

引文格式：

李明珠, 赵习枝, 陈才, 等. 基于聚类的地铁通勤行为时空规律挖掘方法 [J]. 集成技术, 2023, 12(1): 79-90.

Li MZ, Zhao XZ, Chen C, et al. Spatial and temporal law mining method of subway commuting behavior based on clustering [J]. Journal of Integration Technology, 2023, 12(1): 79-90.

基于聚类的地铁通勤行为时空规律挖掘方法

李明珠^{1,2} 赵习枝^{2*} 陈才^{2,3} 张福浩² 朱军¹ 仇阿根²

¹(西南交通大学地球科学与环境工程学院 成都 611756)

²(中国测绘科学研究院 北京 100830)

³(江苏海洋大学 连云港 222005)

摘要 针对目前通勤群体的划分方法较少考虑通勤行程的时间连续性特征的问题, 该文基于上海市一周的地铁刷卡数据, 构建了通勤人群职住识别模型, 并定义了一种通勤行程时间相似度计算方法, 然后提取特征对通勤群体进行层次聚类, 并利用热点分析模型进行空间分析和可视化, 探究通勤人群的时空规律性及上海市的职住空间分布特征。结果表明: (1) 就业单中心模式明显, 不同簇的就业热点均分布在市中心, 居住点呈“西热东冷”的空间组织特征。(2) 主流通勤时段为 7:00—8:30 和 17:00—19:00, 近半数通勤人群在早主流时段通勤。(3) 不同通勤类型的出行时间特征总体上与其职住热点分布一致。该文提出的研究方法揭示了通勤人群的出行时间规律与其职住热点空间分布具有较强的关联性, 可为城市运行管理和城市规划提供参考信息。

关键词 时间相似度; 层次聚类; 通勤人群划分; 时空规律; 通勤行为; 职住关系

中图分类号 P 2 **文献标志码** A **doi**: 10.12146/j.issn.2095-3135.20220909001

Spatial and Temporal Law Mining Method of Subway Commuting Behavior Based on Clustering

LI Mingzhu^{1,2} ZHAO Xizhi^{2*} CHEN Cai^{2,3} ZHANG Fuhao² ZHU Jun¹ Qiu Agen²

¹(Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China)

²(Chinese Academy of Surveying and Mapping, Beijing 100830, China)

³(Jiangsu Ocean University, Lianyungang 222005, China)

*Corresponding Author: zxz@casm.ac.cn

Abstract The current method of dividing commuting groups takes less into account the time continuity characteristics of commuting trips. Based on the one-week subway card swipe data in Shanghai, this paper

收稿日期: 2022-09-09 修回日期: 2022-11-24

基金项目: 国家重点研发计划项目(2019YFB2102503)

作者简介: 李明珠, 硕士研究生, 研究方向为时空数据挖掘; 赵习枝(通讯作者), 博士, 助理研究员, 研究方向为时空数据挖掘, E-mail: zxz@casm.ac.cn; 陈才, 硕士研究生, 研究方向为时空数据挖掘; 张福浩, 研究员, 研究方向为时空数据挖掘; 朱军, 教授, 研究方向为地理大数据; 仇阿根, 研究员, 研究方向为地理国情。

constructs a work-life recognition model for commuters, defines a commuting trip time similarity calculation method, and then extracts the features to classify commuter groups hierarchically, and uses the hot spot analysis model to perform spatial analysis and visual expression for spatial analysis and visualization, and explores the spatiotemporal regularity of commuters and the spatial distribution characteristics of work-housing organization characteristics in Shanghai. The results show that: (1) The employment single center model is obvious, and the employment hotspots of different clusters are distributed in the city center, and the settlements are characterized by the spatial organization of “hot in the west and cold in the east”. (2) The mainstream commute hours are 7:00—8:30 and 17:00—19:00, with nearly half of the commuters commuting during the morning rush hours and 90% leaving the work places before 19:30. (3) The travel time characteristics of the different commuting types are generally consistent with the distribution of their work and housing hotspots. The proposed research method reveals that the travel time law of commuters has a strong correlation with the spatial distribution of work-housing hotspots, which provides reference information for urban operation management and urban planning.

Keywords temporal similarity; hierarchical clustering; commuter crowd division; space and time laws; commuting; job-housing relationship

Funding This work is supported by National Key Research and Development Program Project (2019YFB2102503)

1 引 言

自改革开放以来,我国大城市的城市范围迅速扩展,人口快速增长,使得职住空间组织模式不断演化。适度的职住分离布局,有助于提高城市效率、降低土地利用强度。但随着城市规模的不断扩张,职住分离导致通勤距离显著增加,不仅降低了居民幸福感,还引发了环境污染问题^[1]。通过开展职住关系的研究,深化对中国城市发展规律的认识,对于解决职住空间结构不平衡及其衍生的交通环境住房等问题,具有重要理论和现实指导意义。

在宏观层面上,职住关系研究通常基于一定的范围尺度,如区(县)、街道、交通分析小区、千米尺度的格网等^[2-3],利用出行调查获得的统计数据,通过研究区域内职住比^[4]、通勤效率、通勤效率使用率^[5]等评价因子,对职住平衡进行测

算,或对职住空间关系的影响因素和产生机理进行探究。探究内容包括:(1)探究制度转型和城市空间结构变化对都市区就业空间分异、职住空间错位、就业可达性的影响^[6];(2)基于建成环境因素,讨论土地利用的混合程度^[7]、不同土地利用类型等对居民通勤的影响,探究交通设施和职住情况之间的关系^[4]。在微观层面上,现有研究探究了职住关系的影响因素,主要以个人的社会经济属性为解释变量^[8],讨论就业者收入水平、受教育情况、住房类型、工作性质等因素,对居民居住就业区位选择的影响^[9-10]。此外,还通过分析居民通勤移动行为,反映职住空间结构特征和交互规律^[11-12]。随着信息技术手段的快速发展,轨迹大数据为城市职住关系的研究提供了新的数据思路。相比于出行调查统计存在获取成本高和抽样不均匀等问题^[13-14],移动轨迹大数据则具有样本量大、实时全面、成本较小的优点^[4]。

