

基于多源信息的高空间分辨率人口分布研究*

刘治成¹, 肖东升^{1,2,✉}

(1. 西南石油大学 土木工程与测绘学院, 四川 成都 610500; 2. 西南石油大学
测绘遥感地理信息防灾应急研究中心, 四川 成都 610500)

摘要: 针对已有的人口空间化研究多采用静态数据、时空分辨率较低、在应急救援等方面实用性不高的问题, 提出了一种使用高时空分辨率数据, 结合城市圈层结构理论和主成分分析法的建筑物尺度人口估算方法。以成都市为例, 利用腾讯位置大数据, 通过计算不同城市圈层的定位率, 得到了成都市不同时段 1 km × 1 km 的人口分布数据。在此基础上, 以基于建筑物中心点的泰森多边形为人口分配基本单元, 结合宜出行热力数据和 POI 数据, 分别计算其对人口分布的贡献值并赋予计算权值, 得到了成都市青羊区建筑物尺度人口分布数据。街道尺度统计数据回归分析的决定系数 R^2 为 0.926 4, 总体精度较高, 模拟人口分布符合实际情况。

关键词: 应急救援; 主成分分析; 人口分布; POI; 建筑物

中图分类号: P315.94

文献标识码: A

文章编号: 1000-0666(2021)02-0251-11

0 引言

人口是自然灾害最重要的承灾体, 人口分布数据是应急救援的关键支撑数据 (郭建兴等, 2018)。人口分布与城市的经济发展、基础设施建设以及生态环境保护等都有着紧密联系 (李国平, 陈秀欣, 2009)。目前, 我国人口统计是以行政区划逐级统计和汇总的, 最低统计到乡镇街道一级, 10 年更新一次, 不能满足灾害应急救援等方面的应用要求。探究城市内部更小尺度的人口空间分布是人口空间化研究的重点之一 (李泽宇, 董春, 2019; 贾召亮等, 2020)。

随着移动终端及 GPS 技术的高速发展, 移动终端的位置服务进入了研究者的视野。例如王贤文等 (2017) 利用腾讯位置服务相关数据研究了京津冀地区在短期内的人口流动; 潘碧麟等 (2019) 通过微博签到数据研究了成渝之间的城际人口流动规律; 吴中元等 (2019) 利用腾讯位置

大数据对南京市江宁区秣陵街道进行了小区级别的人口数据预测; Sakaki 等 (2010) 通过提取日本社交软件用户在分享信息的同时携带的位置信息, 研究了地震对 Twitter 用户时空行为造成的影响。

在灾害应急救援中, 灾后第一时间的人口空间分布数据可帮助应急救援部门快速部署救援力量, 为受灾人员的救援赢得宝贵的时间 (郭建兴等, 2018)。腾讯位置大数据可以实时记录使用腾讯位置服务用户的位置信息, 空间分辨率为 1 km, 目前腾讯位置服务的使用量已突破日均 500 亿人次, 所以该数据在研究人口空间分布方面具有一定的代表性。该数据获取简单, 可快速获取灾区震前 1 km 空间分辨率的使用腾讯位置大数据服务的人口空间分布, 为救援力量的初步部署提供辅助数据。要更精确地锁定可能受灾人群的具体位置, 缩小救援人员的搜救范围, 提高搜救效率, 则需要分辨率更高的人口空间分布数据。宜出行热力数据是腾讯地图通过模拟计算出的人口密度

* 收稿日期: 2020-11-03.

基金项目: 国家自然科学基金 (51774250)、四川省软科学研究计划项目 (2019JDR0112)、工程结构安全评估与防灾技术四川省青年科技创新研究团队 (2019JDTD0017) 和西南石油大学测绘遥感地信青年科技创新团队 (2019CXTD07) 联合资助。

第一作者简介: 刘治成 (1996-), 硕士研究生, 主要研究方向为地理信息工程. E-mail: 2371944224@qq.com.

✉通讯作者简介: 肖东升 (1974-), 教授, 主要研究方向为测绘科学与技术, 城市防灾减灾应急和油气遥感.
E-mail: xiaodxsds@163.com.

等级,表达人口密度的相对大小,具有实时更新、实时获取的特性,且该数据空间分辨率达到了 27 m,可为更小尺度的人口空间化研究提供一定的参考。所以本文采用微信宜出行热力数据,结合 POI 兴趣点数据对小范围内建筑物尺度的人口空间化进行研究。

1 研究方法

1.1 基本原理

1.1.1 主成分分析

主成分分析 (Principal Components Analysis, PCA) 是通过将原始、复杂、多维的变量进行线性组合,构造互不相关的综合变量,来降低变量维度的一种数据降维处理方法。数据经主成分分析后可在降低数据维度的同时保持原始数据的大部分特征,减少数据的冗余程度,更利于进行后续处理 (贾俊平, 2008)。

设某数据集 \mathbf{X} 存在 n 个样本,每个样本有 m 个变量,则该数据集的矩阵如下:

$$\mathbf{X}_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

由于在实际问题中,矩阵的维度过大可能会导致后续处理复杂,耗时过久,通常需要对数据进行主成分分析以降低数据维度,其具体步骤为:为了消除由于数据量纲不同引起的误差,对原始矩阵进行标准化处理,得到矩阵 $\mathbf{X}_{m \times n}^*$;计算矩阵 $\mathbf{X}_{m \times n}^*$ 的协方差矩阵;将 $\mathbf{X}_{m \times n}^*$ 的协方差矩阵计算化为相关系数矩阵,公式为:

$$\rho_{ij} = \text{cov}(x_i, x_j) / \sqrt{\sigma_{ii} \sigma_{jj}} \quad (2)$$

计算出特征值 $\lambda_1, \lambda_2, \cdots, \lambda_m, \lambda_i \geq 0, i = 1, 2, \cdots, m$,并从小到大排列,得到相对应的正交单位化特征向量 $\theta_1, \theta_2, \cdots, \theta_m$;计算得到各个主成分的贡献率,公式为:

$$P_i = \lambda_i / \sum_{i=1}^m \lambda_i \quad (i = 1, 2, \cdots, m) \quad (3)$$

