



# Data-driven behavioral analysis and applications: A case study in Changchun, China



Xianghua Li<sup>a</sup>, Yue Deng<sup>b</sup>, Xuesong Yuan<sup>c</sup>, Zhen Wang<sup>a</sup>, Chao Gao<sup>a,b,\*</sup>

<sup>a</sup> School of Artificial Intelligence, Optics, and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China

<sup>b</sup> College of Computer and Information Science, Southwest University, Chongqing 400715, China

<sup>c</sup> Ansteel Company Limited Cold-Rolling Silicon Steel Mill, Anshan 114001, China

## ARTICLE INFO

### Article history:

Received 7 July 2021

Received in revised form 1 January 2022

Available online 4 March 2022

### Keywords:

Mobile data

Functional area identification

Student behaviors

## ABSTRACT

The mobile phone data have become crucial in behavioral analysis to detect habits of human mobility and reveal rules of behaviors. Previously, questionnaires were often used to identify urban functional areas, with vast labor and poor timeliness. To address the issue, this paper applies data-driven behavioral analysis to identify functional areas for governments to construct urban design, offer site selection and manage transportation. Moreover, data-driven behavioral analysis can also be applied in student behaviors to help schools adjust facility arrangements, develop learning efficiency and provide high-quality services. Therefore, based on mobile phone data in Changchun, this paper utilizes a two-stage clustering method combining human mobility to identify urban functional areas, including business, working, residential and low passenger-flow areas. The interesting finding is that local prosperity in Changchun is prominent and the proportion of low passenger-flow areas can reflect the development level. Furthermore, this paper compares student behaviors in three schools, which shows each school varies in distribution features of students. Experiments provide enlightening insights to reveal the spatial structure of cities and comprehend the living state of students.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

City, as a complex system, integrates a range of travel behaviors [1,2]. Mobile phone data hide a huge amount of useful information and can be utilized in behavioral analysis, with advantages of wide coverage, large sample size and strong real-time performance [3–7]. Based on these data, behavioral analysis serves as an important foundation for understanding mobility rules of behaviors, analyzing spatiotemporal characteristics of citizens and managing transportation [8–14]. Behavioral analysis combines spatial attributes and rules of human mobility to provide physical and social information for the identification of urban functional areas [2,7,15,16]. The scientific and accurate functional area planning can effectively ease the pressure of living expenses for residents, such as housing purchases and job accessibility [17–20]. Moreover, data-driven behavioral analysis can also be applied in student behaviors for students to improve their learning efficiency and manage their time wisely [21].

In recent years, many researchers utilize point-of-interest (POI) data or remote sensing technology to identify urban functional areas [16,22,23]. POI data provide an effective method on account of its rich information about places, such

\* Corresponding author at: School of Artificial Intelligence, Optics, and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China.

E-mail address: [cgao@nwpu.edu.cn](mailto:cgao@nwpu.edu.cn) (C. Gao).

**Table 1**

Temporal location records of mobile phone users.

Symbol	Description	Example
n	The number of users following this daily mobility pattern	1343
h##	The location ID of a user in an hour from ##:00 to ##:59	10004

**Table 2**

The location information of cell towers.

Symbol	Description	Example
i	The location ID of cell towers	10008
Longitude	The longitude of cell towers	125.2479616
Latitude	The latitude of cell towers	43.9772117
Address	The address of cell towers	Shengtian Road, Lvyuan District

as physical locations and attribute information [15,24]. For instance, Hu et al. analyze the main function and spatial distribution features of detailed functional areas based on the frequency density and the ratio of POI function types [25]. Zhang et al. propose the hierarchical semantic cognition (HSC), which relies on the geographic cognition and considers four semantic layers, in order to recognize urban functional areas in Beijing with POI data [26]. However, these models based on POI data are too complex and the mechanism behind the development of functional areas needs more sufficient research [25]. Remote sensing technology is also applied in the identification of functional areas [27–30]. For example, Aubrecht et al. introduce a novel top-down approach to identify and distinguish areas of mixed use based on the nighttime light (NTL) images [31]. Yan et al. propose a novel scene classification framework to identify dominant urban land use type at the level of traffic analysis zone by integrating probabilistic topic models and support vector machine [32]. However, the remote sensing does not take human mobility into full account so that it has difficulty in identifying functional areas with interaction patterns.

For student behaviors, most researches focus on the behaviors in the class or library solely and meticulously [33–35]. For example, Shelburne et al. investigate usage behaviors and attitudes toward e-books in the library and reveal their predictions of electronic and print book materials [36]. Hollo et al. find that the whole-group instruction is the most frequent format in teaching, but small-group lessons have higher individual opportunities of response and active engagement [37]. However, there are few global researches on student behaviors considering the full life of a student such as a study, rest and sport.

Therefore, this paper applies data-driven behavioral analysis into two aspects. More specifically, we identify the functional areas and detect the mobility patterns by behavioral analysis. The main contributions are as below.

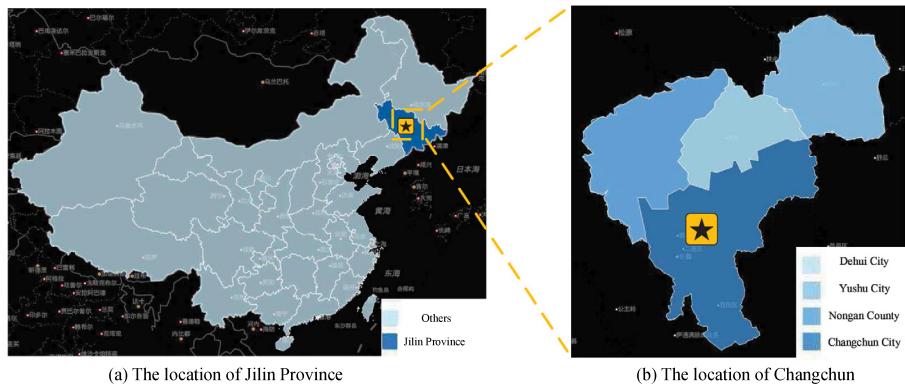
- This paper identifies four kinds of urban functional areas by behavioral analysis through a two-stage clustering method, displaying the prominent local prosperity in Changchun and the relationship of the development level and functional areas. It is helpful to better plan urban development and achieve more precise urban governance.
- This paper reveals mobility patterns of students by behavioral analysis in three schools of Changchun. The distribution features of students in different schools are various. Mobility patterns can provide advice for schools to understand the living state of students, arrange facilities reasonably and improve learning efficiency.

