

# 武汉大学

## 2021 年省级大学生创新训练项目

### 申报书

题目：基于网约车数据和 POI 数据的城市居民出行规律研究

院系：遥感信息系工程学院

项目参与人：陈宇、王宇帆、周佐翼、周祎博、刘子睿

指导老师：秦昆

申报时间：2021 年 1 月 21 日

项目名称		基于网约车数据和 POI 数据的城市居民出行规律研究			
项目创新特色概述		1. 服务理念特色——让生活在城市中的人提高安全感及舒适度 2. 数据来源创新——基于网约车和 POI 数据的居民出行研究 3. 技术手段创新——开发适用性更强的信息提取算法 4. 方法应用创新——建立模型分析网约车用户群体的潜在规律			
项目所属一级学科		工学			
申请经费		3000 元	起止时间	2021 年 2 月至 2022 年 9 月	
申请团队信息	姓 名	学 号	院（系）、专业		E-mail
	陈 宇	2019302130136	遥感信息工程学院		18365253617 984560205@qq.com
	王宇帆	2019302130328			
	周佐翼	2019302130313			
	周祎博	2019302130289			
	刘子睿	2019302130322			
导师信息	姓名	院（系）	职称	联系电话	E-mail
	秦 昆	遥感信息工程学 院	教授	13667287562	qink@whu.edu.cn

## 目录

一、立项依据.....	4
1.1 研究意义.....	4
1.2 国内外研究现状及发展动态分析.....	6
1.3 主要参考文献.....	11
二、研究内容、研究目标及拟解决的关键科学问题.....	12
2.1 研究内容.....	12
2.2 研究目标.....	13
2.3 拟解决的关键科学问题.....	13
三、拟采取的研究方案及可行性分析.....	14
3.1 研究方法.....	14
3.2 技术路线.....	14
3.2.1 总体技术路线.....	14
3.2.2 各部分技术路线.....	15
3.3 可行性分析.....	21
四、项目的特色与创新之处.....	24
4.1 项目创新点.....	24
五、研究计划及预期结果.....	26
5.1 年度研究计划.....	26
5.2 预期研究成果.....	26
六、详细预算.....	26
6.1 小宗研究设备费用.....	26
6.2 文印出版费用.....	27

## 一、立项依据

### 1.1 研究意义

近年来,随着计算机存储和处理能力的不断增强,可获得的数据量与日俱增,使得获得人群长期大规模数据成为可能,同时随着城市生活的日益丰富,人们在城市中的活动种类逐渐增多,城市人群的出行更加复杂和多样,在此背景之下,基于获得的网约车数据来分析和研究城市人群出行规律具有重要意义。

首先,随着互联网及移动终端的普及,普通大众尤其是城市居民,越来越倾向于使用线上的方式来满足日常生活的需求,包括交通和出行。但是随着经济的发展,城市规模的不断扩大,各种基础设施的不断完善,人民生活水平提高,出行更加频繁,出行时间和出行目的地更加灵活多变,传统的城市交通比如公交、地铁和传统出租车,在很多情况下已经不能满足人们日常生活的需要。这种情况下,网约车的出现很好地填补了这一空白。基于网约车数据和 POI 数据研究城市人群的出行规律,能够为网约车服务质量的提升,以及进一步优化城市交通,方便居民出行提供重要帮助。

近年来,网约车作为一种新的行业形态已在全球掀起了巨大浪潮,网约车是互联网技术在汽车共享方面的应用。网约车平台正改变着城市的运输结构和人们的出行方式。网约车的出现改变了传统出租车服务质量差、价格贵、效率低的局面,并且预约服务操作简单,从而满足了乘客的消费需求。网约车的服务模式是一种精准的点对点服务,即有专门车辆随时随地满足乘客的个性化的出行需求,它不仅颠覆了传统出租车行业,而且快速有效地弥补了出行市场的需求,且凭借着便捷舒适的乘坐体验也逐渐获得了市场认可。2020 年 4 月 15 日,国家财政部、人力资源社会保障部、中国人民银行发布《关于进一步加大创业担保贷款贴息力度全力支持重点群体创业就业的通知》,在扩大覆盖范围方面,增加了对网约车司机、网约车企业及其子公司等群体的支持,此次三部委合力推出通知指导,表示 2021 年 1 月 1 日之前的创业担保贷款,利息全部由财政负担,足以说明国家正加大对网约车行业的扶持力度,且从 2015-2018 年我国网约车客运量占出租车客运总量的比重从 9.5%提高到 36.3%,网约车发展的前途是光明的。

