**深圳市出租车出行特征与需求模式分析**

**第一章 绪论**

**1.1 研究背景及意义**

21世纪以来，城市化进程稳步推进，人口高度聚集导致的出行需求增加与规划混乱、效率低下城市交通网络之间的矛盾愈演愈烈。作为城市综合交通运输体系的重要组成部分，出租车具有灵活、便捷、舒适、全天候的特点，其运营状态直接关系到市民的出行便利性和交通拥堵情况。

而随着科学技术的发展，车载GPS设备等新兴技术被广泛应用在交通领域的各个方面，由此产生了大量的出行数据，而出租车GPS数据就是其中的一员。出租车GPS数据作为出租车运营管理的重要信息来源，为深入了解出租车运行状况、分析出行特征、优化调度策略提供了重要的依据。

综上所述，本研究基于出租车GPS数据，进行出租车出行特征与需求模式的分析，旨在挖掘市民的出行规律。研究过程中结合栅格化方法和奇异值分解（SVD），提出了一套完整的出行特征与需求模式分析流程，具有丰富的理论意义。

而本研究的现实应用价值包括如下方面：从出租车运营公司的角度来看，分析出租车出行特征和出租车需求模式，能够了解乘客的出行规律，识别出行热点地区，促进出租车运力资源的合理分配；从司机与乘客的角度来看，分析出租车出行特征和出租车需求模式，有利于司机了解出行需求的变化并预知出行高峰的到来，降低空载率，缩短乘客等待时间，为乘客提供优质服务，改善城市人居出行质量；从交通管理部门的角度来看，有助于揭示市民出行行为规律，掌握市民出行特征，为城市交通规划和运输管理提供科学依据，为其制定交通管理方法提出建议，从而提高城市交通系统运行效率。

**1.2 国内外研究现状**

本研究主要挖掘出租车出行特征和出租车需求模式。事实上包括出租车、共享单车、轨道交通和公共汽车等在内的公共交通工具的出行方式虽各有不同，但它们在出行特征、热点识别、需求模式、需求预测等方面的研究方法是如出一辙的。接下来，本文将阐述国内外不同交通工具研究方法现状，为本文接下来的研究提供参考。

聚类算法是出租车出行热点区域识别、需求预测的重要手段，包括基于密度的聚类算法如DBSCAN聚类算法、基于划分的聚类算法如K-means聚类算法等。胡兰兰[1]使用K-means聚类算法通过对出租车每小时平均收益进行聚类以获取每小时高收益热点区域中心点；Zhao S[2]等人利用一种基于自组织映射和K均值融合算法的聚类模型提高了K-means聚类算法的效率；Loh W K[3]等人提出了CudaSCAN聚类算法借助GPU加速用以解决DBSCAN聚类算法由于查找相邻对象时需要重复扫描与所有对象的距离且聚类过程中保留中间结果而造成的效率低下问题；Shen Y[4]等人使用基于改进的DBSCAN算法提取出行OD热点，并基于密度的距离轨迹聚类算法，以识别具有乘客较多的出行路径；秦昆[5]等人使用基于时空数据场的聚类算法从出租车轨迹数据中提取武汉市出租车热点区域，基于复杂网络理论与方法分析热点区域之间的空间交互作用。

除聚类算法之外，传统机器学习也是研究相关领域的重要方法。Tong Y[6]等人使用海量特征数据集来分析出租车需求模式的影响因素，并借助线性回归模型预测不同时段的出行需求；李岩[7]应用岭回归预测模型、随机森林回归预测模型以及基于这两个模型权重的组合预测模型，对西安市出租车需求进行了预测，并得出模型的性能和研究区域以及评估指标均有关系的结论；Han G[8]提出了一种基于连续隐马尔可夫模型的无监督机器学习方法，用于精确识别不包括出行目的的数据集中每一条OD序列的出行目的。

算力提升使得深度学习逐渐成为研究公共交通出行方面的热点方法。Xiaojian G[9]等人首次使用BP神经网络预测十字路口交通流量，通过构建多层感知机进行预测；Pang J[10]等人使用RNN探究时间步长之间的依赖关系以预测公交车到达时间；Zhang J[11]等人提出了一种基于DNN的时空数据预测模型，并构建了一个名为UrbanFlow的实时人流预测系统，在不同的数据集上效果显著；Liu T[12]等人提出一种卷积神经网络模型，利用局部卷积层和门控单元，简化出租车需求模式多视图时空特征提取过程，便于短时出租车出行需求预测；为降低出租车空载率、减少司机与乘客的等待时间，Xu J[13]等人提出使用长短期记忆神经网络预测各交通小区出行需求量；谢光明[14]使用深圳市福田区的订单数据，基于GAT网络改进的ANS-GLN时空神经网络模型，进行了共享单车社区的构建，并对共享单车的流量进行了预测。

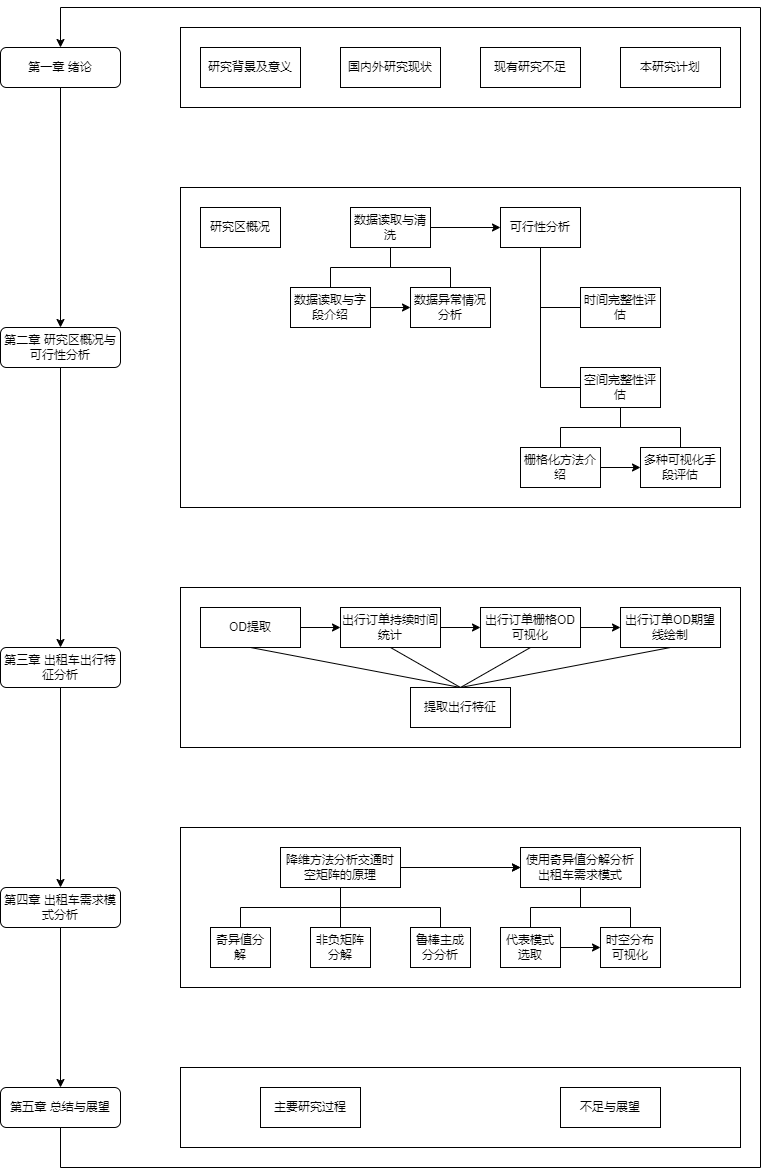
此外，还有许多学者使用了其他研究方法进行分析。徐婷立[15]等人结合多源大数据使用地理加权回归模型（GWR）研究了杭州市共享单车分布的影响机制，并得出了其拟合性能优于普通最小二乘法模型（OLS）的结论；王李轩[16]基于核密度分析方法（KDE）和地理加权回归模型（GWR）研究了西安市城市出租车的载客出行特征，并证明地理加权回归模型优于半对数回归模型与线性模型；孙启鹏[17]等人使用非负矩阵分解算法（NMF）从时空角度挖掘北京市共享单车的工作日骑行数据中的出行规律，并构建基于非负矩阵分解算法的BP神经网络预测模型；高枫[18]等人引入地理探测器，探究广州市主城区中共享单车的骑行目的地的时空特征以及目的地分布影响因素的时间差异；付鑫[19]等人通过构建出租车载客出行行为OD矩阵，提出了基于出租车出行的通勤客流识别模型，并建立了通勤距离和时长的计算模型，用于通勤行为的时空特征分析。

**1.3 现有研究不足**

现有公共交通出行特征与需求模式的研究方法聚焦于聚类、机器学习或深度学习，少部分使用空间统计模型，降维方法的应用较少。降维方法常用于分析高纬度数据，包括主成分分析（PCA）、奇异值分解（SVD）、非负矩阵分解（NMF）、鲁棒主成分分析（RPCA）等，相比于优秀聚类方法的复杂性以及深度学习对算力的要求，降维方法理解简单、计算量小、结果直观。本文尝试阐述降维方法分析交通时空矩阵的原理，并结合栅格化方法使用SVD分析出租车出行模式，为使用降维方法分析其他公共交通工具提供一套完整的流程。

