

目录

1	文献概述	1
1.1	研究背景与问题	1
1.2	研究目标	1
1.3	数据与方法	2
1.4	主要贡献与意义	3
2	文献涉及的 GIS 课程知识点梳理	3
2.1	地理空间对象与要素建模	3
2.2	空间数据模型	4
2.3	空间数据采集与处理	5
2.4	空间数据库与索引优化	5
3	GIS 与计算机专业的结合思考	6
3.1	大数据处理	6
3.2	人工智能	6
3.3	软件工程	7
3.4	数据科学	8
4	课程建议	8
5	结语	9

1 文献概述

1.1 研究背景与问题

本次阅读的文献为《基于多尺度时空聚类的共享单车潮汐特征挖掘与需求预测研究》，该文献于 2022 年发表在《地球信息科学学报》期刊上 [1]，由姜涛等人撰写。

共享单车作为解决城市“最后一公里”出行问题的重要交通方式，在实际运营中面临着严重的潮汐现象。该现象不仅降低了用户体验、增加了运营成本，还导致城市公共空间被过度占用。目前政府和共享单车企业主要采用电子围栏停车站点来规范管理，但由于单个站点的单车流入流出具有随机性和不确定性，基于独立站点的调度难以实现有效管理，因此需要将围栏站点分组成簇，实施区域化管理。

1.2 研究目标

该研究的目标主要包括三个方面。

第一，构建基于时空约束的网络聚类算法：该算法综合考虑空间因素和时间因素作为聚类划分依据，将地理位置相近、用车模式相似的电子围栏站点聚合到同一簇中，实现多尺度群组划分。

第二，在完成站点聚类的基础上，通过对历史订单数据的时空分析，识别出具有明显潮汐特征的区域簇，深入分析潮汐现象的时空分布规律、演变特征以及与城市功能区的关联关系。

第三，基于多尺度时空聚类结果，构建针对不同区域簇的单车需求预测模型：该模型充分考虑时间特征、天气因素、空间位置以及历史需求量等多维度影响因素，采用机器学习方法实现对未来时段单车借还需求的准确预测。

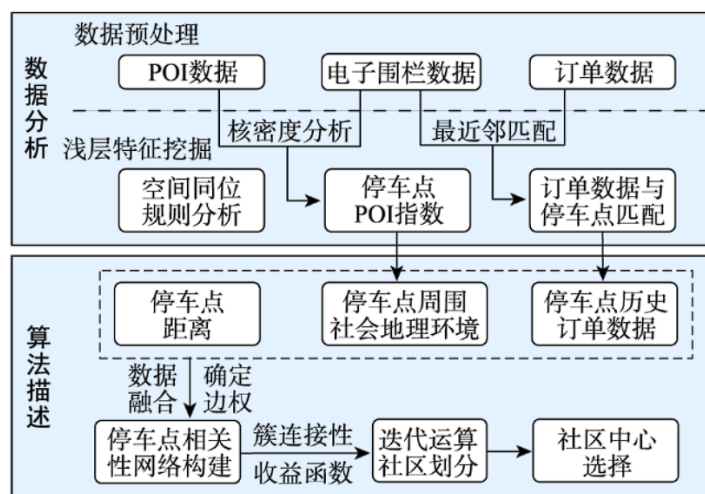


图 1: 技术路线 [1]

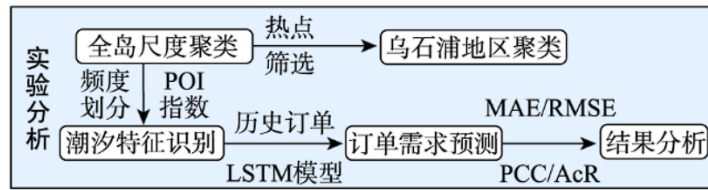


图 2: 技术路线 _ 实验分析 [1]

1.3 数据与方法

研究采用的数据主要包括共享单车订单数据、电子围栏站点数据、POI 兴趣点数据、气象数据和道路网络数据。为了进一步量化 POI 数据对各停车点单车使用的影响, 使用核密度分析方法来计算每个停车点周围各类 POI 对停车点所在位置的影响程度, 量化为停车点 POI 指数 [1]。论文中提到部分计算公示, 比如:

$$D = \frac{1}{(\text{radius})^2} \sum_{i=1}^l \left[\frac{3}{\pi} \left(1 - \left(\frac{\text{dist}_i}{\text{radius}} \right)^2 \right)^2 \right] \quad \text{for } \text{dist}_i < \text{radius}$$

