文献导读（五）：
腾讯位置大数据应用分析：人口分布与城市功能

导读人：李叶增，广东工业大学
陈　斌，美国加州大学戴维斯分校
刘信陶，香港理工大学，

点评人：郭殿升，腾讯地图大数据实验室

北京时间：2020年6月13日上午8点半-9点
U.S. Time: 8:30-9:00 PM, May 29, 2020 (Eastern Time)

议题 Topics

1. 研究背景
2. 数据
3. 方法
4. 结果解读
5. 总结与讨论

文献一：通过集成多源地理空间大数据在建筑物级别绘制精细的人口分布图——研究背景

- Yao, Y., Liu, X., Li, X., Zhang, J., Liang, Z., Mai, K., & Zhang, Y. (2017). Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. International Journal of Geographical Information Science, 31(6), 1220-1244.

➢ 研究背景

- 精细尺度人口分布的数据，尤其是在建筑水平，在许多领域发挥着重要作用，例如流动人口的监测，优化资源配置和城市结构的分析。
- 基于移动位置的服务（LBS）技术在中国得到了快速发展，并且可以从Internet网站或公共服务（包括兴趣点（POI），全球定位系统（GPS）轨迹）中获取大量的多源地理空间大数据：公共交通，移动通信和签到数据。这些数据集可以组合，以确定有关**城市结构，经济活力和交通拥堵信息。**
- **不同的地理空间大数据可以捕获地面真理的不同方面**，以往研究大多数集中在全球人口和环境变化上。但是，因为缺乏可靠的数据和模型，上述研究都无法通过使用多源地理空间大数据在建筑级别分配人口分布。
- 本文重点内容：在考虑建筑物人口分布空间异质性的同时，**通过集成多源地理空间大数据，将普查数据按比例缩小到建筑物的人口分布**。

文献一：通过整合多源地理空间大数据绘制建筑物一级的精细人口分布图——研究数据

➢ 实时腾讯用户密度

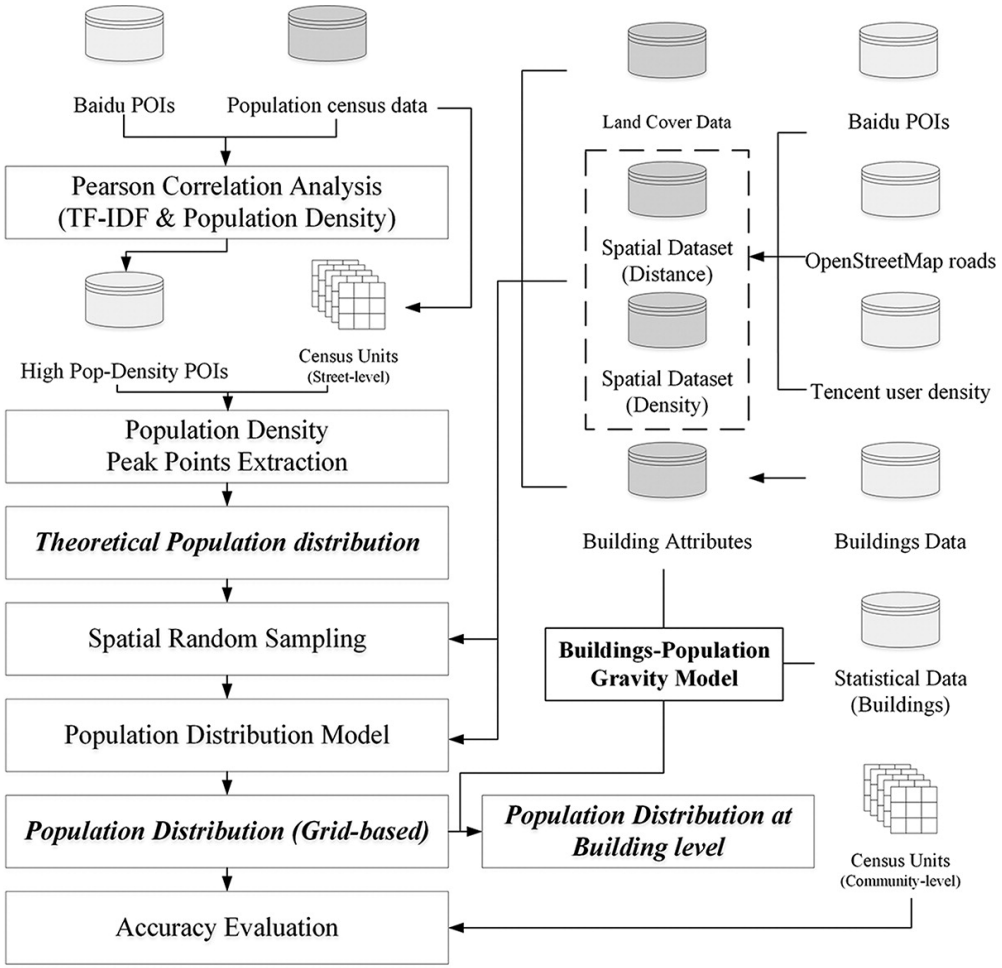
- 实时腾讯用户密度（RTUD）是这项研究中最重要的数据类型，由腾讯（http://www.qq.com）提供，RTUD主要记录了使用腾讯应用程序的智能手机用户的位置，例如腾讯移动应用程序QQ（类似信使的软件），微信（移动聊天软件），Soso Maps（网络地图服务和导航软件）以及一些提供LBS服务的其他移动应用程序。