近年来, 基于移动轨迹数据或社会经济数据等多源数据, 有研究分析了不同性质居民的通勤活动规律^[15-16], 探究了职住关系演变及其影响因素。通过对不同通勤群体之间通勤活动差异进行分析, 有助于更深入地了解通勤群体对于职住空间变化的反应和受影响程度^[17], 从而为规划方案提供建议, 使规划方案尽可能满足通勤群体多样性的通勤需求。目前, 已有学者进行了通勤人群划分的研究, 如付晓等^[18]通过构建居民出行群体画像, 分析不同出行人群的行为特征; 万明等^[19]采用潜在类别分析法对出行数据进行分析, 将出行者划分为 3 类异质群体。基于交通卡数据的研究, 则主要通过指定阈值和利用上下车时刻、出行次数等非连续性通勤特征, 实现通勤个体分类^[20-21], 由于较少考虑通勤行为时间连续性, 难以挖掘更多潜在的时间规律。此外, 基于经验的阈值具有不稳定性, 如会将阈值两端接近的个体划分为两类, 而未考虑到阈值左右的个体具有较强的相似性。因此, 利用合适的指标快速选择准确的阈值较为困难。

在大数据背景下, 聚类算法可根据属性相似性, 快速地将个体划分为不同的组, 近年来, 出现了利用聚类算法挖掘乘客出行特征的趋势^[22-23]。与阈值法相比, 聚类算法更好地考虑到参与聚类个体之间的相似性, 从个体出发, 将特性相似的个体聚为一类。如孙世超等^[24]利用分层聚类方法, 确定了墨西哥城工作丰富和住房丰富的区域。龙瀛等^[25]将共享自行车和码头的相似站点逐层聚类为应用模式, 提取出各站点的时间行为聚类特征。

针对现有研究在划分通勤人群时未充分利用时间特征, 划分方法多采用经验阈值导致准确性不足的问题, 本文进行了一系列研究。首先, 基于上海地铁刷卡数据, 通过构建职住识别模型, 识别通勤群体及其就业居住地; 然后, 采用一种时间相似度计算方法, 即在单向的一维时间空间

中, 将两个个体通勤行程的时间重叠度作为衡量时间相似度的指标, 提取通勤个体的时间相似度特征进行层次聚类, 以研究不同通勤群体的时间规律性; 最后, 利用热点分析模型, 对通勤群体空间规律进行分析, 利用 ArcGIS 平台进行可视化表达, 实现对上海通勤人群的通勤时空规律及职住空间组织特征的探究。

2 研究区域数据

本文选择上海市的整个区域范围作为研究区域, 包括中心城区(黄浦、徐汇、长宁、静安、普陀、虹口、杨浦)、近郊区(闵行、宝山、嘉定)、远郊区(金山、松江、青浦、奉贤、崇明)和浦东新区 16 个行政区, 截至 2020 年 11 月, 总面积约 1 237.85 km², 常住人口为 2 487.09 万人。

研究源数据为 2016 年 07 月 11 日—2016 年 07 月 17 日上海市交通卡连续一周的地铁刷卡数据(交通卡刷卡数据包括地铁、公交、出租刷卡记录, 地铁记录约占 66%, 公交约占 32%, 出租车约占 2%, 由于公交无下车刷卡记录, 本文选择地铁刷卡数据作为研究数据)。选取日期避开了法定节假日和重大活动日, 避免异常情况对研究造成影响。地铁刷卡数据原始字段包括: 交通卡号、刷卡日期、进站刷卡时间、出站刷卡时间、刷卡线路站点、刷卡交易金额、是否优惠。在研究时间段内, 地铁总刷卡次数为 62 968 491 次, 工作日地铁日均刷卡约 994 万次。此外, 研究数据还包括地铁站点和线路空间数据, 共计 323 个地铁站点, 站点数据字段包括站名、经度、纬度和线路号。

3 研究方法

3.1 职住识别模型

从单日通勤特征来看, 工作日通勤活动有两

个刷卡高峰：早高峰与晚高峰，对应上班与下班行为；从多日通勤特征来看，通勤属于满足日常生存需求的出行，工作日期间存在居住地和就业地之间的往返，一周之内满足一定的通勤天数。因此，综合考虑通勤人群稳定性的特点，即工作日高频出行时间段的稳定性、出发地和目的地的稳定性、工作日通勤行为的稳定性，结合已有研究，建立职住识别模型^[26]，从交通卡持卡人中识别出通勤人群和其职住地。

3.1.1 一日通勤出行行为识别步骤

一日通勤出行行为识别步骤为：(1)若持卡人当日首次地铁出行进站刷卡时间在 6:00—10:00 之间，那么将进站点设为居住站点 R (调查统计表示：99.5% 居民每日首次出发点是居住地^[27])；(2)若持卡人当日在 16:00 以后有地铁出行，那么将 16:00 以后的首次进站点设为 W ；(3)若站点 W 与 (1) 中地铁出行的出站点相同，且居住站点 R 与 (2) 中地铁出行的出站点相同，则称持卡人在该日具有通勤出行行为，居住站点为 R ，就业站点为 W 。如图 1 所示，如果持卡人在 6:00—10:00 内首次出行站点为居住站点 R 到就业站点 W ，16:00 后的首次出行站点为就业站点 W 到居住站点 R ，则认为持卡人当日实现由居住站点到就业站点的一次往返，具有通勤出行行为。

可将通勤出行行为识别模型概括如公式 (1)~(3) 所示。

$$6:00 \leq t_1 \leq 10:00 \ \& \ R = e_1 \quad (1)$$

$$p \geq 2 \ \& \ t_{p-1} \leq 16:00 \leq t_p \ \& \ e_p = o_1, o_p = e_1 \quad (2)$$

$$W = e_p \quad (3)$$

其中， t_1 为当日首次地铁出行进站时间； e_1 为当日首次地铁出行进站点； o_1 为当日首次地铁出行出站点； p 为刷卡进站的次序； t_p 为 16:00 以后首次地铁出行进站时间； e_p 为 16:00 以后首次地铁出行进站点； o_p 为 16:00 以后首次地铁出行出站点； R 为一日居住站点； W 为一日就业站点。