订单数据包含用户的借车时间、还车时间、借车位置、还车位置等详细信息, 站点数据包括地理坐标、容量等属性, POI 数据用于分析站点周边的功能区类型, 气象数据包括温度、降水量、风速等要素, 道路网络数据用于计算站点间的网络距离和可达性。通过对这些多源异构数据的整合分析, 研究成功识别出研究区域内的主要潮汐站点和潮汐区域, 发现潮汐现象具有明显的时空规律性, 并且基于多尺度时空聚类的预测模型表现出良好的预测精度, 能够为共享单车的运营调度提供有力支持。

该文章使用到的数据清单如下图:

数据名称	数据时间	数据规模	数据描述	
			字段名称	字段含义
厦门岛共享单车订单数据	2020年12月21—25日6: 00 am—10: 00 am	58万条左右	BICYCLE_ID	加密后的单车ID号
			LATITUDE	纬度/°
			LONGITUDE	经度/°
			LOCK_STATUS	锁状态
			UPDATE_TIME	锁状态更新的时间
厦门岛共享单车电子围栏数据	2020年12月	1.4万个左右	FENCE_ID	电子围栏唯一编号
			FENCE_LOC	电子围栏位置坐标串
厦门岛POI数据	2021年1月	8000条左右	POI_TYPE	POI地物类别
			LATITUDE	纬度/°
			LONGITUDE	经度/°
厦门岛路网数据	2021年1月	8000条道路	Length	道路长度/m
			name	道路名称

图 3: 数据清单 [1]

1.4 主要贡献与意义

该研究的创新点主要体现在方法论和应用两个层面。

在方法论层面,研究将空间约束和时间约束有机结合,提出了适用于共享单车系统的多尺度时空聚类算法,突破了传统聚类方法仅考虑空间距离的局限性,能够更准确地捕捉共享单车出行的时空特征。

在应用层面,研究将聚类结果应用于潮汐特征识别和需求预测,构建了从数据采集、特征提取、模式识别到预测应用的完整技术链条。这项研究不仅具有重要的学术价值,更具有直接的实践意义,研究成果能够应用于共享单车的运营管理实践,通过准确识别潮汐区域和预测需求变化,优化车辆调度策略,提高车辆利用率,改善用户体验,同时减少对城市公共空间的占用,促进共享单车行业的可持续发展 [1]。

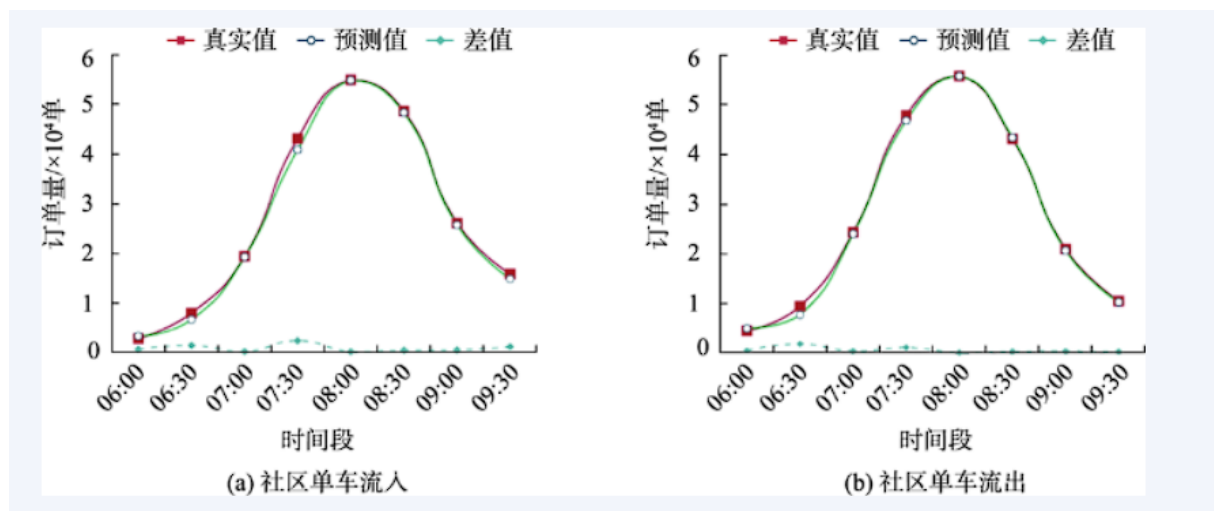


图 4: 全岛尺度下社区单车流入流出时序特征 [1]