但是当下网约车的发展仍旧有待改进,为进一步促进网约车的发展,改进城



市交通系统,满足人民日益增长的需要,有必要进一步加深对于城市人口出行规律的探索。城市人口出行规律主要体现在时间和空间两个方面。但是时间和空间并不是一成不变,城市的格局会随着城市的发展不断变换,在不同的时间,城市的人口流动和分布都有不同的趋势,因此急需研究这些因素影响下的城市居民出行规律,掌握城市人群在不同时间的流向和分布,为网约车服务提供参考,使之能够在不同区域不同时间合理分配资源,避免造成出行资源的不足或冗余,进而造成交通拥堵或乘客的需求难以得到满足。通过研究城市人群出行规律可以进一步发展网约车服务,进一步优化城市交通,方便居民出行。

其次,过去由于科技的局限,海量的个体信息获取成本巨大、时效性低,因此无法有效获取大量的人类行为数据进行研究。而在当今的网络时代人们的出行数据正在不断通过泛在传感器的使用而被完整和系统地记录下来并形成数据,人们的出行数据进一步反映了人们的工作、生活、消费、旅行等行为,因此研究城市人群的出行规律将对城市规划等应用领域产生了积极而深刻的影响。

"城市规划"是规范城市发展建设,研究城市的未来发展、城市的合理布局 and 综合安排城市各项工程建设的综合部署,是一定时期内城市发展的蓝图,城市的复杂系统特性决定了城市规划是随城市发展与运行状况长期调整、不断修订,持续改进和完善的复杂的连续决策过程。在进行城市规划的过程中,动态的获得城市的相关信息尤为重要,而基于网约车数据和 POI 数据研究城市人群出行规律的过程中,我们恰好可以获得相关的,动态的信息,来支持城市规划的顺利进行,处在当下的城市中,由于城市规模的增大,几乎所有的出行都要依靠交通工具,而出行往往带有目的性,通常去往服务设施居多,通过分析数据,获得城市人群的出行规律,在此基础上,可以进一步通过分析规律来挖掘感兴趣的信息,举几个简单的例子,通过对上车点和下车点密度的空间分布,可识别城市主要位置(如火车站、飞机场等;通过对上车点、下车点位置进行聚类划分,得到不同时段上车比下车高或下车比上车高的区域,并推测该区域的成因等等,通过对城市人群出行规律的探索,可以进一步的出城市基础设施的空间分布,进而为城市建造基础设施以及确定设施开放时间等城市规划提供有效信息。

最后与日俱增的网约车数据(点)、网约车 GPS 轨迹数据(线)、及相关街道网络数据等,将构成蕴含城市人群行为模式和活动规律的行为轨迹大数据,城市人

群行为活动的真实记录,与社会因素密切相关。通过行为轨迹大数据的分析和挖掘,可以发现一些隐含的行为模式和活动规律,从而发现和理解个体与环境及其他个体间的交互作用。通过对城市人群行为轨迹数据的特性、时空聚类模式等进行分析,考虑时间空间等因素对城市人群活动规律的影响,并在此基础上对城市人群活动规律的总结方法进行进一步优化,从而为网约车服务合理分配资源,提高服务灵活性提供辅助信息和决策支持,并为 GIS 在人文社会学领域的应用研究奠定基础。

通过聚类算法对网约车数据和 POI 数据处理,结合网约车行为轨迹数据,研究不同因素限制条件下的热点区域,并进行城市人群出行规律的总结,从而为需要了解城市人群出行规律的群体提供参考。本项目将在基于网约车数据和 POI 数据对城市人群出行规律研究产生若干创新性成果,并将为优化群众网约车体验、协助网约车平台合理分配资源、改善城市交通状况做出基础性贡献。

## 1.2 国内外研究现状及发展动态分析

数据密集型的地理空间分析是地理信息科学的学科前沿之一。“大数据”一词首先出现在 2008 年 9 月 Nature 杂志发表的关于 Big Data: Science in the Petabyte Era 的专题讨论。此后,2011 年 2 月,Science 出版专刊 Special Online Collection: Dealing with Data。2012 年 3 月,美国启动了计划投资 2 亿美元的“大数据研究和发展计划”,旨在提高和改进人们从海量和复杂的数据中获取知识的能力。为了迎接大数据时代的挑战和机遇,国际上的专家学者针对大数据处理开展了一系列探索和研究。

