

doi: 10.7621/cjarrp.1005-9121.20191110

· 技术方法 ·

POI 数据辅助下的泸州市城市地块用地类型 现状快速判断与验证^{*}

辜寄蓉^{1**}, 吴修月¹, 黄志勤², 冯义从², 方从刚³

(1. 四川师范大学地理与资源科学学院, 成都 610101; 2. 四川省国土资源厅信息中心, 成都 610072;
3. 成都市国土资源信息中心, 四川成都 610032)

摘要 [目的] 土地利用规划和城市规划的用地地块, 在实际使用过程中, 都会不同程度与规划用途产生差异, 如何把握规划的实施程度是土地管理中的一个重要问题。[方法] 在大数据时代的背景下, 通过对网络大数据可以真实地反映城市地块的真实使用状况。研究采用 OSM (Open Street Map), 对研究区进行城市地块的划分, 形成街区尺度的城市地块。采用 POI (Point of Interest) 数据, 建立了 POI 分类与城市用地分类的映射关系, 计算了不同类型的 POI 数据的频数密度和类型比例, 形成基于 POI 的城市地块用地类型识别方法, 对研究区的城市地块进行土地利用现状识别。识别结果与第二次土地调查分类结果进行对比, 验证方法的可行性。[结果] 从泸州市的实证结果看, 基于 POI 数据的城市地块识别结果与基于地籍数据的城市地块识别结果存在很大的差别。地籍数据主要是表征城市地块中不同实体的土地利用状况, 两种数据采集的标准和表达的意义都不同, 因此造成结果存在较大的差异性。[结论] 研究认为, POI 点有准确的空间位置信息和丰富的属性信息, 主要是表征城市地块中不同实体的真实使用状况, 可以较准确地判定快速发展的城市进程中的城市用地类型, 能较为真实反映城市地块的现实使用状况, 但也要注意 POI 数据在实际应用中存在的缺陷和不足。

关键词 城市地块 用地类型现状 POI OSM 地籍数据 泸州市

中图分类号:F304.5 文献标识码:A 文章编号:1005-9121[2019]11072-08

0 引言

当今是一个大数据时代, 丰富的数据获取渠道和大规模的数据量为城市研究和城市管理带来了新的机遇^[1]。城市地块是城市土地功能和人类活动的载体。城市地块用地类型的判别, 在传统数据环境下主要通过土地供地类型或城市规划类型来确定, 这两种方法都是规划时点对地块未来的利用状况进行的设计与安排。现在城市发展日新月异, 城市地块用地类型的变化也是非常迅速的, 如何快速捕捉城市地块用地类型的变化, 只有辅以对变化敏感的网络大数据, 通过对大数据的挖掘, 建立基于大数据的用地分类方法, 快速精准地对地类进行识别, 实现对城市土地的精细化管理。

网络大数据是指“人、机、物”三元世界在网络空间 (Cyberspace) 中交互、融合所产生并在互联网上可获得的大数据^[2-4]。目前可以通过网络平台中的专业软件、GPS、网络日志等工具来采集诸如兴趣点 (Point of Interest, POI) 数据、智能手机数据、公交智能卡数据、签到数据、出租车轨迹数据等开放式的网络大数据, 其丰富的数据获取渠道和大规模的数据量为智慧城市的研究工作提供了保障。

在此背景下, 国内外一些学者突破传统的用地类型识别方法, 使用开放路网数据、出租车活动的 GPS

收稿日期: 2017-07-27

作者简介: 辜寄蓉 (1968—), 女, 四川成都人, 博士、教授。研究方向: 基于空间分析技术的土地利用与区域发展演化格局研究。
Email: annygu126@126.com

* 资助项目: 国家社会科学基金“乡村振兴战略视阈下川滇藏交界地带特色藏族村寨建设研究”(18BMZ73); 四川省原国土资源厅科技项目“不动产大数据可视化研究”, 基金号: 2018KJ002

