

引用格式: 程静,刘家骏,高勇.基于时间序列聚类方法分析北京出租车出行量的时空特征[J].地球信息科学学报,2016,18(9):1227-1239.
[Cheng J, Liu J J, Gao Y. 2016. Analyzing the spatio-temporal characteristics of Beijing's OD trip volume based on time series clustering method. Journal of Geo-information Science, 18(9):1227-1239.] DOI: 10.3724/SP.J.1047.2016.01227

基于时间序列聚类方法分析北京出租车出行量的时空特征

程 静,刘家骏,高 勇*

北京大学遥感与地理信息系统研究所,北京 100871

Analyzing the Spatio-Temporal Characteristics of Beijing's OD Trip Volume Based on Time Series Clustering Method

CHENG Jing, LIU Jiajun and GAO Yong*

Institute of Remote Sensing and Geographical Information System, Peking University, Beijing, 100871, China

Abstract: Citizens' intra-city trips are often influenced by the allocation of resources and urban functional areas, such as the educational areas, entertainment areas, business areas and residential areas. Therefore, citizens' travelling pattern can reflect the city structure and unveil the urban function zoning. Meanwhile, the widespread of GPS vehicle navigation equipment makes it possible to achieve a vast amount of vehicle trajectory. With the support of the vast vehicle trajectory data, we can analyze citizens' travelling mode and understand the city structure. In this paper, we investigated citizens' travelling pattern and the urban functional structure of Beijing with the taxi trajectory data of one-month period and the information of land parcels divided by major roads. To analyze the citizen's travelling mode, we extracted the trip volume time series in every parcel and adopted a new method which could cover the proximity on both the values and the behavior to cluster the time series data. In the end, we discussed the correlation between citizens' travelling mode and urban functions in different regions, based on Beijing's POI data. The result showed that there were obvious differences in the travelling patterns between the weekdays and weekends. During the weekdays, there were two rush hours, which were different from the ordinary commute rush hours. Looking at the clustering results of the weekday data, the spatial distribution of different clusters basically arranged like concentric circles, and the travelling volume of every circle decreased with respect to the increasing distance to its center. The conclusions made in this research are meaningful for the analysis of citizens' travelling mode and for assisting urban transportation planning.

Key words: trajectory data; spatio-temporal characteristic; urban functional structure; travel mode; time series clustering

*Corresponding author: GAO Yong, E-mail: gaoyong@pku.edu.cn

摘要:受城市资源配置、区域功能分化的影响,城市中居民的出行往往呈现出特定的模式和规律,而这种出行模式的背后反映出城市的功能结构。城市车辆GPS导航的广泛使用,以及车辆轨迹数据的大量获取,为分析城市居民出行模式及理解城市功能结构提供了数据支撑。本文以道路分割城市得到的地块为研究单元,利用北京市一个月的出租车轨迹数据,对北京居民的出行模式及城市功能格局进行分析。在轨迹数据分析中,本文从轨迹数据中提取每个地块的出行量时间序列信息,然后采用结合时间序列距离度量和时间序列自身相关性的聚类方法,对出行量时间序列数据进行聚类分析,从而研究乘客出行的时空分布特征,最后结合北京市POI数据,探讨了不同区域乘客出行规律和区域功能类型的相互关系。结果表明,出租车出行量

收稿日期:2015-09-14;修回日期:2015-12-10.

基金项目:国家自然科学基金项目(41271385,41271386)。

作者简介:程 静(1993-),女,湖北黄冈人,硕士生,研究方向为空间数据挖掘。E-mail: cheng1993jing@163.com

*通讯作者:高 勇(1974-),男,辽宁抚顺人,副教授,研究方向为地理信息检索、空间数据挖掘。E-mail: gaoyong@pku.edu.cn

时间序列模式在工作日和周末间存在明显差异。此外,工作日的2个出行高峰与通常的通勤早晚高峰不同。由出行量所得的区域聚类结构,除具有重要交通枢纽功能的地块外,总体上以市中心为圆心大致呈同心圆分布,且距离市中心越远出行量越小。研究结果对于分析北京市居民出行行为、辅助城市交通规划具有一定的意义。

关键词: 轨迹数据;时空特征;城市功能结构;出行模式;时间序列聚类

1 引言

城市居民的出行和城市功能结构联系密切且相互影响。随着城市的发展,由于城市规划以及居民生活方式的需要往往会形成一些不同的功能区,如商业区、学区、住宅区等。城市区域的不同功能类型结构,影响着人们的出行,使居民出行模式呈现一定规律性。通过对城市居民出行的分析挖掘,能更好地了解城市的功能结构和分布。

城市车辆数目的不断增长、车载GPS的普及,使获取大量实时出租车车辆状态和GPS轨迹数据成为可能。这些车辆轨迹数据为人类出行行为的分析、挖掘、研究提供了大量数据资料。

出租车是城市常见的公共交通工具,与公交车、地铁等公共交通工具不同,出租车的线路不固定,随机性很大,不同地区出租车的密度差异较大,是非通勤性出行、经常出行的主要交通工具,且其出行路径、时间、上下车点信息与人类活动关系密切,能较好地反映城市居民的行为模式。相对于出租车轨迹数据中的路段信息,轨迹数据中上下车点更能反映居民的出行特征,以及居民与区域之间的联系。

目前,已对出租车起始点进行城市居民出行模式分析的研究包括:根据乘客上下车点分析热点区域的出行模式^[1];对出租车起点终点形成的流进行聚类,发现同一类中的流的起点和终点具有相似的社区功能^[2];对出租车时空分布,主要上下车点聚集区域进行可视化,分析城市中出租车运营最繁忙的时段^[3];对基于出租车下车和上车差异的时间序列进行聚类分析,得出聚类空间分布与POI代表的用地类型相吻合^[4];结合层次路网分解,居民出行和各区域的POI点数据来揭示不同区域的城市功能^[5]。

本文以主要道路为界,将北京市划分为地块,以每个地块上下车出行量的时间序列作为地块的属性,对地块进行聚类分析。在聚类分析中,以往的研究一般采用K均值、K中值或其他静态多属性序列样本聚类方法,但静态多属性序列数据将每一个属性看做独立的值,而在时间序列中,不同时刻上的属性值存在关联,因此需要重新定义时间序列

的相似度计算方法。本文考虑到出行量时间序列的连续性,采用结合了时间序列距离度量和时间序列自身相关性的聚类方法,并且加入不同类型POI点来分析不同出行类型区域和城市功能区的相互关系。

2 研究区和数据源

2.1 研究区域和地块的分割

本文研究区域为北京市,采用城市主要道路网络切割城市空间生成的地块为基本空间单元。首先,通过选取城市主要道路切割城市空间,得到初始的地块单元;然后,对初始地块单元进行筛选,必要时通过目视解译的方式对地块单元进行删除、合并等操作,得到较为合理的用地单元矢量数据^[6],最终将北京市分为1911个地块(图1)。由图1可看出,从城区到郊区,地块的面积逐渐增大、路网密度逐渐变稀疏。这与北京实际情况相符,市区人口密度大、交通流量高,因此路网密集;郊区人口相对稀疏、交通流量低,路网也较稀疏。

