

无桩共享单车“潮汐”特征的多尺度时空聚类挖掘

开题报告

张浩怡、潘凌志

2024 年 12 月 2 日

1 研究背景以及意义

共享单车的兴起，“共享单车+地铁”“共享单车+公交”已成为城市通勤的主要接驳方式，但共享单车的“潮汐效应”也成为共享单车管理和资源调配的“痛点”和“难点”。具体而言，自行车的“借还”存在特定时间段的波峰波谷现象，在城市下班的早晚高峰“借不到，停不好”，由此造成了市民生活的不便利以及公共资源的浪费。因此，如何发现并理解共享单车的“潮汐”规律，对于共享单车的有序规范发展，优化用车体验和环境等具有重要意义。

相对于有桩“公共自行车”，无桩共享单车的出现较晚^[1]。目前相关研究大多关注有桩“公共自行车”的资源分配问题。这得益于有桩共享单车较长的历史背景、固定停车桩位带来的问题化简以及长期研究的积累。具体而言，这些研究可分为两类，一类注重于公共自行车用车量的影响因素。不同学者采用不同的方法进行共享单车聚类问题的研究。如 El-Assi 等（2017）基于经验探讨了有桩共享单车系统的使用模式，包括早晚高峰期间的租还车行为^[2]。随着数据收集技术的发展，研究开始利用大数据分析来识别热点区域。（高楹等，2018）^[1]。另一类则探索如何通过合理调度实现公共自行车的供需平衡。主要是结合启发式算法实现路径优化^[3-4]，近年来随着神经网络技术的发展，则更加注重深度学习的应用和跨学科方法的结合。其中图卷积神经网络（GCN）在共享单车潮汐点预测中被广泛应用，这种深度学习方法能够更好地捕捉复杂的时空依赖性（王竞成等，2021）^[5]。

综上所述，已有研究主要关注有桩共享单车的用车影响因素以及站点之间的调度方法。而由于无桩共享单车不存在集中的站点，车辆的起止位置仅与用户个人的出行目的有关，分散于城市中，导致其形成的网络极其复杂；而其流动性更强的特点也导致存车数量难以实时监控。因此，我们希望可以系统研究无桩共享单车的时空分布特征及影响因素，并提出针对性、可行性高的优化策略。

2 研究问题

1. 问题一：在微观尺度上，识别并聚类具有不同“潮汐”特征的源汇区域。结合土地利用分类数据，判断各个土地类型对共享单车使用情况的影响。
2. 问题二：在宏观尺度上，进行网格间的单元调度。基于网格化数据和土地利用类型数据，思考局部共享单车资源调度方法。

表 1: 实验数据清单

数据名称	数据时间	数据规模	字段名称	字段含义
深圳市共享单车订单数据	2021 年 8 月 23-30 日 0:00 - 24:00	7,965,384 条	START_TIME	开始时间
			END_TIME	结束时间
			START_LAT	开始纬度/°
			START_LNG	开始经度/°
			END_LAT	结束纬度/°
			END_LNG	结束经度/°
			USER_ID	用户 ID
			COM_ID	企业 ID
深圳市 POI 数据	2021 年 8 月	3,068,185 条	POI_TYPE	地物类别
			LATITUDE	纬度/°
			LONGITUDE	经度/°

3 数据来源与描述

如表 1 所示, 本文的数据主要来自两部分,

1. 2021 深圳开放数据应用创新大赛”的数据。来源于深圳市政府数据开放平台（下称平台，相关链接：<https://opendata.sz.gov.cn>）。数据包含 2021 年 1 月至 8 月的部分共享单车订单数据，共 8 个数据项，数据类型均为字符串型，全量为 244,638,540 条，为确保用户隐私安全，数据已进行必要的脱敏处理。
2. 深圳市土地使用情况兴趣点（POI, Point of Interest）。来源于公开网络平台 POI 数据（相关链接：<https://www.poi86.com>）。数据包含 2021 年深圳市各区的土地使用情况。共 23 种使用用途，全量 3068185 条。

4 技术路线

如图 1 所示, 本研究方法包括数据分析、建模分析和实验分析 3 个部分。

其中数据分析部分包括数据预处理和浅层特征发掘。数据预处理包括删除深圳市范围外、速度异常、持续时间异常等等噪音数据点。以及对数据集进行日期上的选取与抽样。

建模分析则通过已有特征进行相关性网络构建，识别并聚类出不同的潮汐区域。按照深圳土地使用情况的地理区块划分统计车辆订单的平均流入流出值，对车辆需求做出预测，作为调度的参考。

