

**本科毕业设计（论文）**

基于共享单车骑行数据的上海市通勤模式分析

**Commuting Patterns Unveiled through Shared Bicycle Data**

学 院：测绘与空间信息学院

专业班级：地理信息科学 2 班

姓 名：潘志清

学 号：202001020717

指导教师：牟乃夏

完成日期：待定

教务处制

**学位论文原创性声明**

本人呈交给山东科技大学的学位论文，除所列参考文献和世所公认的文献外，全部是本人攻读学位期间在导师指导下的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

若有不实之处，本人愿意承担相关法律责任。

本人签名： 日 期： 年 月 日

**学位论文使用授权声明**

本人完全了解山东科技大学有关保留、使用学位论文的规定，同意本人所撰写的学位论文的使用授权按照学校的管理规定处理。

作为申请学位的条件之一，学校有权保留学位论文并向国家有关部门或其指定机构送交论文的电子版和纸质版；有权将学位论文的全部或部分内容编入有关数据库发表，并可以以电子、网络及其他数字媒体形式公开出版；允许学校档案馆和图书馆保留学位论文的纸质版和电子版，可以使用影印、缩印或扫描等复制手段保存和汇编学位论文；为教学和科研目的，学校档案馆和图书馆可以将公开的学位论文作为资料在档案馆、图书馆等场所或在校园网上供校内师生阅读、浏览。

（保密的学位论文在解密后适用本授权）

作者签名： 导师签名：

日 期： 年 月 日 日 期： 年 月 日

**摘 要**

中文摘要应具有高度的概括性，语言精炼、明确，扼要叙述论文（设计）的主要内容，包括研究目的与意义、研究内容与方法以及研究结论等，同时需要突出论文（设计）的新论点、新见解或创造性成果。要求用中、英文分别书写，字数不少于400字。英文摘要内容应与中文摘要一致，语句通顺，语法正确，准确反映论文（设计）内容。

关键词是供检索用的主题词条，应采用能覆盖毕业论文（设计）主要内容的通用技术词条（参照相应的技术术语标准），可从标题或正文中选择3-5个最能表达主要内容的词语作为关键词，按词条的外延层次排列（外延大的排在前面）。关键词有中、英文对照，分别附于中、英文摘要后。

关键词：山东科技大学；本科生；毕业设计（论文）

**ABSTRACT**

In order to study……

………….......

**KEY WORDS:** undergraduate; graduation project (thesis); SDUST

**目 录**

(一级标题宋体4号加黑，二级及以下标题宋体小4，1.5倍行距)

从该部分开始，外语专业使用相应语种，格式要求与中文模板一致

**第1章 绪论 1**

1.1 二级题目 1

1.1.1 三级题目 3

**第2章 一级题目 4**

2.1 二级题目 5

2.1.1 三级题目 5

**参考文献 6**

**附录 7**

**致谢 8**

一级标题：中文黑体3号加黑，英文Times New Roman 小2加粗，单倍行距，段前0.8行，段后0.5行。

**第1章 一级标题**

1.1 二级题目

二级标题：黑体4号加黑，行距固定值20磅，段前0.5行，段后0.5行

……………………………………………………………………………………………………………………………………………………………………………………

三级标题：黑体小4号加黑，行距固定值20磅，段前0.5行，段后0.5行。

1.1.1 三级题目

正文内容…………………………………………………………………………

…………………………………………………………………………………………………………………………………………………………………………………………

正文内容：中文宋体小4号，英文Times New Roman 小4号，间距段前段后均为0行，行距固定值20磅，首行缩进2字符。

**公式示例：**

单颗粒球体在无限水体中等速下沉时，其沉速机理可看作对称绕流阻力与颗粒有效重力相平衡[13]，即

正文页码从正文开始至附录（正文-参考文献-附录-致谢）使用阿拉伯数字编排，小五号Times New Roman居中。

（γs-γ）\*π = CD πγ （1.1）

公式为单倍行距，居中，编号右齐。公式中符号字体与正文应一致。公式按章编号，例如 第1章中的第1个公式为 1.1

（建议公式用微软office的公式编辑器输入）

Stokes曾以粘滞性流体的一般性的运动方程式作基础，忽略惯性项的条件下推导出滞留区的阻力系数为

CD=24/Red （1.2）

**表示例：**

表3.1 各模型的计算条件

|  |  |  |
| --- | --- | --- |
| 类 型 | Ⅰ、单一介质 | Ⅱ、多介质 |
| 采用的模拟系统 | 多介质 | 多介质 |
| R１领域的弹性模量/(MPa) | 1000 | 1000 |
| R2领域的弹性模量/(MPa) | 1000 | 10000 |
| R１领域的泊松比 | 0.25 | 0.25 |
| R2领域的泊松比 | 0.25 | 0.25 |
| a/ (m) | 2.0 | 2.0 |
| b/ (m) | 5.0 | 5.0 |
| 内压P/ (MPa) | -1.0 | -1.0 |