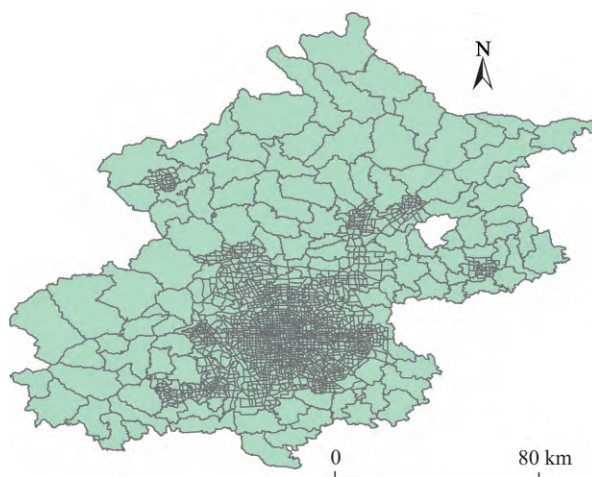


图1 北京市地块划分结果

Fig.1 The division result of the traffic zones in Beijing

2.2 出租车轨迹数据

原始出租车轨迹数据为2012年11月北京市约1.2万辆出租车的运营轨迹,采样间隔为10 s。经预处理后,去掉信息缺失、不完整的数据,随后从处理

后的轨迹数据中提取出租车载客的起始点轨迹数据。其中,每条轨迹包含:出租车编号、起点空载与否的标志(若无乘客为0,有则为1)、起点时刻、终点空载与否标志、终点时刻、起点经纬度、终点经纬度、路程总长度。平均每天非空的载客起始点轨迹数约72万条。经统计得到11月上下车点量随时间波动的时间序列如图2所示。

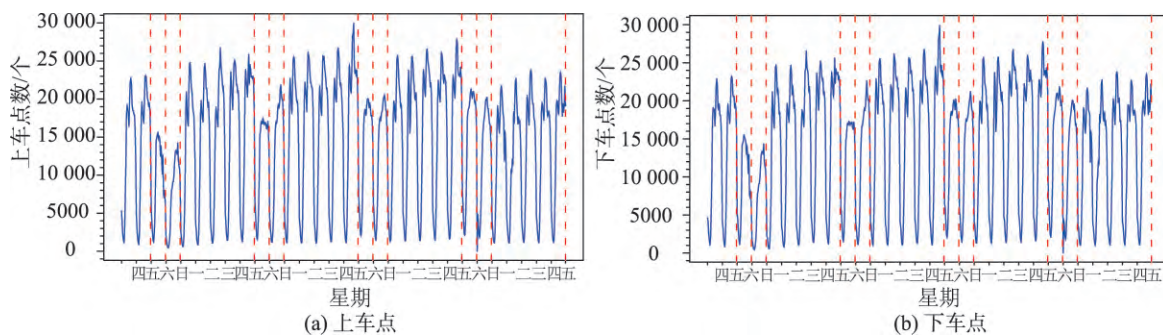


图2 出租车轨迹上下车点数量随时间变化

Fig.2 The time series of taxi pick-ups and drop-offs

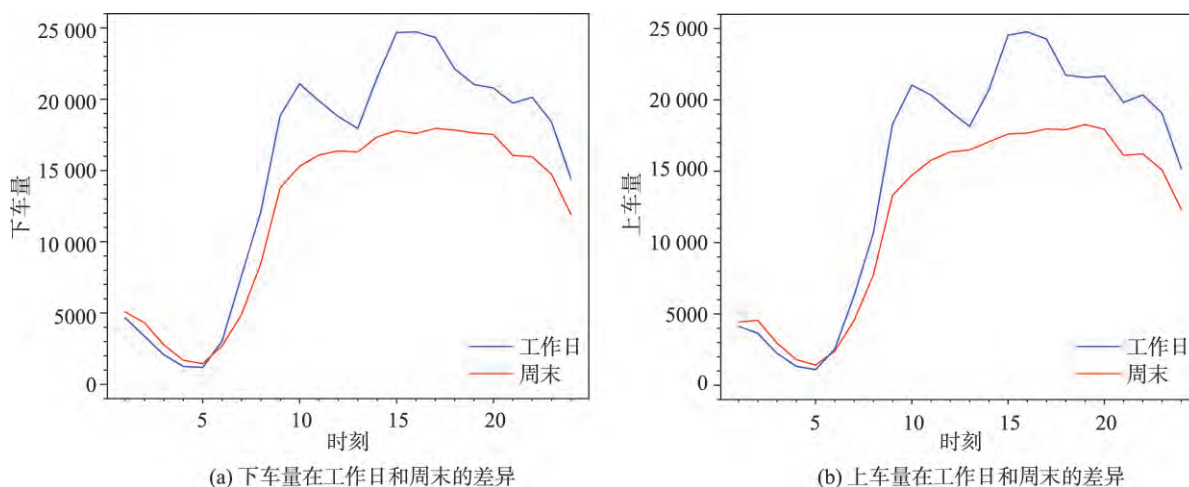


图3 工作日周末每天平均上下车数量对比

Fig.3 The comparison of daily average amount of pick-ups and drop-offs between weekdays and weekends

从图3可看出,除0:00-5:00外,周末上下车量略高于周中上下车数量,周中上下车量整体高于周末上下车数量。周中工作日的上下车量随时间波动较明显,早高峰在8:30-9:00,14:00-16:00为一天最大值,12:00-13:00之间有一个低谷值。出租车出行早高峰与通勤早高峰^[7]相比推迟了半小时。在通勤中乘客一般为市区居民,出租车大多为中短途出行,因此出发时间会稍晚;而非通勤出行,为了避免早高峰堵车,也会岔开高峰期出行,因此出租车早高峰相比早高峰有所推迟。出租车比公交车更灵

周一到周五为工作日,居民通勤的时间弹性和空间弹性较小,因此时间模式相似,而周末居民出行的目的、时间、空间等弹性均较大,造成时间模式与工作日相比差异较大(图2)。将11月所有的工作日上下车量的时间序列和休息日的时间序列平均到一天(图3),可更加明显地看出工作日和周末上下车点时间序列的差异。

活,更快捷,因此事务性出行比例较高,出租车一天的出行峰值出现在下午也正体现这一点,中午则为通常午餐时间,出行人较少,是出行的低谷。周末从8:30-21:30,上下车数目随时间波动变化不大,没有明显的出行时间倾向。

3 研究方法

3.1 数据处理

由上下车人数时间分布特征可知,工作日和周

末的时间序列模式存在较大差异,因此将工作日和周末数据分为2组数据单独处理。首先从轨迹数据中提取出租车上下车点信息,然后统计出研究区域每个地块每小时的上车量,得到每个地块的上下车量时间序列,随后计算工作日22天以及周末8天的每日平均上下车量时间序列,最后对这1911个地块的工作日上车和下车时间序列、周末上车和下车时间序列4组数据进行聚类分析。目前,时间序列聚类的算法一般根据时间序列自身特征对静态数据的聚类算法进行修正^[8-9]。本实验将静态数据常用的相似度计算方法转换为适合时间序列聚类的方法,然后用静态数据的一套聚类算法对时间序列数据进行聚类。