则其累计贡献率为:

$$P_i(\text{sum}) = \sum_{i=1}^m P_i \quad (i = 1, 2, \cdots, m) \quad (4)$$

主成分提取原则通常按累计贡献率达到约 80% 以上,或主成分特征值大于 1 提取主成分。得到提取的主成分后,计算每个主成分的得分,公式为:

$$\omega_{ri} = \theta_{1i} \mathbf{X}_{r1}^* + \theta_{2i} \mathbf{X}_{r2}^* + \cdots + \theta_{mi} \mathbf{X}_{rm}^* \quad (r = 1, 2, \cdots, n) \quad (5)$$

式中: i 为主成分提取个数。

1.1.2 泰森多边形

泰森多边形最初是由荷兰气象学家 Thiessen 提出的用离散气象站点数据计算平均降雨量的一种插值分析方法,由其基点周围的邻点的垂直平分线组成,具有多边形内任意一点距多边形控制点的距离小于到其他多边形控制点距离的特点 (闫庆武等, 2011a)。

1.2 技术路线

本文首先通过腾讯位置大数据结合城市圈层结构理论得到了成都市全域 1 km 人口空间分布数据,再结合宜出行热力数据和建筑物分布数据将人口分配至建筑物中,最终得到了青羊区建筑物尺度人口分布数据,具体技术路线如图 1 所示。

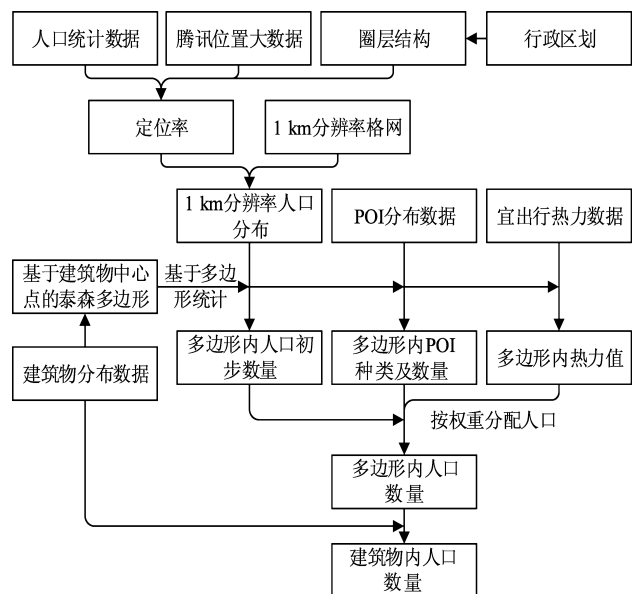


图 1 技术路线

Fig. 1 Technical route

2 研究区概况与数据

2.1 研究区概况

成都市位于四川盆地西部、青藏高原东缘,东北与德阳市相接、东南与资阳市相邻、南与眉山市相连、西南与雅安市接壤、西北与阿坝藏族羌族自治州邻接,地理位置处于(30°05′~31°26′N,102°54′~104°53′E)。成都市包括锦江区、青羊区、成华区、武侯区、金牛区、双流区、龙泉驿区、新都区、郫都区、温江区、青白江区11个市辖区和都江堰市、彭州市、崇州市、邛崃市、简阳市5个县级市以及新津县、金堂县、大邑县、蒲江县4个县。

据成都市发布的2018年统计年鉴数据,截至2018年年底成都市常住人口数为1 633万,其中城镇常住人口数1 194万,常住人口城镇化率达到了73.12%;截至2018年年底成都市户籍人口数为1 478.1万,其中城镇户籍人口数为901.6万,户籍人口城镇化率达到了61%,在西部地区城市中达到了相当高的水平(成都市统计局,2019)。

成都市毗邻龙门山地震带,地震等自然灾害多发,2008年汶川地震造成成都市死亡4 156人,受伤17 802人。成都市作为西部地区人口最多的市级行政单位,对掌握高时空分辨率人口空间分布情况有现实需求,所以本文将成都市作为研究区。

2.2 数据来源及预处理

为了数据的统一,本文所获取的具有位置信息的坐标系数据均采用WGS 1984坐标系,数据单位均为国际单位制标准单位。本文数据来源如下:

(1) 2018年成都市人口统计数据来源于成都市及各个区(县)的2018年统计年鉴。2018年成都市行政区划图来源于GADM网站^①。

(2) 2019年3月11—15日腾讯位置大数据及宜出行热力数据来自腾讯位置大数据官网^②及微信宜出行官网^③,腾讯位置大数据包括经度、纬度、定位人数及定位时间;宜出行数据包括经度、纬度、热力值和采集时间,采集数据格式为Excel表

格。将表格数据加载到ArcGIS 10.2中转换得到矢量点数据。3月11—15日均为工作日,在研究成都市人口分布的普遍规律时较具代表性。

(3) 成都市建筑物数据和POI数据在百度地图开放平台^④申请开发秘钥获取,数据格式为Excel表格,在ArcGIS 10.2中转换为矢量点数据并进行数据清洗,筛选出需要的数据。

(4) 成都市遥感影像来自于Google地图,数据为2018—2019年的卫星影像拼接图,分辨率为2 m。

3 大范围人口数据空间化

人口数据空间化的基本思路是将人口统计数据与其他一些影响人口分布的数据,如土地利用类型、建筑物空间分布等结合起来,建立多元回归等模型,将人口统计数据分配在行政区划、格网或其他尺度上(黄益修,2016)。目前已建立的人口空间化模型主要有核密度估计模型(闫庆武等,2011b;梁亚婷,2015)和多因素融合模型(裴惠娟等,2016;曹伟超等,2012)等。

3.1 研究方法

本文选取腾讯位置大数据作为变量,假设在1小时内,在该区域使用腾讯位置大数据服务的人不重复且分布较为均匀,则可通过某一区域的定位数量推算出该区域的定位率,通过定位率可以算出各计算单元的人数,得到成都市的人口分布。

由于城市人口密度随距城市中心的距离呈衰减趋势(Clark, 1951),按照城市圈层结构理论和各行政区人口、经济等原因,可将成都市分成3个圈层(陈艺,2019),第一圈为金牛区、成华区、锦江区、武侯区和青羊区;第二圈为双流区、龙泉驿区、青白江区、新都区、郫都区以及温江区;第三圈为简阳市、金堂县、彭州市、都江堰市、崇州市、大邑县、邛崃市、蒲江县和新津县。

考虑到后期验证需使用人口统计数据,所以选取每晚21—24时进行计算,因为在成都市春季的该时间段,人们大都处在居所(肖东升等,2016),人口分布情况与常住人口统计数据较为吻

① <https://gadm.org/index.html>.

② <https://heat.qq.com/>.