**1.4 本研究计划**

本研究计划路线图如图1.1所示。



**图1.1 研究计划图**

**第二章 研究区概况与可行性分析**

**2.1 研究区概况**

深圳市地处中国广东省南部、珠江口东岸。截至2022年末，全市下辖9个区，常住人口1766.18万人。城市交通方面以深圳市交通运输局发布的数据[20]为例：2023年全市公共交通客运量38.15亿人次，同比增长39.18%，日均1045.09万人次，其中巡游出租车客运量2.82亿人次，同比增长14.62%。庞大的出行需求使得深圳市出租车出行特征与需求模式分析具有非常强的意义。

**2.2 出租车GPS数据的读取与清洗**

**2.2.1 出租车数据的读取**

通过出租车GPS设备以一定的采样频率追踪记录并保存出租车的GPS地理坐标位置而产生的出租车GPS数据一般包括时间、空间与个体三个维度的信息，采样频率一般为15s一条记录，单个城市的出租车GPS数据产出速率一般为2GB每日，数据量大小取决于运营中的出租车数量与车载GPS设备的采样频率，数据包括经度、纬度、时间、载客状态、车辆ID、方向角、速度、是否在快速路上等字段。

用Python读取深圳市交通运输局提供的一天时间内抽样500辆的出租车GPS数据，各字段含义如表2.1所示。

**表2.1 出租车GPS数据示例**

|  |  |
| --- | --- |
| 字段 | 示例 |
| 车牌号（VehicleNum） | 34745 |
| 时间（Stime） | 20:27:43 |
| 经度（Lng） | 113.806847 |
| 纬度（Lat） | 22.623249 |
| 载客状态（OpenStatus） | 1 |
| 车速（Speed） | 27 |

**2.2.2 异常数据的清洗**

异常数据体现在载客状态字段上。

载客状态字段通过1和0表示出租车载客状态：字段值为1时出租车载客；字段值为0时出租车空载。将数据按照车牌号和时间排序，正常情况下一辆出租车的载客状态序列是连续0序列和连续1序列交替出现，如表2.2所示。当载客状态字段值由A时刻的0转为B时刻的1时，代表乘客在A时刻与B时刻之间上车。当载客状态字段值由C时刻的1转为D时刻的0时，代表乘客在C时刻与D时刻之间下车。由于GPS设备的采样频率一般为15s一条记录，可以认为相邻采样点在时间和空间上都非常接近，A和B两个时刻均为上车点，C和D两个时刻均为下车点，本研究设A时刻为上车点，C时刻为下车点。

**表2.2 正常情况下一辆出租车的载客状态**

|  |  |  |
| --- | --- | --- |
|  | 载客状态（OpenStatus） |  |
|  | 0 |  |
|  | 0 |  |
| A时刻 | 0 |  |
| B时刻 | 1 | 此时乘客上车 |
|  | 1 |  |
|  | 1 |  |
| C时刻 | 1 |  |
| D时刻 | 0 | 此时乘客下车 |
|  | 0 |  |

载客状态字段的异常情况如表2.3和2.4所示。

情况一为连续的0序列中突然出现一个1。该异常数据按照正确情况读取会表现为乘客的上下车记录连续，后续出行特征分析中会表现为乘客在同一地点上下车。

情况二为连续的1序列中突然出现一个0。不排除一些出租车由于口碑、系统安排等原因生意火爆，上一个订单结束后就立即在上一个订单的下车点开始下一个订单。但考虑到出租车GPS设备的采样频率一般为15s一条记录，在如此短的时间内完成上一个订单的结束与下一个订单的开始较难，因此本研究设该情况为小概率事件。

**表2.3 异常情况一**

|  |  |
| --- | --- |
| 载客状态（OpenStatus） |  |
| 0 |  |
| 0 |  |
| 0 |  |
| 1 | 异常 |
| 0 |  |
| 0 |  |
| 0 |  |

**表2.4 异常情况二**

|  |  |
| --- | --- |
| 载客状态（OpenStatus） |  |
| 1 |  |
| 1 |  |
| 1 |  |
| 0 | 异常 |
| 1 |  |
| 1 |  |
| 1 |  |

载客状态字段异常的原因大概率是出租车GPS设备的突发异常，使用Python找出符合下述特征的数据删除即可：某条数据的载客状态字段值与前一条数据和后一条数据的载客状态字段值均不同且三条数据属于同一辆出租车。

**2.3 可行性分析**

可行性分析可以确保分析的有效性，包括数据基本特征、数据完整性与准确性分析。

数据基本特征以出租车GPS设备采样频率为例：如果采样频率为1s一条，可以分析出租车的运行状况；如果采样频率为15s一条，可以分析出租车的出行特征、需求模式；如果采样频率为1h一条，可以分析出租车的热点分布。

数据完整性以出租车GPS数据为例：在个体维度上，数据是抽样数据还是全量数据，数据来源是单源还是多源；在时间维度上，某个时段内缺失数据是正常现象还是其他原因引起的；在空间维度上，某个区域内缺失数据丢失是正常现象还是其他原因引起的。此外还要考虑缺失占比、能否补全、对研究的影响等情况

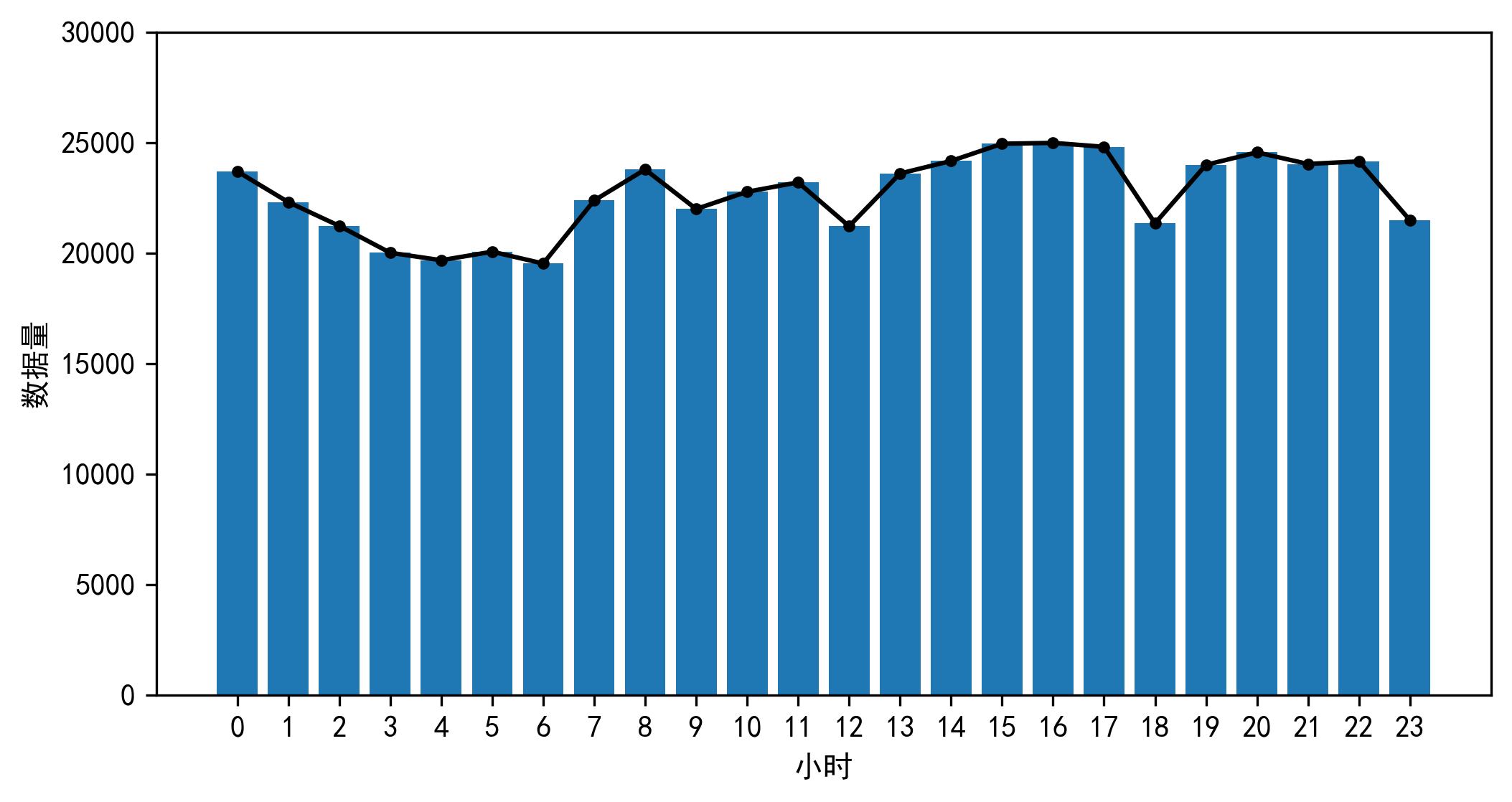
数据准确性以出租车GPS数据为例：轨迹是否超出研究区范围；轨迹是否偏离道路：车速是否异常。

本研究通过时间与空间完整性评估对出租车GPS数据进行可行性分析。

**2.3.1 出租车GPS数据的时间完整性评估**

本研究通过可视化出租车GPS数据的时间分布来评估出租车GPS数据的空间完整性。

本研究从时间字段中提取出小时信息，统计出租车GPS数据一天内各小时数据量，统计结果如图2.1所示，出租车GPS数据的小时分布完整，无异常缺失，并呈现出白天多、夜晚少、有早晚高峰期等规律。综上所述，出租车GPS数据的时间完整性较好。



