山 东 科 技 大 学

本科毕业设计（论文）开题报告

题 目 基于共享单车骑行数据的上海市通勤模式分析

学院（系、部） 测绘与空间信息学院

专业班级 地理信息科学 20-2

姓 名 潘志清

学 号 202001020717

指 导 教 师 牟乃夏

填表时间： 2024 年 3 月 15 日

填表说明

1.开题报告作为毕业设计（论文）答辩委员会对学生答辩资格审查的依据材料之一。

2.此报告应在指导教师指导下，由学生在毕业设计（论文）工作前期完成，经指导教师签署意见、相关系主任审查后生效。

3.学生应按照学校统一设计的电子文档标准格式，用A4纸打印。

4.参考文献不少于10篇，其中应有适当的外文资料（一般不少于2篇）。

5.开题报告作为毕业设计（论文）资料，与毕业设计（论文）一同存档。

|  |  |
| --- | --- |
| 毕业设计（论文）题目 | 基于现代浏览器的GIS系统设计与实现 |
| 题目类型 | □毕业设计 ■毕业论文 |
| 题目来源 | □实验 □实习 □工程实践 □社会调查 ■科研选题等其他 |
| 1. 本课题的研究目的和意义   城市通勤数据反映了城市空间内中小尺度、周期性、密集型、高强度的人员流动情况。人们骑行共享单车属于一种缓慢的匀速移动，且具有较高的灵活性，其轨迹往往能够渗透到城市的每一条毛细血管般狭窄的道路中，相较于固定线路的通勤方式更能反映人的行为特性。结合以上两点，借助共享单车通勤轨迹数据，我们既可以从宏观上把握城市主要交通干线、主要产业热点区域一周七个工作日的人口规律性流动情况，分析总体城市职住分布与产业布局情况；我们也可以从城市的微观结构着手，探究渗入城市细小角落的骑行轨迹究竟反映了怎样的行为与细碎环境的相互作用。  无桩电子围栏式共享单车是一种具有浓厚互联网气息的产品服务，也是共享经济快速发展阶段的标志性产物。共享单车一般依赖 GPS 定位设备与互联网接入，借助电子围栏算法及主动式 GPS 定位圈定借还车区域及使用范围，并结合移动客户端与后台调度算法规范用户骑行行为（如骑行红包奖励及出圈扣除押金惩罚等）以达到维持正常经营的目的。相较于传统的有桩式公共自行车，共享单车具有极大地便利性性优势与极低的准入门槛，因此一经投放市场，共享单车的用户规模便快速扩大，其本身也成为解决城市通勤“最后一公里”、缓解城市污染与交通拥堵等问题的重要解决方案。然而，基于共享单车骑行轨迹的人员与物质流动模型其实是一种耗散模型，与任何自然发生的能量耗散一样，共享单车网络对于城市通勤的贡献也始终随着一天中时间的推移自然下降到低水平，也就意味着原本相对集中的易于使用的共享单车分布会由于通勤等原因变得分散和难以使用。想要维持共享单车的合理分布及单车流动“借-还”循环畅通，就需要依赖基于通勤模式经验指导下的强制调度，这种调度一般发生在非通勤时间，共享单车服务公司使用卡车或平板三轮车搜集散乱在城市各处的共享单车并整齐放置到合理的地点以支撑第二天的骑行通勤需求。因此，由共享单车骑行轨迹数据中挖掘出的通勤模式具有一定的现实指导意义。  综上所述，本研究旨在深入理解共享单车在城市通勤中时空模式，为城市规划和交通管理提供科学参考，同时为共享单车服务的优化提供实践指导。 | |