- 腾讯用户的分布可以看作是总体人口动态分布的一种偏向抽样。

➢ 辅助空间数据集

- 广州的GIS数据集，POI由百度地图服务（http://map.baidu.com）提供

文献一：通过整合多源地理空间大数据绘制建筑物一级的精细人口分布图——研究方法

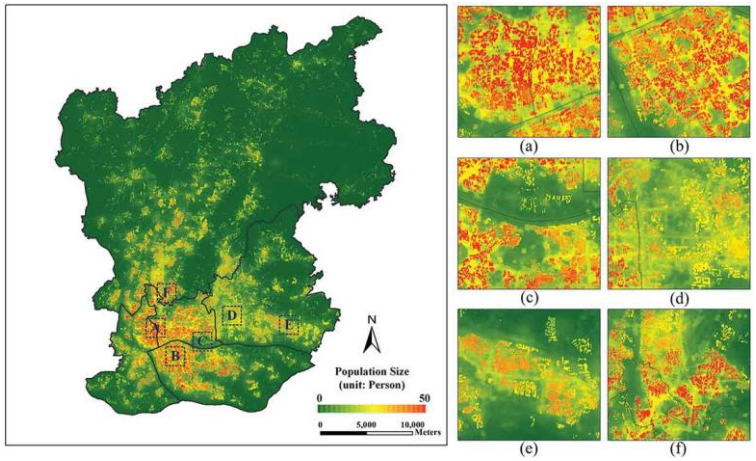


在建筑规模上绘制人口图的拟议方法流程图

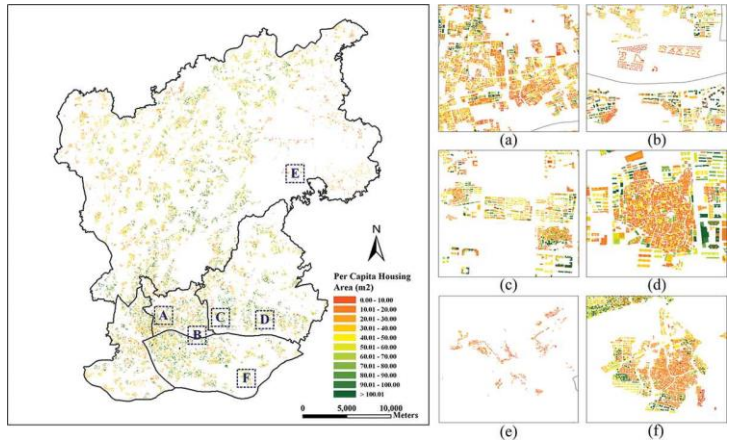
- 三个步骤来估算建筑物中的人口分布：

- 选择可以表明高人口密度的POI，并绘制了每个人口普查单位的初步人口分解图。
- 通过整合RFA和地理空间大数据的多个来源建立了非线性人口模型，以表明研究区域的空间异质性。
- 使用提出的迭代建筑人口引力模型计算了建筑级别的人口分布，并将结果的可靠性与人口普查数据和其他最新方法得出的结果进行了比较。