数据、地理位置标签等居民数字足迹实现对土地利用类型的动态感知^[5]。Toole 等^[6] (2012) 利用波士顿地区大约 60 万匿名用户的移动电话的呼叫详细记录 (CDR) 数据, 分析了居住用地、商业用地、工业用地、公园等其他用的时空特征, 以此来推断城市用地类型。Hollenstein 等^[7] (2013) 通过获取 800 万个 Flikr 网站上公开的有位置、时间等信息的图像数据来确定伦敦、芝加哥悉等 6 个城市的中心城区。龙瀛等 (2014) 提出了利用 OSM 数据和 POI 自动识别来描述地块的方法, 其中利用路网数据划定地块边界, POI 数据来识别生成地块的开发密度、功能、混合使用程度以及人类活动。池娇等^[8] (2016) 利用 POI 数据的类型划分出武汉市都市发展区的城市功能区进行划分及可视化。梁军辉等^[5] (2016) 在分析海量出租车 GPS 数据与城市用地类型之间的相关关系的基础上, 对深圳市福田区城市土地利用类型进行动态感知。大数据的应用较好地避免了因为“数据孤岛”所导致的研究数据缺乏的问题, 使得研究数据不在限于部门或行业的数据; 同时由于大数据来源丰富、更新快速的特点, 也使得研究的角度更丰富, 可以更好地支撑如规划实施验证等问题。当前基于大数据对城市用地分类的理念和方法进行革新, 已成为研究的重要方向^[24]。

1 数据来源及数据处理

泸州是地处长江上游的地级市, 位于四川省东南部, 长江和沱江的交汇处。由于研究区域最好具有丰富的城市道路网和较完整的城市用地规划以及聚集的城市活动这样的特征, 我们选取了泸州市城市规划中所定义的中心城区范围中的核心区域作为研究区。研究区包括泸州市的江阳区和龙马潭区的部分区域。

文章利用开放道路数据 (OSM) 和兴趣点 (POI) 数据对泸州市中心城区的核心区域进行城市用地地块类型的识别与分析。该研究中涉及的大数据包括 ESRI 公司的开放路网数据 Open Street Map (OSM), 腾讯公司的开放兴趣点数据 Point of Interest (POI)。为判断 POI 数据对城市地块用地类型分类结果的准确性, 我们利用地籍数据进行了用地分类的划分, 将两者的划分结果进行对比分析。

1.1 POI 数据处理

互联网地图的兴趣点是面向大众的。如谷歌、百度等网站允许市民提供他们自己感兴趣地点的描述和这些地点的位置信息, 其 POI 信息基本包含人们日常需要了解的信息。由于 POI 数据分类并不支持进行城市用地类型的判别, 因此, 一方面要在对 POI 数据进行清理, 形成与城市用地分类 (GB50137-2011) 的对应关系; 另一方面, 一个城市地块上会有多个不同类型的 POI 点, 研究建立了一种测算方法, 将多类型 POI 点汇总为一种城市土地利用类型。

该研究中收集的 POI 来自腾讯地图 (截止 2016 年 6 月), 并根据腾讯地图的 POI 分类关键词进行了重新分类 (图 1)。共分为 15 大类, 150 个二级类, 二级类中可能包含三级类。每条 POI 记录中包含经度、纬度、名称、地址、电话、类别 6 个属性。

为进行城市用地类型识别, 需要将 POI 分类与城市用地分类标准相匹配, 得到 POI 分类与城市用地分类的映射关系。同时对原始 POI 数据中, 存在类别误分数据进行清理。其中, POI 数据中与仓储物流有关的点大多是一些物流公司、快递服务点, 属于“公司企业”分类, 与城市用地分类标准中所规定的“仓储用地”的定义不符。因此, 取消此城市用地类型。具体分类映射关系图如图 2 所示。

1.2 路网数据处理

路网地图数据是构建城市地块的不可缺少的支撑。该研究中, 城市现状路网数据来源于 ESRI 公司的 OpenStreetMap (<http://www.openstreetmap.org/> 获取时间 2016 年 2 月), 数据精度 1:1 万, 其提供的道路网数据中包括名称、类型、长度等属性。由于泸州是二级城市, 其 OSM 数据只包括了部分主干道、次干道和铁路等, 而城市内部街道的路网数据不完整。研究中, 使用天地图上下载的高清影像图 (1.194 3m, 17 级) 对研究区的城市内部道路进补充和修正, 并将 OSM 数据集中的双线道路处理为单线道路。研究区道路网成果如图 2 所示。