**图2.1 出租车GPS数据每小时数据量统计图**

**2.3.2 出租车GPS数据的空间完整性评估**

本研究通过可视化出租车GPS数据的空间分布来评估出租车GPS数据的空间完整性。

出租车GPS数据的经纬度精度达到了小数点后6位，两条数据即使产生于相近地点，它们的经纬度也不同，因此不能直接使用经纬度统计空间分布，它在所有位置统计的点数量都是1。因此本研究需要通过特定的地理空间统计单元进行空间统计，本研究采用栅格化方法贯穿全程。

栅格化方法即用栅格覆盖研究区，判断矢量图形与栅格的空间关系，当矢量图形非点要素时必然出现多个矢量图形的边界位于同一个栅格，本研究设栅格中心点所属的矢量图形为栅格所属的矢量图形（也可以使用空间插值方法进行判断）。栅格化方法的优势在于空间分析的最小单元是大小形状一样的栅格，便于单元之间属性比较。栅格化方法的劣势在于必然造成精度下降，降低精度损失需要减小栅格大小。特别的，由于GeoPandas中要求输入为矢量数据，本研究的栅格以矢量数据存储，即用多个大小形状一样的矢量正方形表示栅格，后续所提到的栅格如无特殊说明，默认为以矢量数据存储的栅格。

栅格化方法要考虑两个问题：如何依据研究区范围和栅格大小生成栅格；如何将出租车GPS数据点高效对应到栅格上。

对于第一个问题，本研究先基于研究区范围中心点和栅格大小计算每个栅格对应的经度和纬度，再通过每个栅格的编号推算出四个顶点坐标，然后生成每个栅格对应的矢量格子。将地球近似为正球体（实际上地球为椭球体，两两栅格之间的应不同），设地球半径为，研究区范围左下角经纬度为，右上角经纬度为，则中心点坐标如式2.1所示：

设经线的周长为，对应经度的360°，纬线的周长为，对应纬度的360°，栅格长度为，则栅格对应的经度和纬度如式2.2和2.3所示：

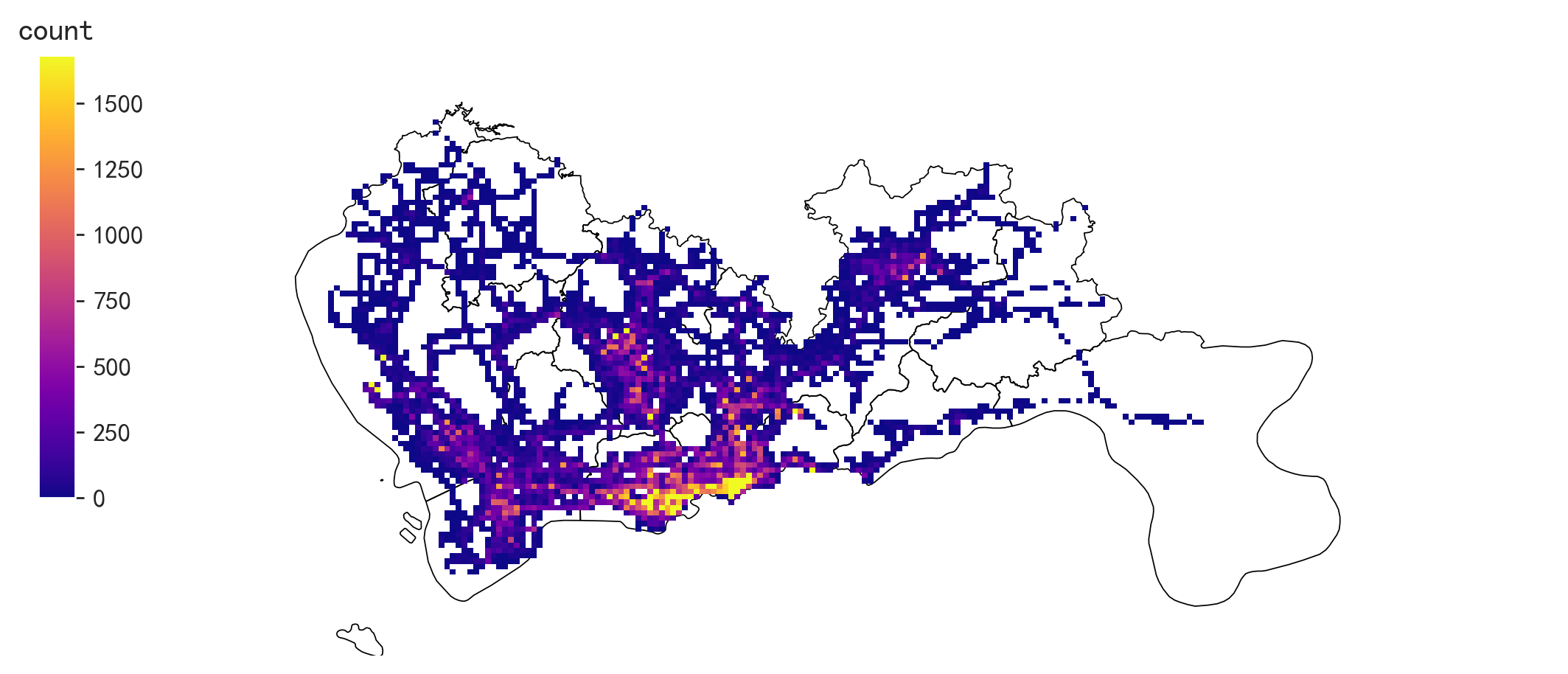
然后以研究区范围左下角作为第一个栅格的中心点，从该点开始将栅格覆盖研究区范围，设栅格的编号为，行列号从0开始，设每个栅格的中心点坐标为，其求解如式2.4和2.5所示：

对中心点坐标上加下减，左加右减求解顶点坐标从数学角度来看是正确的，但由于编程语言小数计算的精度问题，会出现式2.6和2.7的情况，导致栅格之间无法贴齐，对此本研究只计算每个栅格的左下角顶点，剩下的三个顶点使用周围三个栅格的计算结果。

对于单个栅格而言，从左下角开始按逆时针方向的4个顶点坐标计算公式如式2.8~2.11所示。顶点计算完成后循环生成栅格即可。

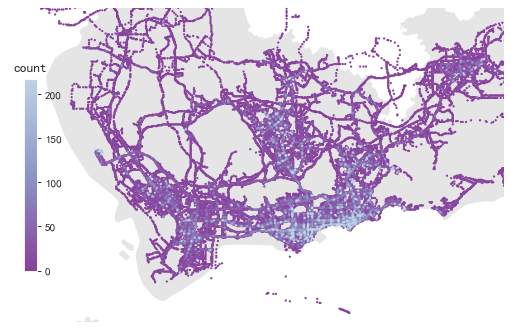
对于第二个问题，由于数据均以经纬度存储且数量级大，如果在第一个问题中使用如ArcMap 10.8中的“创建渔网”工具创建好栅格后，在第二个问题中对每个数据点都用复杂的空间连接匹配到栅格上，效率低，因此本研究通过Python的向量运算加速使用公式将数据点对应至栅格这一过程，效率高。给定研究区范围内的任一数据点的经纬度，只需要计算和距离研究区最左下角的栅格的左下顶点和分别经过了个经度和个纬度即可，计算公式如式2.12和2.13所示：

可视化结果如图2.2所示（本研究使用的栅格大小为）。



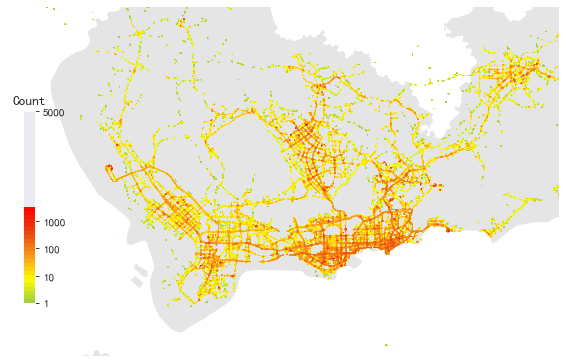
**图2.2 出租车GPS数据空间分布栅格图**

只考虑评估空间完整性，散点图的精度比栅格化方法高，经纬度保留三位小数的绘制结果如图2.3所示。



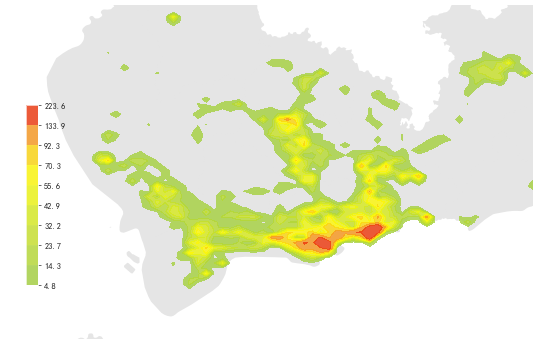
**图2.3 出租车GPS数据空间分布散点图**

热力图也可以评估空间完整性。本研究借助Matplotlib绘制出租车GPS数据空间分布等高线图，通过颜色值展示数据量密度，具有现实物理意义，绘制结果如图2.4所示。



**图2.4 出租车GPS数据空间分布等高线图**

热力图还可以表达出租车GPS数据空间分布的二维核密度分布，赋予每一个数据点一定的影响范围，通过离散数据估计连续概率密度并平滑数据分布，结果为抽象的密度值，无现实意义，使用Seaborn的绘制结果如图2.5所示。