**图示例：**



表题标在标上方，表题用五号宋体加黑，表格内中文用五号宋体，英文用五号Times New Roman字体。表按照章编号，例如第3章中的第1个表编号为表3.1。

图题写在图的下方居中，五号宋体加黑居中，英文用五号Times New Roman字体。图按照章编号，例如第3章的第1个图编号为图3.1

**图3.1 清水明渠水流下w2工况下的三维时均流速图**

**第2章 相关理论基础**

2.1 相关理论基础

2.1.1 行为地理学

行为地理学是人文地理学的一个重要分支[12]，该学科在 20 世纪初期诞生，是地理学、心理学、行为科学的交叉学科，在后来的发展中逐步融合行为主义地理学与时间地理学两大理论基础，形成了一套相对完善的研究范式与理论体系。该学科主要探讨人与地理环境之间的互动，强调“以人为本”，从微观人地关系的研究出发探索宏观时空规律。受到行为经济学中将非理性的经济主体作为微观研究的对象启发，行为地理学所研究的“人”是一种有限理性的个体，这些个体的行为虽然受到地理环境等的客观制约而呈现出一定规律性，但也会出于主观做出一些随机行为。个体可以依据属性汇总为群体、可以依据社交网络与空间活动范围汇总为社群，这些不同类别不同层次的群体同样被归为“人”这一研究对象。行为地理学所研究的“地”指的是一种抽象的、主客观相结合的行为空间，比如城市意向、认知空间等。行为地理学强调人地关系的统一性，人与地相互依赖、相互影响、相互塑造，这二者是不可分割的。

与传统的将研究区域视作静态的“块”不同，行为地理学的研究强调动态性，这种动态性体现在：1）其研究图景是一个具有复杂时空流动模式的网络：个体与个体之间、个体与环境之间均存在复杂的互动关系，甚至这些关系本身也相互制约相互影响，这种有机的网络模型所具有的动态性是传统的静态区域块模型所无法比拟的；2）研究视野综合了宏观与微观并可动态调整：通过汇总个体行为，总结区域规律来不断扩大研究视野。通过不断缩小研究视野，深挖“人”这一复杂个体全生命周期内在认知的流变规律，更好地关怀个体生命质量；3）数据及对应处理方法层面的动态性：移动互联网、GPS 和移动位置服务(location-based services，LBS)等技术提供实动态数据支撑，人工智能、大数据技术等先进的数据处理手段提供快速动态的分析能力。

随着时代的发展，“人本主义”日益得到社会各界的认同，着重研究人的日常行为及其与环境的交互关系的行为地理学也越发受到重视。行为地理学以其独特的研究视角，在可持续发展、社会公平、国土空间规划与社会治理等方面发挥出越来越大的作用。

2.1.2 城市职住平衡与过剩通勤理论

随着中国城镇化水平不断提高，城市交通拥堵问题日益凸显，其背后城市交通巨大供需失衡问题也逐渐被社会各界所认识。对于城市居民而言，住所与工作地点相距甚远往往会带来巨大的通勤压力，这也是城市交通拥堵的一个重要原因。研究并达成职住平衡可以有效缓解城市交通拥堵，职住平衡理论也可以为城市规划提供有用的指导。职住平衡理论（Jobs-Housing Balance）[13]最早由 Howard 在1902年出版的书籍《明日田园城市》中提出，他认为应当建设一种具有“自力性”（Self-contained）的社区，这种社区能够确保居民的就业需求与住房需求一并得到满足，最终由众多具有“自力性”的社区组成“田园城市”。这样的“田园城市”可以有效缓解交通拥堵及环境污染，减轻居民通勤压力，缩减中低收入家庭开支提升幸福感。