③ <http://c.easygo.qq.com>.

④ <http://lbsyun.baidu.com/>.

合, 另外 3 个季节人员流动情况与春季类似, 则分圈层计算每个圈层定位率的公式为:

$$lp(i) = \frac{t_i}{RP_i} \quad (6)$$

式中: lp 为定位率; i 为圈层数; t 为平均每小时定位次数; RP 为常住人口数。由各圈层的定位率计算各圈层中 1 km 分辨率格网中的人数, 计算公式为:

$$SP_j = \frac{t_j}{lp(i)} = \frac{t_j RP_i}{T_j} \quad (7)$$

式中: SP_j 为格网 j 中的模拟人数; t_j 为格网 j 定位次数。

3.2 实验过程及结果分析

2019 年 3 月 11—15 日成都市腾讯位置大数据平均每小时定位量时序如图 2 所示。从图中可以看出, 在 0—5 时定位量呈下降趋势, 大部分人都陆续进入睡眠状态, 但依旧有 100 万左右的定位次数, 原因可能为在夜晚手机未关机, 软件自动请求定位服务以及尚有一部分人没有进入睡眠时间; 5—13 时定位量呈增加趋势, 人们陆续起床, 准备或进入上班、上学状态, 定位量在 13 时达到峰值, 原因可能为午饭后手机使用量增多; 13—15 时定位量急剧下降, 多数人进入到午休或准备上班; 15—17 时定位量基本处于平稳, 多数人回到上班或上学状态; 18 时定位量增加, 多数人在归家途中, 手机的使用率增加。19—22 时定位量平稳, 多数人在家中; 22—24 时定位量呈减少趋势,

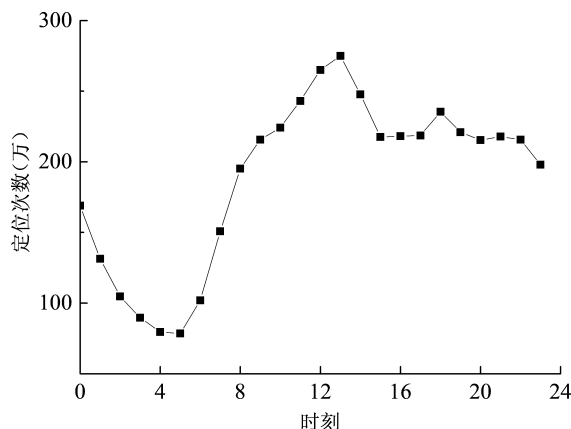


图 2 成都市定位数据时序图

Fig. 2 Time series analysis of positioning data in Chengdu City

人们逐渐进入睡眠状态。由上可知, 定位量的变化基本和当地居民的生活规律相符。

利用式 (6) 对上述定位量式进行计算, 可得到各圈层定位率: 第一圈层的定位率为 22%, 第二圈层定位率为 21%, 第三圈层定位率为 8%。说明在经济越发达、人口越密集的地区, 腾讯位置大数据服务的使用率也越高, 定位率也越高, 符合实际情况。再通过式 (2) 计算出每个格网的人口值, 结果如图 3 所示。

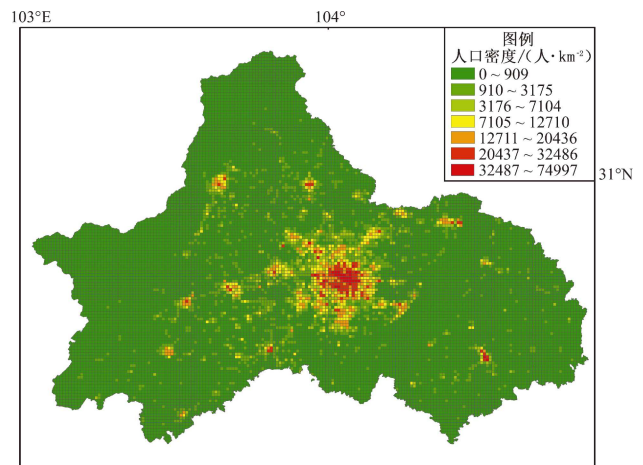


图 3 成都市人口分布

Fig. 3 Population distribution in Chengdu City

从图 3 可以看出成都市人口分布总体较为集中, 其中人口密度最大的区域为成都市几个中心城区, 从中心城区向外, 人口密度呈逐渐下降的趋势, 在各个行政区的中心城区人口呈明显的聚集现象, 在非城区地带, 也存在一些人口密度稍大的区域, 对比成都市卫星影像来看, 主要为乡镇的人口聚居区域, 符合实际情况。

根据实验结果统计成都市各个区 (县) 栅格人口总值与常住人口数, 见表 1。从表中可以得出在区 (县) 级人口数据验证下, 本文的人口分布计算结果与实际结果的平均绝对误差为 90 392 人, 平均相对误差为 12.69%, 准确度较高。总体来说, 计算结果中人口密度较大的区 (县) 误差小于人口密度较小的区 (县), 其中绝对误差最大的为郫都区, 因为该区中包含很多工厂, 本文计算郫都区人口时包含了工厂的工人, 而工人的平均年龄较低, 智能手机使用率高, 导致人口估算值偏高。人口估算值与各区 (县) 常住人口统计数据回归分析如图 4 所示。

| 表1 各个区（县）级人口计算值与统计值 | | | | |
|--|-----------|-----------|---------|---------|
| Tab. 1 Calculated and statistical population in each district and county of Chengdu City | | | | |
| 区（县） | 常住人口数 | 模拟人口数 | 绝对误差 | 相对误差 |
| 成华区 | 953 600 | 963 903 | 10 303 | 0.010 8 |
| 崇州市 | 665 100 | 801 912 | 136 812 | 0.205 7 |
| 大邑县 | 511 200 | 521 181 | 9 981 | 0.019 5 |
| 都江堰市 | 696 900 | 742 014 | 45 114 | 0.064 7 |
| 简阳市 | 1 065 000 | 954 197 | 110 803 | 0.104 0 |
| 金牛区 | 1 216 500 | 1 048 101 | 168 399 | 0.138 4 |
| 金堂县 | 714 000 | 782 611 | 68 611 | 0.096 1 |
| 锦江区 | 708 300 | 864 262 | 155 962 | 0.220 2 |
| 龙泉驿区 | 914 200 | 756 349 | 157 851 | 0.172 7 |
| 彭州市 | 778 000 | 706 201 | 71 799 | 0.092 3 |
| 郫都区 | 857 100 | 1 075 849 | 218 749 | 0.255 2 |
| 蒲江县 | 261 200 | 213 462 | 47 738 | 0.182 8 |
| 青白江区 | 428 721 | 271 447 | 157 274 | 0.366 8 |
| 青羊区 | 847 200 | 857 060 | 9 860 | 0.011 6 |
| 邛崃市 | 618 900 | 535 470 | 83 430 | 0.134 8 |
| 双流区 | 1 466 300 | 1 616 418 | 150 118 | 0.102 4 |
| 温江区 | 522 600 | 567 804 | 45 204 | 0.086 5 |
| 武侯区 | 1 869 000 | 1 861 270 | 7 730 | 0.004 1 |
| 新都区 | 906 000 | 807 078 | 98 922 | 0.109 2 |
| 新津县 | 330 700 | 383 876 | 53 176 | 0.160 8 |