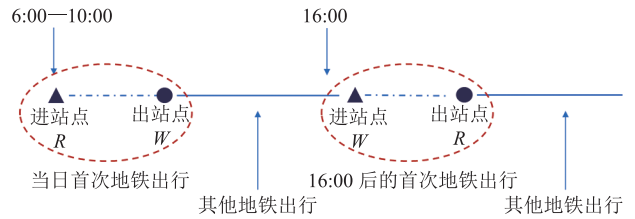


图 1 一日地铁出行记录示意图

Fig. 1 One-day subway trip record schematic diagram

3.1.2 一周通勤个体与职住地识别步骤

一周通勤个体与职住地识别步骤为：(1)持卡人一周具有 3 次及以上的通勤出行行为，则认为该持卡人为通勤个体。(2)若通勤者有且仅有一个居住站点，就将该站点视为其居住地；若有 2 个及以上居住站点，则计算每个居住站点的概率；(3)选择概率最大且次数大于 1 的居住站点作为居住地，若存在两个这样的居住站点，且两居住站点距离 ≤ 1.4 km (步行合理接驳阈值范围为 600~800 m，是较适中且能容忍的最大程度^[28-29])，则将两站点的中间点作为居住地；若两居住站点距离 > 1 km，则剔除该持卡人数据。

就业地识别方法同居住地识别方法。

3.2 时间规律挖掘

3.2.1 时间相似度计算方法

通勤是在时间和空间上的移动行为，为探究通勤群体^[30]在早晚通勤期出行的时间规律，挖掘不同通勤需求并制定更合理的交通规划方案，常对通勤人群进行分类，将具有相似出行时间规律的人划分为一类。常见的分类方法为阈值法，例如，将第一次刷卡记录早于 6:00 或 6:30 的持卡人定义为早出型通勤人群^[21]，但该划分方法可能会将在阈值两端但接近的乘客划分为两类 (如 6:29 和 6:31 出行的通勤人群会被分为两类)。因此，参考 Murtagh 等^[31]计算公交乘客时间相似度的度量方法，本文提出一种通勤行程时间相似性的度量方法：若两个行程在时间上有重叠部分 (图 2(a))，则认为它们相似；若两个行程没有重

叠部分(图 2(b)), 且行程间隔大于较小行程, 说明明显分离趋势, 则认为它们相似度为 0; 若行程间隔小于较小行程, 那么仍然认为通勤行程时间具有相似性。

具体计算方法为: 若行程时间存在重叠, 则两行程的时间相似度为重叠时间长度与较长行程时间长度的比值; 若不存在重叠但时间间隔小于较短行程时间长度, 则两行程的时间相似度为时间间隔与较长行程时间长度的比值。为避免相似度出现负值, 将以上两种情况的计算结果加 1 作为最终相似度值。

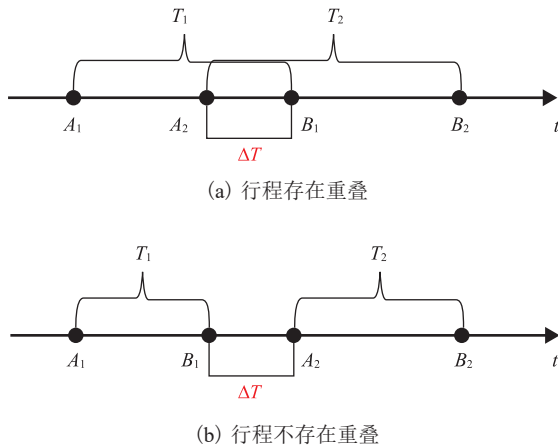


图 2 行程时间关系

Fig. 2 Time relationship of the trips

若两次早高峰行程为 $T_1(A_1, B_1), T_2(A_2, B_2)$, 其中, A_1, A_2 为两行程的上车时间, B_1, B_2 为两行程的下车时间, ΔT 为两行程的时间间隔(图 2), T_1, T_2 的相似度计算公式如公式(4)所示(晚高峰行程时间相似度同早高峰)。

$$TS(T_1, T_2) = \begin{cases} 1 + \frac{\min(B_1, B_2) - \max(A_1, A_2)}{\max(B_1 - A_1, B_2 - A_2)}; (B_1 < B_2 \text{ \& \& } A_1 > A_2) \text{ \& \& } (B_1 > B_2 \text{ \& \& } A_1 < A_2) \\ 1 - \frac{\Delta T}{\max(B_1 - A_1, B_2 - A_2)}; \Delta T < \min(B_1 - A_1, B_2 - A_2) \\ 0; \text{其他} \end{cases} \quad (4)$$

地铁刷卡数据会提供乘客的上车、下车刷卡信息, 在通勤人群一天的刷卡记录中, 分别提取早晚通勤高峰期的通勤出行记录, 代入上式计算

时间相似度。两类通勤人群的时间相似度即为早晚行程时间相似度之和。

3.2.2 时间聚类方法

层次聚类通常被用于时序数据的趋势分析^[32-33], 本文通过层次聚类对时间相似度进行聚类分析, 以揭示通勤群体的层次结构。层次聚类的表现形式有“自上而下”的分裂和“自下而上”的凝聚层次聚类两种。凝聚层次聚类通过距离函数将数据集划分为多类, 以降低簇数量, 不断重复直至形成一个单独的类。

层次聚类主要的步骤在于确定不同层次最接近的簇类数, 集群可通过距离判断数据点之间的相似性, 对于簇间的距离度量有多种方式: 最小值法(Single)、最大值法(Complete)、平均值法(Average)及最小方差法(Ward)。根据已有研究对上述方法的比较^[34], 当无离群值时, Ward 性能较好, 故本文最终选择 Ward (即在并类时, 选择使误差平方和增加最少的两类进行合并)。

对抽取的所有通勤人群进行聚类后, 本文结合轮廓系数^[35-36](Silhouette)进行最佳簇数的选择。其中, Silhouette 值越大, 聚类效果越好。