3.2 时间序列聚类相似度度量方法选取

由于时间序列的动态特性,时间序列相似度计算问题通常十分复杂。静态多属性序列数据的距离函数将每一个属性看做独立的值,而在时间序列中不同时刻上的属性值存在关联,因此需要重新定义时间序列的相似度计算方法。本文选择时间序列间距离度量和时间序列自身相关性相结合的改进CORT方法^[10-11]。2个时间序列自身属性随时间变化特征的相似性由一阶时间相关系数度量。2个时间内序列 X_t, Y_t 间的一阶时间相关系数定义如式(1)所示,其中 t 代表时刻。

$$CORT(X_t, Y_t) = \frac{\sum_{t=1}^{T-1} (X_{t+1} - X_t)(Y_{t+1} - Y_t)}{\sqrt{\sum_{t=1}^{T-1} (X_{t+1} - X_t)^2} \sqrt{\sum_{t=1}^{T-1} (Y_{t+1} - Y_t)^2}} \quad (1)$$

$CORT(X_t, Y_t)$ 的值在区间 $[-1, 1]$ 内,值越大,2个时间序列每时刻变化率的大小和方向越相似;值越接近-1,则2个序列变化率大小相似但方向相反;若为0,则表明2个序列之间线性独立。结合了2个时间序列的一阶时间相关系数后,该相似度的度量定义如式(2)所示。

$$d_{CORT}(X_t, Y_t) = \phi_k[CORT(X_t, Y_t)] \cdot d(X_t, Y_t) \quad (2)$$

式中: $\phi_k(\cdot)$ 是一个调整函数,用时间相关系数来修正传统距离 $d(X_t, Y_t)$ (如欧氏距离,DWT距离等),一般使用指数调整函数(式(3))。

$$\phi_k(u) = \frac{2}{1 + \exp(ku)}, \quad k \geq 0 \quad (3)$$

CORT方法不仅考虑了时间序列之间形状的相似性的,同时通过对时间序列一阶时间相关系数度

量时间序列波动特征之间的相似度。本文参数 K 值取2,传统距离度量 $d(X_t, Y_t)$ 选择动态时间扭曲距离DWT^[12]。DWT方法由Berndt和Clifford(1994)应用于时间序列的模式挖掘中^[13],其主要目的是在2个序列之间找到一个映射 $r = ((X_{a_1}, Y_{b_1}), \dots, (X_{a_m}, Y_{b_m}))$,其中 $a_i, b_i \in \{1, \dots, T\}$,如 $a_1 = b_1 = 1, a_m = b_m = T$,且对于 $i \in \{1, \dots, m-1\}$, $a_i + 1 = a_{i+1}$,或者 $a_i + 1 = a_{i+1}, b_i + 1 = b_{i+1}$,然后度量每对映射中2个时刻属性的距离 (X_{a_i}, Y_{b_i}) ,并使其和最小(式(4))。

$$d(X_t, Y_t) = \min_{r \in M} \left(\sum_{i=1}^m |X_{a_i} - Y_{b_i}| \right) \quad (4)$$

3.3 时间序列聚类算法的选取

相似度度量所得的地块距离矩阵,记录了1911个地块出行量时间序列模式间的相似度。基于该相似度矩阵可进行时间序列数据聚类。

本文选择环绕中心点划分算法(Partitioning Around Medoids, PAM)对时间序列数据进行聚类。PAM算法是K中值算法(K-medoids)的实现。相比K均值算法,K中值结果受簇中的边缘值和噪声的影响更小,且K均值的计算把每一个时刻的值进行平均,不能突出波形的作用。

聚类数目 K 值的确定需要综合考虑数据集的大小、分类的目的以及聚类效果的有效性。对于时间序列,质心的度量会失去时间序列的波动特征,意义不明显,因此应避免包含样本对质心偏移量度量的指标。在本研究中主要采用轮廓系数Silhouette^[14]和Dunn^[15]2种指标来选择最优聚类数目,这2指标综合考虑了聚类结构中不同类别样本的离散性和同一类别样本的凝聚性。由于样本数据中,出行量为0的区域较多,因此将出行量为0的区域单独划分为一类。对其余地块计算其轮廓系数值和Dunn值随聚类个数 K 值的变化如图4所示。2个指标值越大表明聚类效果越好,但考虑到样本数目较大, K 值取值在3以上较为合理。综合考虑2个聚类指标结果和原始数据大小,取聚类个数 K 值为6。

4 出租车出行量的时空特征分析

4.1 出行生成量的时空特征分析

对北京市出租车工作日上车量时间序列进行聚类分析。上车量为0的地块命名为clust0,对6类聚类结果,按上车量由小到大依次命名为clust1-

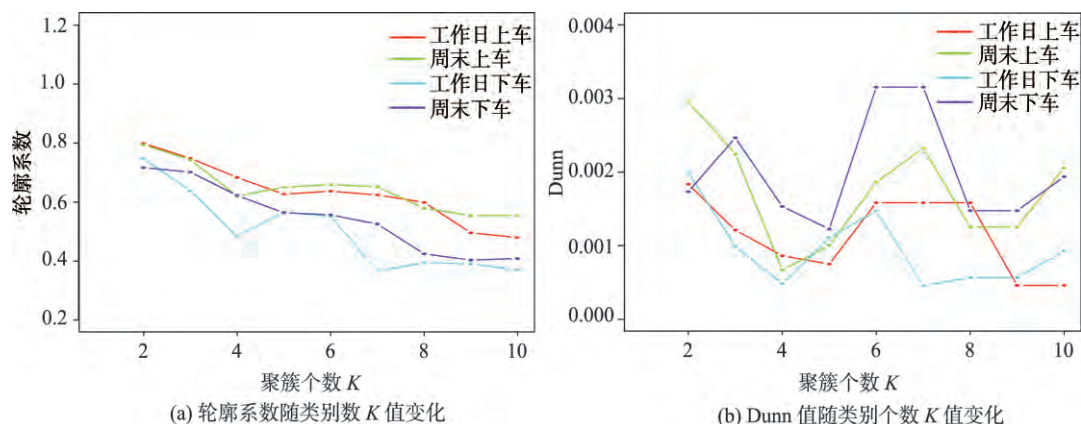


图4 轮廓系数和Dunn值随K值的变化

Fig.4 The changes of Silhouette and Dunn results with respect to different K values

clust6。每一种类别在空间分布图和时间序列曲线图中颜色保持一致,其结果如图5所示。

clust1是样本地块数目最多且总面积最大的类别,郊区地块大多被归为这一类。可看出城区和郊区乘客出行模式有着明显的差异:郊区空间分异较小,出行模式一致;而城区面积虽小但类别众多。clust1上车量极少,且波动不大,没有明显的峰谷特征。可看出郊区乘客打车随机性很大。郊区居民收入偏低,通勤主要以公共交通为主^[7],出租车出行率较低。