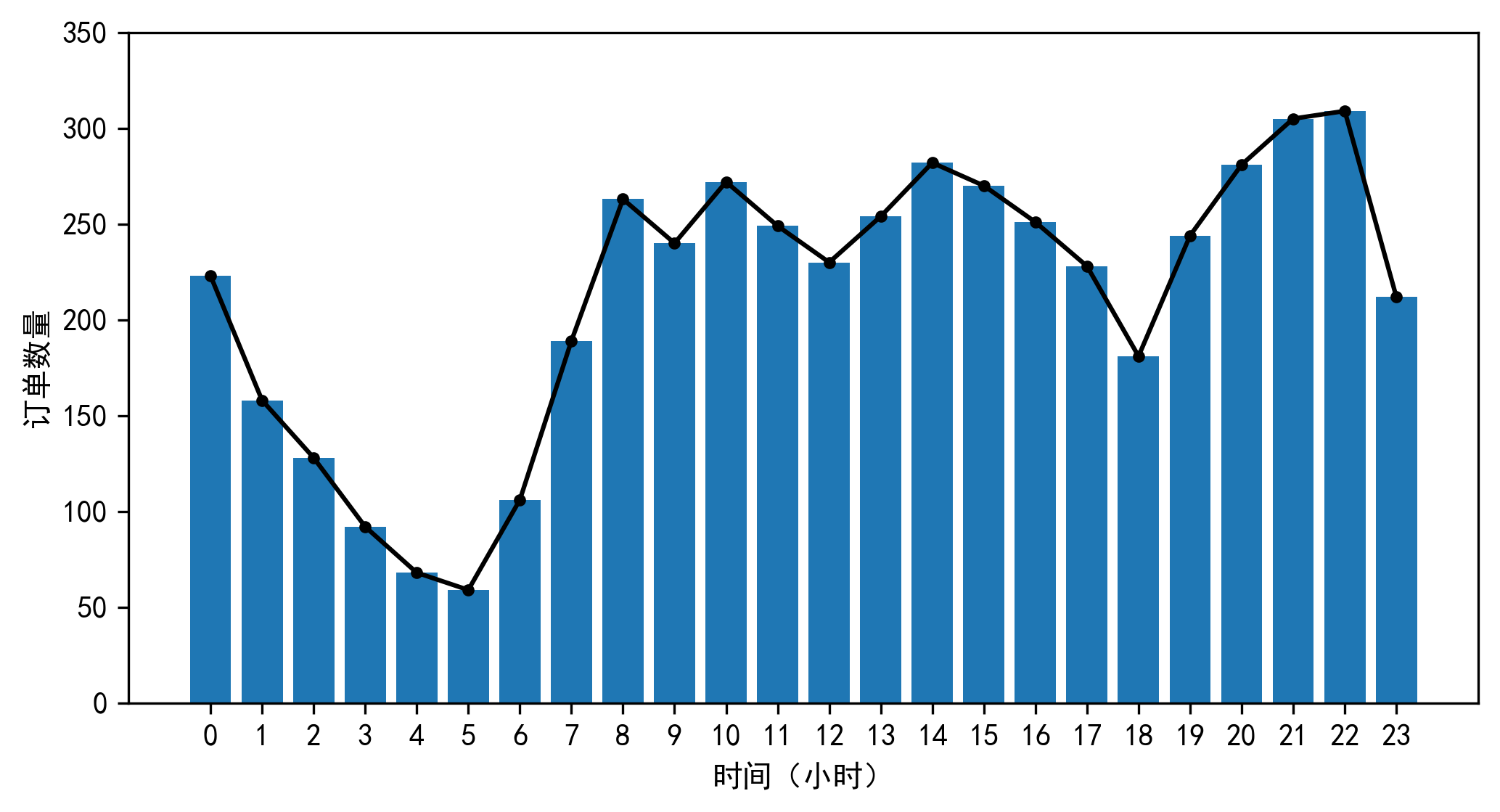
**图2.5 出租车GPS数据空间分布的二维核密度分布图**

四种可视化手段表明数据集中分布于深圳市中心城区，郊区大部分区域无数据，有数据区域中数据的空间分布呈线形形态，判断为与道路分布吻合，是郊区出租车需求小的体现。综上所述，出租车GPS数据的空间完整性较好。

**第三章 出租车出行特征分析**

**3.1 出租车出行订单的OD提取**

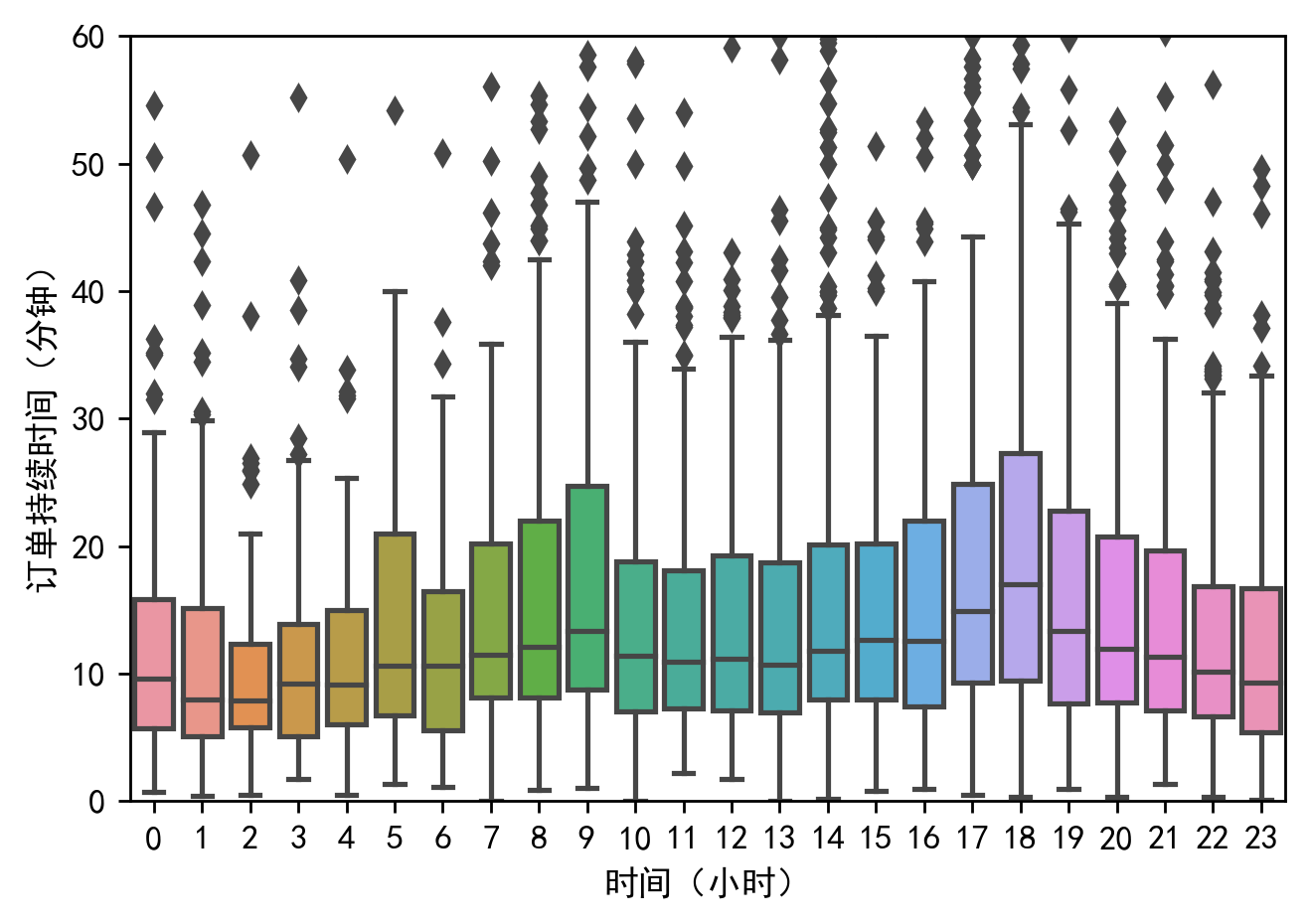
从原数据中生成OD表，每一行代表一个出行订单（OD记录），字段为出行的起点、终点、开始时间、结束时间与车辆信息，然后统计一天内各小时的出行订单量。如图3.1所示，出租车出行订单无异常缺失，有早晚高峰和凌晨低谷存在。综上所述，出租车出行订单的时间完整性高。