3.3 空间规律挖掘

3.3.1 热点分析

热点分析(Getis-Ord G_i^*)是一种利用 Getis-Ord G_i^* 指数衡量观测数据之间空间依赖性的技术, 可用于识别局部的空间自相关现象^[37-38]。基于量测邻近观测值, 可确定热点或冷点区域, 具有统计显著性意义的热点不仅要求该位置自身要素具有较高值, 而且周围要素也要具有较高值。本文以居住站点和就业站点的刷卡数作为属性值, 利用 ArcGIS 热点分析工具, 分析每类通勤人群职住热点和冷点区域, 并进行可视化表达, 用于识别每类通勤者就业和居住密度与整体情况有明显差异的位置。Getis-Ord G_i^* 局部统计表达式如公式(5)所示。

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{x} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}} \quad (5)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - \bar{x}^2} \quad (6)$$

其中, G_i^* 为返回的 z 得分, z 得分越高, 聚集越紧密; x_j 为要素 j 的属性值; n 为总数; $w_{i,j}$ 为要素 i 和 j 之间的空间权重, 表示要素 i 和 j 之间空间区域的邻近关系, 其可根据邻接标准或距离标准来度量; \bar{x} 为 x_1, x_2, \dots, x_n 的均值。

4 职住空间关系与通勤时空特征分析

4.1 上海市总体职住特征分析

核密度估计是一种典型的分析点模式空间分布及集聚变化情况的分析方法。由于本文的分析对象为轨道交通乘客的职住区域, 故本文基于识别的通勤人群的通勤数据, 提取居住站点和就业站点, 利用核密度估计方法, 以分析上海通勤人群居住与就业总体的空间聚集特征并进行可视化分析。

如图 3 所示, 居住高聚集区域主要沿内环线分布, 并在内环以外沿轨道线呈放射状向外扩散分布, 沿交通线呈点状离散分布或呈带状分布, 且总体上浦西的居住密度高于浦东。

由图 4 可知, 就业活动主要在内环以内, 就业聚集点在内环内呈现东-西方向多核心分布, 少量就业聚集地散落在内环以外, 但范围不会延伸到外环以外, 散落的高密度就业地是上海一些典型的产业园区, 如东南方的张江高科技园区以及西边的漕河经济技术开发区。

4.2 通勤活动的时空特征分析

通勤活动直接体现为轨道交通站点客流, 站点客流是站点周边用地形态产生的出行需求体现, 亦是站点设施使用情况的直接反映^[39]。由于

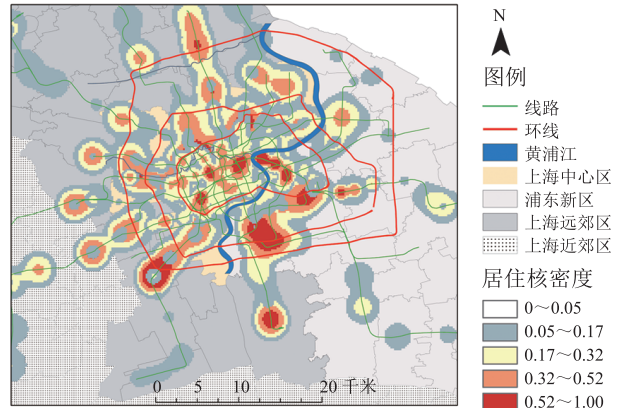


图 3 居住核密度图

Fig. 3 Kernel density of the residence

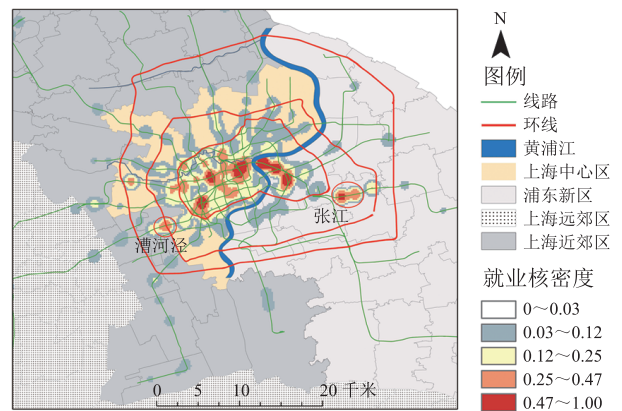


图 4 就业核密度图

Fig. 4 Kernel density of the workplace

通勤人员依据站点进行通勤活动, 故根据构建的职住识别模型识别出 562 268 位通勤人员。随机选取 5% 的通勤人员, 将上述的时间聚类方法应用于抽取的通勤样本。通过计算时间相似度, 可得到一个相似度矩阵实现层次聚类, 表 1 为随机抽取 5 名通勤人计算出的时间相似度示例。由于应用背景是通勤人群的分类, 本实验将聚类数范围初定为 3~7, 计算每个簇数对应的 Silhouette 指数, 并与其他层次聚类进行比较, 结果如表 2 所示。由表 2 可知, Ward 的结果整体较优。此外, 当簇数为 5 时, Silhouette 指数较高, 因此, 为揭示通勤人群的层次结构, 本实验将通勤人群分为 5 簇进行分析。

4.2.1 时间特征分析

根据 5 个簇早晚通勤期 (6:00—10:00、

表 1 时间相似度

Table 1 Temporal similarity

用户	用户				
	1	2	3	4	5
1	4.000	2.562	2.323	2.637	0.000
2	2.562	4.000	3.041	2.334	0.000
3	2.323	3.041	4.000	2.569	0.000
4	2.637	2.334	2.569	4.000	0.000
5	0.000	0.000	0.000	0.000	4.000

表 2 簇数与对应的 Silhouette 指数

Table 2 Number of clusters and Silhouette values

簇数	Silhouette 值			
	Ward	Single	Complete	Average
3	0.302	0.275	0.251	0.256
4	0.305	0.220	0.246	0.293
5	0.306	0.228	0.247	0.299
6	0.276	0.215	0.201	0.284
7	0.271	0.204	0.229	0.256

