

黄瑞克¹ 李力¹

1. 东南大学建筑学院; 101012053@seu.edu.cn

Huang Ruike¹ Li Li¹

1. School of Architecture, Southeast University

国家重点研发计划 2019YFD1100805

基于定位大数据和行动轨迹聚类的乡村居民行为特征挖掘

——以宜兴市丁蜀镇西望村为例

Behavior Feature Mining of Rural Residents Based on Location Big Data and Action Trajectory Clustering: A Case Study for Xiwang Village, Dingshu, Yixing

摘要:乡村振兴离不开乡村建设,当今乡村的建设规划与设计具有自上而下的主观性质,一定程度造成了低效建设和资源浪费,大数据应用的发展为乡村建设指导打开了新思路。本文以宜兴市丁蜀镇西望村为研究对象,运用 Wi-Fi 探针定位技术进行中尺度、无感知的居民定位大数据采集,经过多层级数据清洗及路径可视化处理后得到单日乡村居民个体行动轨迹图。通过卷积自编码神经网络实现行为轨迹图的降维后进行聚类分析和可视化,进一步结合乡村特征,总结出该乡村行为特征并归纳得出 5 种该乡村中典型的行为模式。本研究较为全面反映了乡村居民生活的实际情况,为乡村人居环境研究提供了新方法,能够为乡村规划和改造提供决策支持。

关键词:乡村行为模式;行动轨迹聚类;Wi-Fi 探针定位技术

Abstract: Rural revitalization is inseparable from rural construction, and now the planning and design for rural construction have a top-down subjective nature, resulting in inefficient construction and waste of resources to a certain extent. The development of big data applications has opened up new ideas for rural construction guidance. This paper takes Xiwang Village, Dingshu Town, Yixing City as the research object, and uses Wi-Fi probe positioning technology to collect middle-scale, non-perceptual resident positioning big data. After multi-level data cleaning and path visualization processing, the single-day rural residents are obtained graph of individual movement trajectories. The dimensionality reduction of the behavioral trajectory graph is achieved through the convolutional auto-encoding neural network, and then clustering analysis and visualization is carried out, further combined with the characteristics of the village, five typical behavioral patterns in the village are summarized. This research comprehensively reflects the actual living conditions of rural residents, provides a new method for the study of rural living environment, and can provide decision support for rural planning and reconstruction.

Keywords: Rural Behavior Pattern; Action Trajectory Clustering; Wi-Fi Probe Positioning Technology

推进全面乡村振兴并逐步实现全体人民共同富裕是近年来国内重大议题,而乡村环境建设和改造是乡村

振兴的主力军之一。国内乡村数量极其众多且多样,而直接参与乡村建设的设计师数量相对较少,设计师无法

深入调研每个乡村且缺乏客观、理性而有效的数据支撑,导致设计往往依赖设计师自身经验和主观判断,经常出现“城市视角”的乡村建设,出现一定程度的过度开发和资源浪费。

对环境中的居民的行为特征进行分析能够帮助设计师有效、快速了解该环境的潜在因素与特点,熟悉场地人群的活动方式,从而更好地帮助设计师设计出以人为本、因地制宜地设计方案。国内外已有不少学者从环境行为特征入手探究场地特点。丹麦学者杨·盖尔于1970年就已在丹麦皇家建筑学的专项研究通过PSPL调研法^[1](公共空间-公共生活调研法)对环境中的公共活动模式进行研究。在国内,冯建喜等^[2]以国家统计局的调查数据为基础,研究老年人在城市空间中的出行特征,为城市空间的适老化发展提供有效支撑。随着科技发展,相较于传统需要大量人工与时间进行数据收集的调研方式,新兴大数据采集技术和分析技术的介入能够收集到更全面的数据基础和提供更深度的数据分析。陈仲等^[3]通过对手机信令数据进行分析归纳出三亚市中居民的典型出行模式;徐晓伟等^[4]以城市交通卡数据为基础,探索多元出行数据下居民行为特征的分析方法。