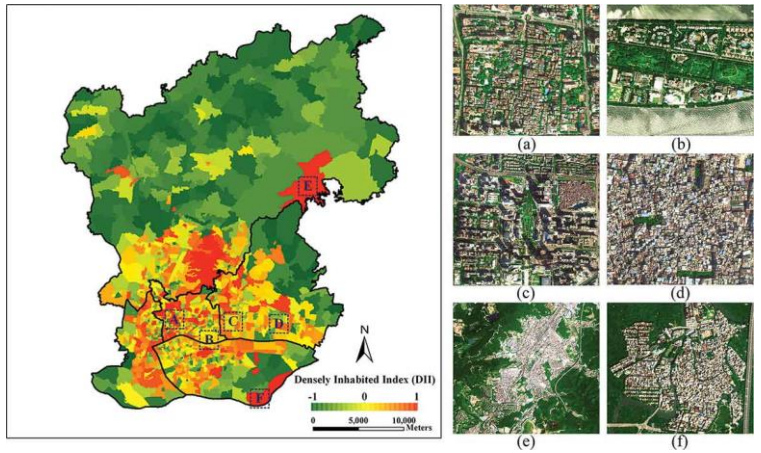
文献一：通过整合多源地理空间大数据绘制建筑物一级的精细人口分布图——结果解读



通过RFA在25 m的局部尺度上模拟的人口分布图



根据拟议的人口建筑引力模型得出的建筑物人均住房面积



根据建筑物级别的模拟人口分布，街道一级每个管理单元中的密集居住指数（DII）

文献一：通过整合多源地理空间大数据绘制建筑物一级的精细人口分布图——总结与讨论

本文主要创新点：

- 通过将官方调查数据中的多源信息融合到地理空间大数据中，本研究建立了多尺度人口模型以缩减人口普查数据的比例，并获得了25 m精细空间分辨率的高精度人口图。
- 提出了住宅建筑与人口之间的引力模型，以首次在建筑物级别生成合理的人口分布。通过与几种流行的人口图绘制方法进行比较，提出的方法在社区一级获得了最高的准确性

本文的可重复与可复制性：

- 本文以广州为例，其他城市人口数据可结合其建筑物分布，运用本文方法进行复制，可拓展到全国其他地区。

尚存局限之处：

- 建筑物类型实际上与与人类活动目的地相关联的各种偏好有关，在建议的重力模型中应考虑不同建筑物类型的吸引因素。

未来建议与可扩展方向：

- 本文以广州为例，其他城市人口数据可结合其建筑物分布，运用本文方法进行复制，可拓展到全国其他地区。

文献二：集成多源大数据以推断建筑功能——研究背景

- Niu, N., Liu, X., Jin, H., Ye, X., Liu, Y., Li, X., ... & Li, S. (2017). Integrating multi-source big data to infer building functions. International Journal of Geographical Information Science, 31(9), 1871-1890.

➢ 研究背景

- 近年来，**从“大数据”样本中提取特定的有用信息已成为一种重要的新趋势**，这可以在估算城市土地使用的研究。大数据的使用在寻求确定给定人群与他们经常出入的地点有关的日常流动方式的研究中尤其重要。这样的模式（例如，通过确认人们通常在工作日早上离开家并在晚上回到家，而与工作场所有关的模式相反），使研究人员可以洞悉给定位置的社会功能。

- 有关城市建筑功能的信息不仅有助于更好地了解城市的运作方式，而且还有助于为决策者建立评估和改善城市规划有效性的基础。尽管有这些优势，但可能**仅仅是由于缺乏可用数据**，迄今为止，几乎没有任何学术研究成功地集成了多源“大数据”以检查建筑级别的城市土地使用情况。

- 论文重点内容：**整合了多源大数据**（微信用户的实时位置记录，出租车GPS轨迹数据，兴趣点（POI）数据和高分辨率Quickbird图像的建筑足迹数据），**提出了一种基于密度的方法**来推断中国广州市天河区城市建筑的功能。

文献二：集成多源大数据以推断建筑功能——研究数据

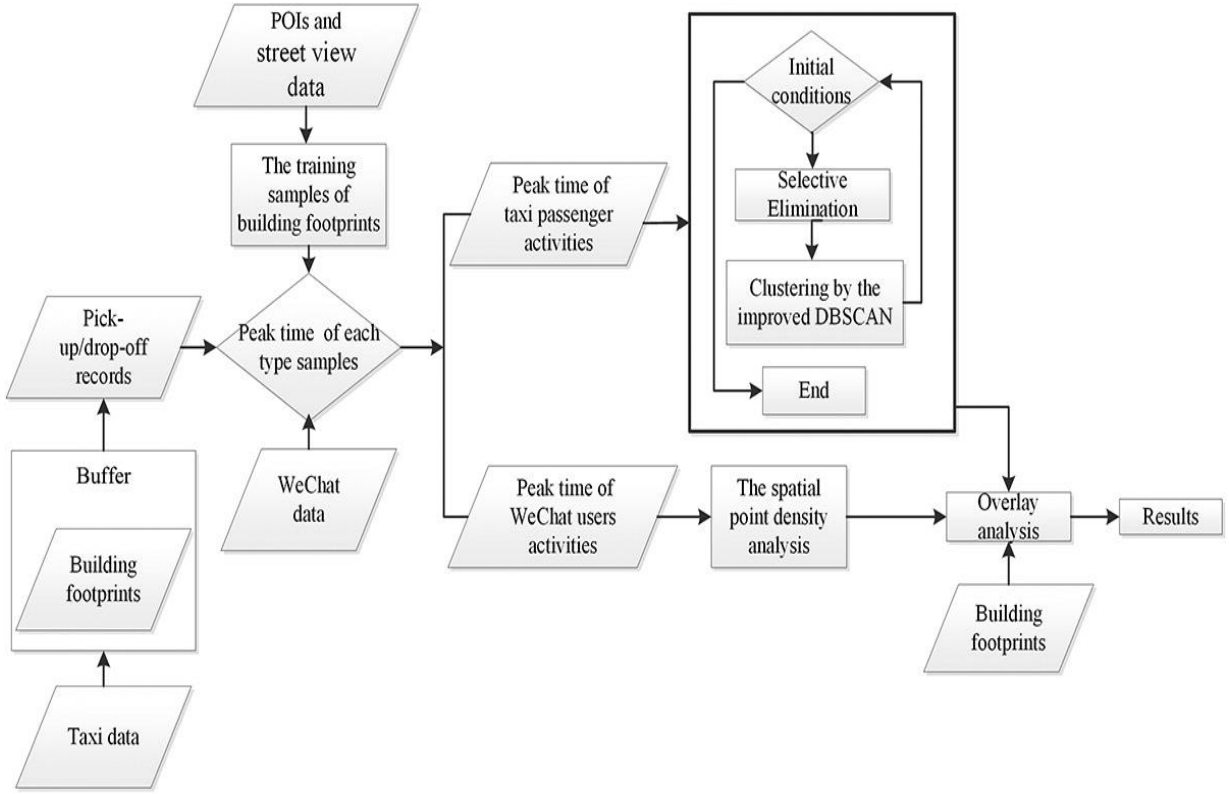
➢ 实时微信用户密度信息:

- **实时微信用户密度信息:**本研究中使用的第三个数据集由实时微信用户密度信息组成。为了获取这些实时数据，基于'Easygo'（http://ur.tencent.com）的API（应用程序接口）开发了一个Web爬网程序，以记录涵盖我们研究区域的实时微信用户数据。它是微信的开放平台，为公众提供周围感兴趣地点的实时拥挤感。通过执行我们的程序，我们收集了2015年6月15日至21日的每小时微信用户密度数据，其空间分辨率为25 m。

➢ 辅助数据集

- 建筑足迹数据集
- 出租车数据集
- POI数据库
- 百度地图获取的街景。

文献二：集成多源大数据以推断建筑功能——研究方法



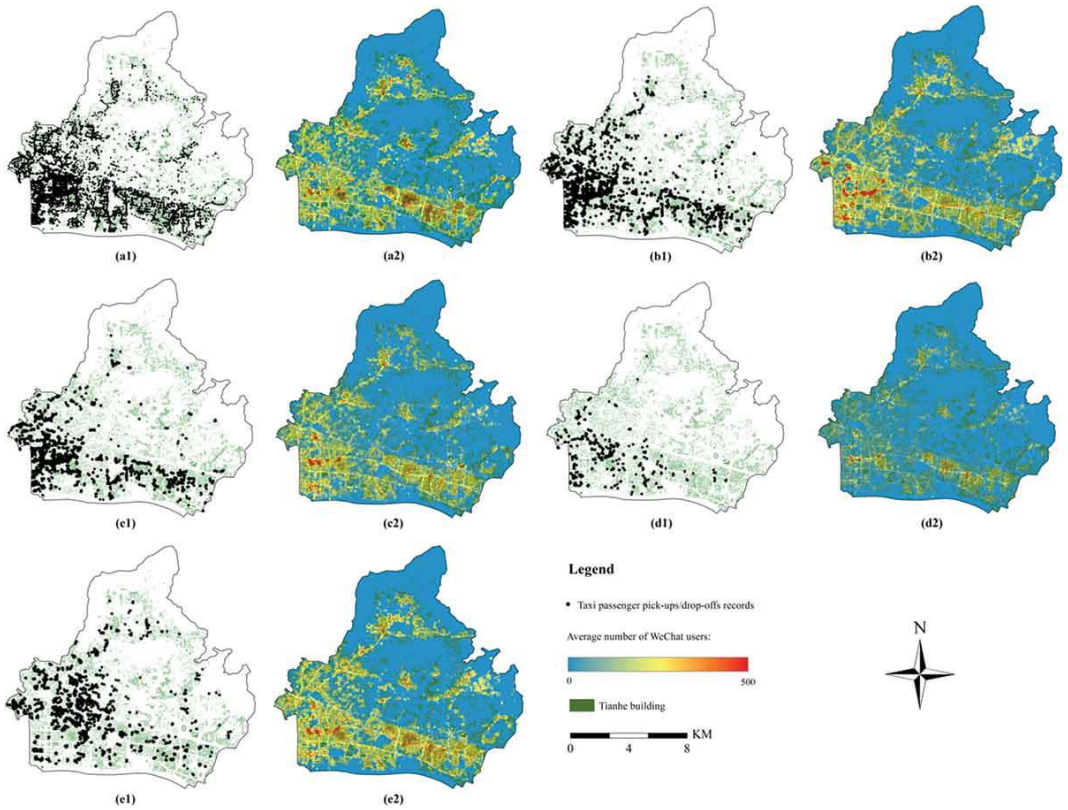
基于密度的方法

➢ 三个步骤：

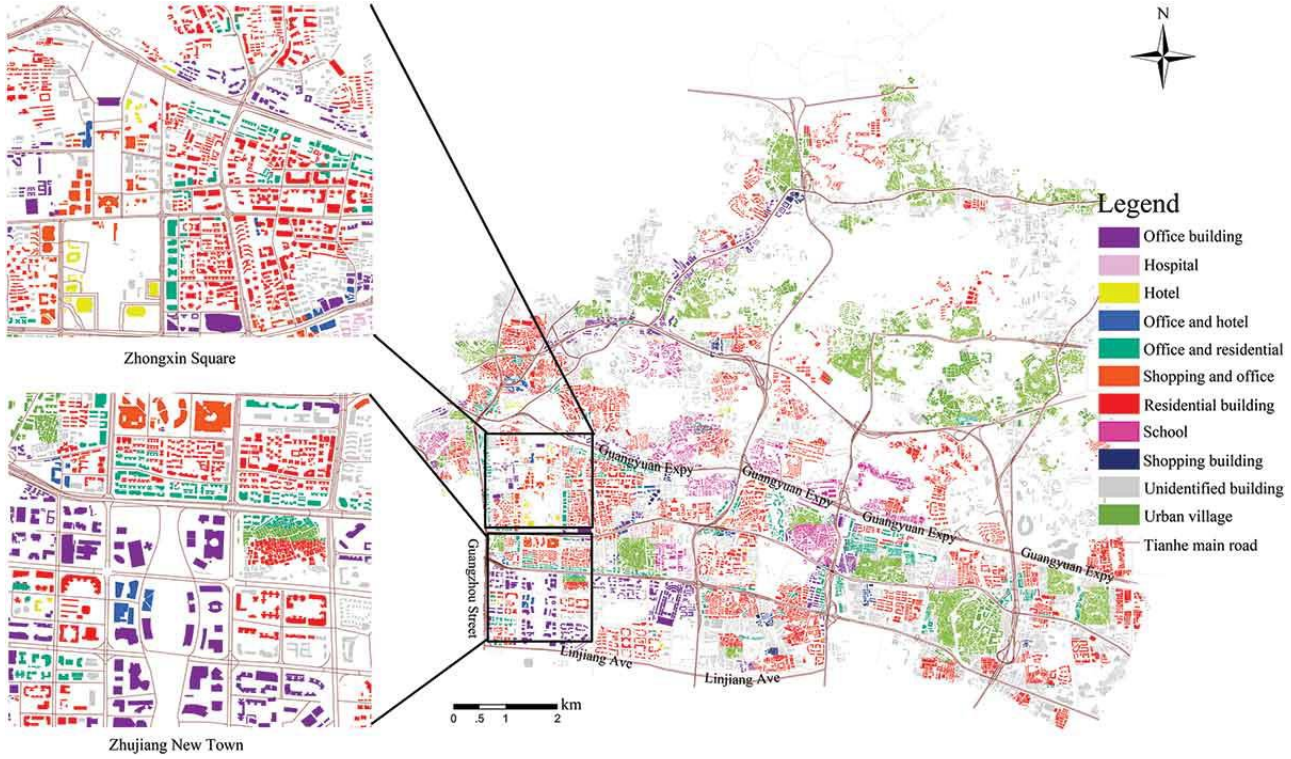
- **个人人行为的指标**（出租车乘客的接送和下车地点）建立城市建筑物之间的关系
- 经过修改的**基于密度的应用空间聚类**将其聚类。使用噪声算法（修改后的DBSCAN）。使用空间点密度分析法对来自微信的社交媒体数据进行了分析。
- 根据一组定义的评估规则推断建筑功能。

将微信数据的空间点密度分析的结果以及天河的建筑足迹相结合，在ArcGIS 10.2中进行了覆盖分析。然后将出租车数据和天河的建筑足迹的聚类结果重叠，并将微信数据的空间点密度分析结果与天河的建筑足迹进行整合。

文献二：集成多源大数据以推断建筑功能——结果解读



住宅建筑物（包括城市村庄），（b1-2）办公大楼，（c1-2）购物建筑物，（d1）高峰时间的空间覆盖分析结果（出租车数据和微信数据）家酒店和（e1-2）学校



天河区（中国广州）的推断建筑功能

文献二：集成多源大数据以推断建筑功能——总结与讨论

本文主要创新点：

- 提出了一种基于密度的方法，通过集成多源“大数据”来推断建筑功能。
- 微信数据的使用比使用出租车数据识别住宅，办公室和城市村庄获得的推断结果更好。相反，使用出租车数据可以更轻松地识别购物中心，酒店和医院。
- 在显示建筑规模上的人类动力与城市综合体之间的关系时，使用集成的多源大数据比单源大数据更有效。