**图3.1 出租车订单数量小时分布图**

**3.2 出租车出行订单持续时间统计**

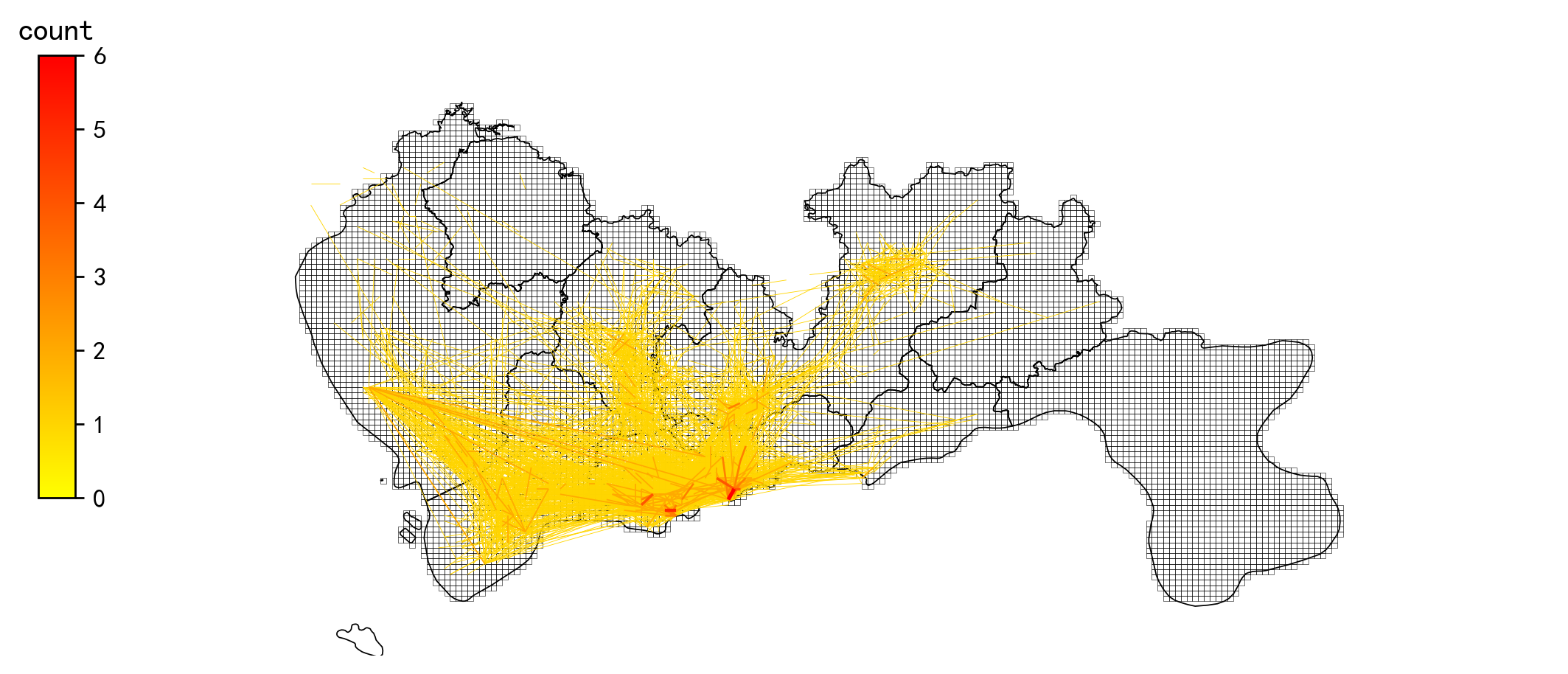
出行订单持续时间应处于合理范围内，先计算每一个订单的持续时间，然后按时间统计，生成一天内出租车订单持续时间按小时分组箱形图。如图3.2所示，一天中不同时段的出租车订单的持续时间存在差异，早晚高峰长而其他时间段短，原因是早晚高峰出行人数多、交通拥挤。



**图3.2 出租车订单持续时间按小时分组箱形图**

**3.3 出租车出行订单的栅格OD可视化**

使用栅格化方法，将OD点与栅格相对应，统计栅格内O点和D点的数量，计算两两栅格之间的出行量与每一条OD记录起终点的经纬度坐标，绘制出租车出行订单的栅格OD分布图。如图3.3所示，中心城区订单多，郊区订单较少。



**图3.3 出租车出行订单的栅格OD分布图**

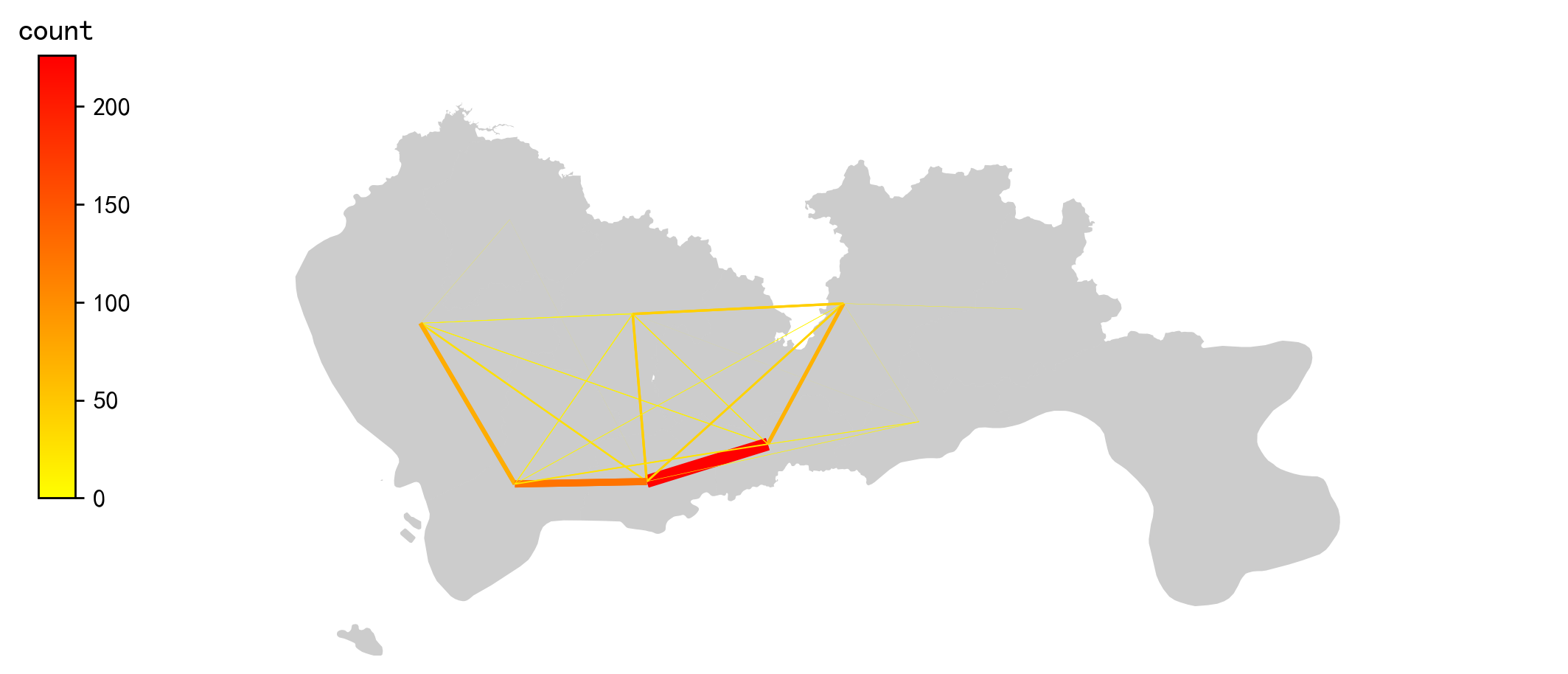
**3.4 出租车出行订单的OD期望线绘制**

OD期望线的绘制方法为：首先将出行订单按照交通小区（不规则的矢量图形，一般以行政区划替代，本研究的交通小区为深圳市各区）分类并匹配至对应的交通小区，然后计算交通小区之间的出行量，接着以交通小区质心作为OD期望线的起终点，最后将出行量映射到线的颜色和粗细上。绘制的关键是将OD点匹配至对应的交通小区，有两种方法。

方法一是直接判断OD点与交通小区边界的空间关系以匹配至交通小区，当数据量级大且交通小区边界复杂的情况下效率低。

方法二为栅格化方法，第一步将OD点匹配至栅格，第二步将栅格中心点匹配至交通小区，此时OD点与交通小区的空间关系转化为栅格与交通小区的空间关系。方法二效率高，但会出现OD点属于交通小区A，其所属栅格中心点属于交通小区B，其被归为交通小区B的错误，可以通过减小栅格面积来减少这种错误。

本研究使用方法二绘制图3.4深圳市出租车出行OD期望线图。如图所示，南山、福田、罗湖三区之间的出行联系最为紧密，南山、罗湖两区与宝安、龙岗两区之间的出行联系较为紧密，其余各区之间的出行联系较少，中心城区出行多、郊区出行少。



**图3.4 深圳市出租车出行OD期望线图**

**第四章 出租车需求模式分析**

**4.1 降维方法分析****交通时空矩阵的原理**

研究出租车需求模式需要研究交通时空矩阵，交通时空矩阵是行和列分别表示时间和空间维度的矩阵，矩阵内蕴含多种规律且与时空特征相关。降维方法包括将交通时空矩阵分解为不同模式的叠加、消除集中相关性强的维度、减少数据存储所需的时间与空间、便于可视化等特点，是分析交通时空矩阵的重要手段。下面介绍奇异值分解、非负矩阵分解与鲁棒主成分分析三种降维方法分析交通时空矩阵的方法。

考虑一个行为小时段，列为栅格，矩阵中每一个值表示某栅格在某小时段中的上车人数的交通时空矩阵，事实上对于任意矩阵总存在一个奇异值分解如式4.1所示，假设是一个大小的矩阵，则：左奇异矩阵是一个大小的正交矩阵，里面的向量为左奇异向量；是一个大小的对角矩阵，其对角线上为奇异值（奇异值一般会被从大到小排列）；右奇异矩阵是一个大小的正交矩阵，里面的向量为右奇异向量。