The remaining of this paper is organized as follows. Section 2 introduces the data records. Following that, Section 3 displays indexes and our clustering method for identifying functional areas. Then, Section 4 shows the identification results in terms of functional areas and human activity, and analyzes the accuracy of identification results. In Section 5, it compares student behaviors in three schools on weekdays and weekends. Eventually, the paper concludes with a brief summary in Section 6.

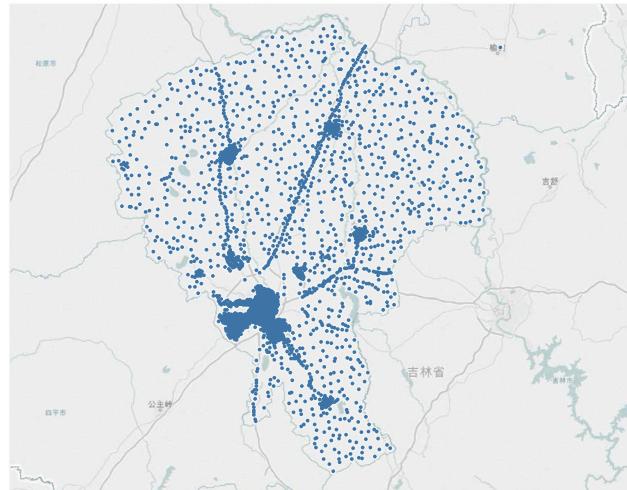
## 2. Data records

Changchun, a large metropolitan city, locates in the middle part of Northeast China and covers the area of 20604 km<sup>2</sup>, with a population of 7.793 million. It is one of the central cities and national comprehensive transportation hubs. Fig. 1 shows the location of Changchun City from the perspective of China and Jilin Province, respectively.

The dataset is composed of two parts, including the mobility records and the GPS coordinates of cell towers [38]. More specifically, the mobility data consist of comma-separated values (CSV) files with the hourly location records of over 2,066,000 anonymized mobile phone users for a week, shown in Table 1. If the mobile phone fails to locate due to shutdown or other reasons, the location record will be lost and recorded as 0. Moreover, GPS data contain the specific locations of cell towers, displayed in Table 2. There are 3,406 cell towers in total, whose spatial distribution is shown in Fig. 2. All cell towers have GPS data, but only 2,141 cell towers own mobility data. In order to avoid potential reidentification of users by illegal attacks, we remove the identification for each user and then only retain daily patterns of human mobility.



**Fig. 1.** The geographical location of Changchun. (a) The location of Jilin Province containing Changchun on the map of China. (b) The location of Changchun on the map of Jilin Province.



**Fig. 2.** The spatial distribution of 3,406 cell towers in Changchun. Each blue node represents a cell tower. Most cell towers are clustered in small areas while a small number of cell towers are scattered throughout Changchun.

### 3. Method

In this section, we introduce a group of vectors that depict the human mobility for the identification of urban functional areas and two indexes for the evaluation of clustering algorithms in Section 3.1. Then, Section 3.2 displays the clustering methods for the identification of functional areas.

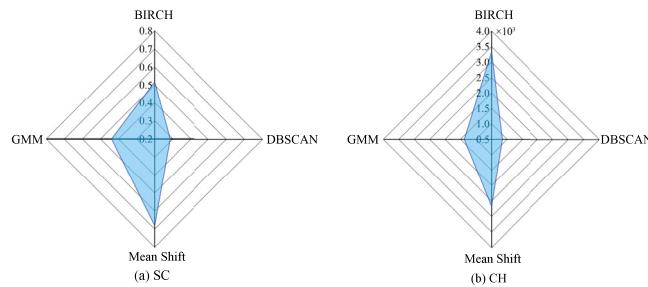
#### 3.1. Metrics

The number of people at different times and locations is a sign of travel needs, which have much to do with regional functionality. Based on this kind of human mobility, the population distribution of cell towers at continuous intervals for the day can reflect the functionality of covered areas. Therefore, this paper takes the temporal distribution feature vector of population (TDFVP) as the index of functional area identification, shown in Eqs. (1) and (2).

$$F_i = (n_1, n_2, \dots, n_t) \quad t \in [1, 24] \quad (1)$$

$$F_c = (m_1, m_2, \dots, m_t) \quad t \in [1, 24] \quad (2)$$

where  $F_i$  is the temporal distribution feature vector of a cell tower  $i$  and  $F_c$  is the temporal distribution feature vector of a clustering  $c$ .  $n_t$  represents the number of users during the time period  $t$  and  $m_t$  represents the average number of users for cell towers in the clustering  $c$  during the time period  $t$ .



**Fig. 3.** The clustering results of four clustering methods are based on (a) Silhouette Coefficient (SC) and (b) Calinski–Harabasz (CH). The mean shift algorithm is a better method in SC and the effect of BIRCH is far better than other algorithms in CH.

Silhouette Coefficient (SC) [39], as an evaluation method of clustering results, can be used to evaluate the effect of different clustering algorithms based on the same original data by combining the cohesion and separation, displayed in Eqs. (3)–(5). The higher the SC value is, the more reasonable the clustering results are.

$$b_i = \min \{b_{i,1}, b_{i,2}, \dots, b_{i,k}\} \quad (3)$$

$$s_i = \frac{b_i - a_i}{\max \{a_i, b_i\}} \quad (4)$$

$$SC = \frac{\sum_1^N s_i}{N} \quad (5)$$

where  $b_{i,k}$  represents the average distance from a sample  $i$  to the remaining samples in the clustering  $k$  whose minimum is  $b_i$ .  $a_i$  means the average distance between the sample  $i$  and other samples in the same clustering.  $s_i$  refers to the Silhouette Coefficient of the sample  $i$ .  $N$  stands for the number of samples.