将多种城市功能理想化地集成在许多相互离散的小块土地上，这种理想化的城市模式在现实中往往难以实现。“职住分离”的现状是市场长期自主选择的结果，一方面企业追求产业聚集带来的规模效益，企业以零散的方式分布于小片地块会降低生产效率提高管理成本，另一方面，求职者更愿意为了高薪前往较远的工作地就职，随着城市公共交通系统的发展，通勤成本不断下降，这种职住距离还会被进一步拉远。另外，对于一个具体的家庭而言，求职并不是选择居住地唯一需求，医疗保健、子女求学及休闲放松等综合需求同样重要，尤其是优质的公共服务资源（重点学校、三甲医院等）往往成为一个家庭选择居住地的主要因素，单纯强调职住一体并不能很好地满足家庭的综合性需求。因此，需要客观看待城市职住分离的现状，着重研究城市通勤现状，过剩通勤可以作为衡量一个城市职住平衡的量化指标，较为客观地反映城市通勤现状。

过剩通勤（excess commuting）指的是实际通勤距离与理论通勤距离之间的差值，该理论最早在 1982 年由 Hamilton 首先提出，经过三十多年的发展逐步形成了一套较为完善的理论体系。在城市建模方面，该领域最早将城市建模为单中心模型，认为城市中心提供了所有的就业机会。后来随着计算机技术及地理信息系统的发展，有学者使用泰森多边形将城市划分为多个职住格网，更精细地建模城市职住空间分布情况。在理论通勤距离计算方面，主要计算方法包括：1）交通问题线性规划最优化函数（Transportation Problem in Linear Programming, TPLP）：采用线性规划的方法，以交通成本最小为优化目标，职住人数分布守恒为约束条件求解理论通勤距离；2）最大熵值法与蒙特卡罗仿真模型：最大熵值法是一种基于信息论的随机模拟方法，计算简单，不需要大量数据。蒙特卡罗模拟是一种基于随机抽样的统计方法，灵活性高，计算量大，结果受随机抽样影响；3）等比例匹配通勤：认为就业岗位分配与企业所占区域在区域中的比例相同。其中，TPLP 的其基本原理是将城市中的就业岗位和居住地视为节点，将通勤视为弧线，并以交通成本最小化为目标，求解所有居民的通勤路径和通勤距离。

其中， 为最小平均通勤距离， 为城市通勤总人口， 为城市居民总数， 为城市就业岗位总数， 为居民 与就业岗位 之间的通勤成本（一般为距离）， 为居民 与就业岗位 之间的通勤量（一般为人数）。

上述公式需要满足如下约束条件：

其中， 为居民 的出发地， 为就业岗位 的目的地。

居民实际通勤情况可以使用问卷调查、GPS 轨迹数据、公交刷卡数据等方式获取，通过计算实际通勤距离 结合理论通勤距离 可以计算出过剩通勤率 ：

2.2 相关研究技术

2.2.1 聚类分析

聚类分析指的是按照一定的分类规则（距离、相似性等）将一组数据划分为多个类簇的过程，是一种无监督的探索性数据分析（数据挖掘）方法，在大数据分析、计算机图形学、模式识别等专业领域应用广泛。聚类分析过程往往涉及许多算法与子任务迭代处理，且并非完全自动化，一般需要根据结果手动调整参数优化。

聚类分析的核心问题是聚类模型，典型的聚类模型可以分为以下几类：1）连接性模型：层次聚类是连接性模型的典型代表。它通过不断合并距离相近的数据对象来构建层次化的簇结构。该方法的优点是易于理解和实现，但对初始度量方法的选择比较敏感。2）质心模型：k均值算法是质心模型的代表，它通过迭代地移动簇的质心来使得簇内数据的距离平方和最小化，最终将数据划分到k个簇中。3）分布模型：使用统计分布对簇进行建模，例如期望最大化算法使用的多元正态分布。4）密度模型：DBSCAN和OPTICS算法是密度模型的代表，它们将数据空间中密度较高的区域定义为簇，并通过密度阈值或可达距离来识别簇中的数据点。5）基于图的模型：HCS聚类算法是基于图的模型的代表，它将数据点之间的相似性关系转换为图结构，并通过图的社区发现算法来识别簇中的数据点。6）神经模型：自组织映射（SOM）是神经模型的代表，它是一种无监督神经网络，将数据点映射到一个二维网格上，并通过相邻神经元的激活模式来识别簇中的数据点。