如果把和理解为列向量和构成的矩阵，那么矩阵可以进一步分解为式4.2，其中是秩为1的矩阵，为奇异值，4.2式表明交通时空矩阵可以分解为个秩为1的矩阵的叠加，每个矩阵的权重为，即交通时空矩阵可以分解为多个以简单规律构成的矩阵叠加，每个代表一种矩阵的分布模式，其中表示各模式在时间维度上的变化规律，表示各模式在空间维度上的变化规律。由于交通时空矩阵变化规律强，用少数分布模式即可代表大部分规律，因此将奇异值分解的结果按照奇异值从大到小排序，选取连续变化较大的前几个奇异值所代表的模式即可分析交通时空矩阵的需求模式。

奇异值分解得到的和向量中均有可能同时出现正负值使得该方法需要解释符号含义，不够直观。而非负矩阵分解是Lee和Seung于1999年在*Nature*上提出的一种矩阵分解方法[21]，其产生的结果都是非负的。NMF即对大小为的任意非负矩阵，其能够找到大小的非负矩阵和大小的非负矩阵使得式4.3成立，其中，其通过构建和后进行迭代来逼近，当且仅当和收敛时计算完毕，原矩阵和两个分解矩阵的乘积可以存在误差。

与奇异值分解同理，如果将和分别看成列向量与行向量构成的矩阵，则可以分解为个秩为1的矩阵叠加，如式4.4所示，每一个矩阵也是一种模式，代表了模式的行分布规律，代表了模式的列分布规律。

奇异值分解与非负矩阵分解的区别在于前者通过确定每一个主成分容纳的信息大小，在规律明显的数据上得到的不同模式之间的奇异值量级差异大，只需要少数模式就可以表示大部分信息，而后者每一种模式之间的大小差异不明显，即前者倾向于先用少量的模式表达出数据中主要的规律，再用剩下的模式去微调规律，模式之间的差异大，后者倾向于将矩阵分解为多个主要模式进行叠加，模式之间的差异小。

主成分分析与奇异值分解易受噪声（异常值）的干扰，鲁棒主成分分析可以解决这一问题。鲁棒主成分分析将原始矩阵分解为式4.5。其中为低秩矩阵，类似于奇异值分解是主要模式的叠加；为稀疏矩阵，蕴含异常信息，提高了的稳定性。

注意到的大小近似于，则定义噪声矩阵如式4.6所示，其量级远小于、与，因此鲁棒主成分分析将分解为、与。

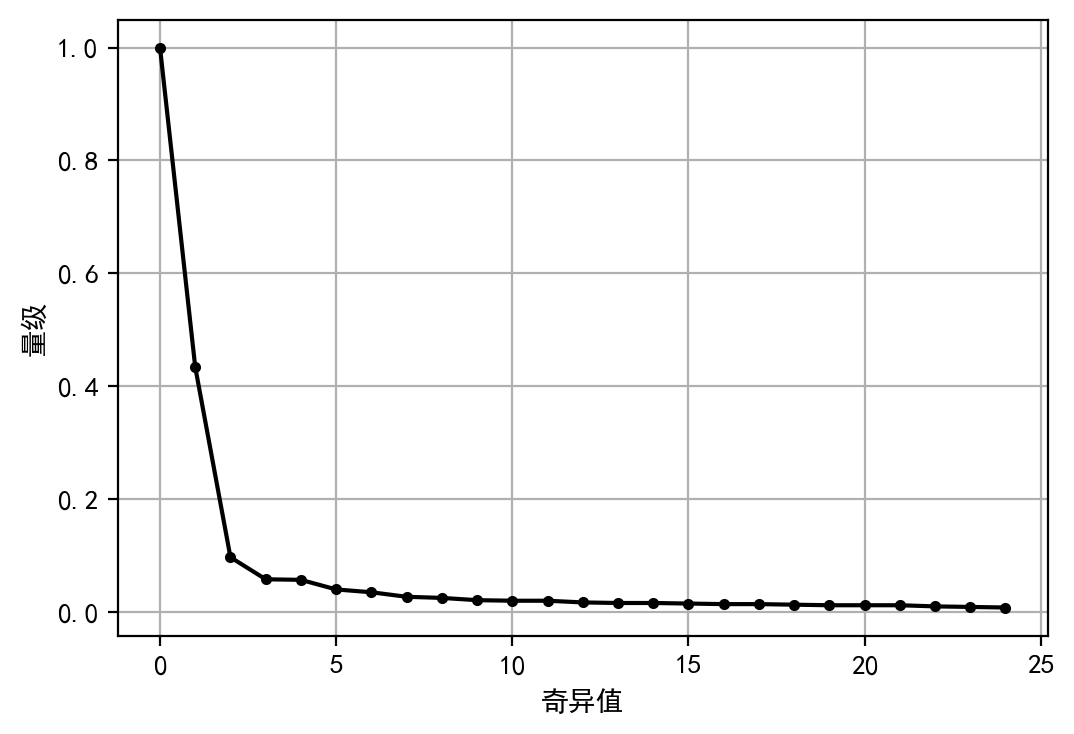
**4.2 使用奇异值分解分析出租车需求模式**

本研究使用奇异值分解分析出租车需求模式。

从第三章整理出的OD数据中提取出时间段数×栅格数大小的交通时空矩阵，时间维度（行）是小时，空间维度（列）是栅格，统计的值为每个时刻每个栅格的上车人数。

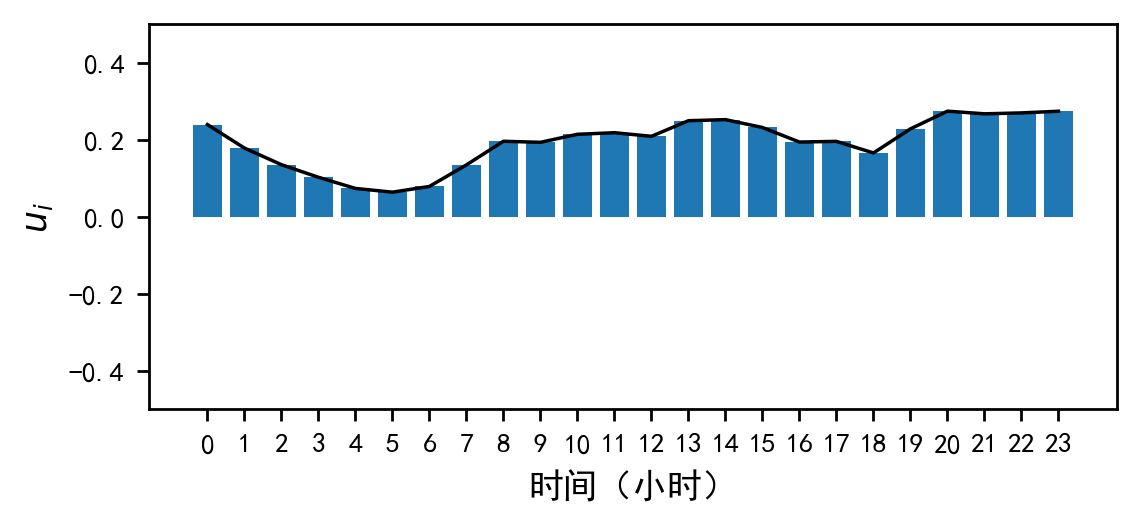
注意到最初生成的中绝大部分的值为空值，原因为深圳全市范围内大部分的栅格都只在部分时段有打车需求，这样的无法用于奇异值分解，需要去除空值，去除方法为：一是将所有空值置0，表示没有打车需求；二是只保留需求较大的栅格作为研究对象，剔除需求较小的栅格，这样中就只剩下少量的空值，对这些空值再使用方法一置0。方法一的劣势为进行奇异值分解后所得的模式会被大量的0影响，且一个栅格如果没有或偶尔有几个打车需求那么研究意义不大，因此本研究选择方法二。

本研究保留空值数小于5的列（即认为一天中至少有20h有打车需求的栅格才值得研究），空值置0，然后进行奇异值分解，在结果中同时调整和的正负并保证向量中绝对值最大的值为正。对于奇异值分解的结果，设最大的奇异值为1，对剩余奇异值进行标准化（标准化方法为总和标准化）以表示各奇异值之间的量级大小，绘制图4.1奇异值变化碎石图。如图所示，曲线在第一个奇异值到第三个奇异值之间的下降速度非常快，从第三个奇异值开始奇异值之间的差距变小且趋于稳定，前两个奇异值的量级较大，因此本研究选择前3个模式来分析出租车需求模式，并对其时间向量分布和空间向量分布进行可视化。

