. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 383–391. [\[CrossRef\]](#)
31. Xi, X.; Keogh, E.; Shelton, C.; Wei, L.; Ratanamahatana, C.N. Fast time series classification using numerosity reduction. In Proceedings of the 23rd International Conference on Machine learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 1033–1040. [\[CrossRef\]](#)
32. Peter, R.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [\[CrossRef\]](#)
33. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *1*, 224–227. [\[CrossRef\]](#)
34. Song, S.; Li, W.; Yang, D. Research on the Methods of Home Identification Based on Mobile Phone Data. *China Transp. Rev.* **2015**, *37*, 72–76.
35. Jian, Z. Research on Distribution Features of Home-Work Location and Commuting Features Based on Multi-Source Data in Shenzhen. Master's Thesis, Wuhan University, Wuhan, China, 2018.
36. Howard, E. *Garden Cities of Tomorrow*; Faber: London, UK, 1946; pp. 9–28.
37. Mumford, L. The Urban Prospect. *Urban Stud.* **1968**, *6*, 246–248.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.