Calinski–Harabasz (CH) [40] calculates the scores by evaluating the inter-class variance and intra-class variance, which is shown in Eq. (6). The higher the CH score is, the better clustering performance the algorithm will get.

$$CH(k) = \frac{\text{tr } B(k)/(k-1)}{\text{tr } W(k)/(n-k)} \quad (6)$$

where  $n$  represents the number of clusters and  $k$  means the current cluster.  $\text{tr } B(k)$  and  $\text{tr } W(k)$  are the trace of the inter-scatter matrix and intra-scatter matrix, respectively.

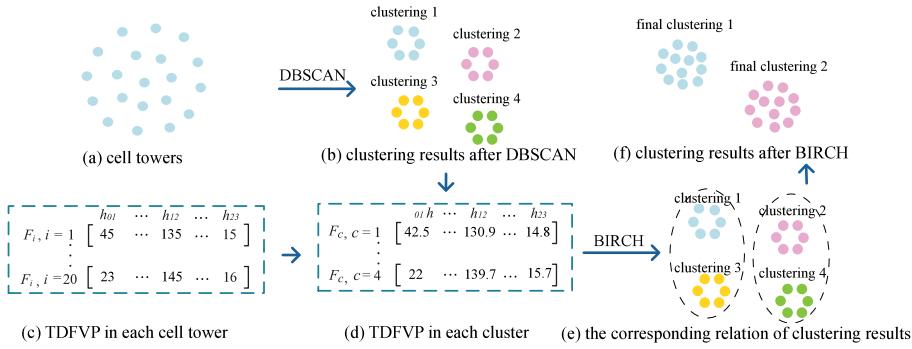
### 3.2. Clustering methods

In order to combine the effect of human mobility, we utilize a two-stage clustering method to identify urban functional areas. More specifically, in the first stage, we apply the DBSCAN algorithm [41] to cluster cell towers only with GSP data. In the second stage, BIRCH algorithm [42] is used to classify cell towers with the temporal distribution feature vector of population (TDFVP).

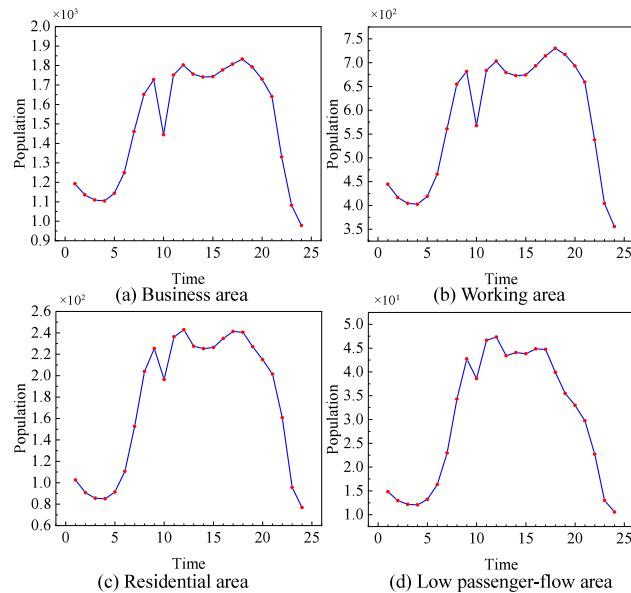
There are some cell towers without mobility data, so we cannot identify their functions by TDFVP. DBSCAN algorithm is developed for clustering spatial points based on the density difference. In most cases, the closer areas are, the more likely they are to display the same functionality. Therefore, for cell towers only with GPS data, DBSCAN algorithm is suitable to identify the functionality.

For the cell towers with TDFVP, we consider four methods to identify functional areas in order to find the more appropriate clustering method. More specifically, BIRCH algorithm [42] is a hierarchical method which is good at clustering large datasets especially. DBSCAN algorithm [41] is a density-based algorithm for discovering clusters in large spatial databases with noises. Mean shift algorithm [43] is a powerful nonparametric technique without the number of clusters and the shape of clusters. GMM algorithm [44] is composed of linear superposition for several Gaussian models. This section compares the four clustering methods based on two indexes (i.e., SC and CH). Fig. 3 shows the BIRCH algorithm performs much better in CH and the mean shift has better performance in SC. However, the mean shift is time-consuming and sensitive to parameters. Therefore, we choose the BIRCH algorithm to cluster cell towers based on TDFVP.

The process of functional area identification is shown in Fig. 4. To begin with, we cluster all 3,406 towers by GPS data using DBSCAN algorithm. Then we regard the averages of TDFVP for cell towers in the same cluster (except the cell towers without TDFVP) as the TDFVP of these clusters. Following that, BIRCH algorithm is utilized to divide functional areas based on TDFVP for clusters. At last, the functionality is identified for each cell tower by the correspondence of clusters. For example, if two clusters divided by the DBSCAN algorithm are classified into one cluster in the BIRCH algorithm, all cell towers in two clusters belong to one cluster eventually.



**Fig. 4.** The flow chart of the clustering process. Firstly, we utilize GPS data in (a) to classify cell towers by the DBSCAN algorithm. By combining clustering results in (b) and the temporal distribution feature vector of population (TDFVP) of each cell tower in (c), we can obtain the TDFVP in each cluster in (d). Following that, BIRCH algorithm is applied to divide the TDFVP of all clusters and the correspondence of two clustering algorithms is shown in (e). The final clustering results are displayed in (f).