2 文献涉及的 GIS 课程知识点梳理

2.1 地理空间对象与要素建模

在地理空间对象方面,研究涉及了多种类型的空间对象建模。电子围栏站点被抽象为点要素,每个站点对象包含几何属性如经纬度坐标、拓扑属性如与其他站点的空间邻近关系、描述属性如站点容量和设置时间,以及动态属性如实时车辆数量和历史借还车频次。POI 兴趣点同样作为点对象用于表征城市功能区,这些点对象的类型和密度反映了区域的功能属性。用户的借车点和还车点构成 OD 对,在空间上可表示为线要素,记录了共享单车的流动方向和流量。

通过时空聚类算法生成的区域簇则是面对象,每个簇包含若干个具有相似特征的站

点，是进行区域化管理和需求预测的基本空间单元。此外，共享单车的需求在空间上呈现连续变化特征，可以用需求强度场来表示，该场的数值反映了不同位置的单车需求密度。通过对这些不同类型地理空间对象的建模和分析，研究实现了对共享单车系统的全面刻画。

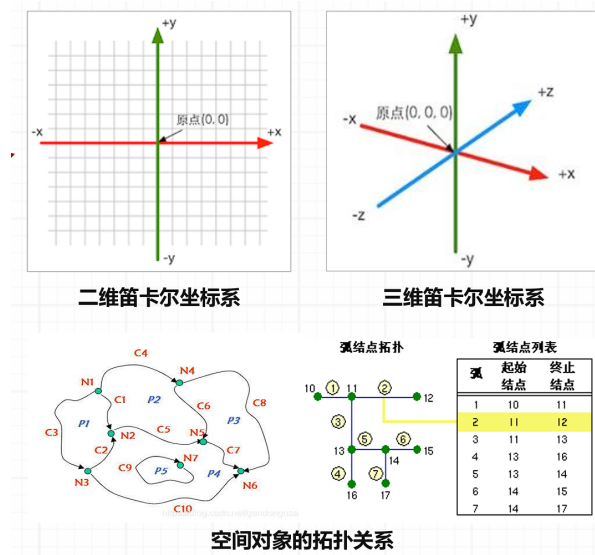


图 5: 空间特征

2.2 空间数据模型

该论文研究主要采用矢量数据模型来组织和管理相关空间数据。电子围栏站点数据采用点矢量模型存储，每个点记录站点的空间位置和属性信息，点矢量模型能够精确记录站点地理位置，便于进行空间查询、距离计算和邻近分析。道路网络数据采用线矢量模型表达，用于计算站点间的网络距离和可达性，OD 轨迹数据也可用线矢量模型表示流动方向和路径。研究区域边界、功能区划分、聚类生成的区域簇等采用面矢量模型，支持区域统计、空间叠加等分析操作。



图 6: 空间对象抽象为多种数据及其组合

更重要的是，考虑到共享单车数据的时空特性，研究实际上构建了时空数据模型。

通过快照模型将不同时间点的空间状态独立存储，便于进行时间断面分析和变化检测。通过事件模型将每次借车和还车行为记录为时空事件，包含时间戳、位置和用户属性，能够完整记录系统的动态变化过程。研究还隐含地使用了网络数据模型，将站点作为网络节点，OD 流作为网络边，流量大小作为边的权重，形成动态的时空网络，支持流量分析、社区发现和中心性计算等操作。

2.3 空间数据采集与处理

空间数据采集是研究的基础环节。共享单车订单数据通过移动互联网技术实时采集，每辆单车配备 GPS 定位模块和智能锁，能够自动记录精确到秒的借还车时间、GPS 坐标位置、匿名化用户 ID 和车辆状态等信息。这种移动感知数据采集方式具有自动化、高频率、大规模和精细化的特点，每天可产生数万至数十万条记录。

电子围栏站点数据通过高精度 GPS 设备实地测量或结合卫星影像进行数字化获取。研究还需要多种基础地理数据作为空间参考，数据质量控制是关键环节，需要进行异常值检测识别和剔除定位异常记录，对缺失值进行处理，将不同来源数据统一到同一坐标系，确保时间基准一致。原始数据还需要经过预处理，包括将 GPS 点位配准到最近的电子围栏站点，将连续时间转换为时间段，按站点和时间段统计借还车数量，以及将单车数据与 POI 数据、气象数据进行空间关联融合。