16:00—22:00)的刷卡量统计可知,各簇的上下班刷卡时间的分布存在一定的差异(图5)。

cluster1 的上班刷卡时间集中在 7:00—8:30, 下班刷卡时间集中在 17:00—19:00; cluster2 的上班刷卡时间分布在 8:00—9:00, 下班刷卡时间集中在 17:30—19:00; cluster3 的上班刷卡时间明显较晚, 出行时间最晚, 主要集中在 8:30—10:00, 下班刷卡时间集中在 16:00—17:30 和 18:30—22:00, 存在加班晚归现象; cluster4 的上班刷卡时间比 cluster1 推迟 0.5 h, 分布在 7:30—8:30, 下班刷卡时间与 cluster1 和 cluster2 相似, 主要集中在 17:30—19:00; cluster5 明显有较多的早出通勤者, 上班刷卡时间集中在 6:00—7:30, 下班时间也偏早, 集中在 16:30—18:30。

利用时间相似的聚类划分, 要求通勤人员不仅在刷卡时刻相似, 还要在出行时段具有一定的相似性, 综合上述的上下班刷卡时间分布, 可将通勤人员汇总成 3 类(表 3): (1)上海大多数通勤人员属于常规型通勤者, 占比为 68%, 在常规型

中的 3 类群体也具有一定规律, cluster1 和 cluster4 上班出行时间更偏早(7:00—8:30), cluster2 上班出行时段偏晚(8:00—9:00); (2)晚出型通勤者占比为 21%, 下班有早归和加班晚归两种情况, 由图 5(c)可知, 19:30 以后下班的通勤人群占比约为 10%; (3)早出型通勤者占比最小, 早出型通勤人员同时也是早归型通勤人员, 占通勤人群的 11%。

4.2.2 空间特征分析

在空间分布上, 基于 ArcGIS 软件, 得到 5 个簇的就业点与居住点冷热空间分布图(图 6)。由图 6 可知: (1)就业热点聚集在城市中心区分布, 5 类就业热点基本分布在中环内, 集中在内环, 结合就业核密度图可知, 就业呈现向心聚集的单中心格局, 但在中心显现就业的多核心结构。(2)各类的居住热点分布存在一定差异, 但与居住核密度图具有一致性。具体地, cluster5 和 cluster4 的居住热点离中心就业区较远, 分布在远郊区的 9 号线末端; cluster3 的居住次热点分布在城南中心区, 离中心就业区最近, 使得其有条件能够晚出行上班, 但需要支付更多的住房成本, 结合时间特征分析, cluster3 可能是工作时间灵活的高收入群体和加班活动多的软件信息就业者; cluster1 和 cluster2 的居住热点主要集中在就业中心区的近郊与远郊区, cluster1 热点分布在 7 号、1 号和 3 号线末端及中心区北边的 11 号线, cluster2 热点分布在 10 号、9 号和 5 号线并向西延伸至远郊区。综上所述, 早出型离就业中心区更远, 晚出型离就业中心区更近。(3)居住的冷点区基本聚集在黄浦江东边及中心区北边, 该情况与浦东与浦西发展时期不同、浦南与浦北发展差异有关。

5 结 语

目前, 通勤群体的划分方法较少考虑通勤行程的时间连续性特征, 针对该问题, 本文进行了

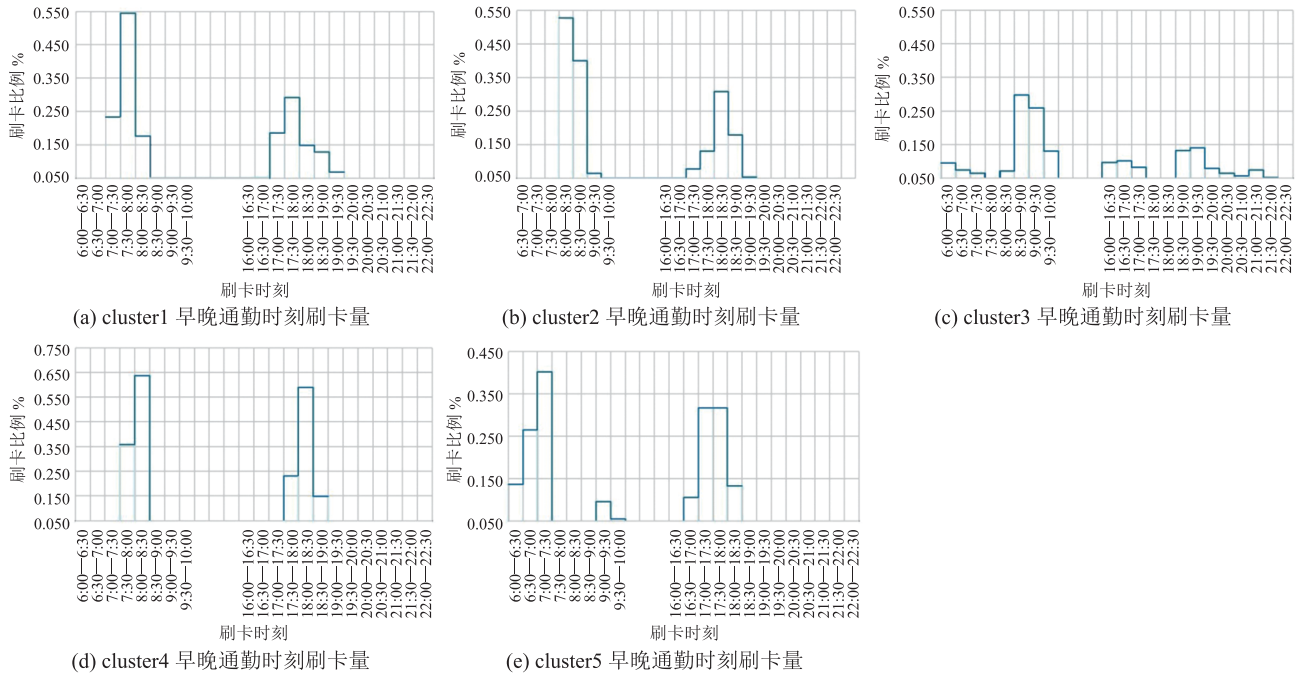


图5 早晚通勤高峰刷卡情况

Fig. 5 Travel rate of morning-evening commuter rush hours

表3 通勤出行时间规律

Table 3 Time patterns of commuting

类型	上班刷卡时间	簇类别	占比	下班刷卡时间
常规型通勤者	偏早 7:00—8:30	cluster1	35%	下班刷卡时间集中在 17:00—19:00
	偏早 7:30—8:30	cluster4	14%	
	偏晚 8:00—9:00	cluster2	19%	
晚出型通勤者	8:30—10:00	cluster3	21%	16:00—17:30, 早归型; 18:30—22:00, 存在加班晚归
早出型通勤者	6:00—7:30	cluster5	11%	下班时间较早, 集中在 16:30—18:30, 属于早出早归型