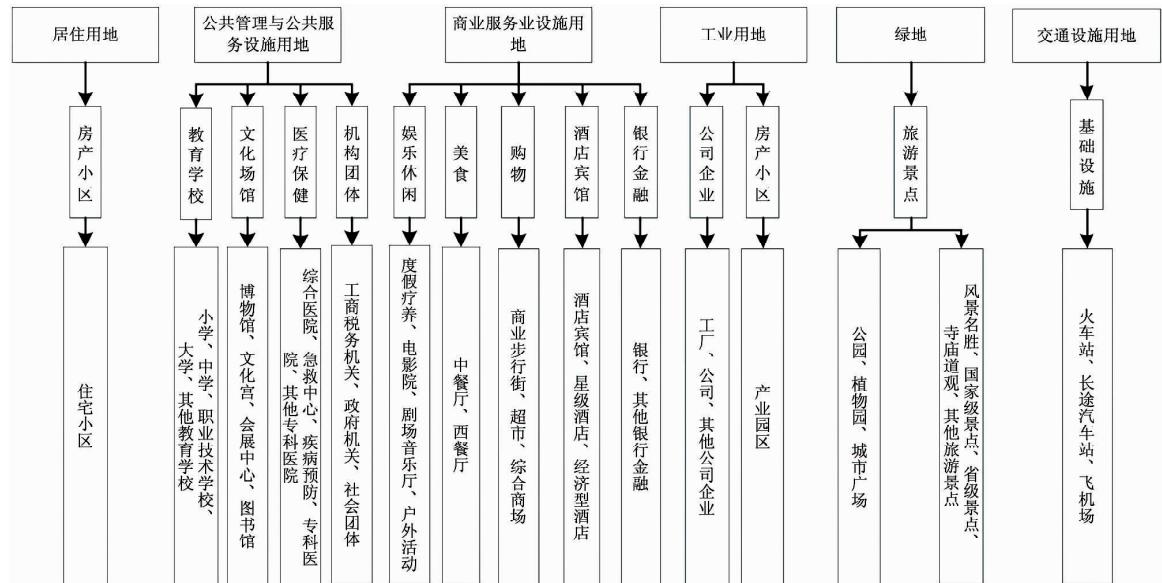


图 1 POI 类型与城市用地分类映射关系图

1.3 地籍数据处理

地籍数据是通过对每一宗地（由界址线包围的土地）的土地权属、位置、界址、数量、质量以及利用状况进行调查和测量后绘制而成的地图数据^[8]。由于城市用地地块大于城镇地籍图斑，研究以城市用地分类标准（GB50137-2011）为准则，建立土地利用分类与之的映射关系（图 3）。研究区内，各类用地的面积汇总见表 1。