通过大数据对居民行为特征进行分析进而支撑环境改造,在城市中已有广泛的应用基础,但在国内数以万计的乡村中,由于基础设施薄弱,大数据技术并未在乡村中得到充分实施。Wi-Fi 探针技术能够有效适应乡村环境的中尺度、弱基础设施的现状,对人群数据进行高效采集。而乡村具有人口基数小、人口流动性弱等特点也使得乡村行为特征研究能够有效指导乡村建设。

本文以宜兴市丁蜀镇西望村为研究对象,通过 Wi-Fi 探针技术采集的定位大数据构建乡村中居民个体行动轨迹。结合乡村特征对轨迹图像进行降维和聚类,挖掘出特有行为特征和5种居民当日典型行为模式,为西望村后续的环境规划与改造提供了坚实的支撑。

1 数据与方法

1.1 Wi-Fi 探针定位技术

常用的定位技术有GPS、蓝牙定位、超宽带(UWB)定位等,GPS范围大但是精度低,不适合乡村中尺度的定位应用;蓝牙定位和超宽带定位虽然能在中尺度上做到高精度定位,但需要被测者额外携带信号发射设备。

Wi-Fi 探针技术基于IEEE 802.11协议,采用WLAN(无线局域网)技术实现对开启或连接Wi-Fi的设备(如智能手机、智能手表、智能家居等)进行MAC信息

采集,每台联网设备具有全球独一无二的MAC地址,结合采集设备的位置信息便可以对智能设备进行定位。随着乡村中智能手机的普及,通过Wi-Fi探针技术对乡村中智能设备携带者进行定位数据采集已可以实现。

本次研究采用由东南大学建筑学院建筑运算与运用研究所研发的Wi-Fi嗅探设备进行数据收集,其具有低环境介入、高便携性、无感知收集和续航持久等特点,仅需被测者手机Wi-Fi开启状态就能收集到其定位信息,能够被轻易布置在乡村各角落进行定位数据收集。无感知的收集方式也能最大程度保证数据的真实性。

结合村落环境在特定空间节点进行Wi-Fi嗅探装置布点(图1),本次研究采集到1300万余条定位数据,巨大的样本量为后续客观理性分析提供有效支撑。

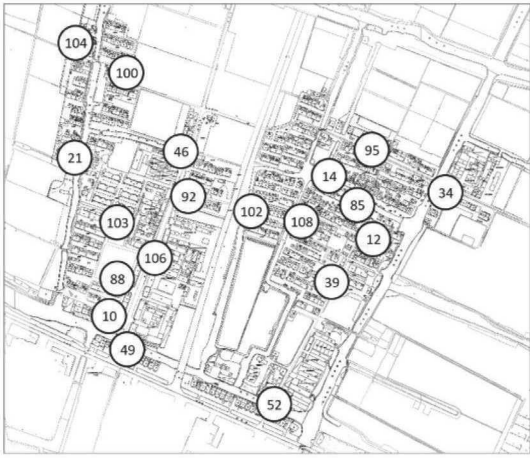


图1 西望村Wi-Fi嗅探装置布点(数字为装置编号)
(图片来源:作者自绘)

1.2 定位数据清洗及整合

采用Wi-Fi探针技术进行的定位数据采集方式缺点是会收集到大量干扰信息:①凡是具有Wi-Fi功能的设备均会被纳入定位对象中;②本次研究采集数据重点是行动轨迹,而部分经过性质的瞬时数据也会对主要数据造成干扰;③部分智能手机具有MAC信息隐私保护功能,会向外发送伪MAC地址而导致设备无法获取正确的定位信息;④受到信号干扰,部分定位信息获取不完全以至呈片段式,会对轨迹数据总体分析造成影响。

运用相应的数学统计算法、关联规则算法及结合伪MAC地址特征标识,能够针对上述问题对数据进行高效清洗,清洗后有效数据有370余万条,极大提高了定位大数据的可靠性。以清洗后具备有效性的单个MAC地址为索引,便可获取该MAC在村庄中一天内各时段定位坐标变化(平均每日180余条)。本研究中由于只需要保留轨迹拓扑关系,故采用村内相对坐标(表1)。