本文的可重复与可复制性：

- 基于此方法推断城市建筑的功能，可研究如何紧密结合城市建筑功能，对COVID-19的传播起到防控作用。

尚存局限之处：

- 私家车增多，出租车数据和微信数据仅占人类活动的一部分，不能全面反映人类活动；多源数据的时间并不总是同步的

未来建议与可扩展方向：

- 可为其他城市研究（例如城市活力）的精细化提供一些帮助。应用于发展中国家和发达国家的更多城市。此外，我们还将整合其他大数据（例如公交和地铁的智能卡数据），并利用多学科知识来推断城市建筑的混合功能。

文献三：基于多源大数据的中国基本城市土地利用制图
——研究背景

- Gong, P., Chen, B., Li, X., Liu, H., Wang, J., Bai, Y., ... & Feng, Y. (2020). Mapping essential urban land use categories in China (EULUC-China): preliminary results for 2018. Science Bulletin, 65(3), 182-187.

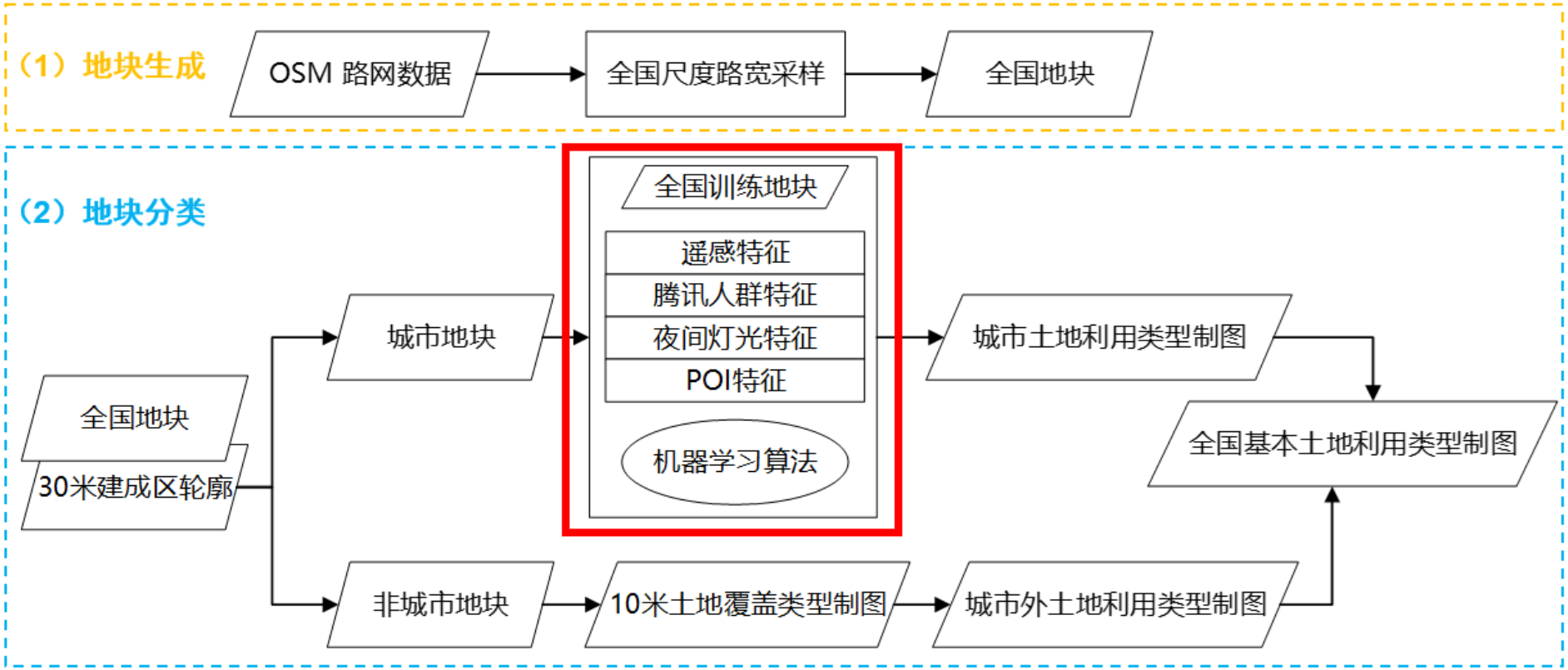
➢ 研究背景

- 城市土地利用是地球上人类最高水平的改变，并且由于人口增加和城市化而在迅速变化。城市用地对当地的水文学，气候，生物多样性和粮食生产有广泛的影响。为了应对由城市化引起的各种国家和全球环境挑战，重要的是要**在国家和全球范围内使用相同或一致的数据源**，相同或兼容的分类系统和方法得出城市土地利用制图。
- 尽管在全球范围内存在许多城市土地利用制图，但在**大空间范围内不存在详细的城市土地利用制图**。即使在美国，中国和欧盟等大国或地区级别，也很少有一致的土地利用制图工作。
- 中国的城市用地发展迅速，以至于2015年取代美国成为建成区面积最大的国家。但是，我们**对中国城市中详细的土地利用类型的分布，模式和组成并不完全了解**。
- **论文主要内容：**本文提出了涵盖居住-休闲-交通-工业-办公（RETIO）五大类用地的“基本城市土地利用类型（EULUC)”的概念，首次实现了全国范围内城市土地利用制图。