该文将利用泸州市第二次土地调查形成地籍数据与 POI 数据得到的结果进行对比，检验 POI 数据得到的城市用地类型精度。

2 用地类型判定

该文以街区为研究对象，利用网络大数据中的 POI 数据和开放路网数据，对城市地块进行划分，判定城市地块用地类型，并进行验证。具体的技术流程如图 4 所示。

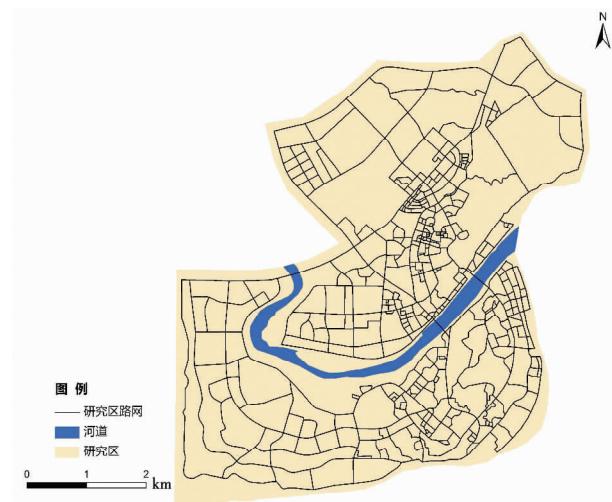


图 2 研究区道路分布

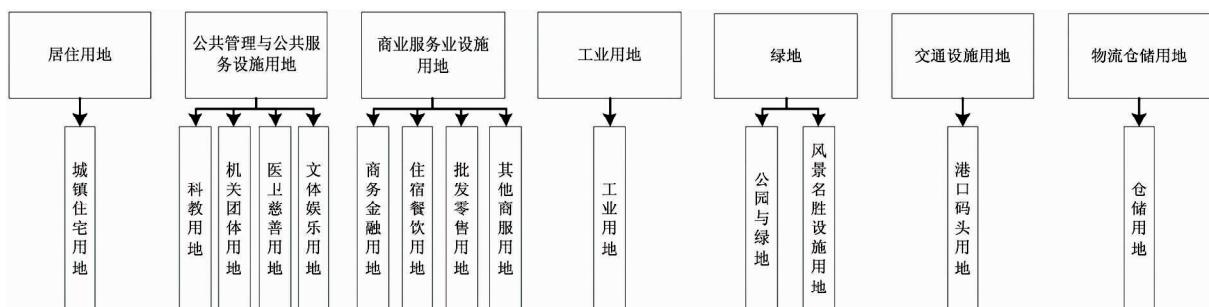


图 3 地籍数据与城市用地分类映射关系

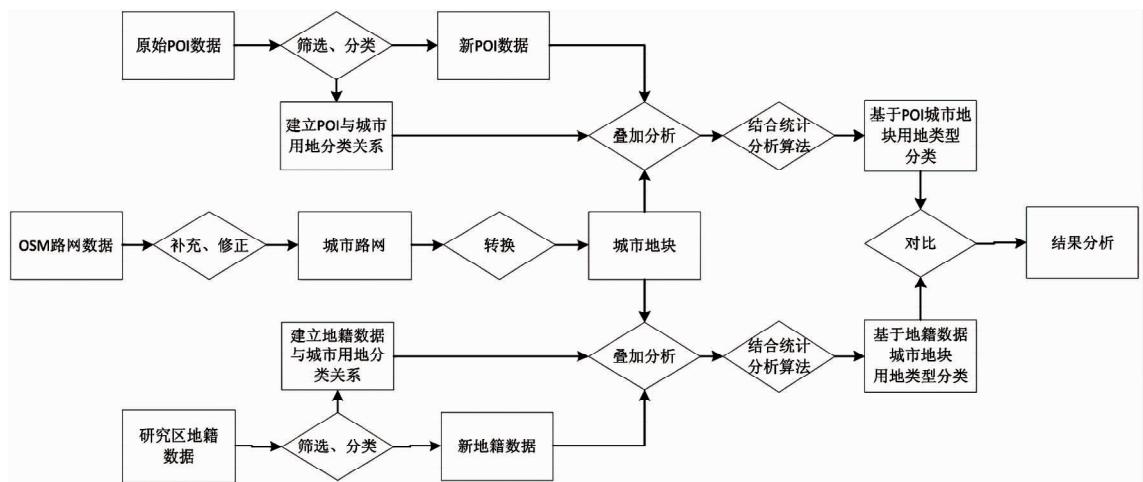


图 4 技术流程

2.1 提取城市地块

地块是指可辨认出同类属性的最小土地单元，即内部特性和区位条件相对均匀的土地区域，是评定和划定土地级别的基本空间单位^[9]。在不同的尺度上，地块表现为不同的形式，其面积可大可小^[11]。该文中的地块是指在街区尺度上，经由城市道路所划分出来的城市地块。城市地块划分步骤为：①在 OSM 数据的基础上绘制并创建研究区路网线图层；②对创建的线图层进行数据检查和道路修剪，去除没有与其他道路相连接的街道，确保道路数据图层具有完整的拓扑连接关系，具体结果如图 5 所示；③根据城市道路设计规范设定道路缓冲宽度：主干路（35m）、次干路（25m）、支路（15m），从而形成道路的面图层；④利用道路的面图层对研究区进行擦除处理，其形成的剩余空间则为城市地块，共计 372 个。具体的城市地块划分结果如图 6 所示。

2.2 基于 POI 数据的用地类型判别

考虑到某些 POI 点分散在城市街道上，未处于相邻的地块内，因此对 2.1 中获取的城市地块，设立 10m 的缓冲区，以便能囊括更多的 POI 点。

分别统计每个地块及缓冲范围内，不同类型的 POI 数量。对每个地块单元，构建指标频数密度 (frequency density, FD) 和类型比例 (category ration, CR) 来识别地块用地类型^[12-14]，计算公式为：

$$F_i = \frac{n_i}{N_i} \quad (i = 1, 2, \dots, 6) \quad (1)$$

$$C_i = \frac{F_i}{\sum_i F_i} \times 100\% \quad (i = 1, 2, \dots, 6) \quad (2)$$