表 1 单个 MAC 地址坐标信息示例
(来源:作者自绘)

MAC	探针编号	相对坐标	时间	信号强度
180,251,xxx,xxx,1,124	108	(71.972,42.254)	2021-04-17 06:06:17	-34
180,251,xxx,xxx,1,124	14	(78.651,51.368)	2021-04-17 07:55:54	-33
...
180,251,xxx,xxx,1,124	108	(71.972,42.254)	2021-04-17 21:42:01	-36

1.3 行动轨迹可视化

经过数据清洗后,整理得到 8000 余个 MAC 地址的坐标变化信息。MAC 地址坐标变化信息经过处理后能够得到该 MAC 在当日的行动轨迹。受限于 Wi-Fi 嗅探设备的铺设密度,其初始数据仅能以“点到点”形式呈现,但在布置设备时综合考虑了设备之间路径的唯一性,同时,结合 Dijkstra 寻径算法计算出任意设备间的最短路径,便可以批量化地尽可能真实还原每个被测目标在村落中真实的行动轨迹(图 2)。

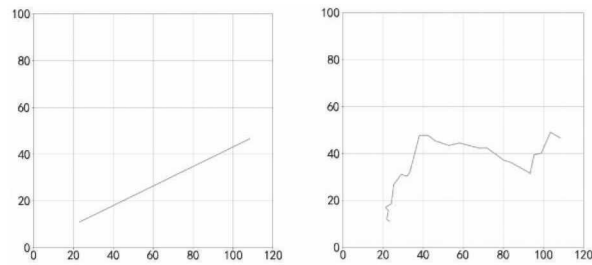


图 2 相对坐标下还原后轨迹
(图片来源:作者自绘)

活规律性强,一天中的行动轨迹如仅在二维平面上表示则会有大量重合,导致时间信息无法呈现。将时间作为第三维度,并以不同色彩表示行动轨迹在一天中的时间分布,相应的时间信息就能体现在所绘制的轨迹图中。由图 3 可以看出某 MAC 地址在 2021 年 4 月 16 日中在村落中的行动轨迹时空分布,结合西望村实际环境可做以下推断:该被测对象 8 点左右从住处前往工作地点,12 点返回住处,午休 1 小时左右后返回工作地点,之后于 18 点结束工作回到住处。

1.4 基于轨迹图像聚类

MAC 轨迹数据数量庞大,且每个轨迹数据所记录的坐标变化规律不尽相同,直接对坐标数据进行聚类效果较差。

1.4.1 卷积自编码神经网络降维

卷积自编码神经网络(convolutional auto-encoder,

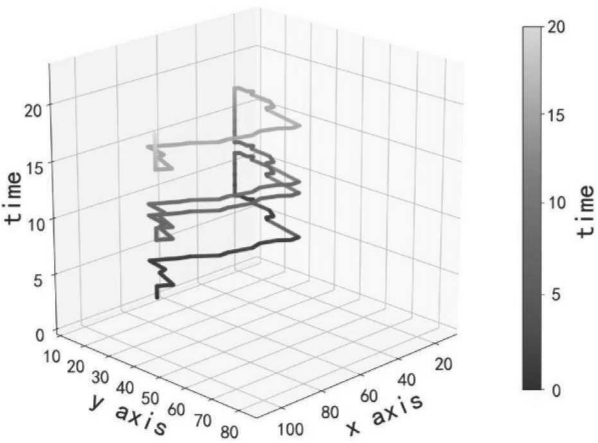


图 3 2021 年 4 月 16 日某被测对象行动轨迹时空分布
(图片来源:作者自绘)

CAE)综合了卷积神经网络和自编码神经网络两者特点,其不具有类型标签,是一种无监督学习,目前在图像特征提取方面已有广泛应用。对其进行训练学习,经过一系列卷积、池化和反卷积操作后,卷积自编码神经网络能够提取出尽可能包含图像所有信息的特征向量集,从而实现图像信息降维,以得到的特征向量集为基础便能实现较为精确的聚类。