在上述聚类模型的基础上，可以进一步设计用于区分不同类别的规则，这些规则一般依赖一种相似性度量来划分，例如欧几里得距离、曼哈顿距离、闵可夫斯基距离、皮尔逊相似度、豪斯多夫距离等。根据分类规则对于分类算法的约束程度（是否允许重复分类、是否允许存在未分类目标等），又可以分为硬性聚类与模糊聚类（软聚类）。

2.2.2 时空轨迹聚类分析

时空轨迹建模为某一对象按时间顺序记录的空间位置序列，对其的聚类分析就是根据轨迹间的时空相似性度量划分为不同类簇的过程。时空轨迹数据具有一定的特殊性，首先，它是以时间顺序记录的移动对象位置序列，内含时间因素代表其具有动态性，同一个对象在不同时刻具有不同位置属性，这与一般聚类分析的静态数据（如带有诸多静态属性数据的人或物）具有较大差别。其次，它有明确的线状空间结构，相较于一般的（多维空间中）点状数据模型，轨迹数据本质上为连续的同质顺序点列，轨迹间的相似性度量无法直接套用一般聚类的概念为团块的聚类模型已有的度量方法，而是需设计与之适应的特殊算法。

考虑到时空轨迹的特殊性，无法直接套用一般的聚类方法，而是需要根据轨迹的时空间结构设计特定的相似性度量算法，部分典型的时空轨迹相似性度量算法：1）轨迹间欧氏距离：将待分类轨迹视作随时间变动的点集，在每一个时间断面计算轨迹点间的欧氏距离并在全局时间轴方向上累积。对噪声敏感，需要采样频率一至，否则会有较大误差。2）最小外包络距离：首先使用轨迹特征点（方向快速变化的点）切分轨迹，然后提取这些子段的外包络矩形并组成顺序序列作为轨迹的概化表征，最后累计矩形间的距离以度量轨迹间的相似性。该方法对原始轨迹平滑去噪，但也会丢失一部分信息。3）子轨迹相似性度量[11]（TRACLUS）：该算法首先根据最小描述长度原则将轨迹划分为多个子轨迹，然后计算子轨迹间的三种距离（垂直距离、平行距离、角度距离）并加权计算的综合相似性量值。4）历史最近距离：在一段给定的时间范围内，检查各个时刻两个轨迹之间的距离（欧氏距离或其他），选取最小距离作为历史最近距离。

2.2.3 用户行为分析

用户行为分析（User behavior analytics）是一种通过分析用户使用数据来了解用户行为模式的方法。企业可以收集用户点击次数、登录时间、浏览时长等数据，并通过数据挖掘、模式分析等手段进行分析。用户行为分析可以帮助企业在不打扰用户的情况下（相较于发放用户问卷）更好地了解用户群体，从而提升服务质量。此外，高效及时的用户行为分析还可以防范恶意攻击（例如DDoS攻击），过滤异常账号（刷单诈骗）。一些常用的用户数据挖掘方法包括统计用户使用频率并划分用户等级、计算并绘制用户贡献累计分布曲线（CDF曲线）、计算用户留存率等。其中，用户留存率（User Retention Rate）是一个用于衡量产品吸引力及用户粘性的重要指标，它指的是在某一时间段内开始使用产品或服务的用户，经过一段时间后仍然继续使用该产品或服务的用户的比例。

夏琼燕等人[14]下载并研究了2005年到2007年 Open Street Map （OSM）开源地理信息网站中志愿者用户数据（记录了用户的贡献行为），通过计算用户贡献累计分布曲线，他们发现该网站的贡献者符合“二八定律”，即该网站两成的贡献者贡献了约80%的数据，并根据贡献率将用户分为多个活跃等级。通过合理选取时间窗口，计算用户留存率，结合用户活跃度分类，他们发现OSM社区中低活跃度的用户留存率偏低，而高活跃度用户总体保持稳定，社区稳定存续主要依靠中高活跃度用户。

2.3 本章小结

本章以行为地理学、城市职住平衡理论和过剩通勤理论为理论基础，并探讨聚类分析、用户行为分析和时空轨迹聚类分析等相关研究技术。行为地理学强调以人为本，探讨人与地理环境的互动关系，并在可持续发展和社会公平等领域发挥重要作用。城市职住平衡理论指出有效缓解城市交通拥堵的关键是合理规划就业与住房地点，而过剩通勤理论提供了衡量城市通勤状况的量化指标。在相关研究技术方面，聚类分析和用户行为分析为数据挖掘和模式分析提供了重要手段，特别是时空轨迹聚类分析针对轨迹数据的特殊性提出了一系列相似性度量算法。这些理论和技术的研究与应用有助于更好地理解和解决城市发展中的诸多问题，促进城市可持续发展和社会治理的进步。