表 1 地籍数据面积统计

城市用地类型	土地利用类型	面积 (万 m ²)
居住用地	城镇住宅用地	225.124 856
绿地	风景名胜设施用地	1.057 924
	公园与绿地	23.077 97
工业用地	工业用地	49.385 335
商业服务业设施用地	批发零售用地	14.891 968
	商务金融用地	11.394 046
	住宿餐饮用地	4.271 772
	其他商服用地	1.749 226
公共管理与公共服务设施用地	机关团体用地	26.162 760
	科教用地	59.240 621
	文体娱乐用地	4.73 5000
	医疗卫生用地	12.005 004
物流仓储用地	仓储用地	11.378 811
交通设施用地	港口码头用地	1.060 968



图 5 城市地块

式(1)、(2)中, i 表示 POI 类型; n_i 表示地块单元内第 i 种类型 POI 数量; N_i 表示第 i 种类型 POI 总数; F_i 表示第 i 种类型 POI 占该类型 POI 总数的频数密度; C_i 表示第 i 中类型 POI 的频数密度占单元内所有类型 POI 频数密度的比例。

研究中将类型比例值是否超过 50% 作为判断单元用地类型的标准。当某一地块单元中某种类型的 POI 比例占到 50% 及以上时, 根据图 1 所确定的 POI 与城市用地类型的映射关系, 判定该地块单元的城市用地类型; 而当单元内所有类型的 POI 比例均没有到达 50% 时, 判定该地块单元为混合地类; 当单元内不包含 POI, 即类型比例为空值时, 判定该地块单元为无数据区。分类结果如图 6 所示, 统计结果如表 2 所示。

2.3 基于地籍数据的用地类型判别

将该区域的地籍数据与 2.1 提取的城市地块进行叠加, 计算研究区不同地块上不同类型地籍数据的面积占比 (A) 和类型比例 (R), 计算公式为:

$$A_i = \frac{s_i}{S_i} \quad (i=1, 2, \dots, 7) \quad (3)$$

$$R_i = \frac{A_i}{\sum_i A_i} \quad (i=1, 2, \dots, 7) \quad (4)$$

式(3)、(4)中, i 表示地籍数据的类型; s_i 表示某地块单元内地块中第 i 种类型地籍数据的面积, S_i 表示 i 种类型地籍数据的总面积; A_i 表示第 i 种类型地籍数据占该种类型地籍数据的比例; R_i 表示第 i 中类型地籍数据的频数密度占单元内所有类型地籍数据频数密度的比例。

将研究中类型比例值是否超过 50% 作为判断标准。当某一地块单元中某种类型的地籍数据比例占到 50% 及以上时, 根据图 2 所确定的土地利用分类与城市用地类型的映射关系, 判定该地块单元的城市土地用地类型。当单元内所有类型的地籍数据比例均没有到达 50% 时, 判定该地块单元为混合地类。当单元

表 2 POI 数据地块类型

地块类型	地块个数
公共管理与公共服务设施用地	34
住宅用地	69
商业服务设施用地	83
工业用地	44
绿地	20
交通设施用地	3
混合用地	74
无数据区	45