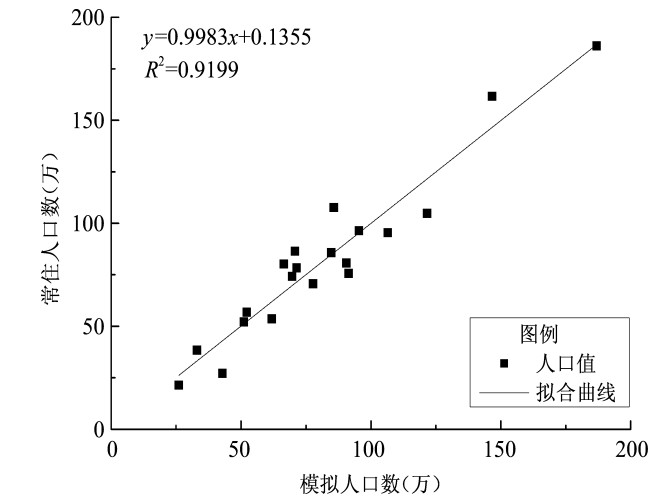


图4 区（县）尺度模拟人口与常住人口回归分析
Fig. 4 Regressive analysis of simulated population and permanent population on district/county scale

由图4可看出，通过计算不同圈层的定位率得出的各区（县）人口数与常住人口数偏差较小，其中线性回归拟合直线的斜率为0.9983， R^2 为0.9199，说明该模型的计算结果总体精度较高。

据上述原理可得各时间段成都市人口分布情况，如图5所示。

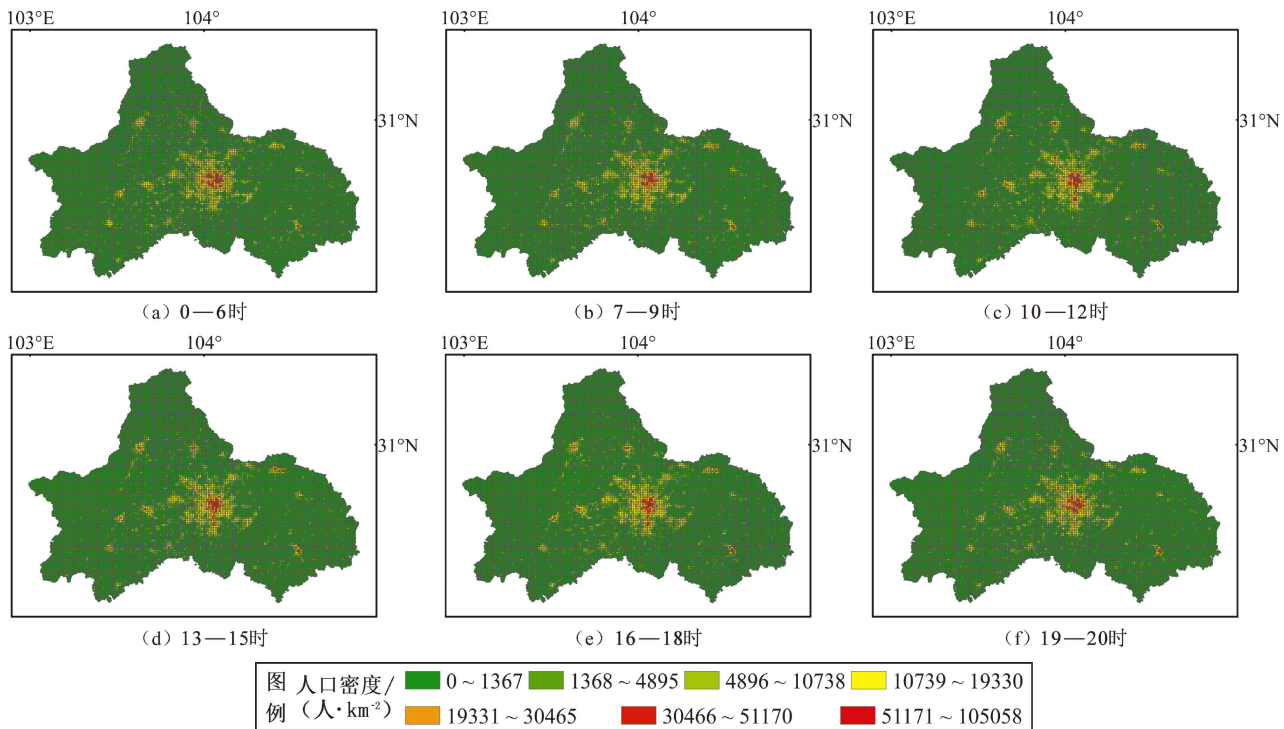


图5 成都市不同时段人口分布
Fig. 5 Population distribution in different periods in Chengdu City

4 小范围建筑物尺度人口空间化

在灾害发生后, 1 km 空间分辨率的人口数据可帮助救援力量快速到达待救援区域, 但是在进行更细致搜救时, 则需要参考更高空间分辨率的人口分布数据。发生在城市的灾害中, 建筑物的结构损坏、倒塌, 是造成人员伤亡的主要因素(李媛媛等, 2015), 所以建筑物尺度的人口空间分布对应急救援具有一定的现实意义。在已有的研究中(胡杨等, 2020; 董春等, 2019; 赵真等, 2019), 建筑物尺度的人口分布多是基于建筑物功能分类, 将人口分配至居住建筑上。但是在实际情况中, 建筑物功能往往不是单一的, 人口并非只分布在居住建筑, 所以在进行人口分配时, 首先要判断建筑物内是否有人口分布。