大数据时代的到来,给测绘地理信息创新发展带来了良好机遇。当前,海量的地理信息数据远未被充分利用,测绘地理信息部门、地理信息企业逐步认识到大数据技术带来的深刻影响。加大对地理信息数据开发应用的研究,充分挖掘地理信息数据这一“金矿”,促进地理空间数据的广泛、深入应用,将极大地促进地理信息行业的发展。与日俱增的行为轨迹数据构成了典型的地理时空大数据之一,为地理时空大数据的理论与方法研究提供了典型的实验研究领域和丰富的数据源。

结合本项目的研究重点,从以下方面介绍国内外研究现状及发展动态。

### (1) 城市人群出行热点区域的聚类研究

根据韦氏词典中的定义，热点区域是指一个比其他区域具有更多的兴趣点、人类活动更加频繁的地理区域。在现代城市中，城市中的热点区域往往代表着人们出入次数较多、交通流量较大，出行需求较高的区域，是人们频繁、密集出行的直接体现。因此，在研究城市人群出行规律的过程中对于热点区域的研究具有非常重要的价值。从网约车数据中发现城市居民出行热点区域更是成为了一个新的研究关注点。

对于出行热点区域的研究，比较传统的方法包括问卷调查法，但是从获得数据的途径来说，这种方法往往会消耗大量的人力物力，从获得数据的质量来说，这种方法所获得的数据往往相对滞后，存在时效性差，准确度低等问题，获得的结果往往参考意义不大<sup>[2]</sup>。

近年来，随着 GPS 定位、卫星导航、无线通信等技术的快速发展，民用 GPS 等定位设备不断普及和广泛应用。这些 GPS 定位设备以及基于位置信息服务（LBS）的各种应用产生了大量的来自移动对象的时空轨迹数据。时空数据挖掘技术、地理信息技术的不断成熟也使得研究这些大量移动对象的轨迹数据成为可能。这些移动对象轨迹数据根据来源可以分为两大类：一种轨迹数据是出租车、公交车等公共交通工具产生的，另一种轨迹数据来自个人活动，如手机定位数据、私家车轨迹数据等。由于这些数据中每一个轨迹点反映空间上的一个位置以及时间上的一个事件，而多个轨迹点的集合记录着该移动对象在空间和时间上的移动路径，代表着人们日常活动与出行行为特征。轨迹聚类的目的是识别相似的轨迹运动模式，所以通过移动对象轨迹聚类研究可以发掘人类群体整体移动的趋势，识别人们感兴趣的热点路径和热点区域以及挖掘居民通过某种交通方式出行的移动时空特征<sup>[4]</sup>。因此当下国内外的研究主要偏向与使用聚类分析的方法。所谓聚类分析，又称群分析，它是研究（样品或指标）分类问题的一种统计分析方法，同时也是数据挖掘的一个重要算法，聚类分析是由若干模式组成的，通常，模式是一个度量的向量，或者是多维空间中的一个点。聚类分析以相似性为基础，在一个聚类中的模式之间比不在同一聚类中的模式之间具有更多的相似性。其中在国内外相关研究中使用比较广泛的有 DBSCAN 算法、OPTICS 算法等，DBSCAN 算法把邻域范围内所包含的空间聚类对象的最小数目作为空间密度，而空间密度相连对象的最大集合就是空间聚类簇。其在聚类过程中需要提前设置  $\varepsilon$  和