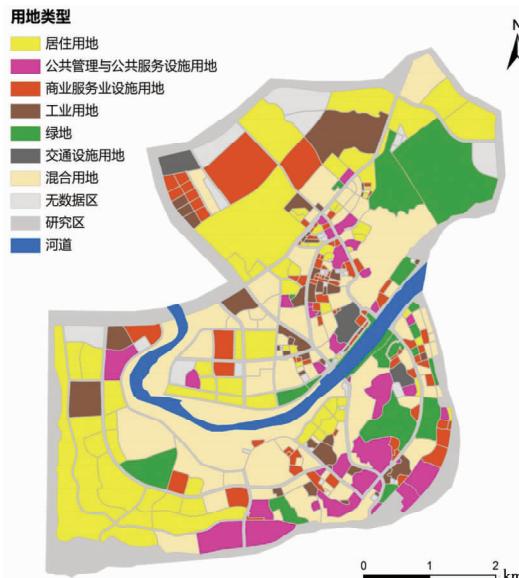


图 6 POI 地块类型识别图

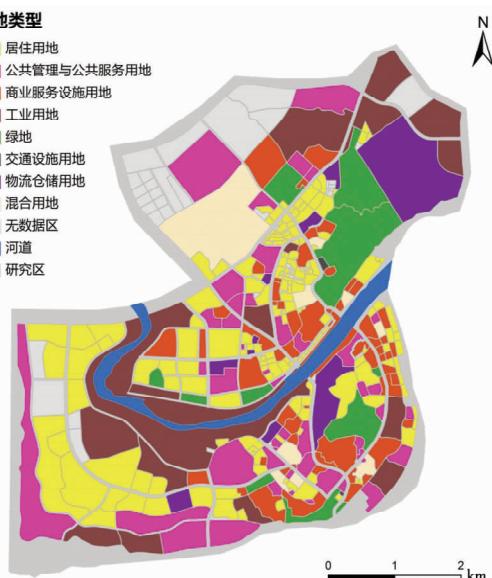


图 7 地籍数据地块类型识别图

内不包含地籍数据, 即类型比例为空值时, 判定该地块单元为无数据区。分类结果如图 7 所示, 统计结果如表 3 所示。

3 判定结果验证

将 2.2 和 2.3 中两类判别结果进行对比, 统计分析其分类结果的一致性和差异性。

3.1 一致性验证

分类结果一致性的统计是以 POI 分类结果为基准, 统计地籍数据分类结果中与 POI 分类结果相一致和不一致的地块个数和类型, 建立地块用地类型一致性判断矩阵统计, 具体统计结果如表 4 所示。

表 3 地籍数据地块类型

地块类型	地块个数
公共管理与公共服务设施用地	46
住宅用地	175
商业服务设施用地	60
工业用地	17
绿地	11
交通设施用地	2
物流仓储用地	10
混合用地	9
无数据区	42

表 4 地块用地类型一致性判断矩阵

POI 判别结果 \ 地籍判别结果	无数据区	居住用地	公共管理与公共服务设施用地	商业服务设施用地	工业用地	绿地	交通设施用地	物流仓储用地	混合用地
无数据区	10	30	3	1	1	0	0	0	0
居住用地	8	38	5	3	6	1	0	2	1
公共管理与公共服务设施用地	1	8	13	7	1	2	0	1	1
商业服务业设施用地	13	48	5	11	0	1	0	1	3
工业用地	6	18	6	8	1	0	1	2	2
绿地	1	5	2	5	2	3	1	1	1
交通设施用地	1	1	0	1	0	0	0	0	0
混合用地	2	27	12	19	6	4	0	3	1

表 4 中, 每一相同字段名的纵行和横向形成一个数据对照组。如地籍判别结果中“无数据区”(纵行)与 POI 判别结果中“无数据区”(横行)进行判读时, 分别表示: 1、地籍判别结果中“无数据区”(纵行)中, 被判断为无数据区的地块共有 43 个, 其中与 POI 判别结果对比, 有 10 个同样判别为“无数据区”, 有 8 个在 POI 判别结果中为“居住区”, 依次类推。2、POI 判别结果中“无数据区”(横行)中, 被判断为无数据区的地块共有 45 个, 其中与地籍差别结果对比, 有 10 个同样判别为“无数据区”, 有 30 个在地籍判别结果中为“居住区”, 依次类推。

3.2 典型区域验证

由表 4 可以看出: ①POI 识别的无数据区地块中有 67% 的地块被地籍数据识别为居住用地的地块; ②POI 识别的居住用地共有 69 个, 而地籍数据识别的居住用地地块共有 175 个; ③POI 识别的公共管理与公共服务设施用地与地籍数据识别的同类地块有一定的差距。POI 将一部分地籍数据判定的公共管理用地识别为混合用地, 而 POI 判定的公共管理与公共服务设施用地被地籍数据多数判定为居住用地和商业用地等; ④POI 识别的商业服务设施用地地块中, 绝大多被地籍数据识别的为居住用地地块; ⑤POI 识别的类型为绿地的地块比地籍数据识别的地块多出接近一半, 且相同的地块也较少, 不一致的地块主要表现在地籍数据判定为居住用地和商业用地的地块中; ⑥POI 识别的交通设施用地地块与地籍数据识别的交通设施用地地块完全不重合; ⑦POI 识别的混合用地地块有 74 个, 而地籍数据识别的混合用地地块只有 9 个; ⑧POI 数据中关于物流仓储的数据多为快递公司或物流公司等点位信息, 且网络中 POI