青羊区位于成都市中心地带, 面积 67.78 km², 常住人口约 85 万, 是成都人口密度最大的区(县)之一, 城镇化程度高, 辖区内建筑类型丰富(成都市统计局, 2019), 非常适合进行建筑物尺度的人口空间化实验。

4.1 研究方法

热力数据表示周围人口密度的相对大小, 假设人口只分布在存在热力值的范围, 则可以用热力值区分建筑物是否需分配人口。由于建筑物损坏会影响该建筑物周围人的安全, 所以将建筑物内的人口和其周围的人口均分配至该建筑物。本文采用以建筑物中心点为基准点建立泰森多边形, 使用泰森多边形内的人口计入多边形内的建筑的方法将人口全部归算至建筑物中。由于在实际计算中可能会出现少数泰森多边形内并没有包含热力数据点的情况, 会造成少数存在人口的建筑没有被分配到人口的问题。为了减少这类问题的影响, 本文引入 POI (Point of Interest) 数据进行改善, 相关研究表明, POI 兴趣点数据与人口分布数据存在显著的相关性(邱歌, 2019)。由于 POI 数据种类较多, 为了减少变量之间的互相影响, 需对 POI 数据进行主成分分析, 降低数据维度。根据主成分的贡献率可得到综合主成分得分, 即 POI 贡献值:

$$S_{\text{POI}_j} = \frac{P_1}{\sum_{i=1}^n \lambda_i} \text{PCA}_1 + \frac{P_2}{\sum_{i=1}^n \lambda_i} \text{PCA}_2 + \cdots + \frac{P_n}{\sum_{i=1}^n \lambda_i} \text{PCA}_n \quad (8)$$

式(3)可简化为:

$$S_{\text{POI}_j} = \sum_{i=1}^n (p_i \times \text{PCA}_{ij}) / \sum_{i=1}^n p_i \quad (9)$$

式中: S_{POI_j} 为第 j 个泰森多边形的 POI 贡献值; p_i 为第 i 个主成分的贡献率; PCA_{ij} 为第 j 个泰森多边形的第 i 个主成分的计算值; n 为提取的主成分个数。

通过对计算出的兴趣点贡献值与热力值进行联合计算, 则可根据 POI 密度进行分区, 计算出每个泰森多边形内的人口:

$$\text{POP}_l = (a \times S_{\text{POI}_l} / S_{\text{POI}_k} + b \times \text{TV}_l / \text{TV}_k) \times \text{SP}_l \quad (10)$$

式中: a 、 b 为两计算因子的权值, 满足 $a + b = 1$; POP 为泰森多边形内计算人口值; TV 为宜出行热力值; SP 为前文基于腾讯位置大数据模拟的人口数据; l 为泰森多边形序号; k 为 POI 密度分区编号。由于 POI 多分布于街道两侧, 为了减少 POI 对街道两侧人口过高估计, 故在确定 a 、 b 取值时, 令 $a < b$ 。

4.2 实验过程

以宜出行数据点为中心建立 30 m 缓冲区, 取与宜出行缓冲区面有交集的建筑物作为研究时段有人员分布的建筑物, 研究区局部建筑物分类如图 6 所示。

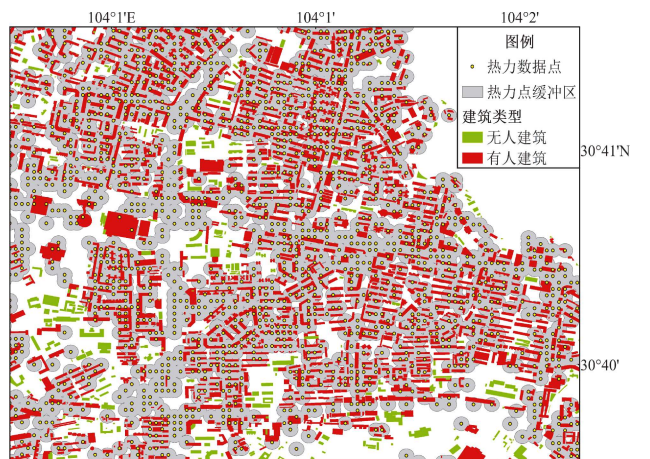


图 6 研究区局部建筑物分类示意图

Fig. 6 Schematic diagram of local building classification in the study area

因为在进行人口统计时多是对居住建筑进行统计，所以为了实验的准确性，避免部分建筑物如公厕、岗亭、变压器室等干扰实验结果，通过将建筑物数据与卫星影像叠加分析，发现城市中少数小于 100 m² 的建筑多为公厕、岗亭、变压器室等非居住建筑及数据源采集误差，所以将研究区内有人建筑物的面积阈值设置为 100 m²，则按有人建筑物中心建立的泰森多边形，如图 7 所示。

如图 7 所示，本文将基于有人建筑物中心点建立的泰森多边形作为人口分配基础单元，统计多边形内的宜出行热力值和 POI 总数，作为人口分配的依据。根据百度地图对 POI 数据的分类，在百度地图开放平台获取了 2019 年成都市青羊区的 11 类与人类活动相关的兴趣点数据，对街道尺度上各类兴趣点与人口进行相关性分析，结果见表 2。

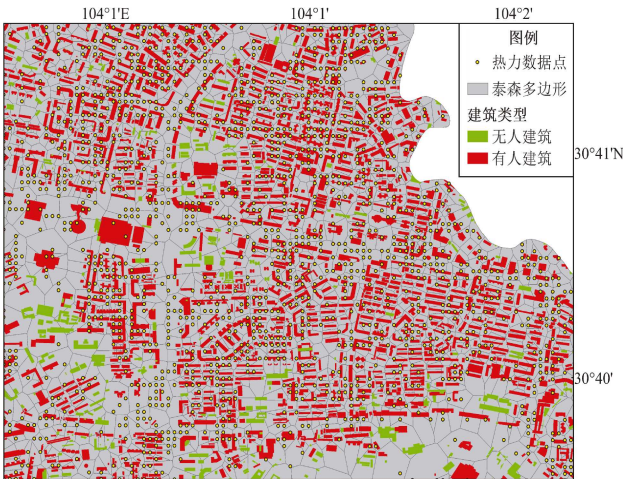


图 7 研究区局部泰森多边形分布示意图
Fig. 7 Schematic diagram of local Tyson polygon distribution in the study area