MinPts 两个重要参数，所以聚类结果容易受到这两个全局参数的影响，并且难以适应密度分布不均匀的空间数据。而 OPTICS 算法是在 DBSCAN 基础上的改进算法，它引入了核心距离与可达距离的概念对基于密度的数据结构进行排序，生成聚类结果的可达距离图，可以发现聚类对象中的不同密度的空间簇。在此基础上，国内外的研究提出了更多改进的算法，致力于使得聚类算法更加适应于对网约车数据的处理，比如一些文献中提出的对 OPTIC 算法的改进算法——TR-OPTICS 算法<sup>[4]</sup>。传统 OPTICS 算法在计算核心对象时是以该对象为中心，阈值  $\epsilon$  为半径的圆形搜索区域。而本文研究对象是载客轨迹线段，由于不同的载客轨迹线段之间形状、长度不一，圆形搜索领域不再适合载客轨迹的聚类，因此本文采用轨迹的外包矩形作为核心轨迹对象的搜索邻域，在聚类过程中，可达距离越小代表空间密度越大，算法中用一个可达距离升序排列的种子序列来储存待扩张的子轨迹，通过搜索邻域依次向空间稠密的方向扩张。OPTICS 算法相对于经典的 DBSCAN 算法有两个新的概念，就是核心距离与可达距离。由于本文研究对象为载客轨迹，核心距离与可达距离不再是简单的点与点之间的欧几里德距离（Euclidean distance），而是载客轨迹线段之间的轨迹距离（水平距离、垂直距离、角度 距离）。因此对聚类算法中的核心距离与可达距离概念进行了重新定义。还有一些文献中提到受基于密度的聚类算法以及区域网格化方法的启发，提出了一种基于网格密度的热点区域探测方法：GScan 算法<sup>[2]</sup>。该方法通过划分网格单元将大量空间 GPS 数据离散化，再利用密度阈值筛选热点网格单元，并对临近可达的热点网格进行合并，从而挖掘出城市中的热点区域。还有一些文献提出了其他解决方法：考虑时空相关性的时空聚类：考虑时空相关性，提出顾及时空相关性的时空聚类方法，主要包括时空谱聚类和时空密度聚类。这两种方法都同时考虑了轨迹数据的属性特征和时空相关性特征，对这两种特征分别进行相似性度量和接近性度量。接着利用时空聚类方法对行为轨迹数据进行时空聚类挖掘，挖掘其时空聚类模式，如客流热点区域、出租车接客模式、拥堵模式、出行模式等，并进一步考虑其时间特征，分析其时空演化模式<sup>[20]</sup>。

在目前，该方面研究中所遇到的问题，就是聚类算法无法很好的满足人们需要。而今后随着人工智能，机器学习，模式类别和数据挖掘等领域传统方法的不断发展和新技术的出现，聚类算法的发展主要有以下方面：1.分类模式 2.关联规



则 3.决策树 4.序列模式 5.聚类分析模式 6.神经网络算法。

此外,还有国内近年来发展的数据场理论,也在热点区域的检测中大放异彩:出租车轨迹数据中涵盖着乘客的出行目的和移动模式。若某区域聚集的轨迹点越多,说明该区域对乘客的“吸引力”越大,且其相邻一定范围内的区域也会聚集着较多的乘客。这种区域对轨迹点的影响以及区域之间相互影响的现象符合数据场中数据质点相互“吸引”的假设理论,因此可用数据场中的势函数来量化这种影响。若某区域具有较高势值,便表明该区域聚集了较多的轨迹点,则可视作城市热点区域。因此,利用网格划分法将轨迹数据进行区域单元划分,然后基于数据场理论,产生了一种新的探测轨迹点聚集区域的方法<sup>[19]</sup>。传统的数据场势值计算中没有考虑时间对场的影响,仅以距离作为轨迹点相似性度量的标准,不适用于时空数据的相似性度量。因此,利用数据场理论进行出租车轨迹的聚集性分析首先需要对势函数进行时空扩展<sup>[21]</sup>。

## (2) 城市人群出行规律的 OD 调查方法及结果分析

OD 调查即起讫点调查,是对一个调查区域内出行个体的出行起点和终点的调查,该技术最早应用于国外的城市交通规划中。OD 调查一般从描绘调查区域的略图开始,调查区域分为直接影响区域和间接影响区域,直接影响区域和间接影响区域均可分为一系列小区,各小区编号,并假定所有出行的起点和讫点均在其形心,区域划分的目的,一是将区域交通需求产生与区域经济、人口等联系起来;二是将交通需求流动状况用跨区域的方式表现出来。在对网约车数据的研究中,一般用到的是 OD 流模式的研究,OD(Origin—Destination)流指轨迹中起点到终点的流动性特征,其中 O 表示出行起点,D 表示出行终点。OD 流从统计的角度反映移动对象的活动和群体轨迹的动态特征。当前研究主要集中在 O、D 位置和 OD 流的移动性两方面。①O、D 位置的研究关注轨迹起止点位置特征。网约车数据 OD 点在乘客的角度相当于上车点、下车点。在国内外的研究中 Veloso 等(2011)发现上车点和下车点密度的空间分布,可识别城市主要位置(如火车站、飞机场等)。Guo 等 (2012)将上车点、下车点位置进行聚类划分,得到不同时段上车比下车高或下车比上车高的区域,并推测该区域的成因。童晓君等(2012)对网约车上车点、下车点进行聚类分析,发现了节假日与工作日的上车、下车热点区域。②OD 流移动性研究。包括 OD 流的可视化、基于 OD 流的社区结构发现