| 1. 本课题的主要研究内容（提纲）   1. **数据搜集与数据预处理**  1. 共享单车轨迹数据清洗：异常值剔除（范围异常、骑行时间异常、骑行速度异常）、重排序（按照骑行日期重新排序）、以天为单位切分为独立文件、某一天的独立文件再以 24 小时制排序。  2. 共享单车轨迹数据预处理：根据公式计算轨迹点 GeoHash 值作为新的属性保存在文件中、删除原先的用户及车辆ID属性以时间顺序重新编码记录索引号。  3. 乱序共享单车轨迹的重建与简化：根据 GeoHash 值采用最邻近法重建骑行轨迹。  2. **基于共享单车轨迹的时空通勤模式分析**  1. 统计用户通勤距离、骑行时间等参数随时间的变化规律。  2. 根据用户骑行数据绘制热力图，提取骑行热点区域，分析原因。  3. 使用线密度分析提取骑行轨迹与各级别路网的线密度栅格。  4. 结合路网数据与骑行轨迹分析上海市道路建设对居民骑行通勤模式的影响。  5. 基于均匀时空格网统计借还车分布状况，并据此进行核密度分析。  3. **结论分析与制图**  1. 结合上述分析步骤的结果，使用相关软件制作统计图表、热力图、频次空间分布图等图表用于辅助分析与表达。  2. 研究居民骑行通勤模式、结合上海市路网数据提出合理的道路建设优化建议。 | |
| 1. 文献综述（国内外研究情况及其发展）   公共自行车系统（Bicycle-sharing system）最早可以追溯到1965年阿姆斯特丹城市街头的“白色自行车（White Bicycle Plan）”计划[1]，该计划将自行车刷成醒目的白色并免费提供给人们随意使用。由于缺乏有效的监管，多数自行车在很短的时间内遭到偷窃或蓄意破坏，致使该项目难以为继。后来，一些项目通过引入带有监控设备的停靠点并发放收取了一定押金的智能卡来约束使用者的行为，例如英国朴茨茅斯大学于1995年启动的“智能单车”项目等。然而，固定的停靠站点难以充分满足用户灵活的骑行需求，反而削减了自行车这一灵活交通工具的可达性优势[2]。随着移动互联网的发展及移动支付手段的普及，一种新兴的无桩公共自行车系统[3]取代了原先依赖停靠点的公共自行车系统。该系统使用电子围栏技术及移动端应用程序来规范用户用车行为，这种带有鲜明的互联网及共享经济时代特色的最新一代公共自行车系统，即当今人们所熟知的“共享单车”系统。共享单车出行数据蕴含丰富的时空信息，通过对这些数据的分析、整理及挖掘可以为城市治理及服务优化提供重要依据。  采用合适的时空可视化手段可以有效发现隐藏在数据背后的规律。全雨霏[4]从宏观大数据角度，综合考虑骑行轨迹的起点、终点及起-终连结的时空特征，采用统计分析、问卷调查、实地调研等方法研究了南京市共享单车通勤模式，分析出共享单车使用时段的峰值规律、空间热点分布规律及其影响因素（绿地分布、公共交通换乘站点分布等）并结合此为城市规划提出自己的建议。Corcoran等人[5]基于澳大利亚布里斯班市的 "CityCycle" 公共自行车系统出行数据，采用多种可视化方法着重研究天气及日历事件（例如公共假期、周末、节假日等）对骑行时空模式的影响。共享单车出行数据中若存在明确的始末点位置记录，对其采用合适的空间可视化手段可以有效揭示用户骑行的O-D模式。Yang Xu等人[6]通过对新加坡主要无桩自行车共享运营商提供的四个月GPS数据进行特征分解分析，采用合适的时空可视化手段研究了不同地点的骑行活动时间波动模式，并将无桩自行车使用的时空模式关联到多个关键的建筑环境指标。杨永崇[7]等人基于城市骑行大数据，使用地理可视化手段研究城市骑行热点区域及用户出行OD模型。高楹[8]等人基于北京市摩拜单车运营数据，采用空间可视化手段研究该区域共享单车“源”、“汇”点的时空分布模式，并基于此提出一种局地优化的调度模型以优化共享单车小区域内的调度服务。  聚类分析可以有效识别具有相似模式的站点，这些站点可以进一步构建为网络以分析社区划分及流动模式等深层次内容。王若萱等人[9]使用随机森林模型，结合一套基于兴趣点数据的目的地区域建模指标体系重点分析了上海市区域性共享单车通勤模式，为后续该领域内的模拟仿真系统提供了研究基础。Jiménez 等人[10]基于用户出行行为模式数据，采用特定数据挖掘工具基于所提出的三个指标（利用率、累计行程比率和周转站比率）对都柏林公共自行车共享系统的站点进行聚类分析。