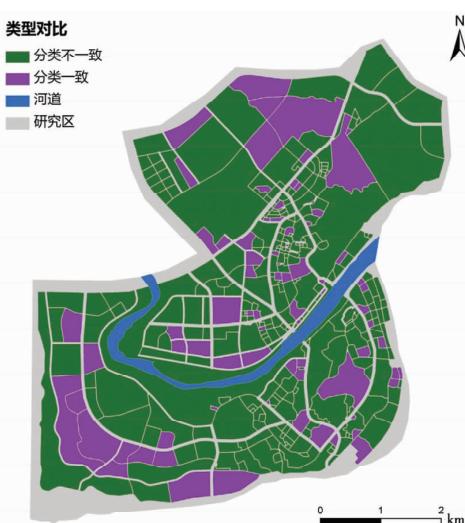


图 8 空间分布差异图

的分类体系中没有明确的物流仓储这一分类，因此城市用地类型中物流仓储用地缺失。通过地籍数据可识别出该类用地的地块。一致性对比统计结果显示，地块类型一致的地块仅有 72 处。将这 72 处用地进行空间对比，可以发现，分类结果一致的地块在研究区中分布的比较零散，研究区中的绝大部分地块的分类结果都不一致，分布对比结果如图 8。对研究区中的以下两个区域（红色区域和黄色区域）进行对比分析。

由图 9 和图 10 对比可以看出，分类不一致的地块主要集中在研究区的中部、东部以及河道以南区域。中部区域（红色圈）是泸州市主城区的核心区域，这里人口分布密度高，经济活跃，经历历史的积累聚集了大量的商业服务设施，这一区域的商业繁华，是大型购物中心、商业街区的聚集地。从 POI 的分类结果中可以看出，该地区主要是由商业服务设施用地和工业用地组成，但是地籍数据的分类结果中该地区的绝大部分地块均为居住用地。

从图 11 和图 12 对比可以看出，黄色区域在 POI 识别结果中显示多为混合用地类型的地块以及少量的居住用地地块；而在地籍数据的识别中，这一区域的地块类型多为单一类型的地块，主要以工业用地、居住用地和公共管理与公共服务设施用地为主。



图 9 红色区域 POI 分类结果



图 10 红色区域地籍数据分类结果



图 11 黄色区域 POI 数据分类结果



图 12 黄色区域地籍数据分类结果

4 总论与建议

土地利用规划和城市规划的用地地块，在实际使用过程中，都会不同程度与规划用途产生差异。因此分析结果中，基于 POI 数据的城市地块识别结果与基于地籍数据的城市地块识别结果存在很大的差别，这个结果是正确的，正说明了实际用地过程中形成的土地利用类型与规划时的预判出现了较大的差异。一方面说明了规划的刚性有余，而弹性不足；另一方面也说明在城市用地监督执法中的执法力度不够，监控方法落后，无法快速捕捉到用地类型超越规划的变化。

POI 点有准确的空间位置信息和丰富的属性信息，主要是表征城市地块中不同实体的真实使用状况，可以较准确地判定快速发展的城市进程中的城市用地类型，能较为真实反映城市地块的现实使用状况，可为规划实施的效果、效益评价提供决策依据。

通过该次研究，发现 POI 数据在实际的应用中存在的缺陷和不足，主要体现在以下几方面：（1）POI 数据作为点状数据源，不同的 POI 点表征的实体对象面积不同，而实体对象的占地面积对该地块的用地类型有着重要的影响。在该次研究中，仅使用 POI 的数量来判定地块的用地类型，未能考虑到 POI 点所代表