clust2分布在五环和六环之间,为近郊地带,此处存在单一大型居民区,由于职住分离,在工作日这些地区的出行一般为通勤性出行^[16]。该区域上车人数也偏少,在8:00有明显峰值,为通勤早高峰,12:00为一个低谷,13:00以后相比上午出行量减少,波动较小。

clust3所处区域多小村镇。这类地块上车量整体比clust1、clust2多,同样在8:00有一个通勤小高峰,12:00有一低谷,之后13:00–21:00峰谷特征不明显,但上车量相比上午并未减少,该地区可能为混合型居民区。

clust4和clust5主要分布在四环内的市中心。2类时间序列变化相似,在9:00和14:30有2个峰值,且下午上车量高于上午。该类地区多为混合商圈,市中心居民通勤距离短,因此出发时间较晚,而事务性出行则一般在上班之后出发,一般娱乐性出行则会避开早高峰,因此早通勤峰值推迟了1 h左右。下午的出行一般为工作、事务性出行或娱乐性出行。

clust6样本量和面积最小,但上车量最大。由

时间序列波形图可看出,白天上车量随时间整体呈上升趋势,局部有波动,在夜晚21:00–22:00达到一天中上车量的最大值。这类地块分别包含,中关村-人大商圈、北京北站-西直门附近、芍药居-对外经贸大学附近、朝阳区燕莎商场附近、北京站附近、北京西站-宣武门附近,北京南站和首都国际机场附近。这些地区主要为人流量巨大的火车站和机场,因此在晚上达到一天出行量最大值。

整个聚类结果的空间分布大体呈现类同心圆状分布,clust1在最外圈,clust4、clust5、clust6在中心,各类地块随着距市中心距离的增加上车量逐渐减少,聚类结果按出行量由小到大的顺序命名为clust1–clust6,各类别之间的最大差异为出行量大小(图5(c))。郊区空间分异小,郊区地块几乎都被划分为同一种类型,而城区分异较大,出行多种地块类型。但城区clust4、clust5类时间序列曲线相似,主要差别为出行量大小间差异。

周末上车点的时间序列聚类结果如图6所示。

与工作日出行特征相比,在空间分布上,仍类似同心圆状分布,且市区空间分异大,郊区空间分异小。类别的空间分布位置也与周中类似,但许多地块在周末相比工作日被划分为出行量较小的地块如:由工作日clust5变为clust4、clust4变为clust3,出行量为0的地块也增多。每一类别平均上车量也相对工作日下降。这体现了周末休息日,出租车出行量整体的减少。clust6是唯一一类工作日、周末地块数目和分布完全相同的一类。由于此处为火车站、飞机场所在地,其出行模式不受工作日影响。