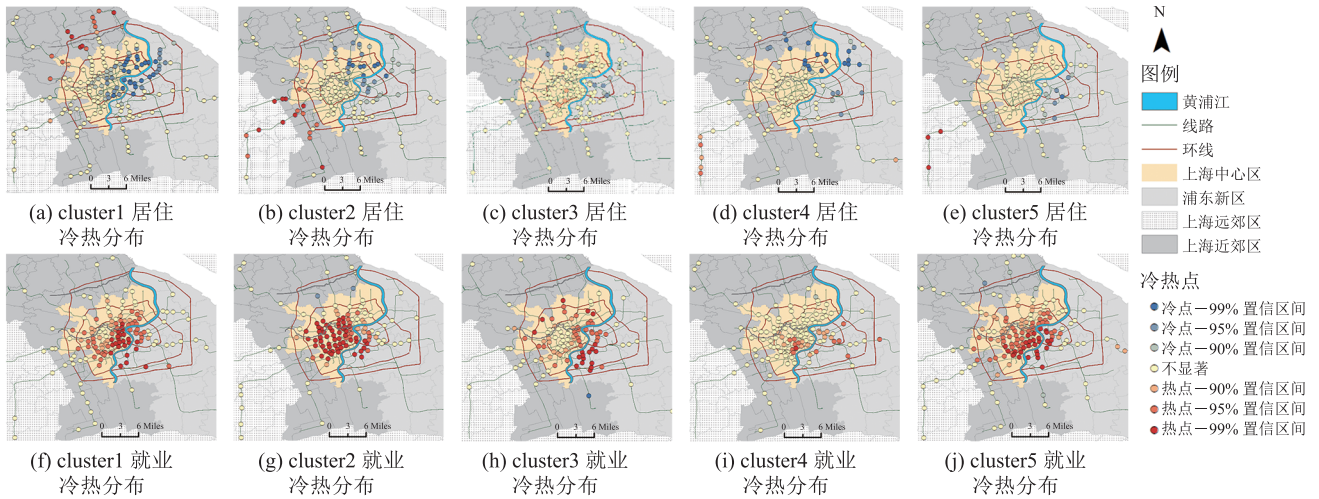


图6 居住与就业冷热空间分布

Fig. 6 Spatial hot-cold distribution map of residence and workplace

进一步的研究。本文基于上海市 288 个轨道交通站点和连续一周的交通卡地铁刷卡数据, 构建了职住识别模型, 用于识别通勤人群和职住地, 定义了一种通勤行程时间相似度的度量方法, 基于层次聚类的结果, 对通勤人群进行细化分析, 并结合热点分析模型, 探究各类型通勤人群居住与就业的热点区域, 挖掘上海市通勤人群潜在的时空规律及职住空间组织特征。

研究发现: (1) 上海市的职住空间呈环状分布, 就业热点区分布在城市中心区域, 具有强就业吸引力; 居住郊区化明显, 居住热点区大多分布在市中心以外, 与总体居住和就业的核密度图一致, 即高密度居住点在内环外沿轨道线分布、高密度就业地在中心区聚集。(2) 按照通勤出行时间规律, 可将上海市通勤人员划分为常规型、晚出型和早出早归型, 与居住热点分布相对应。总体上, 越早出行的类型, 其居住热点离就业中心区越远, 在常规型中偏早出行的两个簇, 其居住热点离就业中心区比偏晚的簇更远, 从侧面验证了上海的单中心结构。此外, 上海大多数的通勤方式属于常规型通勤, 即在 7:00—9:00 进行上班通勤活动, 17:00—19:00 进行下班通勤活动, 占比达 68%。在常规型通勤中, 上班通勤出行时间主要聚集在 7:00—8:30。(3) 各类型的就业热点区差异不明显, 大多集中在内环; 但上海市中心黄浦江东西侧的居住热点分布不一致, 居住热点分布在北部和西部的近郊和远郊区, 居住冷点区主要分布在中心区的东部, 并向浦东新区延伸。

尽管 Song 等^[37]和 Goulet^[38]试图利用 POI 赋予通勤人员属性, 但因 POI 存在多尺度疏密不一的问题, 难以构建有效的属性连接。由于缺少通勤人群的社会经济数据, 故本文未能在不同群体差异和职住空间分布的成因方面进行深入探讨。此外, 职住空间的形成是一个多因素作用和长期动态演化的过程, 尹芹等^[39]学者目前也仅对单一

时间段内的居民职住变化进行研究。可根据居民的通勤模式对职住空间规划提供科学建议, 在后续研究中, 将多年的通勤数据和社会经济数据纳入研究, 从更广的时间度揭示通勤人群层次结构和职住空间分布的变化规律及影响因素, 有助于更好地了解不同通勤群体的需求。