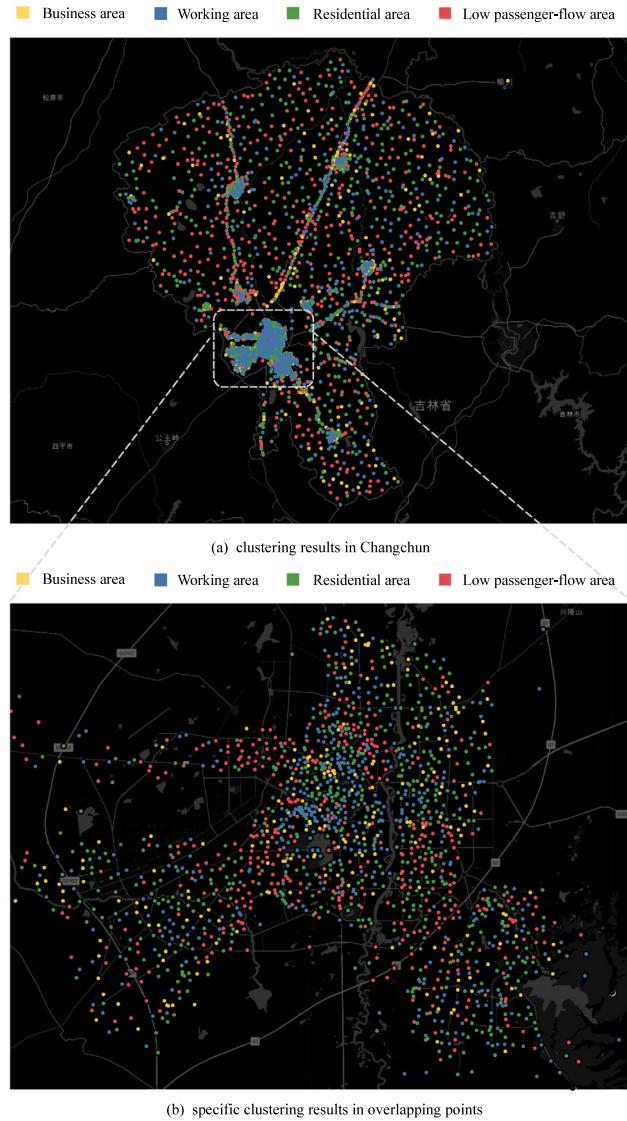
**Fig. 5.** The average temporal distribution population in four kinds of functional areas including (a) business area (b) working area (c) residential area and (d) low passenger-flow area. The x axis represents the time period and the y axis stands for the population of communication by phones.

#### 4. Identification of functional areas by behavioral analysis

##### 4.1. The pattern of functional areas

After applying the two-stage clustering method, there are four clusterings in total. In order to identify the functionality of four clusters, this paper analyzes the temporal distribution feature vector of population (TDFVP) among these cell towers in the same clustering, as shown in Fig. 5. Combined with GPS data of these cell towers, we find business areas, working areas, residential areas and low passenger-flow areas, whose characteristics are shown below.

- Business area: there are thousands of residents in the whole day. The population gradually increases from morning to afternoon except the decrease during 10:00–11:00, keeps slightly stable during 13:00–16:00, peaks around 18:00–19:00 and obviously decreases after 21:00.
- Working area: there is a large population during working time from almost 8:00 to 19:00 with the abrupt decline during 10:00–11:00.
- Residential area: the population from 18:00 to 20:00 is stable compared with other three types maybe because residents constantly come back from companies during the period.
- Low passenger-flow area: the population is small in the whole day.



**Fig. 6.** The map of clustering results. There are four clusters in Changchun including business areas, working area, residential area and low passenger-flow area.

The final clustering results of all cell towers are shown in Fig. 6 with four kinds of functional areas. There are some overlapping points on the map, so we magnify the part with the most overlapping points in Fig. 6(b). More specifically, business areas account for 11.19% and working areas occupy 24.54% of all areas. Then the proportion of residential areas is 30.80% and that of low passenger-flow areas is 33.47%. The specific statistical results of functional areas in each district are displayed in Table 3. The results show that there is a negative correlation between the proportion of low passenger-flow areas and the developmental level in each district. For example, Erdao District, Chaoyang District, Nuanguan District, Kuancheng District and Lvyuan District are the main urban areas of Changchun. The proportion of low passenger-flow areas in these five districts is lower than the other four districts. Moreover, Nong'an County, in the relatively remote areas of Changchun, owns the lowest proportion of business areas and the largest proportion of low passenger-flow areas, which is consistent with the actuality of Nong'an County. Therefore, the proportion of low passenger-flow areas can reflect the developmental level of districts to some extent.

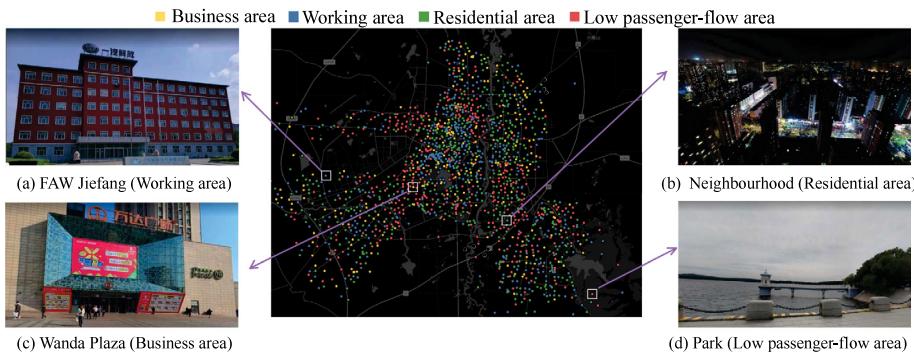
#### 4.2. Accuracy assessment

In order to verify the accuracy of identification results, we compare our results with the urban planning in Changchun. It shows that the identification results are in line with the urban planning in Changchun. Since there is no specific land use

**Table 3**

The statistics of functional areas in each district. BA, WA, RA and LPFA mean business areas, working areas, residential areas and low passenger-flow areas, respectively. Each integer stands for the number of functional areas and each percentage represents the proportion of functional areas.

District	BA	WA	RA	LPFA	Total					
Erdao District	43	11.85%	97	26.72%	140	38.57%	83	<b>22.86%</b>	363	100%
Chaoyang District	58	14.50%	105	26.25%	117	29.25%	120	<b>30.00%</b>	400	100%
Nanguan District	56	8.85%	176	27.80%	202	31.91%	199	<b>31.44%</b>	633	100%
Kuancheng District	55	15.07%	100	27.40%	92	25.20%	118	<b>32.33%</b>	365	100%
Lvyuan District	25	11.21%	57	25.56%	68	30.49%	73	<b>32.74%</b>	223	100%
Dehui District	42	11.60%	86	23.76%	115	31.77%	119	<b>32.87%</b>	362	100%
Shuangyang District	40	17.54%	50	21.93%	62	27.19%	76	<b>33.34%</b>	228	100%
Jutai City	33	10.57%	78	25.00%	88	28.21%	113	<b>36.22%</b>	312	100%
Nong'an County	29	5.58%	87	16.73%	165	31.73%	239	<b>45.96%</b>	520	100%
Total	381		836		1049		1140		3406	
Ratio	<b>11.19%</b>		<b>24.54%</b>		<b>30.80%</b>		<b>33.47%</b>		100%	



**Fig. 7.** The comparison of identification results and reality. We select randomly one sample in each type of functional areas to make comparison with the real functionality. It shows the functionality of these cell towers is in line with the reality.

