山 东 科 技 大 学

本科毕业设计（论文）中期检查报告

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级2班

姓 名 潘 志 清

学 号 202001020717

指 导 教 师 牟 乃 夏

填表时间： 2024 年 4 月 18 日

**填表说明**

1. 请学生根据学校、学院（系、教研室）检查的要求认真进行自查，及时发现毕业设计（论文）研究过程中存在的问题，分析原因，并提出解决思路和措施，明确下一阶段任务。
2. 每位学生应根据项目实施情况认真、实事求是填写。填写字体请用宋体小四号，并用A4纸打印，于左侧装订成册。
3. 该表填写完毕后，须请指导教师审核，并签署意见。
4. 《山东科技大学本科毕业设计（论文）中期检查报告》将作为答辩资格审查的主要材料之一。
5. 本表格不够可自行扩页。

|  |  |  |
| --- | --- | --- |
| **毕业设计（论文）**  **题目** | 基于共享单车骑行数据的上海市通勤模式分析 | |
| **题目是否有变化** | 有。原题目技术难度过大，且不具备实际意义。 | |
| **整体进度情况** | 师生双选 | 完成进度： 100 % |
| 任务书下达 | 完成进度： 100 % |
| 开题报告 | 完成进度： 50 % |
| 开题审核表 | 完成进度： 50 % |
| 文献综述 | 完成进度： 50 % |
| 外文译文 | 完成进度： 100 % |
| 毕业设计（论文）正文 | 完成进度： 30 % |
| **毕业设计(论文)研究进展情况** | 1. 已撰写第一章（绪论）章节及第三章（研究数据与方法）的数据预处理章节。  2. 已编写并测试坐标系转换算法与路段长度量测算法。  3. 已实现并测试 GeoHash 空间索引编码算法。  4. 找到 TRACLUS 聚类算法的有关代码资料，逐步开展代码测试整合工作。 | |
| **已取得的阶段性成果** | 1. 已搜集的数据：包括 2016 年 8 月上海地区的共享单车骑行轨迹数据、2020 年 8 月共享单车骑行起止点数据、上海市主要道路矢量数据、上海市主要道路等级及人行道面积数据（政府公开文件，无位置信息）。  2. 实现坐标系转换算法与路段长度量测算法：实现 WGS 84 经纬度坐标转平面坐标的球面墨卡托投影变换及后续的平面距离量测算法、实现半正弦公式直接量测弧段距离。  3. 已实现并测试 GeoHash 空间索引编码算法，并验真其对于查找最邻近点的效果。 | |
| **存在的问题及解决思路：** | 1. 骑行轨迹重建与简化：由于网络延迟等因素，数据集中记录的轨迹信息存在一定的乱序现象，需要寻找一种排序算法来重建用户的骑行轨迹。另外，轨迹数据包含太多轨迹点，且一部分轨迹点间空间差异过小属于骑行路径中的冗余节点，需要寻找一种有效的轨迹简化算法。以上两个步骤可以合并在一个算法中解决。经分析，轨迹数据中的顺序点其实占大多数，乱序现象一般出现在用户启动车辆与交还车辆时且仅影响个别点，完全可以将骑行轨迹视作多个有序子轨迹段的重组问题，而不必从点的角度重建整个骑行轨迹。而子轨迹间的连接点往往是特征点，这样以一定的阈值函数简化子轨迹并重组即可获得简化后的顺序骑行轨迹。  2. 空间索引构建：空间索引可以用来提高空间查询等操作的效率。例如，在骑行轨迹重建中需要寻找某一点的最邻近点，这时合理的空间索引就可以用来估算距离并快速检索最邻近点。拟采用 GeoHash 来对经纬度坐标进行编码，构建空间索引值。GeoHash 递归地将地球表面分为多层级的网格区域，索引值的比特数越长所代表的格网越小，定位也就越精确，并且得益于多层级的格网划分与编码方式，空间上临近的位置经过 GeoHash 算法编码后得到的索引值也具有较高的相似度。  3. 基于骑行轨迹数据的空间聚类分析方法：拟采用 TRACLUS 聚类算法归纳并提取某一时段的公共骑行轨迹。TRACLUS 算法是一种轨迹聚类算法，最早用于归纳飓风运动的公共轨迹。该算法采用分段归组框架，首先利用最小描述长度（MDL）原理将轨迹分段成线段，以最小化描述所需的信息量，从而更准确地捕获轨迹中的变化和模式。随后，利用基于密度的线段聚类算法将相似线段聚集成簇，通过设计距离函数来定义线段的密度，使得算法能够有效处理任意形状的线段簇并过滤噪声。此外，TRACLUS还提供了简单而有效的启发式方法来确定算法参数，提高了算法的灵活性和通用性。综上所述，TRACLUS算法的分段归组框架结合了MDL原理、基于密度的线段聚类方法和参数优化策略，能够在轨迹数据中准确发现子轨迹，并具有较强的鲁棒性 | |
| **下一阶段的工作计划和研究内容：** | 1. 数据清洗与预处理：采用上文提到的解决思路，清洗数据，重建骑行轨迹并进行空间索引编码。  2. 从时间和空间角度对（有效的）骑行轨迹进行分析：对点集采用聚类分析，提取出骑行热点地区。使用 TRACLUS 聚类算法归纳出不同时段、不同工作日的上海市骑行公共轨迹，用于通勤模式的研究。结合上海市主要道路的矢量数据，分析热点地区、骑行公共轨迹与主要道路的重合情况，分析城市骑行热点道路。  3. 制图与论文撰写：从时间角度，根据骑行日期、骑行时间、骑行长度等对轨迹数据进行分类统计，绘制多种图表用于揭示用户通勤的时间规律。从空间角度，考虑到数据量比较大，使用对大数据显示优化过的百度地图 API 绘制热力图、使用 ArcGIS 制作不同工作日的骑行公共轨迹，并将其与城市主要道路的缓冲区进行求交以制作城市热点道路图。 | |
| **指导教师意见** | 指导教师（签名）：  年 月 日 | |