张芳等人[11]使用最邻近查找法将共享单车轨迹点聚类到最近的城市兴趣点上，以实现空间划分及骑行网络节点生成。刘冰等人[12]使用聚类及社区发现算法（鲁汶算法）研究基于共享单车的生活圈，他们发现骑行生活圈的空间范围相较于步行生活圈更大也更复杂，是居民日常通勤的基础性组成部分，合理地布置骑行网络对于便利居民生活具有重要意义。  回归分析可以有效探究用户行为与多种影响因素之间的关联。常新[13]对深圳市龙岗区时间跨度为半个月的共享单车轨迹数据进行回归分析，总结出骑行行为模式并反向推求城市环境对骑行行为的支撑作用，最后使用空间自相关理论分析原因，指出道路的繁忙程度对于骑行频次具有影响，并根据研究结果提出了合理的城市规划建议。谢国微[14]从天气与城市建成对骑行行为的影响出发，综合采用统计方法、地理加权回归模型(GWR)、核密度分析法、多元线性回归法等多种研究手段探究共享单车使用情况的时空模式及其影响因素。高枫[15]等人使用地理探测器研究广州市主城区共享单车时空分布模式及其影响因素。Shen等人[16]通过对新加坡无桩自行车共享服务的GPS数据进行收集和空间自回归模型分析，研究了自行车保有量、周边建筑环境、公共交通便利性、自行车基础设施和天气条件等因素对无桩式共享单车使用量的影响。  大量的真实出行数据可以有效支撑模拟仿真系统，基于这些数据也可以实现智能预测。随着数据驱动的智能算法逐步发展，也有一部分研究者注意到共享单车轨迹中蕴含的智能，他们使用新兴的数据驱动智能算法直接将数据本身转化为解决问题的能力，为交通预测、道路规划、大数据感知等应用场景提供助力。谢光明[17]为了解决共享单车时空供需不平衡问题，基于大量骑行轨迹数据研究出一套单车流量预测方法。他首先使用两级聚类及 K-means 算法基于密度找到虚拟站点，然后根据用户访问频次结合鲁汶(Louvain)算法得到虚拟站点间的社区网络最后使用图注意力网络(GAT)及LSTM网络并融合气象参数完成单车流量预测。该方法对于共享单车的动态调度，缓解单车分布时空不平衡具有重要的指导意义。Jie Bao 等人[18]基于上海市共享单车大数据记录开发出一套城市自行车道规划算法，该算法具有一个用于描述轨迹长度与收益关系的启发式函数，能够根据项目预算、施工周期等因素灵活调整自行车道设计结果，具有较高的工程实践价值。王俊等人[19]使用带有卷积层的长短期记忆网络（ConvLSTM），结合南京地区共享单车需求量，天气，季节等参数训练深度学习预测模型，对于需求量的预测精度相较于单一网络（卷积网络或长短期记忆网络）取得一定提升。  共享单车带来的环境效益及问题也是许多学者研究的方向。李文翔等人[20]从环保的角度，使用问卷调查法、统计分析法及核密度分析法等方法结合减排因子计算公式评估了2016年上海市全年共享单车对于六种主要污染物的减排效益及空间分布模式，指出上海市人口密集区域具有更大的减排潜力。虽然共享单车是一种新兴互联网经济下发展起来的环境友好型便利交通设施具有诸多优势，共享单车市场快速膨胀也带来了许多城市治理难题，最为显著的问题之一就是大量损坏、报废共享单车的该如何处置。刘泉宏[21]等人重点考虑了共享单车生命周期末环的回收处理问题，基于武汉市共享单车报废点位采用先求取重心后进行 K-means 聚类的处理方法，计算出合理的共享单车回收站部署位置。该研究对于共享单车生命周期中后管理维护站点的选址提供了一定的技术支持。  上文提到的大多研究都是将共享单车轨迹视作一个整体，轨迹点视作点云，往往采用基于密度的空间聚类与统计方法，忽略了共享单车骑行轨迹作为矢量线段集合所内含的空间结构信息。任丹[22]使用 TRACLUS 算法船舶轨迹进行聚类，提取出公共航行轨迹，与一般的基于点密度聚类方式不同，该方法考虑到了轨迹数据本身作为向量的几何特征，对于共享单车骑行轨迹研究领域具有启发意义。  