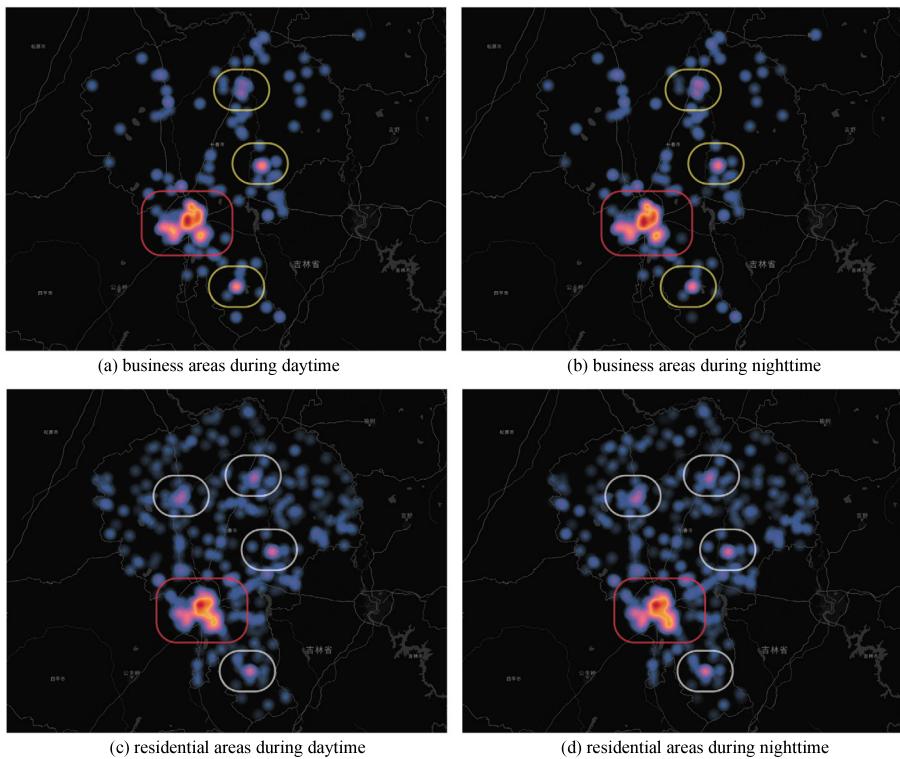
in 2017, we utilize the profit-oriented land supply table in 2018. More specifically, the table only provides the acreage of business areas and residential areas, so we calculate the proportion of business areas and residential areas, respectively. The proportion of business areas is 53.41% and that of residential areas is 46.59% in the table. In our identification results, business areas and residential areas account for 53.71% and 46.29%, respectively. It proves our results are close to the urban planning in Changchun.

The accuracy of our identification results is also proved in Fig. 7. More specifically, we select one cell tower at random in each kind of functional areas and compare its functionality with visual results by using Google Street View. Fig. 7 shows that identification results take Wanda Plaza, FAW Jiefang company, communities and Moon Lake Park as business areas, working areas, residential areas and low passenger-flow areas, respectively. Therefore, our identification results of functional area in this paper are feasible and reliable, and have the potential to provide scientific support for urban planning and policymaking.

#### 4.3. The pattern of human activity

The identification of urban functional areas helps detect the patterns of human activity, which has a profound impact on urban morphology due to the interaction of functional areas and human beings. Based on an existing survey, the morning peak (08:00–09:00) and evening peak (18:00–19:00) are helpful to divide the patterns of human activity into daytime part (8:00–18:59) and nighttime part (19:00–7:59) since residents usually do different kinds of things in a day such as work and entertainment [45]. Therefore, we analyze the population flows during the daytime and nighttime in business and residential areas.

The heat map of population flows is shown in Fig. 8. It shows the junction of Kuancheng District, Lvyuan District, Chaoyang District and Erdao District (i.e., areas circled by the red box) has the most population flows regardless of time and functional area types. Moreover, areas with a large population flows in the daytime also reach high population levels in the nighttime. For business areas, popular areas are the junction mentioned above, central regions of Dehui District, Jutai District and Shuangyang District (i.e., areas circled by yellow boxes). For residential areas, preferred areas locate in the junction, areas around Nong'an station and central regions of Dehui District, Jutai District and Shuangyang District (i.e., areas circled by white boxes). Therefore, the phenomenon of local prosperity is obvious in Changchun. The government can make policies to motivate prosperous areas to drive the development of surrounding areas, thus enhancing the economy of all.



**Fig. 8.** The heat map of population flows in business areas and residential areas during the daytime and nighttime. The finding is that the junction of Kuancheng District, Lvyan District, Chaoyang District and Erdao District (i.e., areas circled by the red box) is prosperous during the whole day.

**Table 4**

The proportion of students in study areas, rest areas and sport areas on Monday. JU, NNU and CUT mean Jilin University, Northeast Normal University and Changchun University of Technology, respectively. \* represents no record since there is no cell tower around sport areas in CUT.