本文以定位大数据得到的 8000 余张像素为 $128 * 128$ 的 RGB 轨迹图像集为训练基础,设定相应的卷积自编码神经网络卷积层、池化层参数后,最终将每个轨迹图像降维至由 24 个特征向量构成的集合。

1.4.2 降维后数据聚类

本文使用基于 K-means 改良的 K-means++ 算法对神经网络降维后的 24 维数据进行聚类。K-means++ 有效解决了 K-means 初始化随机种子的不确定性,能够实现更好的聚类结果。

与 K-means 相同,K-means++ 需要在聚类前人为确定聚类数量,轮廓分析(silhouette analysis)能够对聚类质量进行定量分析,其将聚类样本与最近簇中所有点之间的平均距离——分离度,与聚类样本与簇内其他点之

间的平均距离——内聚度相减,再除以二者中的较大值,从而得到轮廓系数。轮廓系数取值范围为 $-1\sim 1$,越接近于1表示聚类质量越好。

本文设定 k 值范围为 $2 \sim 20$, 针对每个 k 值多次计算求得轮廓系数的平均值, 根据轮廓系数最大值所对应的 k 值来确定最终聚类数。当 $k=5$ 时, 轮廓系数达到范围内最大值 (图 4), 故将聚类数设定为 5。

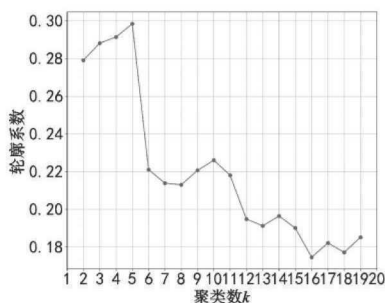


图 4 聚类数 k 与轮廓系数分布图
(图片来源:作者自绘)

采用 t-SNE 算法可将高维数据集映射到二维平面以观察聚类效果(图 5),不同簇按照颜色区分,可见降维后相互重叠较少,聚类结果较为明晰。

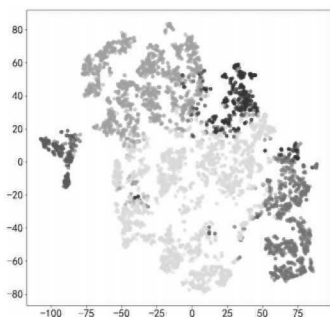


图 5 t-SNE 算法降维后的可视化聚类效果
(图片来源:作者自绘)

2 轨迹特征分析

2.1 总体特征

将所有轨迹路径分解成排布在平面上的点阵,当个体经过或停留在其中一点时,点的大小随之变大,由此可反映居民整体在场地中的行动和停留状况。如图6所示,深色区域越大即代表居民在此停留时间越长,图中序号为探针编号。可以发现西望村最大量的居民活动集中在108、102、92号探针处,这也是西望村居住区的中心区域。其次是西边蜀古路沿线(46、92、103、106、49号探针)——西望村核心商业区和研发生产区,和东北角(95、34号探针)——西望村较小的研发生产区,两地

行动轨迹密度仅次于西望村中心居住区。

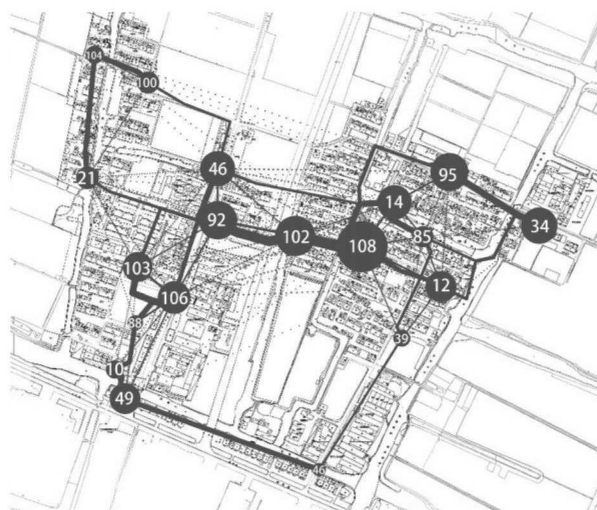


图6 西望村总体行动轨迹(数字为设备编号)