在时间序列模式上,与工作日相比早高峰现象几乎消失,且12:00–13:00的低谷也消失,一天的时

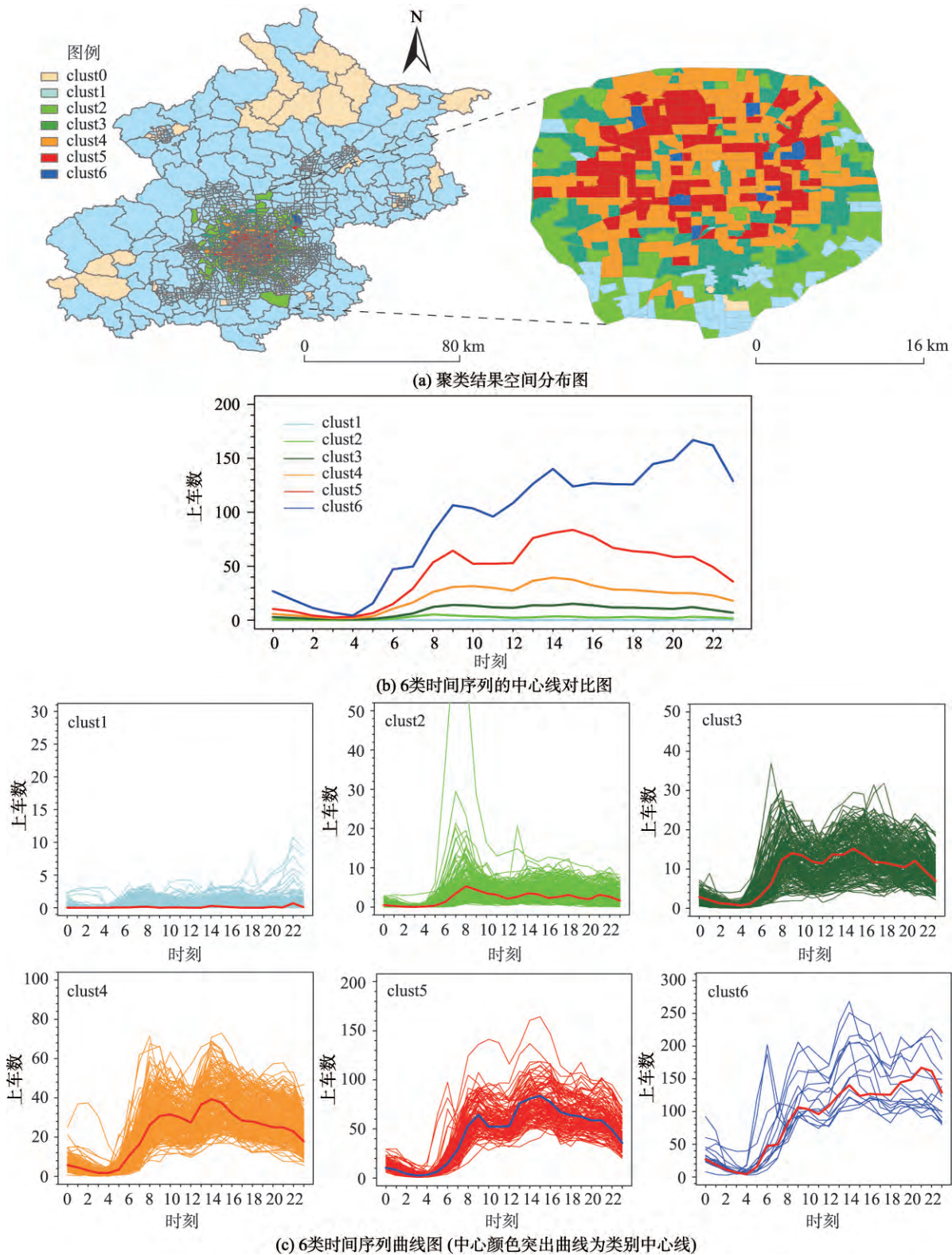


图5 工作日上车量聚类结果

Fig.5 The clustering results of weekday pick-ups' time series

间序列波动不明显,上下午出行量差别不大。可见,周末人们的出行随意性更强,而中午外出就餐的概率也增大。

4.2 出行吸引力的时空特征分析

工作日下车点的时间序列聚类结果如图7所示。同样按照下车量由小到大排列 clust1-clust6,

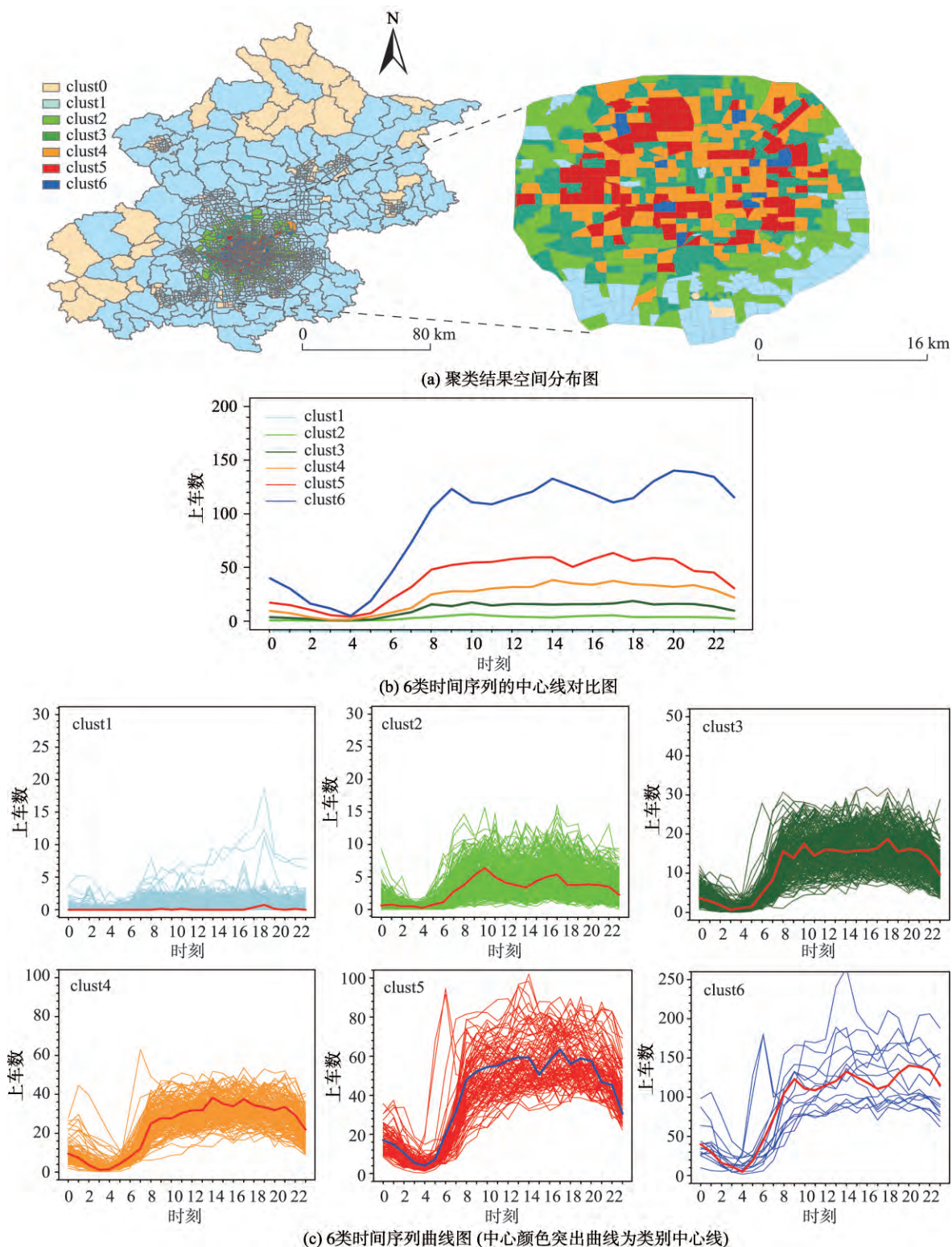


图6 周末上车量聚类结果模具费

Fig.6 The clustering results of weekend pick-ups' time series

且每种类别采用和出行量分析中相同的颜色表示。

地处北京边缘郊区的clust1仍然是样本量最大的类别,其在8:00附近有一个小峰值,应为通勤早高峰,但峰值仍极小,此后没有明显峰谷特征。

clust2中地块分布很零散,大多分布在郊区,市区也有零散分布。在8:00为一个不显著的小高峰,可能由通勤造成,在15:00–17:00达到一天下车量最大值,没有明显晚高峰特征。

clust3类地块主要在城区,所在地有较多高级酒店、小区,没有明显规律性。

clust4类地块在市中心,其波形在早上9:30-10:00左右,下午15:00-16:00,以及晚上19:00有3个小高峰,早峰值与通勤高峰相比推迟了1h左右,原因同样和工作日上车分析类似,19:00高峰与晚高峰吻合,下午的高峰,多为事务性出行,可见该类地块应为商务区吸引了不少事务性出行。在12:00、17:00出现2个局部极小值。

clust5类地块主要在四环内,该类地块多为一些酒店、大学、写字楼、商业大厦聚集地。其时间序列模式与clust4类似,但下车量是clust4的2倍多。

clust6类中地块是下车量最多的一类地块,与上车量clust6类别重复度很高,但数量减少了,分别是北京西站-宣武门附近,中关村,北站-西直门附近,朝阳区燕莎商场附近,首都国际机场附近。

整体上看,工作日下车量聚类结果与上车量聚类结果相比,地块空间分布没有明显的规律,各类别地块零散分布。且下车量结果中郊区空间分异增大:在上车量聚类结构中,郊区几乎都被划分为clust1,但下车量聚类结果中,郊区地块包含clust2、clust3、clust4、clust5多种类型。不少郊区地块,工作日上车量很低,但下车量却较大。这些地区吸引了大量来自城区的乘客,可能为度假村等娱乐区。

城区的clust2、clust3时间序列模式相似,而郊区的clust4、clust5时间序列曲线也相似。即城区和郊区出行模式之间存在明显差异,而城区和郊区内部保持了大致的规律性。

在下车序列中没有明显的晚高峰,即下班打车回去的乘客较少。上车量是出行“源”的空间分布,在一定程度上可反映北京市居民的密度和经济收入状况,而下车量的空间分布则反映地区对人群的吸引力。

周末下车点的时间序列聚类结果如图8所示。

周末下车量聚类结果同样在空间分布上十分零散。周末各类别间的空间分布与工作日相比差异较大,例如,clust6只剩首都国际机场附近地块,城区的几个在上车量聚类分析中被归为clust6的商圈,被分为clust5,可见clust5类型所代表的混合型商圈地块在周末出行吸引力在增大,与clust6类别中火车站所在地块出行差异减小。郊区的分异更大了,且部分地块下车量增加,为一些景点、度假村、采摘园等所在地。这些郊区地块在周末体现

出休闲娱乐功能。

从时间序列看,从clust1-clust5全天下车量随时间波动不大。clust6随时间波动较明显,下车量峰值出现在早上6:00左右,6:00以后整体呈下降趋势。

4.3 聚类结果的用地类型分析

为了更直观地分析每一类地块的功能特征和出行特征之间的关系,本文加入北京市POI数据对聚类结果进行分析。该POI数据包含111 751个POI点,共30多种类别,从中挑选15种具有代表性的POI类型,分别为:咖啡厅或茶馆1012个、电影院122个、商务大厦1882个、银行2684个、公园463个、餐厅12 305个、超市或便利店5509个、旅店宾馆3391个、公司企业18 331个、学校4044个、居民地10 735个、别墅366个、垂钓126个、度假村833个、火车站和飞机场6个。