2.4 空间数据库与索引优化

空间数据库技术为研究提供了强大的数据管理和查询分析能力。虽然文献未明确说明具体数据库系统，但从研究涉及的数据规模和分析复杂度来看，必然使用了空间数据库技术。数据需要进行多层次组织，包括存储行政区划、道路网络等静态基础数据的基础图层，存储电子围栏站点、功能区划分等业务相关数据的业务图层，以及存储海量订单事件数据的动态数据层。研究涉及的典型空间数据库查询包括空间关系查询如查询站点周边一定范围内的 POI，空间聚合统计如统计每个区域簇内的总借车量，时空范围查询如查询特定时段和区域内的所有订单，以及空间连接操作如将订单数据与站点数据进行空间连接分配到最近站点。

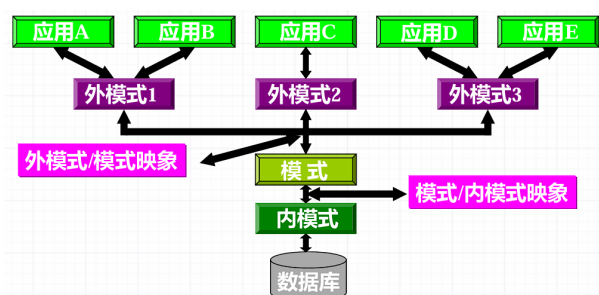


图 7: 数据库系统模式分级

空间分析是研究的核心内容,涉及多种空间分析方法的综合应用。空间聚类分析是研究的核心方法之一,研究创新性地提出了时空约束的网络聚类算法。该算法在空间约束方面计算站点间的距离和邻近性,考虑站点地理位置关系,利用 POI 数据分析站点周边功能属性。在时间约束方面分析站点的时序借还车模式,计算站点间需求模式的相似度,识别具有相似时间变化规律的站点。通过调整聚类参数可以实现多尺度分析,得到不同粒度的空间划分,既能进行宏观区域管理,也能实现精细化站点调度。研究还应用了空间统计分析方法,利用热点分析识别单车需求的热点区域和冷点区域,使用空间自相关分析需求的空间集聚特征,通过空间插值估算未设置站点区域的单车需求。缓冲区分析用于确定站点服务范围,创建以站点为中心的缓冲区代表步行可达范围,统计缓冲区内的人口和 POI 数量评估潜在需求,识别服务盲区,以及分析不同功能区对共享单车需求的辐射影响。

3 GIS 与计算机专业的结合思考

作为计算机科学与技术专业的学生,通过学习本文深刻认识到 GIS 与计算机科学具有天然的密切联系和广阔的融合前景。计算机科学为 GIS 提供了技术支撑和创新动力,而 GIS 为计算机科学提供了重要的应用场景和研究方向。该论文提出的多尺度时空聚类算法本质上是一个复杂的图算法问题,需要综合运用图论、聚类算法、相似度计算等计算机科学的核心知识。传统的聚类算法如 K-means、DBSCAN、层次聚类等在处理空间数据时需要考虑空间约束和拓扑关系,而加入时间维度后问题变得更加复杂,需要设计高效的时空相似度度量方法和聚类优化算法。这要求我们不仅掌握经典的机器学习和数据挖掘算法,还要理解空间数据的特殊性,能够针对具体应用场景进行算法改进和创新。

3.1 大数据处理

共享单车系统每天产生海量的时空轨迹数据,这些数据具有高维、动态、分布式的特点,对数据存储、管理和分析提出了挑战。计算机科学中的分布式计算技术如 Hadoop、Spark,流式处理技术如 Kafka、Flink,以及 NoSQL 数据库如 MongoDB、Cassandra 等(虽然这些大部分我都还不会),都可以应用于 GIS 海量时空数据的处理。同时,并行计算、GPU 加速等技术可以显著提升空间分析的效率,这需要我们具备扎实的系统架构设计和性能优化能力。

3.2 人工智能

本文的需求预测问题本质上是一个时空序列预测问题,这是机器学习和深度学习的重要研究方向。传统的统计学方法如 ARIMA、VAR 等在处理简单时间序列时表现良好,但对于复杂的时空依赖关系建模能力有限。近年来,深度学习技术为时空预测带来