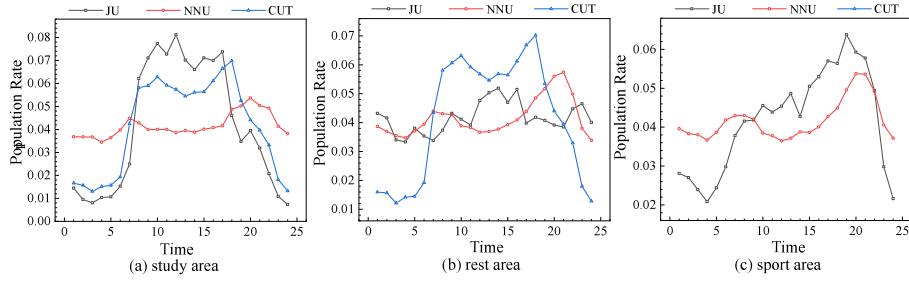
	JU	NNU	CUT
Study areas	38.29%	57.30%	55.18%
Rest areas	27.00%	34.03%	45.67%
Sport areas	38.88%	27.68%	*

## 5. Detection of mobility patterns by behavioral analysis

Mobility data and modern information technology are also helpful to reveal student behaviors in college, detect mobility rules and provide high-quality services for schools and students [46]. In this paper, we analyze student behaviors in terms of study, sport and rest in Section 5.1. Moreover, student behaviors on Monday and Saturday are compared to identify the difference in Section 5.2.

### 5.1. Comparison of student behaviors in three schools

There are three famous schools in Changchun (i.e., Jilin University, Northeast Normal University and Changchun University of Technology). In order to find cell towers located in schools, we first project all cell towers onto the map based on GPS data. Student behaviors in college mainly include study, sport and rest. Therefore, based on these cell towers in schools, then we calculate the proportion of students in study areas (i.e., teaching buildings and libraries), rest areas (i.e., dormitories) and sport areas (i.e., playgrounds and gyms) in each school. Table 4 shows that the proportion of students in study areas in Northeast Normal University is higher than that of other universities. The students in Changchun University of Technology are more likely to relax. \* means there is no cell tower located in sport areas in Changchun University of Technology. For other two schools, Jilin University comes first in the proportion of students in sport areas. The total of proportion for three types is not equal to 1 because some students go to two or three kinds of areas in one day. Moreover, the three proportions of Northeast Normal University have great difference while those in other universities are closer.



**Fig. 9.** The trend of population in (a) study areas (b) rest areas and (c) sport areas on Monday. The x axis is the time period. The y axes are on behalf of the population rate. The results show that students in each school have their own behavior styles.

**Table 5**

The proportion of students in study areas, rest areas and sport areas in three schools on Saturday.

	JU	NNU	CUT
Study areas	35.01%	58.62%	53.65%
Rest areas	19.81%	17.55%	46.94%
Sport areas	46.64%	44.55%	*

Student behaviors vary in different types of areas and different universities. Therefore, in order to find student behaviors more clearly, we visualize the trend of population in three kinds of areas in Fig. 9. Since each area has a different number of students, we utilize population rate to show the trend. Fig. 9(a) displays that the population increase at around 6:00 and decrease at almost 21:00 in study areas. The students in Jilin University and Changchun University of Technology are more likely to study between 8:00 and 18:00 while those in Northeast Normal University prefer studying in the evening. In Fig. 9(b), students in Jilin University relax more at noon while those in other two schools rest more in the morning and afternoon. Students in Northeast Normal University more like morning exercise compared with Jilin University, but on the whole people prefer to work out in the evening, as shown in Fig. 9(c). Therefore, student behaviors of different schools show various styles. The college can adjust the opening and closing hours of the library and change the course arrangement according to student behaviors.

## 5.2. Comparison of student behaviors between monday and saturday

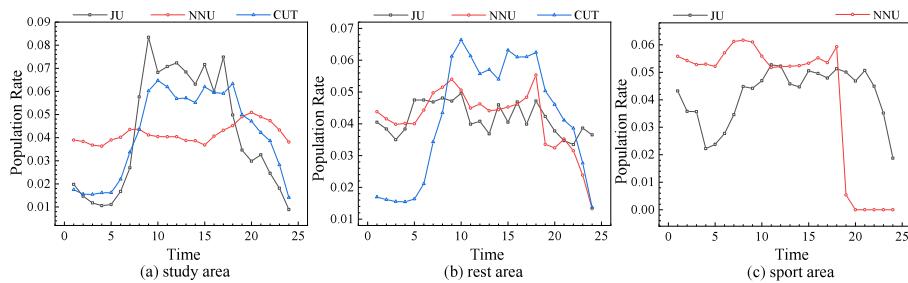
In the previous work, this paper analyzes student behaviors on Monday and those on Saturday also deserve attention. Therefore, this section focuses on the differences of student behaviors between weekdays and weekends.

Compared with Monday in Table 4, Table 5 shows that the proportion of students in study areas decreases 3.28% in Jilin University and 1.53% in Changchun University of Technology on Saturday. Besides, the proportion of students in sport areas increases 7.76% in Jilin University and 16.87% in Northeast Normal University. Therefore, people work out more to relax and balance their study and life on weekends. Moreover, it deserves attention that there is the highest proportion of students in study areas in Northeast Normal University on both Monday and Saturday. Therefore, on the whole, students in Northeast Normal University study harder than other schools.

Although the trend of student behaviors on Saturday bears resemblance to that on Monday, there are still some differences. In comparison with Fig. 9 on Monday, Fig. 10 shows students come back to dormitories earlier since the population rate of rest areas at 24:00 on Saturday equals almost to 0.01 while on Monday equals to 0.03. Moreover, people are less willing to exercise on weekends compared with weekdays. Therefore, students can make full use of spare time on weekends to relax and study according to their own situations. The colleges can hold some activities on weekends to help students adjust emotions and release pressure.

## 6. Conclusion

Through mobile phone data in Changchun, this paper identifies urban functional areas in order to provide suggestions for urban planning and policymaking, including business areas, working areas, residential areas and low passenger-flow areas. We find the development of Changchun focuses too much on certain areas. In each district, low passenger-flow areas almost occupy the largest regions compared with the other three kinds of functional areas. The more prosperous a district is, the smaller the proportion of low passenger-low areas is. On the whole, the layout of urban functional areas in Changchun conforms to the overall requirements of urban planning. Moreover, phone data can reflect student behaviors in schools to some extent, so as to help schools and management departments to choose a suitable timetable for the opening and closing of facilities, such as libraries and stadiums. The interesting finding is that most students are more likely to relax on weekends compared with weekdays.