参 考 文 献

- [1] Modarres A. Commuting, energy consumption, and the challenge of sustainable urban development [J]. *Current Opinion in Environmental Sustainability*, 2017, 25: 1-7.
- [2] 朱玮, 梁雪媚, 桂朝, 等. 上海职住优化效应的代际差异 [J]. *地理学报*, 2020, 75(10): 2192-2205.
Zhu W, Liang XM, Gui Z, et al. The inter-generational differences in the effects of job-housing optimization in Shanghai [J]. *Acta Geographica Sinica*, 2020, 75(10): 2192-2205.
- [3] 刘耀林, 陈龙, 安子豪, 等. 基于公交刷卡数据的武汉市职住通勤特征研究 [J]. *经济地理*, 2019, 39(2): 93-102.
Liu YL, Chen L, An ZH, et al. Research on job-housing and commuting in Wuhan based on bus smart card data [J]. *Economic Geography*, 2019, 39(2): 93-102.
- [4] 申犁帆, 张纯, 李赫, 等. 城市轨道交通通勤与职住平衡状况的关系研究——基于大数据方法的北京实证分析 [J]. *地理科学进展*, 2019, 38(6): 791-806.
Shen LF, Zhang C, Li H, et al. Relationship between urban rail transit commuting and jobs-housing balance: an empirical analysis from Beijing based on big data methods [J]. *Progress in Geography*, 2019, 38(6): 791-806.
- [5] 逯琳, 邵世维, 刘辉. 利用浮动车数据挖掘城市职住空间关系 [J]. *测绘地理信息*, 2021, 46(3): 110-113.
Lu L, Shao SW, Liu H. Urban jobs-housing spatial relationship exploration with floating car data [J]. *Journal of Geomatics*, 2021, 46(3): 110-113.
- [6] 张学波, 宋金平, 陈丽娟, 等. 北京都市区就业空

- 间分异与职住空间错位行业识别 [J]. 人文地理, 2019, 34(3): 83-90.
- Zhang XB, Song JP, Chen LJ, et al. Employment spatial differentiation and the identification of industries affecting on jobs-housing spatial mismatch in Beijing Metropolitan Area [J]. *Human Geography*, 2019, 34(3): 83-90.
- [7] 党云晓, 董冠鹏, 余建辉, 等. 北京土地利用混合度对居民职住分离的影响 [J]. 地理学报, 2015, 70(6): 919-930.
- Dang YX, Dong GP, Yu JH, et al. Impact of land-use mixed degree on resident's home-work separation in Beijing [J]. *Acta Geographica Sinica*, 2015, 70(6): 919-930.
- [8] 张艳, 刘志林. 市场转型背景下北京市中低收入居民的住房机会与职住分离研究 [J]. 地理科学, 2018, 38(1): 11-19.
- Zhang Y, Liu ZL. Access to housings and home-work separation of moderate to low-income residents in Beijing under the market-oriented transition [J]. *Scientia Geographica Sinica*, 2018, 38(1): 11-19.
- [9] 吴翔华, 陈昕雨, 袁丰. 南京市住房困难人群职住关系及影响因素分析 [J]. 地理科学进展, 2019, 38(12): 1890-1902.
- Wu XH, Chen XY, Yuan F. Job-housing relationship of people with housing difficulties and influencing factors in Nanjing City [J]. *Progress in Geography*, 2019, 38(12): 1890-1902.
- [10] 塔娜, 柴彦威. 基于收入群体差异的北京典型郊区低收入居民的行为空间困境 [J]. 地理学报, 2017, 72(10): 1776-1786.
- Ta N, Chai YW. Spatial dilemma of suburban low-income residents: an analysis of behavior space among different income groups [J]. *Acta Geographica Sinica*, 2017, 72(10): 1776-1786.
- [11] 王德, 李丹, 傅英姿. 基于手机信令数据的上海市不同住宅区居民就业空间研究 [J]. 地理学报, 2020, 75(8): 1585-1602.
- Wang D, Li D, Fu YZ. Employment space of residential quarters in Shanghai: an exploration based on mobile signaling data [J]. *Acta Geographica Sinica*, 2020, 75(8): 1585-1602.
- [12] 田金玲, 王德, 谢栋灿, 等. 上海市典型就业区的通勤特征分析与模式总结——张江、金桥和陆家嘴的案例比较 [J]. 地理研究, 2017, 36(1): 134-148.
- Tian JL, Wang D, Xie DC, et al. Identifying the commuting features and patterns of typical employment areas in Shanghai using cellphone signaling data: a case study in Zhangjiang, Jinqiao and Lujiazui [J]. *Geographical Research*, 2017, 36(1): 134-148.
- [13] 罗吉, 黄亚平, 赵丽元, 等. 城市低收入群体二元居住需求与住房供给研究——基于微观实证调查 [J]. 城市规划, 2015, 39(3): 86-93.
- Luo J, Huang YP, Zhao LY, et al. A study on dualistic housing demand of urban low-income groups and housing supply: based on micro-empirical investigations [J]. *City Planning Review*, 2015, 39(3): 86-93.
- [14] 朱洪, 陈欢, 范瑛, 等. 移动互联网背景下的上海市人员出行调查关键技术实践与创新 [J]. 交通与港航, 2022, 9(2): 77-81.
- Zhu H, Chen H, Fan T, et al. Practice and innovation of key technology of Shanghai travel characteristics investigation in the context of mobile internet [J]. *Communication & Shipping*, 2022, 9(2): 77-81.
- [15] 林书亭, 塔娜. 转型期不同住房性质居民的职住关系演变——以上海郊区为例 [J]. 人文地理, 2021, 36(2): 35-45.
- Lin ST, Ta N. The evolution of job-housing relationship among residents of different housing properties in the transition period: a case study in suburban Shanghai [J]. *Human Geography*, 2021, 36(2): 35-45.
- [16] Yang XP, Fang ZX, Xu Y, et al. Revealing temporal stay patterns in human mobility using large-scale mobile phone location data [J]. *Transactions in GIS*, 2021, 25(4): 1927-1948.
- [17] Kieu LM, Bhaskar A, Chung E. Passenger segmentation using smart card data [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(3): 1537-1548.
- [18] 付晓, 陈梓丹, 黄洁. 基于手机信令数据的城市居