对每一类地块,计算该类地块中每种POI点的平均密度: $D_i = \frac{N_i}{\sum_{m \in i} S_m}$,即将该第*i*类地块中POI总

数除以该类地块总面积,然后,对每种POI数据进行归一化。采用最普通的归一化方法, $D_{nor}(i) = \frac{D_i - D_{min}}{D_{max} - D_{min}}$,其中 D_{max} 、 D_{min} 分别为同一种POI在7种类型地块中的最大密度值和最小值。归一化值可以对比POI在7种地块类型中的相对大小,同时避免POI密度值过大或过小差异过大。由于上车量为0的类别clust0中,POI值为也为0,因此在图中没有显示clust0。

结果显示这15种POI在7种类型地块中的密度呈现一定规律性,根据各POI之间的相似性的分为3组,其结果如图9所示。

(1)第1组POI,包括咖啡店/茶馆、电影院、商务大厦、银行一般出现在消费水平、居民收入较高、人口稠密的繁华商圈。一般为高收入人群的工作地,而依托高收入人群的消费需求,咖啡、茶馆和电影院等消费场所以及银行也较密集。可以注意到,各类地块的密度变化相似,说明POI之间有密切联系,经常在同一类用地类型中共同出现。

(2)第2组POI,包括餐厅、超市或便利店、旅店宾馆、公司企业/工厂(这里是指需要工人数较多的大型工厂,如造纸厂、食品厂、水电厂等)、学校、居民地。一般出现在居住或活动人口稠密的地区,而