**Fig. 10.** The trend of population in (a) study areas (b) rest areas and (c) sport areas on Saturday. The x axis is the time period. The y axes are on behalf of the population rate. The results show that students are more willing to study on weekdays while relax on weekends. But students in Northeast Normal University study harder than usual and improve exercise on weekends.

Our future work may combine multi-source data, such as personal cars data, metro trips data and bus trips data, to more accurately identify urban functional areas and detect changes in urban spatial structure. Moreover, we can extend the useful range of mobile phone data to infer the relationships of people and reveal the rule of COVID-19 pandemic epidemiology.

#### CRediT authorship contribution statement

**Xianghua Li:** Conceptualization, Methodology, Writing – review & editing. **Yue Deng:** Data curation, Visualization, Writing – original draft. **Xuesong Yuan:** Resources, Validation. **Zhen Wang:** Project administration, Formal analysis. **Chao Gao:** Supervision, Software, Writing – review & editing.

#### Declaration of competing interest

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

#### Acknowledgments

This work was supported by the Natural Science Fund of Distinguished Young Scholarship of China (No. 62025602), the National Natural Science Foundation of China (Nos. 61976181, U1803263, 11931015), the Natural Science Foundation of Shaanxi Province (No. 2022JM-325), the Fundamental Research Funds for the Central Universities (Nos. D5000210738, D5000210827), in part by the Science and Technology Support Program of Guizhou (No. QKHZC2021YB531) and the Scientific Research Platform Project of Guizhou Minzu University (No. GZMUSYS[2021]04).

#### References

- [1] Junyu Lv, Chen Zhao, An Zeng, Bursty visitation of locations in human mobility, *Physica A* 567 (2021) 125674.
- [2] Fengli Xu, Yong Li, Depeng Jin, Jianhua Lu, Chaoming Song, Emergence of urban growth patterns from human mobility behavior, *Nat. Comput. Sci.* 1 (2021) 791–800.
- [3] Xiaonan Yu, Cesunica Ivey, Zhijiong Huang, Sashikanth Gurram, Vijayaraghavan Sivaraman, Huizhong Shen, Naveen Eluru, Samiul Hasan, Lucas Henneman, Guoliang Shi, Hongliang Zhang, Haofei Yu, Junyu Zheng, Quantifying the impact of daily mobility on errors in air pollution exposure estimation using mobile phone location data, *Environ. Int.* 141 (2020) 105772.
- [4] Danya Bachir, Ghazaleh Khodabandou, Vincent Gauthier, Mounim El Yacoubi, Jakob Puchinger, Inferring dynamic origin-destination flows by transport mode using mobile phone data, *Transp. Res. C* 101 (2019) 254–275.
- [5] Feilong Wang, Cynthia Chen, On data processing required to derive mobility patterns from passively-generated mobile phone data, *Transp. Res. C* 87 (2018) 58–74.
- [6] Fangxia Zhao, Huayan Shang, Role of transportation network on population distribution evolution, *Physica A* 577 (2021) 126076.
- [7] Mariem Fekih, Tom Bellmans, Zbigniew Smoreda, Patrick Bonnel, Angelo Furno, Stéphane Galland, A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of Lyon region (France), *Transportation* 48 (2021) 1671–1702.
- [8] Yimin Wu, Liang Wang, Linghui Fan, Ming Yang, Yu Zhang, Yongheng Feng, Comparison of the spatiotemporal mobility patterns among typical subgroups of the actual population with mobile phone data: A case study of Beijing, *Cities* 100 (2020) 102670.
- [9] Shengjie Lai, Andrea Farnham, Nick W Ruktanonchai, Andrew J Tatem, Measuring mobility, disease connectivity and individual risk: A review of using mobile phone data and mhealth for travel medicine, *J. Travel Med.* 26 (3) (2019) taz019.
- [10] Sihui Guo, Ci Song, Tao Pei, Yaxi Liu, Ting Ma, Yunyan Du, Jie Chen, Zide Fan, Xianli Tang, Yong Peng, Yanbin Wang, Accessibility to urban parks for elderly residents: Perspectives from mobile phone data, *Landsc. Urban Plan.* 191 (2019) 103642.
- [11] Xiping Yang, Zhixiang Fang, Ling Yin, Junyi Li, Yang Zhou, Shiwei Lu, Understanding the spatial structure of urban commuting using mobile phone location data: A case study of Shenzhen, China, *Sustainability* 10 (5) (2018) 1435.
- [12] Umberto Fugiglido, Emanuele Massaro, Paolo Santi, Sebastiano Milardo, Kacem Abida, Rainer Stahlmann, Florian Netter, Carlo Ratti, Driving behavior analysis through CAN bus data in an uncontrolled environment, *IEEE Trans. Intell. Transp. Syst.* 20 (2) (2018) 737–748.