表 2 各类 POI 与人口相关性
Tab. 2 Correlation between various types of POIs and population

| 类型 | 运动健身 | 医疗 | 休闲娱乐 | 生活服务 | 美食 | 旅游景点 | 酒店 | 金融 | 培训机构 | 交通设施 | 房地产 |
|--------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|
| 与人口相关性 | 0. 67 | 0. 81 | 0. 76 | 0. 78 | 0. 82 | -0. 34 | 0. 01 | 0. 27 | 0. 71 | 0. 49 | 0. 66 |

表 2 中酒店和金融与人口的相关性较低，故将其排除。为了减少变量之间的相关影响以及自变量个数，对 9 类 POI 数据进行主成分分析，降低数据维度。首先对数据进行检验，结果见表 3。表中 KMO (Kaise – Meyer – Olkin, 检验统计量) 数值为 0. 648，Bartlett 检验对应 *P* 值小于 0. 001，数据适合进行主成分分析，各变量的方差解释率见表 4。

表 3 KMO 值及 Bartlett 检验
Tab. 3 KMO value and Bartlett test

| KMO 值 | Bartlett 球形度检验 | | |
|--------|----------------|---------|------------|
| | 近似卡方 | df | <i>P</i> 值 |
| 0. 648 | 148. 042 | 36. 000 | 0. 000 |

表 4 方差解释率表
Tab. 4 Interpretation rate of variance

| 编号 | 初始特征根 | | | 主成分提取 | | |
|----|--------|-----------|---------|--------|-----------|---------|
| | 标准化特征根 | 方差解释率 (%) | 累积 (%) | 特征根 | 方差解释率 (%) | 累积 (%) |
| 1 | 6. 318 | 70. 201 | 70. 201 | 6. 318 | 70. 201 | 70. 201 |
| 2 | 1. 307 | 14. 524 | 84. 724 | 1. 307 | 14. 524 | 84. 724 |
| 3 | 0. 638 | 7. 086 | 91. 81 | — | — | — |
| 4 | 0. 413 | 4. 586 | 96. 396 | — | — | — |
| 5 | 0. 159 | 1. 762 | 98. 158 | — | — | — |
| 6 | 0. 083 | 0. 917 | 99. 075 | — | — | — |
| 7 | 0. 069 | 0. 767 | 99. 842 | — | — | — |
| 8 | 0. 008 | 0. 09 | 99. 933 | — | — | — |
| 9 | 0. 006 | 0. 067 | 100 | — | — | — |

按照特征根大于 1 的原则提取得到两个主成分, 其中主成分 1 的方差解释率为 70.57%, 主成分 2 的方差解释率为 13.14%, 累计方差解释率达到 84.09%, 保留了大部分变量信息。提取的两个主成分得分矩阵见表 5。

表 5 主成分得分矩阵

Tab. 5 Component score matrix

| 名称 | 主成分 1 | 主成分 2 |
|------|--------|--------|
| 运动健身 | 0.136 | -0.236 |
| 医疗 | 0.135 | 0.022 |
| 休闲娱乐 | 0.145 | -0.069 |
| 生活服务 | 0.156 | -0.032 |
| 美食 | 0.141 | -0.143 |
| 旅游景点 | -0.017 | 0.74 |
| 培训机构 | 0.153 | 0.095 |
| 交通设施 | 0.132 | 0.305 |
| 房地产 | 0.123 | 0.179 |

从表 5 中可以得到各主成分的得分表达式, 主成分 1 = $0.136 \times \text{运动健身} + 0.135 \times \text{医疗} + 0.145 \times \text{休闲娱乐} + 0.156 \times \text{生活服务} + 0.141 \times \text{美食} - 0.017 \times \text{旅游景点} + 0.153 \times \text{培训机构} + 0.132 \times \text{交}$

通设施 + $0.123 \times \text{房地产}$; 主成分 2 = $-0.236 \times \text{运动健身} + 0.022 \times \text{医疗} - 0.069 \times \text{休闲娱乐} - 0.032 \times \text{生活服务} - 0.143 \times \text{美食} + 0.740 \times \text{旅游景点} + 0.095 \times \text{培训机构} + 0.305 \times \text{交通设施} + 0.179 \times \text{房地产}$ 。结合公式 (9) 可得到每个泰森多边形的兴趣点贡献值为:

$$S_{\text{poi}_i} = 0.843 \times \text{PCA}_1 + 0.157 \times \text{PCA}_2 \quad (11)$$

在进行 POI 密度分区时, 计算出文家街道、黄田坝街道、苏坡街道、金沙街道、东坡街道、草堂街道、光华街道为低密度区, 平均 POI 密度为 549 个/km²; 汪家拐街道、府南街道、新华西路街道、西御河街道、草市街街道、少城街道、太升路街道为高密度区, 平均 POI 密度为 1 460 个/km²。经实验, 式 (10) 中的权值, 当 $a = 0.25$, $b = 0.75$ 时, 结果较好, 则每个泰森多边形内的人口计算公式为:

$$P_i = (0.25 \times S_{\text{poi}_i} / S_{\text{poi}_k} + 0.75 \times TV_i / T_k) \times SP_i \quad (12)$$