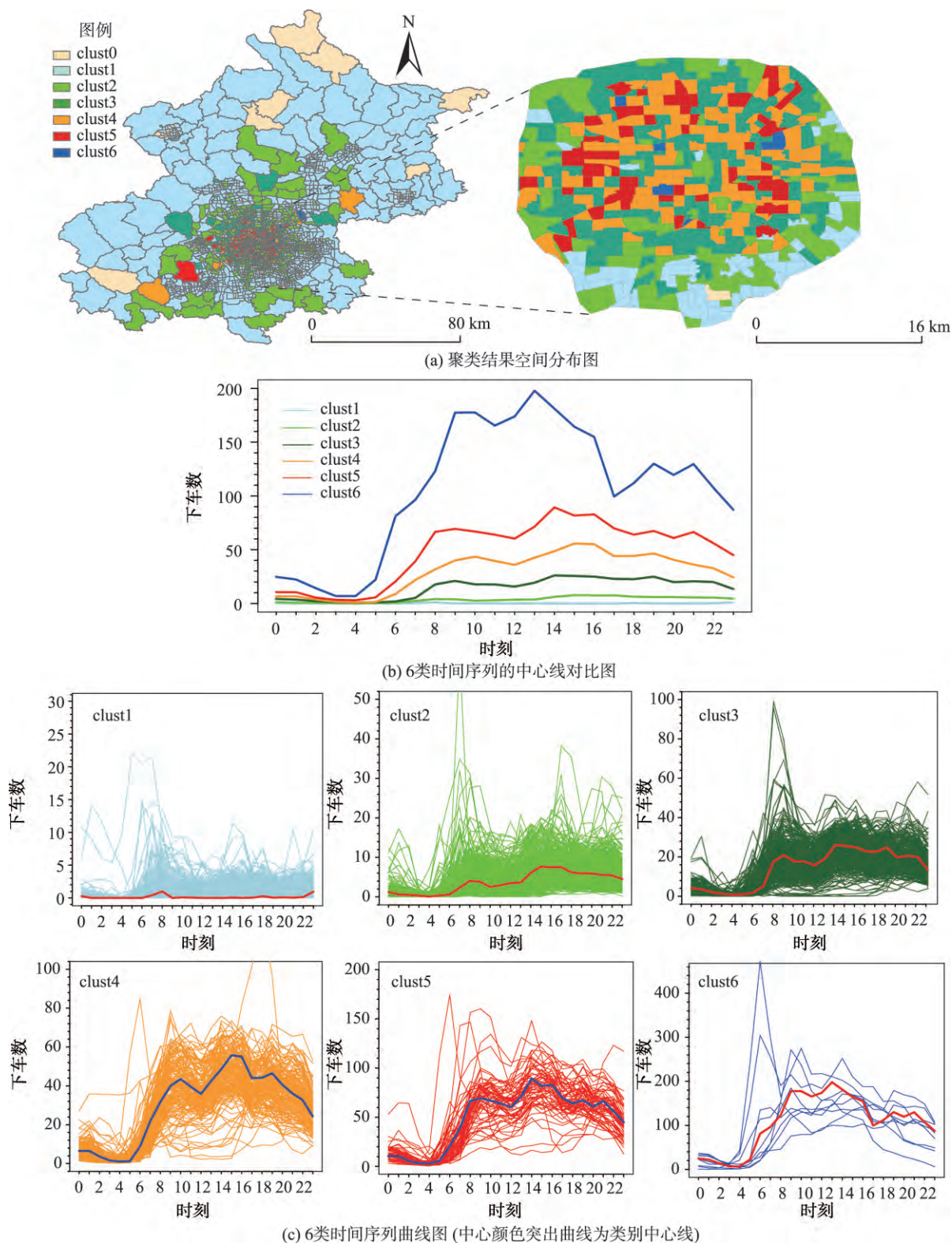


图7 工作日下车量聚类结果

Fig.7 The clustering results of weekday drop-offs' time series

周边的超市便利店、餐厅等场所也较多。第2组的各类地块也具有相似的密度变化。

(3)第3组POI点,包括别墅、垂钓、度假村、公

园、火车站/飞机场。其在每一类地块中密度变化均不同,且彼此差异很大。

北京整体上是一个中心型环状发展的城市。

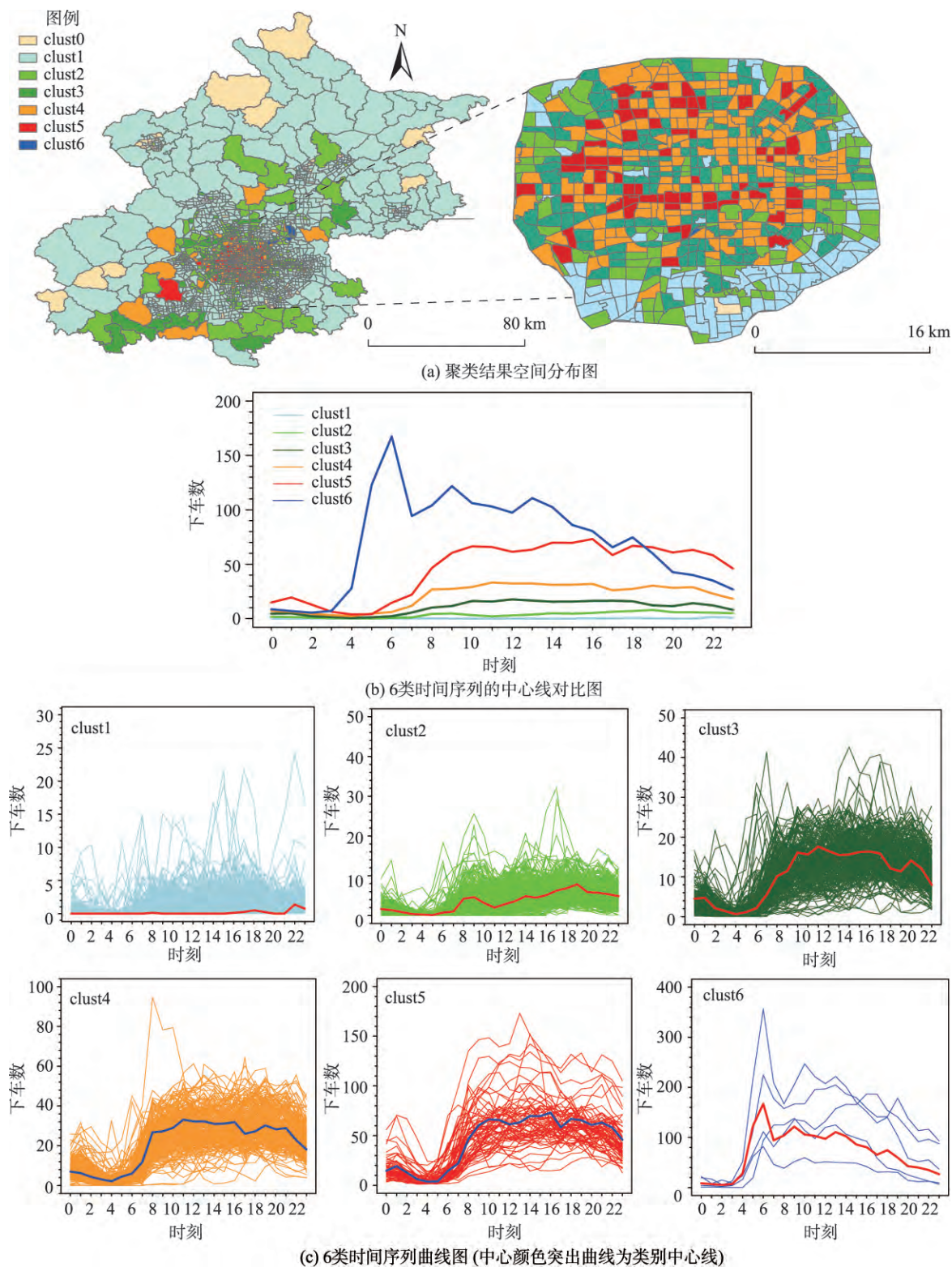


图8 周末下车量聚类结果

Fig.8 The clustering results of weekend drop-offs' time series

市中心(五环以内)经济水平高,人口稠密,商业大厦、写字楼等高级工作区,咖啡店、酒店餐厅、电影院等娱乐休闲场所,以及百货大楼、购物中心等消

费场所和居民小区等都十分密集,集购物、休闲娱乐、工作、居住等功能于一体。

从POI分布可看出市中心功能划分类型众多,

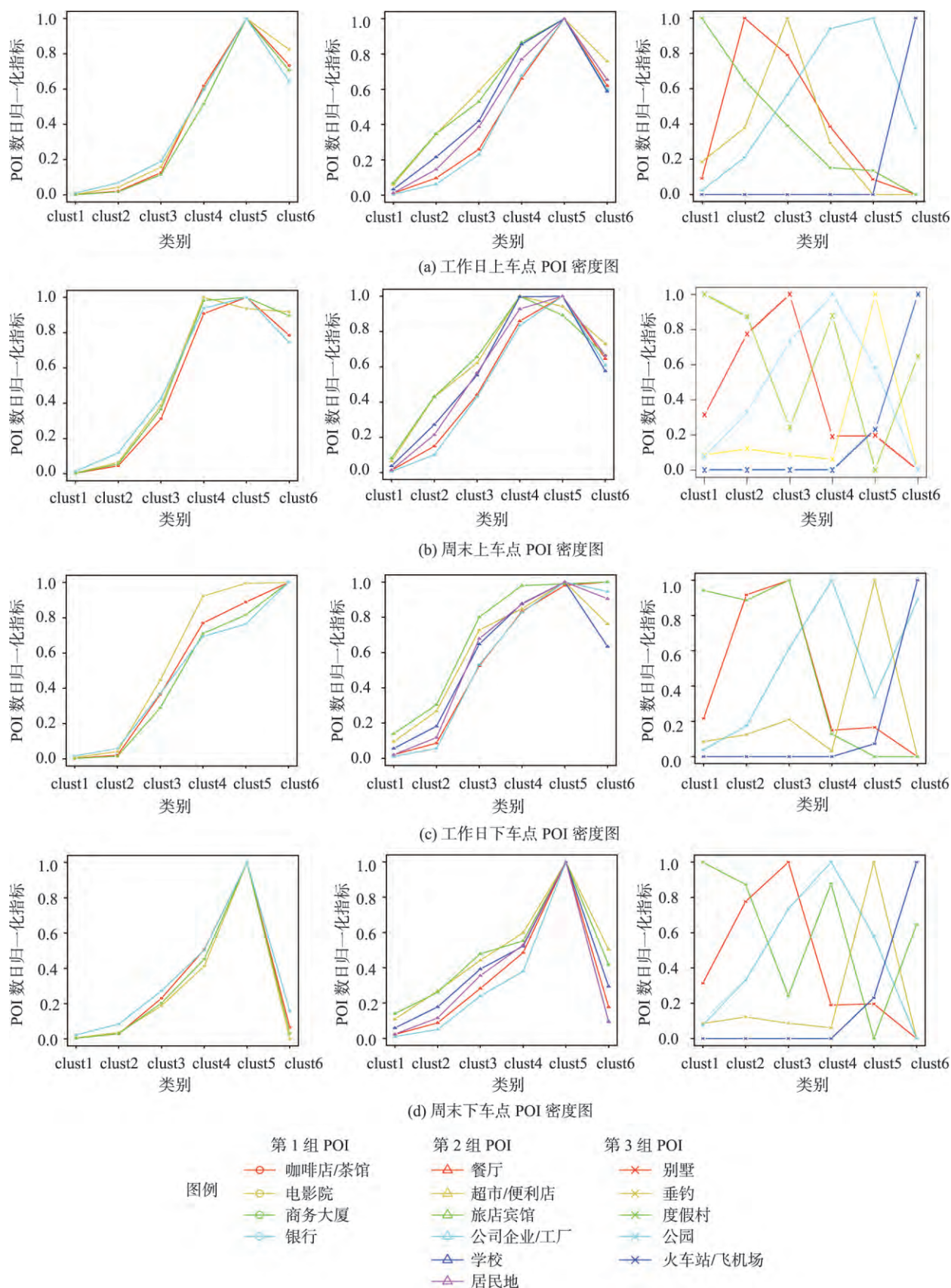


图9 每类地块归一化POI密度值(每组图中第1至第3幅图分别对应第1至第3组POI)