通过分析城市共享单车领域的国内外研究成果，本文作如下总结与思考：  （1）研究细分方向：大致可分为污染物与环境、城市道路规划、模拟仿真算法、社交网络分析、热点路段、时空模式分析及城市道路及空间规划质量评估等。可以发现，共享单车骑行轨迹中蕴含极其丰富的信息，是研究城市、自然、人及其关系的绝佳窗口。  （2）研究数据及方法：除用户骑行轨迹数据外，研究者也会根据研究的细分方向补充部分数据，包括居民出行问卷调查数据、城市基础地理数据（公交站台、绿地分布等）、天气数据、共享单车服务设施数据（报废处理站、虚拟停车桩等）以及其他类型的轨迹数据（包括步行、出租车等）。研究方法大致可分为统计法、聚类法（基于点、基于轨迹线）、地理统计法、及数据驱动的智能算法等。可以发现，共享单车轨迹数据具有丰富的时空结构信息，需要酌情选择合理的分析方法。  （3）思考：骑行轨迹数据往往与城市路网分布、主要基础设施分布以及生活设施分布具有一定的关联性，尤其需要考虑其与主要道路间的关联。骑行轨迹目的地暗含了用户的真实意愿，也可以根据目的地类别对轨迹进行分类，实现更进一步的分析。用户的行为特征及单车流动特征也可以成为重要的分析抓手，例如调查一段时间内某一用户的骑行频率，某段时间内单车的周转率等。共享单车通勤也会存在一定的早晚高峰及潮汐现象，寻找共享单车骑行量时间波动模式及共享单车潮汐区域空间分布模式具有重要意义。 | |
| 1. 拟解决的关键问题   1. **骑行轨迹复原与简化：**由于与服务器的通讯存在延迟等原因，原数据集记录的轨迹中存在个别乱序点，这导致整体轨迹无法直接使用。并且，匀速骑行的轨迹中含有大量值得剔除的冗余点，比如，某段沿平直道路前进的轨迹只需保存起点和终点，而中间点均可视为冗余点。  **2.** **轨迹终点最邻近兴趣点匹配**：轨迹数据只有位置信息而无属性信息，无法进行更进一步的分析，研究。考虑根据兴趣点数据类别来赋予轨迹终点属性，并凭此来分析轨迹终点组成状况。考虑到兴趣点数据数量庞大，需要设计一种能够快速搜寻最近轨迹点的算法。  3. **设计并实现轨迹时空模式分析框架**：拟采用固定分辨率的时空格网进行统计分析，例如统计O-D格网并据此进行通勤潮汐分析。 | |
| 1. 研究思路和方法   **1. 研究思路：**轨迹模式的时空分析  1. 时间变化分析：数据集中记录了2016年8月及2020年8月这两个整月的上海市某品牌共享单车骑行数据，最起码可从年际、月度、逐小时这几个尺度来分析时间上的用户骑行数据变化。  2. 空间分布分析：同时对数据集使用线密度、核密度分析，并结合上海市路网矢量数据进行空间分布的分析。重点考察，热点区域的实际分布，骑行轨迹与上海市各级别路网之间的关联关系。  **2. 研究方法：**  1. 骑行轨迹复原与简化：轨迹复原与简化可以合并到一个步骤中实现。首先，使用 GeoHash 算法对每一个轨迹点建立空间索引并将改索引作为集合的健值（Key），利用 GeoHash 编码长度对应了空间索引定位的精度也就是所表示地理范围的大小对轨迹点进行去重合并，譬如在某一精度参数下（例如六比特位），某两个点的 GeoHash 计算值相同，则根据集合自动去重的特性同一 GeoHash 编码所代表的地理范围内只会保留一个有效轨迹点，这样也就同时实现了轨迹点去重与轨迹简化。执行完以上步骤后，再逐步使用最邻近点算法由起点开始复原骑行轨迹。  2. 轨迹长度计算：采用半正弦公式算法来计算轨迹子段长度，再累加得到总弧段长度。  3. 规则时空格网统计：采用固定分辨率的格网，针对确定的时间窗口统计落入格网内部的轨迹点并基于此进行诸如潮汐分析等的进一步分析。  4. 密度分析：计算轨迹数据、路网数据的线密度栅格，并据此进行进一步分析。也可以使用核密度分析方法来探究骑行起点及终点的热点分布状况。  5. 统计方法：依据时间顺序分别统计用户骑行时间、骑行距离及骑行频次，研究出共享单车通勤模式随时间变化的情况，绘制相关图表，得出结论。 | |