(图片来源:作者自绘)

2.2 行动轨迹聚类及特征分析

通过卷积自编码神经网络对 $128 \times 128 \times 3$ 维度图像降维至 24 维特征向量后,结合 K-means 十聚类算法,本文最终在 8000 余个行动轨迹基础上得到 5 种乡村典型行为模式。

2.2.1 空间特征

图 7 为不同模式的行动轨迹,其中区域深色面积越大代表居民个体在此停留时间越长。可以看出,占比 44.35% 的模式 1 居民行动轨迹主要分布在中心居住区和村西南边商业区,但一直在一点原地停留的活动方式占绝大多数,居民流动性较弱;占比 27.53% 的模式 2 主要分布在蜀古路沿线,蜀古路北段具有更多行动轨迹分布,且此处是蜀古路区域与西望村蜀古路以东区域之间的主要交通口;模式 3 在西望村东北角具有更为活跃的行动轨迹,但其总体占比只有古蜀路沿线的一半左右,为 15.47%;模式 4 与模式 1 相似,主要活动区域为中心居住区,但更少涉足于西南的商业区,其占比仅有模式 1 的五分之一左右,为 8.32%,但模式 4 在中心居住区的流动性更大;模式 5 占比仅有 4.33%,主要分布在西望村西北角,且在西望村的活动范围大部分局限在蜀古路以西。

归纳后可得出,模式1与模式4居民行动轨迹能够代表大部分西望村居民的行为模式,即主要在中心居住区和西南边商业区活跃;模式2可代表西望村的商业行为模式,其以蜀古路为交通依托;模式3活跃区与中心居住区相近;模式5体现出希望村西北角居民的行动轨迹,其活跃性远不及其他类别。



图7 不同行为模式的空间分布

(图片来源:作者自绘)

2.2.2 时间特征

大部分居民行动轨迹特征呈现为一直在同一地点不变,其行动轨迹在一天内的时间属性不具有研究意义。将不同模式群体中具有空间变化特征的个体行动轨迹单独绘制,能够进一步对不同模式群体行为的时间属性进行分析。

图8为所有模式中有移动属性的行动轨迹在时间作为第三维度下的轨迹图,模式1的行动轨迹无论从时间分布还是空间分布相较于其他模式都具有更强的不规律性,且活动范围遍及西望村大部分区域,主要体现在

西南角和东北角之间;主体分布在蜀古路的模式2主要沿蜀古路进行南北向移动,偶尔东西向进入居住区,这种行动模式在一天中几乎任何时间都有发生;模式3主要由西望村东北角进入村内,这种行动模式从6点左右开始发生,一直持续到晚8点左右;与模式1活动范围相似的模式4表现出更强的规律性,从早6点左右开始至晚9点左右表现为在中心居住区东西向道路活动;模式5从西望村西北边进入村内,从6时开始产生进入村内的移动轨迹并持续到晚8点左右。

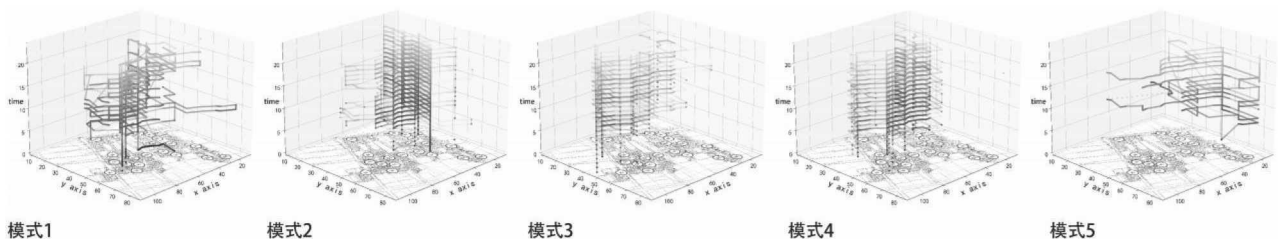


图8 不同行为模式的时空分布

(图片来源:作者自绘)