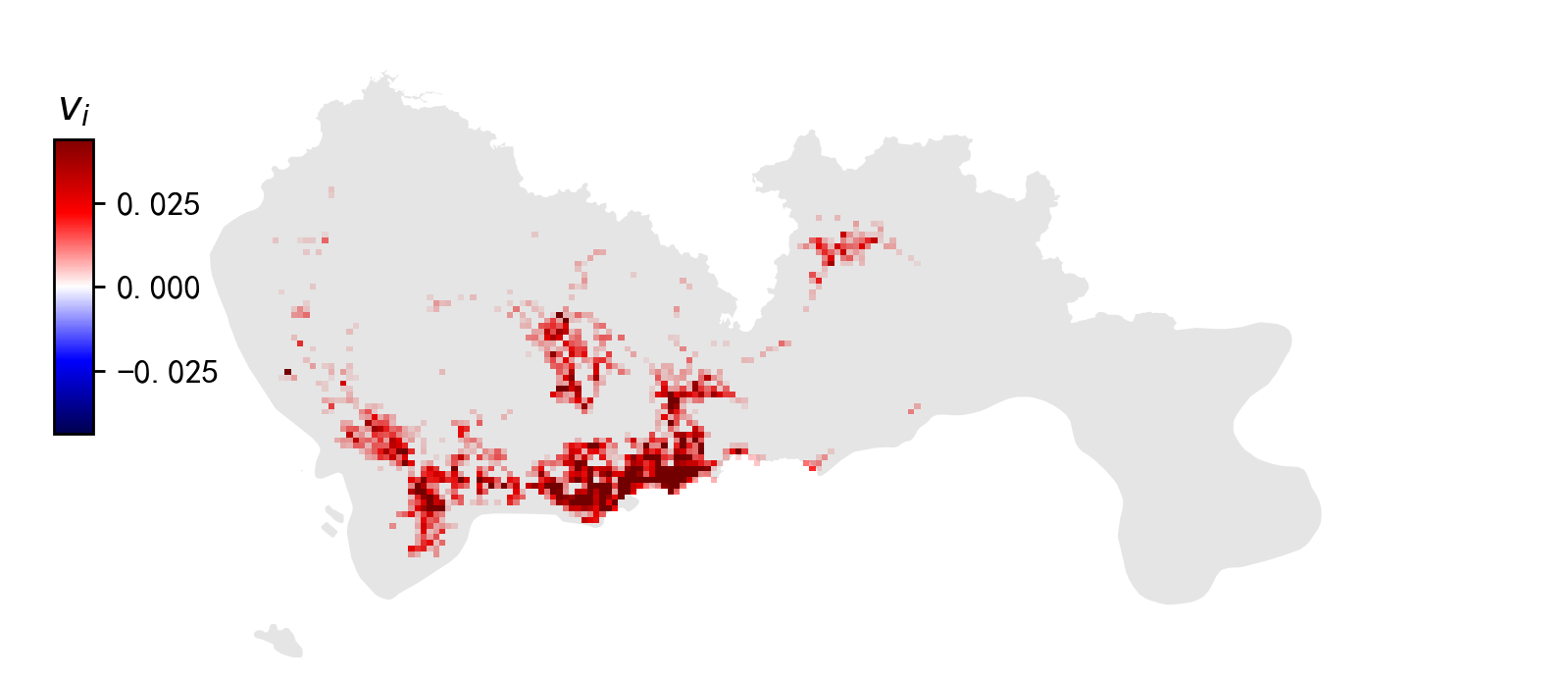


**图4.1 奇异值变化碎石图**

模式一的时间分布如图4.2所示，空间分布如图4.3所示。模式一中与均为正，模式一揭示的需求模式为白天出行需求大、晚上出行需求少，中心城区出行需求大、郊区出行需求少。

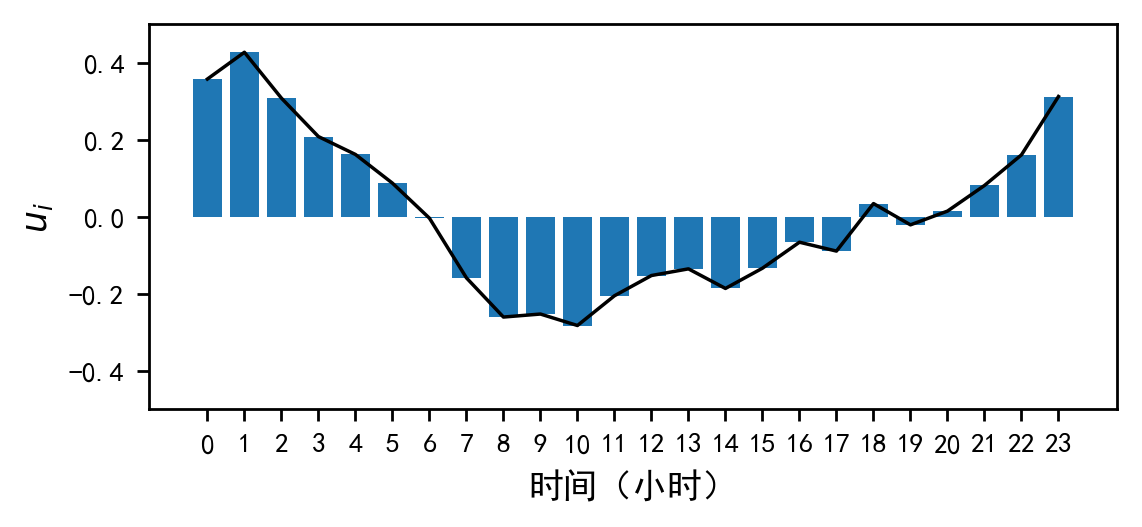


**图4.2 模式一时间分布图**

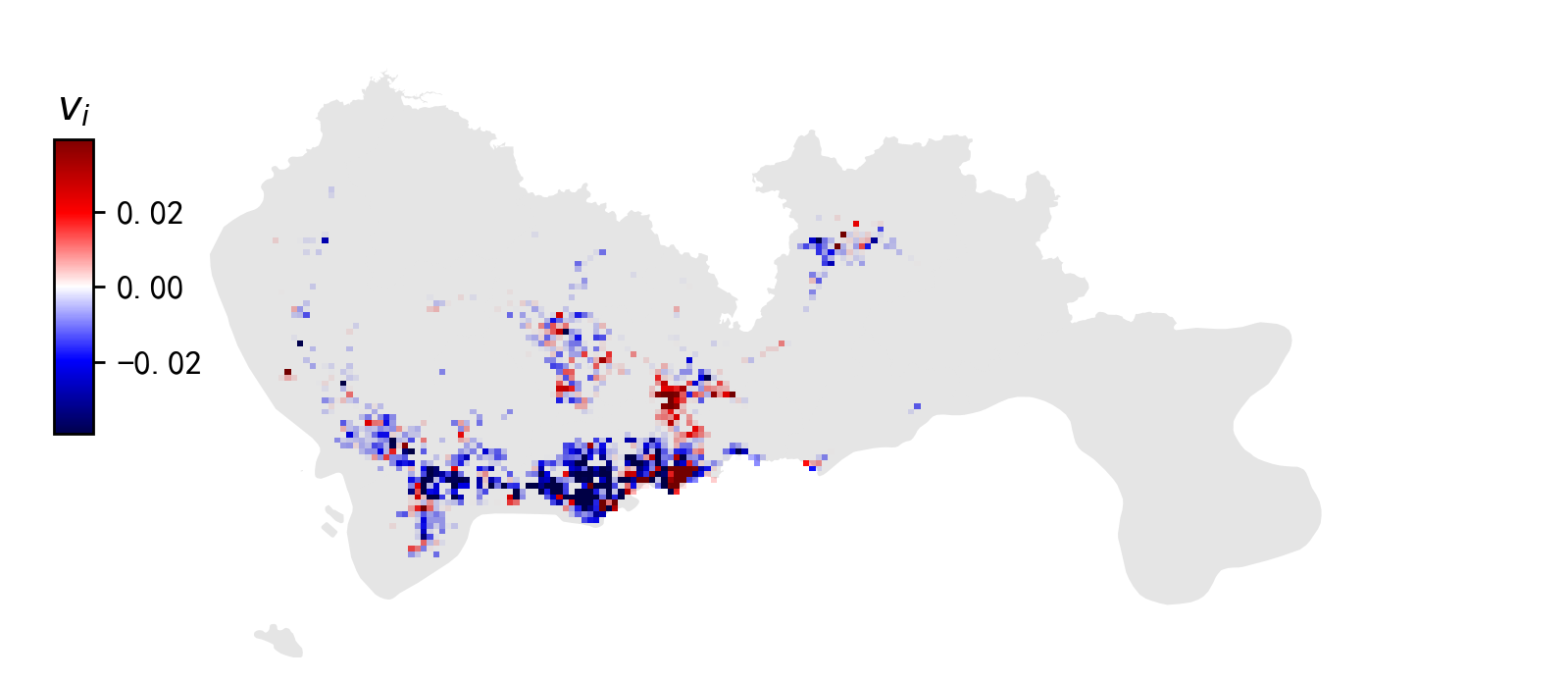


**图4.3 模式一空间分布图**

模式二的时间分布如图4.4所示，空间分布如图4.5所示。模式二在时间上呈现出夜间为正、白天为负的分布，在空间上呈现出老城区（罗湖）为正、新城区（福田、南山）为负的分布。由于时间的负值与空间上的负值相乘负负得正，因此模式二揭示的需求模式为老城区在夜间出行需求多、新城区在白天出行需求多的特点，这是职住分离的一种体现。虽然这两种需求值反方向变化，但意义相同，因此被归为同一种模式。