了突破，循环神经网络 RNN 及其变体 LSTM、GRU 能够捕捉时间序列的长期依赖关系，卷积神经网络 CNN 可以提取空间特征，图神经网络 GNN 则能够建模不规则空间拓扑结构。特别是时空图卷积网络 ST-GCN、时空注意力网络等模型，在交通流量预测、共享单车需求预测等任务上取得了优异性能。这些前沿技术的应用需要我们深入理解深度学习原理，掌握 TensorFlow、PyTorch 等主流框架，具备模型设计、训练调优和部署的实践能力。此外，强化学习技术可以应用于共享单车的智能调度问题，通过构建马尔可夫决策过程建模调度策略，利用 Q-learning、策略梯度等算法学习最优调度策略。

3.3 软件工程

构建一个完整的共享单车智能管理系统需要综合运用软件工程的多个知识领域。系统架构设计需要采用微服务架构实现模块解耦，使用 RESTful API 实现服务间通信，利用消息队列实现异步处理，采用负载均衡和容器化技术保证系统的高可用性和可扩展性。前端开发需要使用 React、Vue 等现代框架构建用户界面，利用 Leaflet、Mapbox 等地图库实现空间数据的可视化，通过 WebSocket 实现实时数据推送。后端开发需要使用 Python、Java、Go 等语言实现业务逻辑，集成 PostGIS 等空间数据库，调用机器学习模型 API 进行预测。

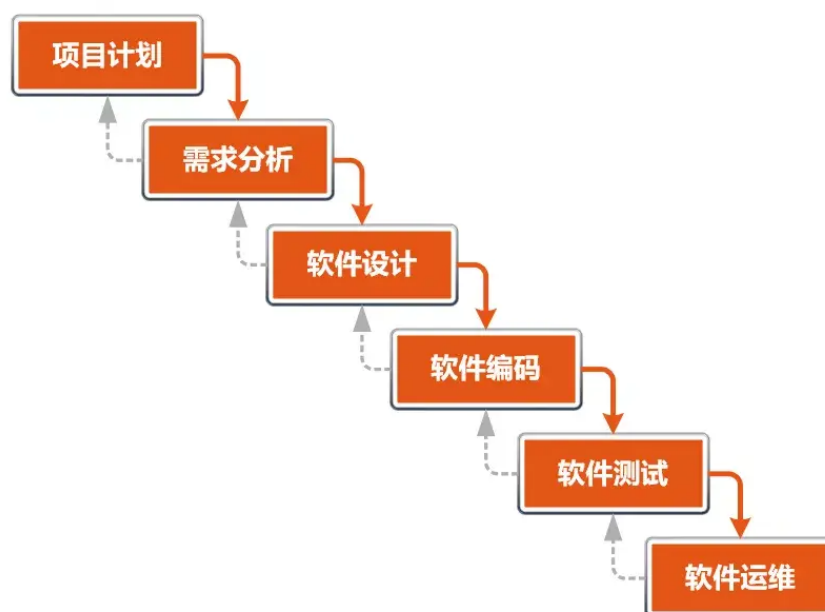


图 8: 软件设计模型

移动端开发需要利用 GPS 定位、推送通知等原生功能，实现用户端的借还车操作和导航服务。系统还需要考虑安全性设计如用户认证授权、数据加密传输、API 访问控制等，以及测试和运维如单元测试、集成测试、性能测试、日志监控、故障排查等。

3.4 数据科学

GIS 应用是数据科学的典型场景，需要完整的数据处理流程。数据采集阶段涉及爬虫技术获取 POI 数据、利用 API 接口获取气象数据、设计数据采集系统实时收集订单数据。数据清洗需要处理缺失值、异常值、重复值，进行数据格式转换和标准化。特征工程是提升模型性能的关键，需要从原始数据中提取有意义的特征，如从时间戳提取小时、星期、是否节假日等时间特征，从 GPS 坐标计算距离、方位角等空间特征，从历史数据统计平均值、方差、趋势等统计特征，从 POI 数据计算周边设施密度、功能区类型等环境特征。数据可视化是探索性数据分析和结果展示的重要手段，需要使用 Matplotlib、Plotly 等工具绘制时间序列图、热力图、流向图、三维地图等多种图表。

4 课程建议

老师讲课讲得很好，课程安排也很合理。前面先讲基础知识，后期再让学生做自己的 GIS 项目。我特别喜欢课程后期的小组展示。每个组都上去讲解自己对于 GIS 应用的项目，虽然大家的项目都看起来不那么完善，有点纸老虎的感觉，但是其实还是很有趣。我印象很深的是比如把地图的概念拓展到游戏地图的一个项目，还有一个徐州本地 GIS 项目，那位同学的 PPT 做得很好。我们组做了一个矿大美食地图的项目，如图9，虽然看起来非常不完善，很多功能都没能实现，比如说放在小程序上让大家都可以来分享美食等等，真的要做起来会非常难，也非常花时间。本以为会受到老师的批评，但是没想到老师很夸赞我们，讲到自己也希望有这样一个小程序，讲到老师自己的经历，听到这里真的心头一热。