最后在深圳全市和热点地区两个尺度下进行聚类划分实验，并在全市聚类划分的基础上对单车潮汐特征挖掘进行研究，用已得到的强源、强汇点的潮汐特征，进行并结合网格间单车调度。

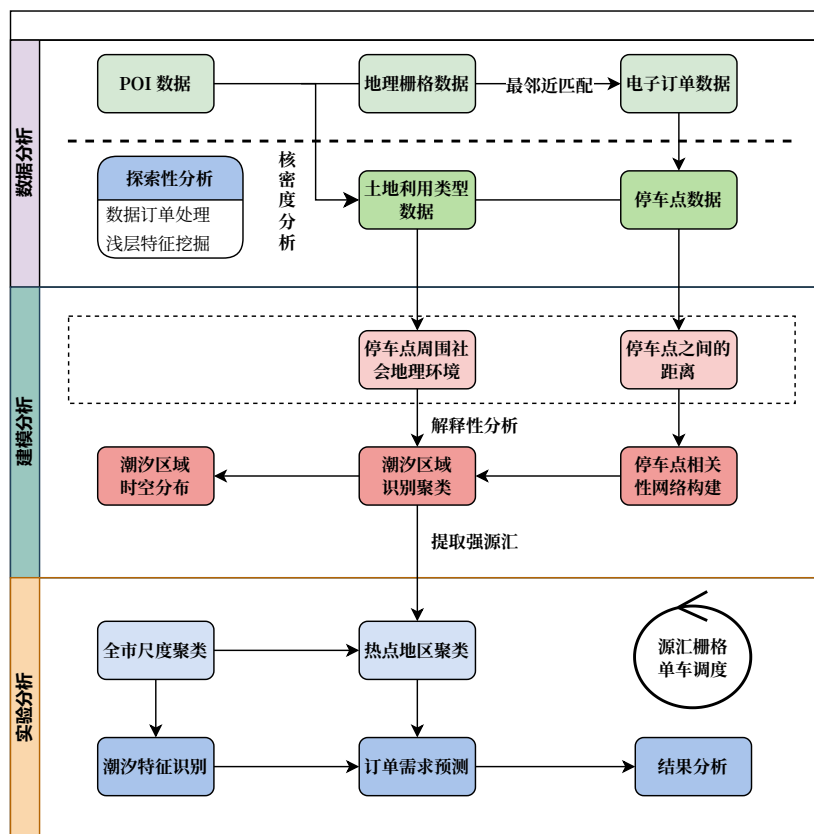


图 1: 技术路线

5 困难点分析

在我们预实验的过程中，发现如下问题：

1. 基于大量数据点的聚类计算与其他分析对运行时间以及运行内存都有较高的要求。而减少数据点以及减少划分的区域数量又会造成结果的偏差。
2. 若划分的区域范围较大，容易出现多数地区的时间潮汐情况相似度很高的情况，难以得出理想结果。
3. 某一地区的源汇可能受到多种因素的影响，例如星期，天气，月份等等。较难判断周期性因素与偶然性因素。
4. 基于经纬度的数据点与不同的地图数据在运算过程可能出现较大的误差。

具体思路及方案将在最终实验报告中进一步讨论。

参考文献

- [1] 高楹, 宋辞, 舒华, 等. 北京市摩拜共享单车源汇时空特征分析及空间调度 [J]. 地球信息科学学报, 2018, 20(8): 1123-1138.
- [2] EL-ASSI W, SALAH MAHMOUD M, NURUL HABIB K. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in toronto [J]. Transportation, 2017, 44: 589-613.
- [3] 王海星, 王德占, 申金升. 蚁群算法解决有时间窗的车辆优化调度问题研究 [J]. 物流技术, 2006 (11): 37-40.
- [4] JUN L. Generalized assignment heuristics for vehicle scheduling[J]. System Engineering Theory & Practice, 1999, 17(1): 27-33.
- [5] 王竟成, 张勇, 胡永利, 等. 基于图卷积网络的交通预测综述 [J]. 北京工业大学学报, 2021, 47(8): 954-970.