Fig.9 The normalized density of POI in every cluster (the order of the three graphs in every sub-graph set corresponds to the order of the POI groups)

不同类型地块中各POI密度不同,从每一类地块中密度较大的POI类型可以识别该类用地,通过每一类别地块的空间分布图,可以观察每类用地的空间分布。如第1组和第2组POI,从clust1-clust5呈上升趋势,在出行量较大的clust5、clust4中密度达到最大值,这2类地区应为混合商圈,与出行量模式所得结论一致。而出行量最大的clust6地块火车站和飞机场POI密度最大。clust2类型别墅密度最大,其次是clust3,这2地块在近郊且出行模式相似。clust1密度最大的为度假村,并且在上车点聚类结果中,度假村的密度在clust1最大,从clust2-clust6大幅度减小(图9-10);而在下车点结果图中,工作日下车结果图中clust2、clust3周末下车结果中clust2、clust4密度都较大(图11-12),即该地区下车量高于上车量且周末高于工作日。结合前面下车量分布结果可知,郊区在下车量密集的几块区域为度假村,因此吸引了城区乘客,且这种现象在周末更明显。

结合POI对地块功能类型的分析结果与上文出行规律中分析相吻合,出行量类别空间分布呈环状分布,城区clust4、clust5以混合型商圈为主,因此城区出租车出行一般为事务性和娱乐休闲性质的出行,会在下午14:00-15:00会出现出行的高峰,早出行高峰通常会比早通勤高峰晚1h左右。郊区有较多娱乐休闲场所,因此会吸引城区居民,导致下车量的空间分异,而这种现象在周末更为明显。

5 结论

本研究对北京市出租车的出行量的时空分布进行分析。不论上车量还是下车量,出租车出行时间序列都存在周期性,周期为一周,在工作日(周一至周五)波动特征相似,而工作日、周末之间存在差异。根据实验结果可看出,出租车体现出具有一定规律性的居民出行模式。

(1)空间分布特征

北京市出租车上车量在空间上的分布大体呈同心圆分布,距市中心越远的地区乘客出行生成量越小,出行量最大的地区则集中在北京几大火车站和首都国际机场附近。出行生成量在郊区和城区之间存在较大差异,郊区出行生成量空间分异小,而城区出行模式空间分异强。出租车上车量的空间分布受工作日和非工作日影响较小。相比上车点分布,出租车下车量空间分布不均匀,郊区的空

间分异性变强,出现一些下车量较高的地区,这种现象在周末更加明显。对比POI分布,这些地区一般为度假村、农家乐等所在地,这体现出郊区的户外休闲娱乐功能。

(2)时间波动特征

上车量时间波动特征从整体看,工作日城区的时间模式相似,存在2个明显峰值,早高峰比通勤早高峰晚1h左右,而下午的高峰在14:00-15:00,处于通勤低谷期;郊区出行量小,波动不明显,与城区差异较大。周末出行无论城区还是郊区,出行都整体减少,峰谷变得不明显。下车量的时间分布特征与上车量时间波动特征相似。

(3)结合POI数据对居民出行与城市功能结构的分析

城区出租车出行量最大的是大型火车地铁站、飞机场等交通枢纽,其次是娱乐消费场所和商务大厦聚集的繁华商圈,最后是活动人口密度的居民区、学区、工厂区。郊区出租车出行量整体比城区低,其出行量最大的是垂钓场所公园较多的区域,其次是别墅较多的区域,最后是度假村、农家乐等区域。同时,虽然郊区整体出行量很低,但度假村等休闲娱乐区会吸引城区居民,导致下车量的空间分异,而这种现象在周末更为明显。通过POI对城市功能区的分析结论与城市区域出行规律相一致。

本文分析了出租车上下车点的时空分布,对了解北京市各地区居民的出行特征具有重要意义;同时,基于出租车出行量的研究探究了北京市城市功能区的空间分布,对理解北京市城市结构布局以及功能区的分异具有实际意义。

参考文献(References):

- [1] Wang H, Zou H, Yue Y, *et al.* Visualizing hot spot analysis result based on Mashup[C]. Proceedings of the International Workshop on Location Based Social Networks, 2009:45-48.
- [2] Zhang W, Li S, Pan G. Mining the semantics of origin-destination flows using taxi traces[C]. Proceedings of the Workshop of Ubiquitous Computing, 2012:943-949.
- [3] Veloso M, Phithakkitnukoon S, Bento C. Urban mobility study using taxi traces[C]. Proceedings of the International Workshop on Trajectory Data Mining and Analysis, 2011:23-30.
- [4] Liu Y, Wang F H, Xiao Y, *et al.* Urban land uses and traffic 'source-sink areas': evidence from GPS-enabled taxi data in Shanghai[J]. Landscape and Urban Planning,

- 2012,106(1):73-87.
- [5] Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs[C]. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012: 186-194.
- [6] 康朝贵. 基于个体时空轨迹数据的居民移动模式和城市空间结构分析方法[D]. 北京: 北京大学, 2015. [Kang C G. Sensing urban space from human activity and its spatio-temporal characteristics[D]. Beijing: Peking University, 2015.]
- [7] 孟斌, 郑丽敏, 于慧丽. 北京城市居民通勤时间变化及影响因素[J]. 地理科学进展, 30(10):1218-1224. [Meng B, Zheng L M, Yu H L. Commuting time change and its influencing factors in Beijing[J]. Progress In Geography, 30(10):1218-1224.]
- [8] Fu T. A review on time series data mining[J]. Engineering Applications of Artificial Intelligence, 2011, 24:164-181.
- [9] Liao T. Clustering of time series data - a survey[J]. Pattern Recognition, 2005, 38:1857-1874.
- [10] Montero P, Vilar J A. TSclust: an R package for time series clustering[J]. Journal of Statistical Software, 2014, 62(1):1-43.
- [11] Chouakria A D, Nagabhushan P N. Adaptive dissimilarity index for measuring time series proximity[J]. Advances in Data Analysis and Classification, 2007, 1(1):5-21.
- [12] Sankoff D, Kruskal J. Time warps, string edits, and macro molecules: the theory and practice of sequence comparison[M]. Boston, MA: Addison Wesley, 1983.
- [13] Berndt D J, Clifford J. Using dynamic time warping to find patterns in time series[A]. In: Fayyad U M, Uthurusamy R (eds.). Knowledge Discovery in Databases: Papers from the 1994 AAAI Workshop[M]. Palo Alto, CA: AAAI Press, 1994:359-370.
- [14] Rousseeuw P J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis[J]. Journal of Computational and Applied Mathematics, 1987, 20:53-65.
- [15] Dunn J. Well separated clusters and optimal fuzzy partitions[J]. Journal of Cybernetics, 1974, 4:95-104.
- [16] 孟斌, 于慧丽, 郑丽敏. 北京市大型居住区居民通勤行为对比研究——以望京居住区和天通苑居住区为例[J]. 地理研究, 2012, 31(11):2069-2079. [Meng B, Yu H L, Zheng L M. The analysis of commuting behavior in the huge residential districts: a case study of Wangjing and Tiantongyuan in Beijing[J]. Geographical Research, 2012, 31(11):2069-2079.]