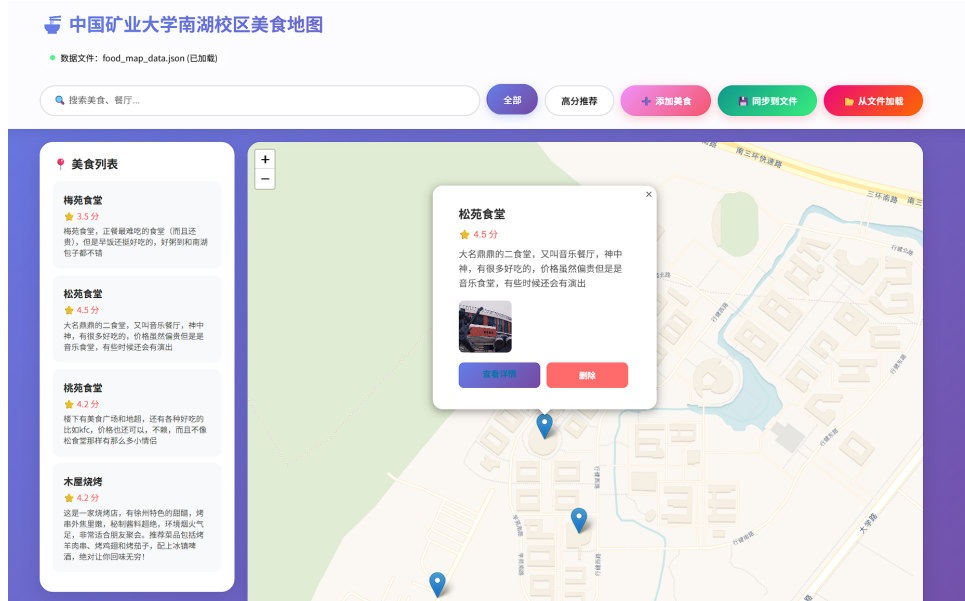


图 9: 矿大美食地图截图

非要说建议的话，大概是希望老师可以给我们讲一些别人做的案例，比如像我阅读的这篇共享单车的论文，通过动手操作和团队协作，能够更好地理解空间数据处理、分析和可视化的核心知识。同时，课程可以适当引入机器学习与大数据技术，帮助学生掌握 GIS 在现代智能城市中的应用。

总体而言，本门 GIS 课程为我们系统学习地理信息系统提供了良好平台，帮助我们掌握了 GIS 的基本概念、原理和方法，了解了其广泛的应用领域，培养了初步的空间思维能力。通过阅读分析这篇共享单车研究论文，我深刻认识到 GIS 不仅是一门技术，更是一种分析和解决问题的思维方式，GIS 在实际应用中需要与专业知识深度融合，掌握 GIS 技能对计算机专业学生的职业发展具有重要价值。

5 结语

通过对《基于多尺度时空聚类的共享单车潮汐特征挖掘与需求预测研究》一文的深入阅读和分析，我系统梳理了 GIS 课程核心知识点在实际研究中的应用，认识到 GIS 作为空间分析工具和思维方法的强大力量。共享单车潮汐现象是城市管理面临的实际问题，而 GIS 技术从地理空间对象建模、空间数据模型选择、多源数据采集整合、空间数据库管理，到多样化空间分析方法的综合应用，为问题解决提供了科学有效的手段，充分展现了其在智慧交通和智慧城市建设中的应用价值。

这次阅读也让我认识到自己在 GIS 知识和技能方面的不足，激励我在今后的学习中更加努力。我将继续深入学习 GIS 开发技术和空间算法，关注时空数据挖掘和深度学习的前沿进展，积极参与 GIS 相关的实践项目和竞赛，努力将 GIS 与计算机专业知识深度融合，培养跨学科思维 and 创新能力，为成为一名既懂空间分析又精通计算机技术的复合型人才而不断进步。

参考文献

- [1] 姜晓, 白璐斌, 楼夏寅, 李梅, and 刘晖. 基于多尺度时空聚类的共享单车潮汐特征挖掘与需求预测研究. 地球信息科学学报, 24(6):1047–1060, 2022.