| 1. 本课题的进度安排   第一阶段（2024.1.12 - 2024.3.15）: 确定研究方向，阅读国内外相关文献，总结论文思路、技术要点及待解决的实际问题。获取相关数据，研所可能会用到的技术框架，熟悉数据处理流程。  第二阶段（2024.3.16 - 2024.4.20）: 筛选、清洗、整合数据，数据处理算法实现与验证。  第三阶段（2024.4.21 - 2024.5.19）: 撰写论文初稿，提交初审，根据建议修改完善论文。  第四阶段（2024.5.20 - 2024.6.19）: 论文答辩，交流讨论。 | |
| 1. 参考文献   [1]李文翔,唐桂孔,刘博,等.基于摩拜骑行数据的上海市共享单车减排效益时空分析[J].环境科学学报,2021,41(11):4752-4759.DOI:10.13671/j.hjkxxb.2021.0213.  [2]王若萱,吴建平,奇格奇. 基于上海市数据的共享单车用户通勤模式研究（英文）[C]中国仿真学会.第三十三届中国仿真大会论文集.2021:16.DOI:10.26914/c.cnkihy.2021.025005.  [3]全雨霏. 南京市共享单车使用的时空特征及其骑行环境评估[D].东南大学,2024.DOI:10.27014/d.cnki.gdnau.2022.001097.  [4]常新. 基于共享单车轨迹数据的城市街道可骑行性研究[D].哈尔滨工业大学,2021.DOI:10.27061/d.cnki.ghgdu.2020.002483.  [5]Jie Bao, Tianfu He, Sijie Ruan , Yanhua Li, and Yu Zheng. 2017. Planning Bike Lanes based on Sharing-Bikes’ Trajectories[c]. In Proceedings of KDD’17, August 13–17, 2017, Halifax, NS, Canada., , 11 pages.  [6]刘泉宏,唐福星.基于K-means聚类算法与重心法的故障共享单车回收中心选址优化[J].运筹与管理,2023,32(07):85-91.  [7]谢光明. 基于改进时空图神经网络的共享单车流量预测[D].华东师范大学,2023.DOI:10.27149/d.cnki.ghdsu.2023.004430.  [8]刘冰,王舸洋,朱俊宇,等.基于共享单车大数据的骑行生活圈识别及其活动网络模式分析[J].城市规划学刊,2023(04):32-40.DOI:10.16361/j.upf.202304005.  [9]王俊,于爱荣.基于ConvLSTM的南京地区共享单车需求预测研究[J].软件工程,2024,27(02):55-59.DOI:10.19644/j.cnki.issn2096-1472.2024.002.011.  [10]谢国微. 天气及建成环境对共享单车出行需求的影响研究[D].南京林业大学,2024.DOI:10.27242/d.cnki.gnjlu.2022.000493.  [11]任丹. 基于TRACLUS算法的船舶轨迹分析系统的设计与实现[D].辽宁师范大学,2021.DOI:10.27212/d.cnki.glnsu.2020.001211. | |
| 指导教师意见  指导教师（签名）：  年 月 日 | |
| 所在系（教研室）意见  负责人（签章）：  年 月 日 | |