图9为不同聚类中个体发生移动行为的次数总和在一天内的时间分布,其横轴越宽代表发生活动的次数越多。除模式5从6点才开始有活动发生外,其余行为模式均从5点就开始有活动发生。其中占比多但流动性较弱、位于中心居住区的模式1在晚6点达到活动高峰,其余时间活动次数并无明显变化;靠近蜀古路、位于商业区的模式2在早9点达到活动高峰,且活动发生次数随时间波动性较大;分布在西望村东北角的模式3活动次数在中午1点左右出现次高峰,在饭前即下午5点左右,和饭后晚9点左右达到活动高峰;与模式1分布相近的模式4在上午7点、中午12点和下午4点达到活动高峰,且在晚11点仍然有较多活动发生;模式5活动在中午12点后出现低谷,在晚6点到达活动高峰,但在晚10点便无活动。

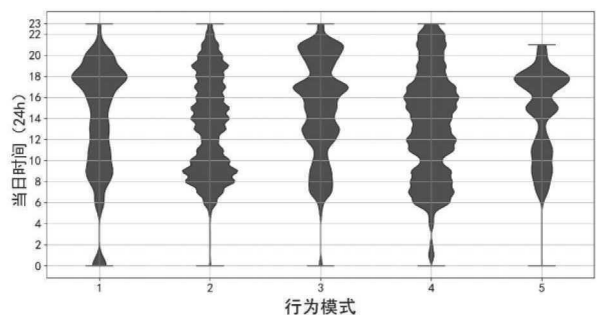


图9 不同模式中个体发生移动行为次数总和的时间分布

(图片来源:作者自绘)

总体而言,不同的行为模式表现出了差异性较大的活动时间分布特点,5种模式整体活动发生次数随时间节点变化而变化的规律性较强。

3 结论与展望

本文依托 Wi-Fi 探针技术收集的宜兴市丁蜀镇西望村居民定位大数据进行了行动轨迹可视化和聚类分析,挖掘出西望村居民的行为特征和 5 种行为模式,且 5 种行为模式均表现出了不同的空间分布和时间分布特征。相较于传统人工走访调研的村落环境的方式,本文采用的大数据采集、处理与分析方式能够有效帮助设计师更快速、更全面了解村内生活环境,为乡村设计和改造提供数理性支撑。

然而,本文的研究仍然具有一定局限性。本文所研究的时间跨度局限在单日,不能体现长时间跨度下的行为特征。另外,受限于 Wi-Fi 探针对被测对象所携带智能设备的要求,部分没有智能设备的小孩、老人无法被纳入分析对象,导致分析结果不具有全面性。今后,更多相关数理关系算法的应用能够让定位大数据得到更细致的挖掘,随着自研 Wi-Fi 标签的应用,老人和小孩也

能够进一步被纳入探测对象,使研究更具全面性。伴随相关大数据采集设备应用进一步成熟以及数据处理、分析方式更加系统,在未来会有更多乡村大数据的研究探索。

参考文献

- [1] GEHL J,GEMZOE L. Public Space-Public Life [M]. Copenhagen: The Danish Architectural Press,1996.
- [2] 冯建喜,杨振山. 南京市城市老年人出行行为的影响因素[J]. 地理科学进展,2015,34(12): 1598-1608.
- [3] 陈仲,杨克青. 居民个体出行行为聚类及出行模式分析——以三亚市为例[J]. 上海城市规划,2020(5):30-35.
- [4] 徐晓伟,杜一,周国春. 基于多源出行数据的居民行为模式分析方法[J]. 计算机应用,2017,37(8): 2362-2367.