

# 重庆大学“大学生科研训练计划”

## 项目中期检查报告

项目学院（公章）：

教务处制表

项目名称	出租车轨迹数据分析		
立项经费	1700(元)	起止时间	2020 年 11 月至 2021 年 11 月
项目组成员	姓名	学号	年级专业
	张翀	20194159	2019 级计算机科学与技术(卓越)
	钟昱琦	20194176	2019 级计算机科学与技术
	闫若华	20194122	2019 级计算机科学与技术
指导教师	孙天昊		
项目计划进度	2020.06：规划、设计项目实施详细方案； 2020.07-09：深入学习和掌握数据清洗、聚类方法理论、可视化方法； 2020.10-12：处理 GPS 数据，同时进一步掌握聚类分析算法、可视化方法，用所学知识处理部分数据； 2021.01-02：对出租车数据进行聚类分析，如上下客热点，交通热点路线； 2021.03：进行可视化呈现和结果分析，并调整优化算法参数和方法，得出 一套城市热点地区、拥堵路段分析系统； 2021.04：撰写论文，整理文档、准备结题答辩。		

<p><b>项目实际进展情况（包括已取得的成果）</b></p>	<p>我们首先学习了是图形板块，学习了图形的颜色填充、使用模块 patches 绘制几何图形（包括圆、椭圆、举行、圆弧、楔形等）、组合展示统计图形的实现方法（包括判别分析示意图、时间序列图、概率密度曲线等）。在元素板块，我们学习了图形元素的设置方法（包括文本内容、计量单位、刻度线、刻度标签和轴脊等），具体内容包括这只文本内容的样式和布局，调整计量单位和计量方法，调整刻度线和刻度标签，以及轴脊的展示效果。交互板块，我们学习了具有交互效果的图形的实现方法，包括绘制动态图形（动画）的方法，以及实现 GUI 效果和时间处理效果的方法。最后，我们学习了从外部导入凸显加在到绘图区域的实现方法。绘制 3D 图形和地图的方法，以及结合前面学习的内容了解综合交叉的应用场景。还学习了使用 LaTeX 和 matplotlib 自带的 TeX 功能渲染文本内容的方法，以及用 matplotlib 书写数学表达式的方法和技巧。</p> <p>同时，项目组成员也深入学习和掌握了数据清洗、聚类方法理论、可视化方法。</p> <p>在数据清洗方面，项目组取得了一定进展。源数据为由逗号分隔符分割的纯文本文件，未标注数据信息。在对照已有论文核查出租车 GPS 数据格式后，确定文件内包含出租车 GPS 编号，时间戳，经度信息，纬度信息，行驶速度，行驶方向，载客状态共七种数据信息，删除了车牌号等敏感信息。确定数据格式后，针对每个数据文件进行了清洗，检测了时间信息是否合法，经度纬度信息是否正常（是否在上海市内），行驶速度是否异常等，根据对应标准进行了筛查，并清除所有非法数据。现已将所有 4816 个数据进行合法性分析，得到 2822 辆有效出租车信息，并以 csv 格式文件存储，标注好对应数据表头信息。</p> <p>在经历数据清洗后，项目组针对时间段将数据进行拆分，并提取出出租车上、下课地点，便于后续按照时间段对数据进行分析，得出城市功能场所激活顺序及不同时间段城市热点区域。</p> <p>在聚类分析方面，目前处于理论学习阶段，学习了 DBSCAN 聚类分析算法，计划于将来进行相关进展。</p>
<p><b>存在问题及拟采取的解决措施</b></p>	<p>由于数据总量较少，经过数据清洗后仅余 2822 辆出租车有效信息，我们计划从网站上下载数据以补充样本并丰富数据类型。</p> <p>针对 GPS 数据点与地图匹配问题，需要进一步学习有关论文，提出一套合适的解决方案。</p>

下阶段主要任务及进程安排	2021 年 1 月至 2 月进行对出租车数据进行聚类分析，如上下客热点，交通热点路线等；3 月进行可视化呈现和结果分析，并调整优化算法参数和方法；得出一套城市热点地区分析系、拥堵路段等分析系统。			
经费使用情况	使用日期	金额（元）	主要用途	合计（元）
	2020. 07. 08	52. 20	《智能聚类分析方法及其应用》书籍购买	236. 2
	2020. 07. 08	56. 60	《Python 数据可视化之matplotlib 实践》书籍购买	
	2020. 07. 08	66. 20	《Python 数据可视化之matplotlib 精进》书籍购买	
	2020. 07. 08	61. 20	《Python 数据分析与可视化从入门到精通》书籍购买	
指导教师意见（建议）	<p>综合评定：</p> <p style="text-align: right;">指导教师（签名）：</p> <p style="text-align: right;">年    月    日</p>			
学院专家评审意见	<p>评定等级（优秀、良好、中等、合格、不合格）</p> <p style="text-align: right;">专家组组长签字：</p> <p style="text-align: right;">年    月    日</p>			

学院意见	<div>主管领导（签字）：</div> <div>年 月 日</div>
------	---------------------------------------

说明： 1、要求用计算机打印，纸张为 A4 大小。  
2、中期检查报告纸质文档由学院保存。