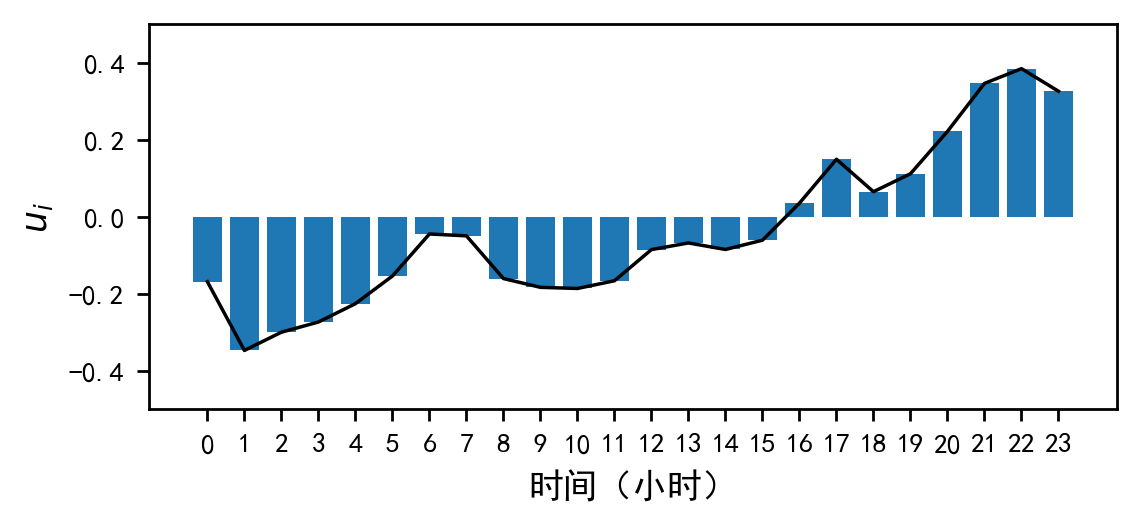


**图4.4 模式二时间分布图**

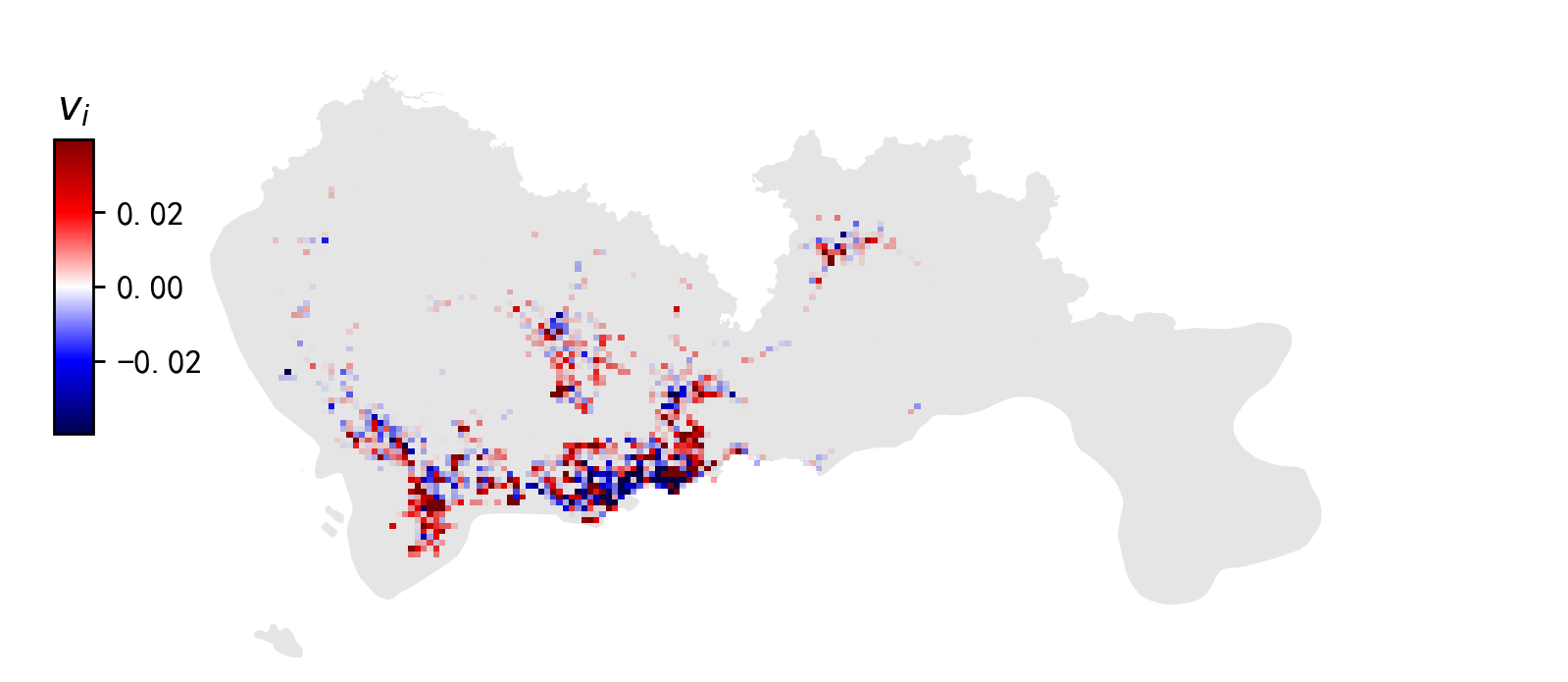


**图4.5 模式二空间分布图**

模式三的时间分布如图4.6所示，空间分布如图4.7所示。模式三在时间上呈现出凌晨至下午为负、晚高峰至夜间为正的分布，空间上呈现出中心城区为负而罗湖区田贝、南山区深圳湾与后海商圈附近为正的分布，模式三揭示的需求模式为前者白天出行需求大，后者夜晚出行需求大。不同于模式一和模式二，模式三表明的是小尺度地段上的出行需求。



**图4.6 模式三时间分布图**



**图4.7 模式三空间分布图**

**第五章 总结与展望**

本研究主要使用了栅格化、奇异值分解等方法，首先通过时空完整性评估与空间完整性评估确保了数据的质量以及研究的可行性，然后通过OD提取、栅格OD与OD期望线可视化等手段揭示了深圳市各行政区划之间的出行特征，最后通过奇异值分解分析了深圳市出租车不同需求模式的时空分布，展示了一套完整的分析流程，对其他公共交通工具的分析具有借鉴意义。

本研究的不足之处在于以下几点，未来可以进一步探索：数据源单一，仅使用出租车GPS数据，可采用多源大数据（POI、土地利用类型等）进行研究；栅格化方法导致了必然的精度损失，且直接基于地理坐标系进行栅格化的精度低于先将地理坐标系转化为投影坐标系再进行栅格化；奇异值分解将个体数据统计为群体，得到的需求模式无法追溯到个体，且需要根据其时空分布规律来推测现实的需求；研究方法单一，可以同时使用奇异值分解与非负矩阵分解进行分析并比较两者的结果，或使用地理加权回归等模型深入研究需求模式的形成因素。

**参考文献**

1. 胡兰兰. 基于GPS出租车高收益热点区域推荐[D].温州大学,2020.
2. Zhao S, Xiao Y, Ning Y, et al. An optimized K-means clustering for improving accuracy in traffic classification[J]. Wireless personal communications, 2021, 120: 81-93.
3. Loh W K, Yu H. Fast density-based clustering through dataset partition using graphics processing units[J]. Information Sciences, 2015, 308: 94-112.
4. Shen Y, Zhao L, Fan J. Analysis and visualization for hot spot based route recommendation using short-dated taxi GPS traces[J]. Information, 2015, 6(2): 134-151.
5. 秦昆,周勍,徐源泉,等.城市交通热点区域的空间交互网络分析[J].地理科学进展,2017,36(09):1149-1157.
6. Tong Y, Chen Y, Zhou Z, et al. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms[C]//Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. 2017: 1653-1662.
7. 李岩. 基于热点区域的出租车出行需求预测模型研究[D].长安大学,2020.
8. Han G, Sohn K. Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model[J]. Transportation Research Part B: Methodological, 2016, 83: 121-135.
9. Xiaojian G, Quan Z. A traffic flow forecasting model based on BP neural network[C]//2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS). IEEE, 2009, 3: 311-314.
10. Pang J, Huang J, Du Y, et al. Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(9): 3283-3293.
11. Zhang J, Zheng Y, Qi D, et al. DNN-based prediction model for spatio-temporal data[C]//Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems. 2016: 1-4.
12. Liu T, Wu W, Zhu Y, et al. Predicting taxi demands via an attention-based convolutional recurrent neural network[J]. Knowledge-Based Systems, 2020, 206: 106294.
13. Xu J, Rahmatizadeh R, Bölöni L, et al. Real-time prediction of taxi demand using recurrent neural networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 19(8): 2572-2581.
14. 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.
15. 徐婷立,王纪武,胡雪薇.基于GWR的杭州共享单车分布影响机制研究[C]//中国城市规划学会.人民城市，规划赋能——2022中国城市规划年会论文集（05城市规划新技术应用）.浙江大学建筑工程学院;塔里木大学水利与建筑工程学院;,2023:10.DOI:10.26914/c.cnkihy.2023.047560.
16. 王李轩. 基于KDE与GWR的城市出租车载客出行特征研究[D].长安大学,2021.DOI:10.26976/d.cnki.gchau.2020.002206.
17. 孙启鹏,曾开邦,张锴琦,等.北京市共享单车出行的时空规律与需求预测研究[J].交通运输系统工程与信息,2022,22(01):332-338.DOI:10.16097/j.cnki.1009-6744.2022.01.035.
18. 高枫,李少英,吴志峰,等.广州市主城区共享单车骑行目的地时空特征与影响因素[J].地理研究,2019,38(12):2859-2872.
19. 付鑫,孙茂棚,孙皓.基于GPS数据的出租车通勤识别及时空特征分析[J].中国公路学报,2017,30(07):134-143.DOI:10.19721/j.cnki.1001-7372.2017.07.017.
20. <http://jtys.sz.gov.cn/gkmlpt/content/11/11130/post_11130513.html#1508>
21. Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. nature, 1999, 401(6755): 788-791.