- [13] Gang Zhong, Tingting Yin, Jian Zhang, Shanglu He, Bin Ran, Characteristics analysis for travel behavior of transportation hub passengers using mobile phone data, *Transportation* 46 (5) (2019) 1713–1736.
- [14] Chao Gao, Yi Fan, Shihong Jiang, Yue Deng, Jiming Liu, Xianghua Li, Dynamic robustness analysis of a two-layer rail transit network model, *IEEE Trans. Intell. Transp. Syst.* (2021) <http://dx.doi.org/10.1109/TITS.2021.3058185>.
- [15] Xiaoyi Zhang, Wenwen Li, Feng Zhang, Renyi Liu, Zhenhong Du, Identifying urban functional zones using public bicycle rental records and point-of-interest data, *ISPRS Int. J. Geo-Inf.* 7 (12) (2018) 459.
- [16] Yuanxin Jia, Yong Ge, Feng Ling, Xian Guo, Jianghao Wang, Le Wang, Yuehong Chen, Xiaodong Li, Urban land use mapping by combining remote sensing imagery and mobile phone positioning data, *Remote Sens.* 10 (3) (2018) 446.
- [17] Jinchao Song, Tao Lin, Xinhua Li, Alexander V Prishchepov, Mapping urban functional zones by integrating very high spatial resolution remote sensing imagery and points of interest: A case study of Xiamen, China, *Remote Sens.* 10 (11) (2018) 1737.
- [18] Pengfei Lin, Jiancheng Weng, Yu Fu, Dimitrios Alivanistos, Baocai Yin, Study on the topology and dynamics of the rail transit network based on automatic fare collection data, *Physica A* 545 (2020) 123538.
- [19] Laura Alessandretti, Ulf Aslak, Sune Lehmann, The scales of human mobility, *Nature* 587 (7834) (2020) 402–407.
- [20] Senya Yang, Jianping Wu, Yanyan Xu, Tao Yang, Revealing heterogeneous spatiotemporal traffic flow patterns of urban road network via tensor decomposition-based clustering approach, *Physica A* 526 (2019) 120688.
- [21] Najya A. Attia, Lubna Baig, Yousef I. Marzouk, Anwar Khan, The potential effect of technology and distractions on undergraduate students' concentration, *Pak. J. Med. Sci.* 33 (4) (2017) 860.
- [22] Wei Tu, Zhongwen Hu, Lefei Li, Jinzhou Cao, Jincheng Jiang, Qiuping Li, Qingquan Li, Portraying urban functional zones by coupling remote sensing imagery and human sensing data, *Remote Sens.* 10 (1) (2018) 141.
- [23] Hao Wu, Lingbo Liu, Yang Yu, Zhenghong Peng, Evaluation and planning of urban green space distribution based on mobile phone data and two-step floating catchment area method, *Sustainability* 10 (1) (2018) 214.
- [24] Beibei Yu, Zhonghui Wang, Haowei Mu, Li Sun, Fengning Hu, Identification of urban functional regions based on floating car track data and POI data, *Sustainability* 11 (23) (2019) 6541.
- [25] Yunfeng Hu, Yueqi Han, Identification of urban functional areas based on POI data: A case study of the Guangzhou economic and technological development zone, *Sustainability* 11 (5) (2019) 1385.
- [26] Xiuyuan Zhang, Shihong Du, Zhihia Zheng, Heuristic sample learning for complex urban scenes: Application to urban functional-zone mapping with VHR images and POI data, *ISPRS J. Photogramm. Remote Sens.* 161 (2020) 1–12.
- [27] Yimin Chen, Xiaoping Liu, Xia Li, Analyzing parcel-level relationships between urban land expansion and activity changes by integrating Landsat and nighttime light data, *Remote Sens.* 9 (2) (2017) 164.
- [28] Bo Huang, Bei Zhao, Yimeng Song, Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery, *Remote Sens. Environ.* 214 (2018) 73–86.
- [29] Jinchao Song, Xiaoye Tong, Lizhe Wang, Chunli Zhao, Alexander V Prishchepov, Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach, *Landsat. Urban Plan.* 190 (2019) 103580.
- [30] Shisong Cao, Deyong Hu, Wenji Zhao, You Mo, Chen Yu, Yang Zhang, Monitoring changes in the impervious surfaces of urban functional zones using multisource remote sensing data: A case study of Tianjin, China, *GISci. Remote Sens.* 56 (7) (2019) 967–987.
- [31] Christoph Aubrecht, José Antonio León Torres, Evaluating multi-sensor nighttime earth observation data for identification of mixed vs. residential use in urban areas, *Remote Sens.* 8 (2) (2016) 114.
- [32] Jingli Yan, Weiqi Zhou, Lijian Han, Yuguo Qian, Mapping vegetation functional types in urban areas with WorldView-2 imagery: Integrating object-based classification with phenology, *Urban For. Urban Green.* 31 (2018) 230–240.
- [33] Rachel Applegate, The library is for studying: Student preferences for study space, *J. Acad. Librariansh.* 35 (4) (2009) 341–346.
- [34] Constantinos M Kokkinos, Georgia Panayiotou, Aggeliki M Davazoglou, Correlates of teacher appraisals of student behaviors, *Psychol. Schools* 42 (1) (2005) 79–89.
- [35] Vicki Tolar Burton, Scott A. Chadwick, Investigating the practices of student researchers: Patterns of use and criteria for use of Internet and library sources, *Comput. Compos.* 17 (3) (2000) 309–328.
- [36] Wendy Allen Shelburne, E-book usage in an academic library: User attitudes and behaviors, *Libr. Collect. Acquis. Tech. Serv.* 33 (2–3) (2009) 59–72.
- [37] Alexandra Hollo, Regina G. Hirn, Teacher and student behaviors in the contexts of grade-level and instructional grouping, *Prev. School Fail.: Altern. Educ. Child. Youth* 59 (1) (2015) 30–39.
- [38] Zhanwei Du, Yongjian Yang, Chao Gao, Liping Huang, Qiuyang Huang, Yuan Bai, The temporal network of mobile phone users in Changchun Municipality, Northeast China, *Sci. Data* 5 (1) (2018) 1–7.
- [39] Andreas Hotho, Alexander Maedche, Steffen Staab, Ontology-based text document clustering, *KI* 16 (4) (2002) 48–54.
- [40] Ujjwal Maulik, Sanghamitra Bandyopadhyay, Performance evaluation of some clustering algorithms and validity indices, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (12) (2002) 1650–1654.
- [41] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996, pp. 226–231.
- [42] Tian Zhang, Raghu Ramakrishnan, Miron Livny, BIRCH: an efficient data clustering method for very large databases, *ACM Sigmod Record* 25 (2) (1996) 103–114.
- [43] Yizong Cheng, Mean shift, mode seeking, and clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 17 (8) (1995) 790–799.
- [44] Carl Edward Rasmussen, The infinite Gaussian mixture model, in: *Advances in Neural Information Processing Systems*, 1999, pp. 554–560.
- [45] Yue Deng, Jiaxin Wang, Chao Gao, Xianghua Li, Zhen Wang, Xuelong Li, Assessing temporal-spatial characteristics of urban travel behaviors from multiday smart-card data, *Physica A* 576 (2021) 126058.
- [46] Naoyuki Takeuchi, Takayuki Mori, Yoshimi Suzukamo, Shin-Ichi Izumi, Integration of teaching processes and learning assessment in the prefrontal cortex during a video game teaching-learning task, *Front. Psychol.* 7 (2017) 2052.