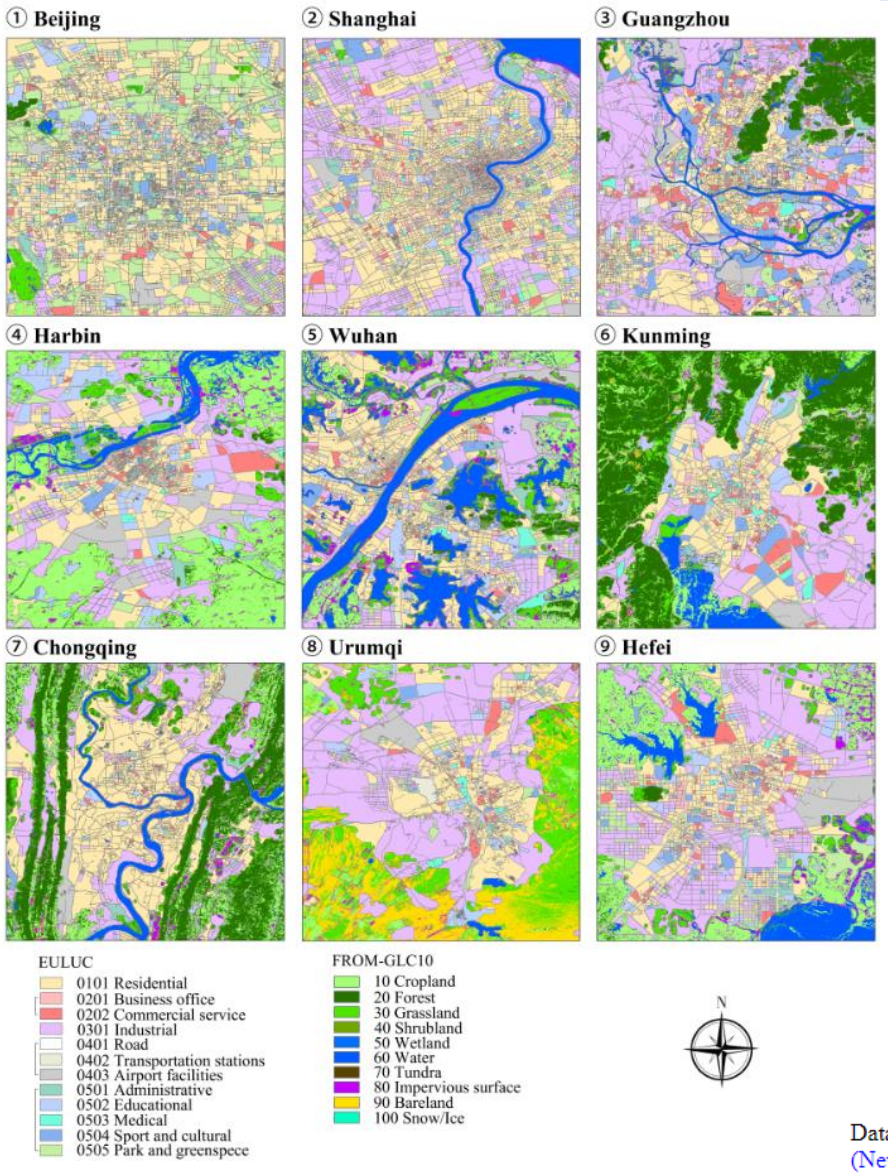
文献三：中国基本城市土地利用制图——研究数据与方法

➢ 主要数据

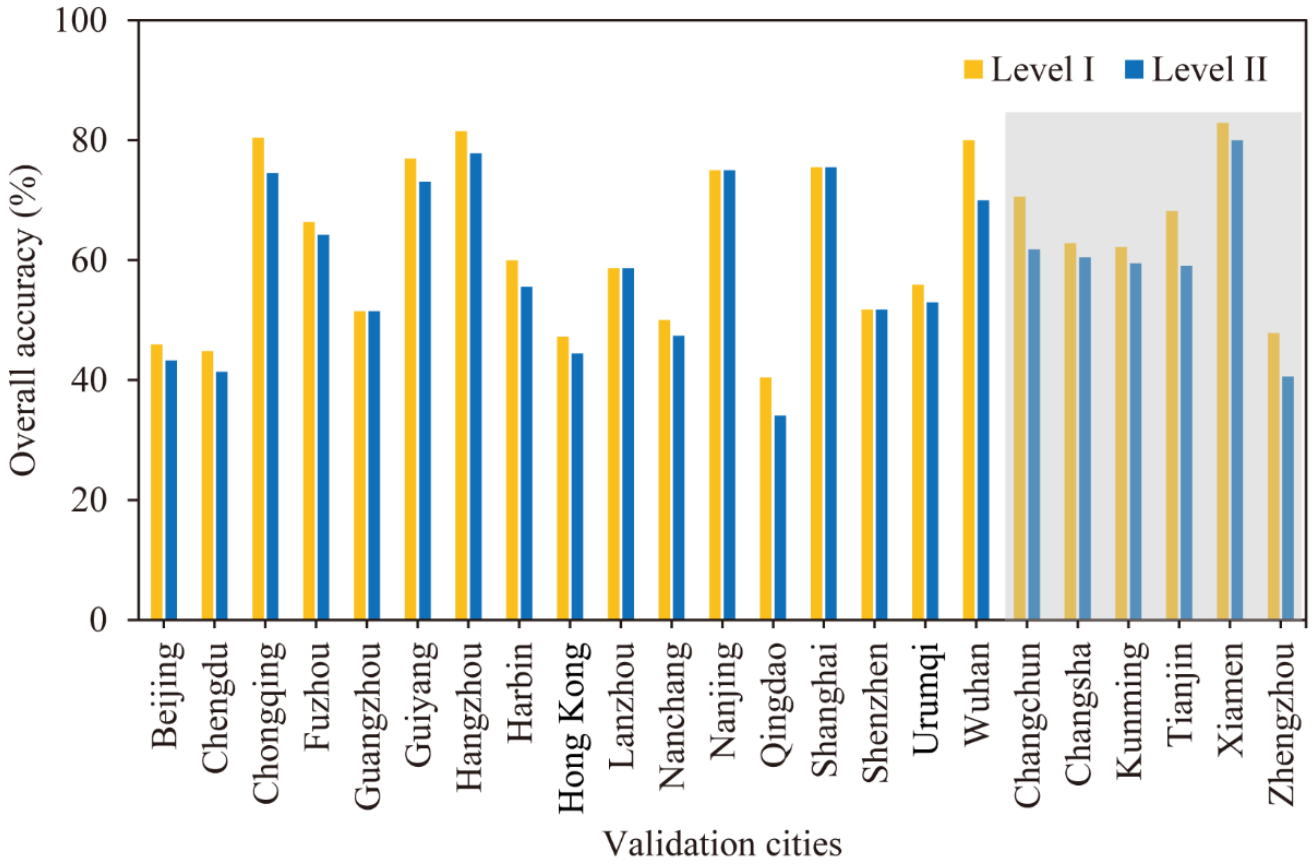
- 10米哨兵遥感数据
- OpenStreetMap数据
- 珞珈一号夜间灯光数据
- 腾讯移动定位
- 高德兴趣点（POI）等



文献三：中国基本城市土地利用制图——结果解读



主要城市制图结果



不同城市的精度验证（一级类和二级类）

数据下载：http://data.ess.tsinghua.edu.cn/

Data Download
(New)
Gong, P., Chen, B., Li, X., et al., 2019. Mapping Essential Urban Land Use Categories in China (EULUC-China): preliminary results for 2018. Science Bulletin. https://doi.org/10.1016/j.scib.2019.12.007, Data download (EULUC-China-2018)

文献三：中国基本城市土地利用制图——结论

本文主要创新点：

- 首套结合卫星遥感和众源大数据，首次实现了全国范围内城市土地利用制图。
- 第一版中国基本城市土地利用制图结果共包括440798个地块（除道路用地），一级总体分类精度为61.2%，二级总体分类精度为 57.5%。
- 以城市地块尺度生产的面向对象的高分辨率城市土地利用制图, 以期与世界共享中国的城市土地利用制图经验和智慧。

本文的可重复与可复制性：

- 可对城市的土地利用进行复制分析

尚存局限之处：

- OSM路网数据在中国不完整，导致城区外围地块划分较大、混合度较高；
- 目前采用的是单一地块单一用地类型，没有考虑混合用地类型；

未来建议与可扩展方向：

- 提高地块的质量；系统地测试样本、分类特征和算法；硬分类和软分类策略。如何在地块尺度进行空间上和垂直方向上对土地使用比例进行软分类核算是对土地利用制图的挑战。