**第3章 一级标题**

3.1 研究范围

本文选取上海市作为研究区域。上海市北界长江，南枕杭州湾，西接江浙两省，东临东海，全市面积 6340.5 平方公里，现辖16个市辖区：黄浦区、徐汇区、长宁区、静安区、普陀区、虹口区、杨浦区、闵行区、宝山区、嘉定区、浦东新、金山区、松江区、青浦区、奉贤区、崇明区。参考第七次全国人口普查主要数据，上海市常住人口为 24870895 人。近几十年来经济飞速发展的同时，上海市逐步形成了多中心、高人口密度及高混杂度的城市格局，这也进一步引起人口拥挤、交通拥堵和环境污染等大城市病。这些城市特征也同时存在于中国大多数大型、超大型城市中，选取作为中国经济中心的上海为研究区域具有一定的典型意及泛化意义。

3.2 数据来源与预处理

本文的研究数据包括骑行轨迹数据与上海市基本地理信息数据，其中骑行轨迹包括由互联网采集整理得到的 2016 年 8 月整月 102361 条上海市某品牌共享单车骑行轨迹数据及 2020 年 8 月整月 12793 条骑行记录，上海市基本地理信息数据包括由 Open Street Map 开源地理信息数据网站获取上海市路网矢与兴趣点分布矢量数据及由高德地图开放平台获取的上海市矢量边界数据。

2016 年骑行记录数据较为详细，除了常见的骑行起止时间及起止点位置字段外，还包含了一个以“#”分隔的骑行轨迹点集字段，但是由于网络延迟、设备故障等原因该字段内的轨迹点为乱序排列，需要设计算法以恢复正确的骑行轨迹（具体方法见 3.2.2），并且骑行实际距离也需要根据正确的轨迹数据计算（计算方法见 3.2.4）。2020 年骑行记录并未包含详细的骑行轨迹点集，但提供骑行距离。对于骑行记录数据，按照如下规则清洗：1）去除骑行时长小于1分钟大于 8 小时的骑行记录；2）对于有轨迹点集的数据，去除记录到的轨迹点数量小于 3 个的记录；3）去除存在字段缺失及异常符号的记录；4）去除骑行终点位于上海市之外的记录，该记录超出研究区域。

3.2.1 数据来源与数据清洗

正文

3.2.2 轨迹重排序及长度量测轨迹简化

正文

3.2.3 轨迹简化

正文

3.3 研究方法

正文

3.3.1 轨迹终点最近邻匹配算法

正文

3.3.2 改进的 TRACLUS 算法

正文

3.4 本章小结

正文

**第4章 一级标题**

4.1 二级题目

正文

4.1.1 三级题目

正文

4.2 二级题目

正文

4.2.1 三级题目

正文

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.

[12]塔娜,柴彦威.行为地理学的学科定位与前沿方向[J].地理科学进展,2022,41(01):1-15.

[13]杨超,汪超.城市过剩通勤与职住平衡模型[J].同济大学学报(自然科学版),2013,41(11):1712-1716.

[14]夏琼燕,罗冠,张翔,等.OpenStreetMap志愿者贡献与留存分析[J].测绘与空间地理信息,2021,44(02):90-93+97.

附录1

附录是对于一些不宜放在正文中，但有参考价值的内容，可编入毕业设计的附录中，例如重要数据、表格、公式、图纸、程序等。附录的篇幅一般不要超过正文。

附录编号依次编为附录 1、附录 2。附录标题各占一行，按 一级标题编排。每一个附录一般应另起一页编排，如果有多个较短的附录，也可接排。

标题：黑体小2加黑居中，单倍行距，段前0.5行，段后0行

内容：宋体小4号，行距固定值20磅，英文用Times New Roman，小4号

致谢

致谢是对在毕业论文（设计）工作中给予各类资助、指导、协助以及提供各种有利条件的单位、指导教师或其他人员表示感谢，语言应实事求是，切忌浮夸之词。

ХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХХ

标题：黑体小2加粗居中，单倍行距，段前0.5行，段后0行

内容：宋体小4号，行距固定值20磅；

英文用Times New Roman，小4号