- 民非通勤出行群体画像——以苏州市为例 [J]. 地理科学, 2022, 42(10): 1727-1734.
- Fu X, Chen ZD, Huang J. Non-work travel group profiles of urban residents based on mobile phone signaling data: a case of Suzhou [J]. *Scientia Geographica Sinica*, 2022, 42(10): 1727-1734.
- [19] 万明, 梁莹, 周涂强, 等. 考虑动态交通信息的异质出行行为分析 [J]. 华东交通大学学报, 2022, 39(5): 61-69.
- Wan M, Liang Y, Zhou TQ, et al. Heterogeneous travel behavior analysis considering dynamic traffic information [J]. *Journal of East China Jiaotong University*, 2022, 39(5): 61-69.
- [20] Long Y, Liu XJ, Zhou JP, et al. Early birds, night owls, and tireless/recurring itinerants: an exploratory analysis of extreme transit behaviors in Beijing, China [J]. *Habitat International*, 2016, 57: 223-232.
- [21] 郭文露, 刘艳芳, 刘子靖, 等. 基于地铁刷卡数据的乘客特别出行行为研究——以武汉市地铁1、2、4号线为例 [J]. 测绘地理信息, 2019, 44(2): 29-33.
- Guo WL, Liu YF, Liu ZJ, et al. Study on special travel behaviors using subway smart card data: a case study the lines 1, 2 and 4 of Wuhan Metro [J]. *Journal of Geomatics*, 2019, 44(2): 29-33.
- [22] Faroqi H, Mesbah M. Inferring trip purpose by clustering sequences of smart card records [J]. *Transportation Research Part C: Emerging Technologies*, 2021, 127: 103131.
- [23] Zhou XL. Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago [J]. *PLoS one*, 2015, 10(10): e0137922.
- [24] 孙世超, 庄斌, 黄伟. 基于机器学习的公交卡数据中通勤人群辨识方法 [J]. 交通工程, 2017, 17(1): 58-64.
- Sun SC, Zhuang B, Huang W. An approach to the identification of commuters based on machine learning of smartcard data [J]. *Journal of Transportation Engineering*, 2017, 17(1): 58-64.
- [25] 龙瀛, 张宇, 崔承印. 利用公交刷卡数据分析北京职住关系和通勤出行 [J]. 地理学报, 2012, 67(10): 1339-1352.
- Long Y, Zhang Y, Cui CY. Identifying commuting pattern of Beijing using bus smart card data [J]. *Acta Geographica Sinica*, 2012, 67(10): 1339-1352.
- [26] 严海, 杨荣荣, 熊文, 等. 步行和自行车接驳轨道交通时空阈值研究 [J]. 城市交通, 2013, 11(2): 83-90.
- Yan H, Yang RR, Xiong W, et al. Critical accessing time and distance for pedestrian and cyclists to urban rail transit [J]. *Urban Transport of China*, 2013, 11(2): 83-90.
- [27] 李辉, 王东炜, 赵湘育, 等. 不同类型轨道交通站点步行与电动自行车接驳分析 [J]. 铁道科学与工程学报, 2015, 12(6): 1493-1499.
- Li H, Wang DW, Zhao XY, et al. Access analysis for pedestrian and electric bicycle to different types of rail transit stations [J]. *Journal of Railway Science and Engineerin*, 2015, 12(6): 1493-1499.
- [28] 翁小雄, 吕攀龙. 基于GBDT算法的地铁IC卡通勤人群识别 [J]. 重庆交通大学学报(自然科学版), 2019, 38(5): 8-12.
- Weng XX, Lv PL. Subway IC card commuter crowd identification based on GBDT algorithm [J]. *Journal of Chongqing Jiaotong University (Natural Science)*, 2019, 38(5): 8-12.
- [29] Faroqi H, Mesbah M, Kim J. Comparing sequential with combined spatiotemporal clustering of passenger trips in the public transit network using smart card data [J]. *Mathematical Problems in Engineering*, 2019, 2019: 5070794.
- [30] Pasupathi S, Shanmuganathan V, Madasamy K, et al. Trend analysis using agglomerative hierarchical clustering approach for time series big data [J]. *The Journal of Supercomputing*, 2021, 77(7): 6505-6524.
- [31] Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview, II [J]. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2017, 7(6): e129.
- [32] 张冬梅. 基于轮廓系数的层次聚类算法研究 [D]. 秦皇岛: 燕山大学, 2009.
- Zhang DM. Research on hierarchical clustering

- algorithm based on silhouette [D]. Qinhuangdao: Yanshan University, 2009.
- [33] 王德, 李丹, 傅英姿. 基于手机信令数据的上海市不同住宅区居民就业空间研究 [J]. 地理学报, 2020, 75(8): 1585-1602.
- Wang D, Li D, Fu YZ. Employment space of residential quarters in Shanghai: an exploration based on mobile signaling data [J]. *Acta Geographica Sinica*, 2020, 75(8): 1585-1602.
- [34] He SW, Yu S, Wei P, et al. A spatial design network analysis of street networks and the locations of leisure entertainment activities: a case study of Wuhan, China [J]. *Sustainable Cities and Society*, 2019, 44: 880-887.
- [35] 彭诗尧. 基于土地利用的城市轨道交通客流时空特性分析 [D]. 北京: 北京交通大学, 2020.
- Peng SY. Spatial and temporal characteristics analysis of urban rail transit passenger flow based on land use [D]. Beijing: Beijing Jiaotong University, 2020.
- [36] 李亮, 赵星, 张海燕, 等. 基于时空维度变量的杭州市轨道交通站点聚类研究 [J]. 北京交通大学学报, 2022, 46(4): 31-42.
- Li L, Zhao X, Zhang HY, et al. Clustering research on Hangzhou metro station based on spatio-temporal variables [J]. *Journal of Beijing Jiaotong University*, 2022, 46(4): 31-42.
- [37] Song ZS, Wilhelmsson M, Yang Z. Constructing segmented rental housing indices: evidence from Beijing, China [J]. *Property Management*, 2022, 40(3): 409-436.
- [38] Goulet-Langlois G. Exploring regularity and structure in travel behavior using smart card data [D]. Massachusetts: Massachusetts Institute of Technology, 2015.
- [39] 尹芹, 孟斌, 张丽英. 基于客流特征的北京地铁站点类型识别 [J]. 地理科学进展, 2016, 35(1): 126-134.
- Yin Q, Meng B, Zhang LY. Classification of subway stations in Beijing based on passenger flow characteristics [J]. *Progress in Geography*, 2016, 35(1): 126-134.