的占地面积的多少对用地类型识别结果的影响。(2) POI 数据分布不均,且存在无 POI 的地区。在该次研究中,研究区中的某些边缘区域中,由于公众兴趣度低,没有采集相关的兴趣点,导致分类结果中存在较多的空值区。另外,城市中心区、商业发达区域的 POI 点数据多,且分布密,又造成判定的混合用地类型比较多的问题。(3) 研究区域本身的社会、经济发展状况也对数据收集产生了比较大的影响。泸州是四川省的二级城市,其数据的丰富度、准确度,都赶不上成都市,甚至绵阳市、德阳市,因此前期数据收集工作是非常重要的。

参考文献

- [1] 龙瀛,吴康,王江浩,刘行健. 大模型:城市和区域研究的新范式. 城市规划学刊, 2014, (6): 52–60.
- [2] 王元卓,靳小龙,程学旗. 网络大数据:现状与展望. 计算机学报, 2013, 36 (6): 1125–1138.
- [3] 蔺旭东,周军锋,刘佳. 资源关联性大数据分析在农业生态环境保护中的应用. 中国农业资源与区划, 2016, 37 (2): 62–65.
- [4] 刘炜,陈景新,薛楠,刘二亮. 区域一体化背景下京津冀循环经济协同发展研究. 中国农业资源与区划, 2015, 36 (2): 78–83.
- [5] 梁军辉,林坚等. 大数据条件下城市用地类型辨识研究——基于出租车 GPS 数据的动态感知. 上海国土资源, 2016, 37 (1): 28–32.
- [6] Toole J L, Ulm M, González M C, et al. Inferring land use from mobile phone activity. Proceedings of the Acm Sigkdd International Workshop on Urban Computing, 2012: 1–8.
- [7] Hollenstein L, Purves R. Exploring place through user-generated content: Using Flickr to describe city cores. Journal of Spatial Information Science, 2010, 1 (1): 21–48.
- [8] 池娇,焦利民,董婷,等. 基于 POI 数据的城市功能区定量识别及其可视化. 测绘地理信息, 2016, 41 (2): 68–73.
- [9] 杜海平,詹长根,李兴林. 现代地籍理论与实践. 深圳: 海天出版社, 1999.
- [10] 王利群,海军,张德群,等. 土地调查工作方法、经验及成果应用研究——以呼伦贝尔市第二次土地调查工作为例. 中国国土资源经济, 2016, 29 (5): 66–69.
- [11] 王崇愚. 多尺度模型及相关分析方法. 复杂系统与复杂性科学, 2004 (1): 9–19.
- [12] 赵卫锋,李清泉,李必军. 利用城市 POI 数据提取分层地标. 遥感学报, 2011, 15 (5): 973–988.
- [13] 张玲. POI 的分类标准研究. 测绘通报, 2012 (10): 82–84.
- [14] 龙瀛. 街道城市主义新数据环境下城市研究与规划设计的新思路. 时代建筑, 2016 (2): 128–132.

RESEARCH ON THE IDENTIFY METHOD OF URBAN BLOCK USE TYPES IN LUZHOU CITY BY POI DATA^{*}

Gu Jirong^{1*}, Wu Xiuyue¹, Huang Zhiqin², Feng Yicong², Fang Conggang³

(1. The Faculty Geography Resource Science of Sichuan Normal University, Chengdu, Sichuan 610101, China;
2. The Information Center of Sichuan Land and Resource, Chengdu, Sichuan 610072, China;
3. The information Center of Bureau of Land and Resources of Chengdu, Chengdu, Sichuan 610032, China)

Abstract The land use types are defined by land use planning or city planning. The types are changing with the actual use processing, and different from the planning. This research aims to identify the actual use types by the internet big data. First, we divided the urban block by OSM (Open Street Map) data, and established a mapping relation between the POI (Point of Interest) data classification, then calculated the frequency density and classification proportion to identify the urban block use types. Finally, we established a mapping relation between the POI data classification and land use classification of cadastral data, and analyzed the two identify results. It showed that the two results were lower consistency. It was because that the land use data confirmed the use types at land supplying, the current use types was changing all the time. The internet big data identified the present situations of land use. The different of two results showed the inconformity between land supplying use types and current use types. In summary, POI data which has accurate spatial location information and rich attribute information, which can identify the land use timely and fast, but POI data has its defects and deficiencies in use.

Keywords urban block; actual use types; POI; OSM; cadastral data; Luzhou