4.3 实验结果及分析

按式 (12) 计算得到每个泰森多边形的人口数, 并将其分配至对应的建筑物中, 其结果 (局部) 如图 8 所示。

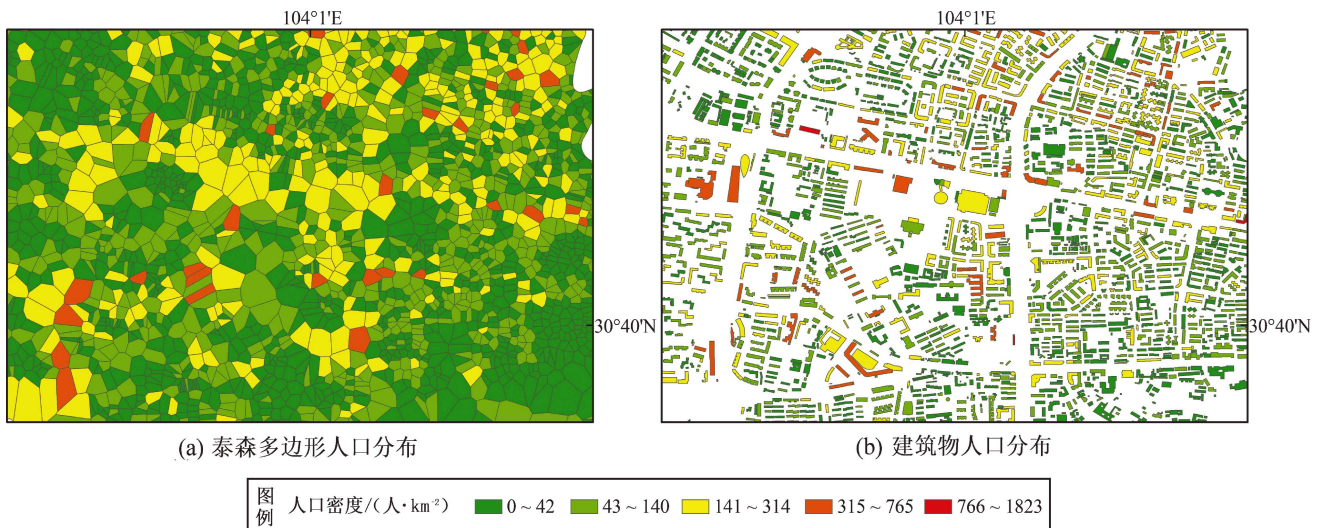


图 8 青羊区建筑物人口分布 (局部)

Fig. 8 Population distribution in buildings in Qingyang District (partial)

图 8 所示区域截自府南街道和光华街道, 可以看到在 21—24 时, 人口主要分布在居住建筑内, 但是也有少部分人口分布在非居住建筑内, 右下角区

域人口稀少, 该区域为金林半岛小区以及杜甫草堂景区。金林半岛小区为多为别墅和 3 层左右的小洋房, 人口密度较低, 实验结果符合实际情况。

由于本文将所有种类的建筑物参与人口分配,且将建筑物周围的人口都归算至建筑物内,单独验证某建筑物内的人口数与实验所得人口数较为困难,且不准确,所以笔者在街道尺度上对实验结果进行验证,青羊区街道模拟人口数与常住人口统计数见表6。

表6 各街道人口实验模拟值与统计值分析表
Tab. 6 Analysis of experimental and statistical values of the population in each sub-district

| 街道 | 常住人口数 (万) | 模拟人口数 (万) | 绝对误差 | 相对误差 |
|--------|--------------|--------------|---------|---------|
| 草市街街道 | 3.200 0 | 3.520 6 | 0.320 6 | 0.100 2 |
| 草堂街道 | 5.145 2 | 5.646 0 | 0.500 8 | 0.097 3 |
| 东坡街道 | 6.788 4 | 6.483 3 | 0.305 1 | 0.044 9 |
| 府南街道 | 12.389 5 | 11.972 4 | 0.417 1 | 0.033 7 |
| 光华街道 | 7.452 7 | 6.295 7 | 1.157 0 | 0.155 2 |
| 黄田坝街道 | 5.300 0 | 4.702 1 | 0.597 9 | 0.112 8 |
| 金沙街道 | 4.695 6 | 5.473 4 | 0.777 8 | 0.165 6 |
| 少城街道 | 8.700 0 | 7.273 6 | 1.426 4 | 0.164 0 |
| 苏坡街道 | 16.790 0 | 16.746 1 | 0.043 9 | 0.002 6 |
| 太升路街道 | 4.486 5 | 5.808 2 | 1.321 7 | 0.294 6 |
| 汪家拐街道 | 3.100 0 | 4.542 2 | 1.442 2 | 0.465 2 |
| 文家街道 | 6.706 6 | 8.323 8 | 1.617 2 | 0.241 1 |
| 西御河街道 | 7.700 0 | 6.775 9 | 0.924 1 | 0.120 0 |
| 新华西路街道 | 5.554 3 | 4.571 5 | 0.982 8 | 0.177 0 |

从表6可以看出实验模拟人口数与常住人口统计数基本一致,平均绝对误差为0.845 3万人,平均相对误差为15.53%,实验结果准确度较高。其中绝对误差最大的街道为文家街道,实验模拟人数较人口统计数少1.617 2万人,绝对误差最小的街道为苏坡街道,误差仅为0.043 9万人。虽然实验结果总体令人满意,但是部分街道出现了较大误差,分析原因可能为:

(1) 本文计算的为21—24时的建筑物内人口平均分布情况,与验证的常住人口统计数据可能会产生一定的差异。

(2) 不同的街道,建筑物的类型各有差异,如部分街道主要为居住建筑,因为本文的研究时间段大部分人都已回到家中,所以该街道的人口误差就相对较小;部分街道分布着大量商业建筑或工厂等,在夜晚商业建筑和工厂等对

人口存在一定的吸引力,导致相应街道的误差相对较大。

通过统计实验得到的各街道建筑物人口数之和与各街道常住人口统计数进行线性回归分析如图9所示。从图中可以看出,模拟结果与人口统计数的回归线斜率为1.024 9, R^2 为0.926 4,实验结果总体可靠性较高,能准确反映大部分建筑物在研究时段真实的人口分布情况。

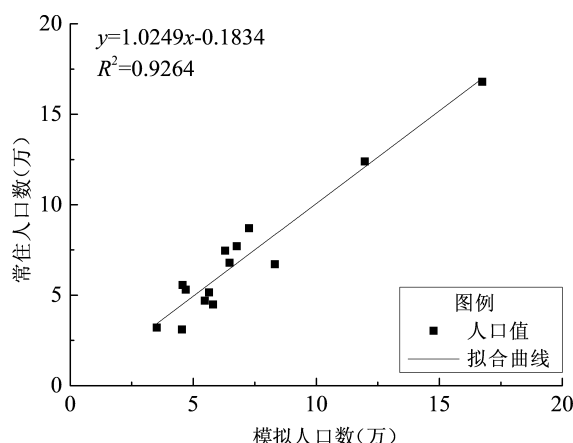


图9 各街道人口统计值与模拟人口线性回归分析
Fig. 9 Linear regression analysis of demographic values and experimental values of the population in each sub-district

5 结论

本文利用多日的腾讯位置大数据,参考城市圈层结构理论,计算出成都市各区(县)手机的定位率,通过定位率得到了不同时段成都市1 km空间分辨率的人口分布数据,经过精度验证,其决定系数 R^2 为0.919 9,总体精度较高。基于已得到的1 km空间分辨率成都市人口分布数据,利用宜出行热力数据和POI数据与人口的相关性,分别计算其对人口的贡献率,实现了青羊区建筑物尺度的人口空间化,并得到了青羊区21—24时的建筑物尺度的人口空间分布数据,在街道尺度的人口统计数据验证下,其决定系数 R^2 为0.926 4,总体结果可信度高。基于本文研究,得出以下结论:

(1) 本文研究数据获取简单,时效性高,1 km空间分辨率下的人口分布估算方法简单准确,

可提高地震应急救援的效率。

(2) 使用泰森多边形作为人口分布的基本单元,可更准确地将建筑物周边的人口分配至建筑物内。

(3) 通过热力数据范围而非建筑物类型区分建筑物是否分配人口,可得到更符合研究时段实际人口分布情况。

虽然本文提出的人口空间化模型有较好的效果,但是也存在以下不足:

(1) 在进行大范围人口分布估算时,仅从定位数量因素考虑了定位率的问题,人口老龄化率、青少年率及城镇化率等可能影响定位率的因素暂未加入计算。

(2) 在建筑物尺度没有到达实时人口分布的情况下的计算。

今后的研究中我们会参考更多影响因素进行分析,以达到高时空分辨率的建筑物尺度人口分布计算,为应急救援、城市建设等做出贡献。

参考文献:

- 曹伟超,陶和平,谭理,等. 2012. 基于多源空间数据的山区人口分布模拟[J]. 国土资源遥感,24(2):61-67.
- 陈艺. 2009. 成都城市经济圈的区域特征及其带动效应[J]. 城市发展研究,16(5):13-17.
- 成都市统计局. 2019. 成都年鉴(2019)[M]. 成都:成都年鉴社.
- 董春,尹诗画,张玉. 2019. 多智能体的城市精细尺度人口估算模型[J]. 测绘科学,44(8):113-119.
- 郭建兴,姬建中,和朝霞,等. 2018. 面向地震应急的山区人口百米格网分布——以麟游县为例[J]. 地震地磁观测与研究,39(6):150-157.
- 胡杨,张小咏,肖迪. 2020. 基于手机大数据的城市建筑物人口时空估算[J]. 系统仿真学报,32(10):1814-1883.
- 黄益修. 2016. 基于夜间灯光遥感影像和社会感知数据的人口空间化研究[D]. 上海:华东师范大学.
- 贾俊平. 2008. 统计学(第三版)[M]. 北京:中国人民大学出版社.
- 贾召亮,郑川,李袁欣,等. 2020. 面向地震应急的人口空间化方法研

究——以云南省大理白族自治州为例[J]. 地震研究,43(4):751-758.

- 李国平,陈秀欣. 2009. 京津冀都市圈人口增长特征及其解释[J]. 地理研究,28(1):191-202.
- 李媛媛,陈建国,张小乐,等. 2015. 基于建筑结构破坏的地震伤亡评估方法及应用[J]. 清华大学学报(自然科学版),55(7):803-807,814.
- 李泽宇,董春. 2019. 引入兴趣点的地理加权人口空间分布模型研究——以天津市为例[J]. 遥感信息,34(2):113-117.
- 梁亚婷. 2015. 基于遥感和GIS的城市人口时空分布研究:以上海市静安区为例[D]. 上海:上海师范大学.
- 潘碧麟,王江浩,葛咏,等. 2019. 基于微博签到数据的成渝城市群空间结构及其城际人口流动研究[J]. 地球信息科学学报,21(1):68-76.
- 裴惠娟,周中红,孙艳萍,等. 2016. 甘肃省人口空间分布特征分析[J]. 西北师范大学学报(自然科学版),52(5):95-100.
- 邱歌. 2019. 基于随机森林模型的高精度人口数据空间化——以郑州市区为例[D]. 内蒙古:内蒙古师范大学.
- 王贤文,王虹茵,李清纯. 2017. 基于地理位置大数据的京津冀城市群短期人口流动研究[J]. 大连理工大学学报(社会科学版),38(2):105-113.
- 吴中元,许捍卫,胡钟敏. 2019. 基于腾讯位置大数据的精细尺度人口空间化:以南京市江宁区秣陵街道为例[J]. 地理与地理信息科学,35(6):61-65.
- 肖东升,常鸣,苏勇,等. 2016. 基于时空菱形的人员在室率评估方法[J]. 世界地震工程,32(1):43-49.
- 闫庆武,卞正富,王红. 2011a. 利用泰森多边形和格网平滑的人口密度空间化研究——以徐州市为例[J]. 武汉大学学报(信息科学版),36(8):987-990,1010-1011.
- 闫庆武,卞正富,张萍,等. 2011b. 基于居民点密度的人口密度空间化[J]. 地理与地理信息科学,27(5):95-98.
- 赵真,郭红梅,张莹,等. 2019. 基于天地图高精度建筑物的四川宝兴县人口估算[J]. 地震研究,42(2):204-209.
- Clark C. 1951. Urban population densities[J]. Journal of Royal Statistical Society,114(4):490-496.
- Sakaki T, Okazaki M, Matsuo Y. 2010. Earthquake shakes twitter users: real-time event detection by social sensors[C]//Proceedings of the 19th International Conference on World Wide Web. New York: ACM:851-860.

Research on Population Distribution with High Spatial Resolution Based on Multi – source Information

LIU Zhicheng¹, XIAO Dongsheng^{1,2}

(1. School of Civil Engineering and Surveying and Mapping, Southwest Petroleum University, Chengdu 610500, Sichuan, China)

(2. Disaster Prevention and Emergency Research Center for Surveying, Mapping and Remote Sensing, Southwest Petroleum University, Chengdu 610500, Sichuan, China)

Abstract

The current spatialized data of population are normally static and of low resolution in time and space, and not so practical for emergency and rescue. To solve this problem, according to the theory of urban circle structure and principal component analysis, we proposed a method of estimating population on building-scale using high resolution data in time and space. Taking Chengdu City as an example, we calculated the location rate of different urban circles using the Tencent Location Big Data, and obtained the population distribution in the grid of $1\text{ km} \times 1\text{ km}$ of Chengdu in different periods of a day. Further, we used the Tyson Polygon based on the center of the building as the basic unit of population distribution, and calculated and weighted the contribution of the data from Tencent Easygo and the POI data from Baidu Map to the population distribution. In this way we obtained the distribution of population on building-scale in Qingyang District of Chengdu. R^2 , the coefficient of determination for the regression analysis of the street-scale statistics, is 0.926 4. Our results are more accurate and conform to the actual distribution of population in this area.

Keywords: emergency and rescue; principal component analysis; population distribution; POI; buildings