等。Guo(2009)对美国人口迁移数据进行分析得到可视化流,揭示人口迁移规律,发现隐含的社区结构。Gao 等(2013)从通话数据中提取通话流与人群移动流可视化,并进行社区划分。Zhang 等 (2012)对出租车数据进行 OD 流聚类分析,找出聚类中心,添加起始点语义信息,发现 OD 流的语义,如车站一车站、商场一车站等。目前 OD 流模式的研究偏向于网约车轨迹数据、人口迁移数据,需更多针对手机通话数据的 OD 流研究。OD 流的可视化研究方面存在不足,如何获得更加直观、动态的可视化 OD 流值得研究<sup>[5]</sup>。

### (3) 城市居民规律的预测

通过轨迹数据挖掘人类的活动模式,结合一些辅助信息可以预测用户轨迹。当前针对轨迹预测的研究主要包括两个方面:①研究轨迹可预测性条;②预测用户轨迹。很多学者关注人类活动轨迹是否可预测。Gonzalez 等(2008)发现个体轨迹表面上是随机的,但人类轨迹在时空中具有规律性,即个体完成一次独立出行后,会以显著概率往返经常到访的地点,居民日常活动模式隐藏着高度的可预测性潜力。Song 等(2010)发现人类活动有 93%的可预测性。Zhao 等(2011)使用离散熵来衡量离散轨迹的可预测性(越规律离散熵越低)。发现轨迹的可预测性潜力后,人们通过轨迹数据挖掘用户的兴趣点、活动模式或结合历史轨迹预测用户轨迹或即将到访的地点。Cho 等(2011)利用全球社交网络数据和手机数据研究人类活动模式,发现短距离时空中 Et 常周期运动与人类经验结合紧密,而长途运动则与社交网络联系密切,因此开发了一种结合日常活动与社交网络的人类活动动力学模型,预测用户即将到访的地点。Di 等(2012)利用个体和群体历史出行轨迹和区域特征(土地利用、兴趣点)预测个体轨迹,发现将群体活动行为整合在个体轨迹预测中能显著提高预测精度。Zhao 等(2011)则把个人轨迹与群体活动轨迹相结合,使用动态贝叶斯网络来预测群体轨迹。Veloso 等 (2011)利用出租车历史数据在给定时间、天气以及前一个下客点类型的情况下,预测下一个上客点。齐观德等(2013)使用出租车历史轨迹数据,利用空车到达时间间隔的概率分布推断乘客等候时间的概率分布,从而预测乘客在某时某地等候出租车的时间,准确率平均为 84.7%<sup>[5]</sup>。

随着网约车的日益广泛和发展,可供研究的网约车数据也越来越丰富,网约车数据和 POI 数据构成了典型的地理时空大数据,并且与人文和社会因素密切

相关。如何利用这些典型的地理时空大数据来分析不同因素影像下的城市人群出行规律,为网约车服务提供进一步优化的可能,是很有研究价值的问题。基于此,本项目对基于网约车数据和 POI 数据对城市人群出行规律进行分析研究,通过聚类算法,OD 调查及分析等方法对网约车数据进行处理,研究多因素限制条件下的城市人群出行规律,进一步对城市人口的出行进行预测从而为进一步推动网约车行业的发展,优化城市交通[21]。

### 1.3 主要参考文献

- [1]张俊涛,武芳,张浩.利用出租车轨迹数据挖掘城市居民出行特征[J].地理与地理信息科学,2015,31(6):104-108.
- [2]郑林江,赵欣,蒋朝辉,邓建国,夏冬,刘卫宁.基于出租车轨迹数据的城市热点出行区域挖掘[J].计算机应用与软件,2018,35(1):1-8.
- [3]童晓君,向南平,朱定局.基于出租车 GPS 数据的城市居民出行行为分析[J].电脑与电信,2012(1):56-59.
- [4]杨树亮,毕硕本,Nkunuzimana A,黄铜,万蕾.一种出租车载客轨迹空间聚类方法[J].计算机工程与应用,2018,54(14):249-255.
- [5]李婷,裴韬,袁焯城,宋辞,王维一,杨格格.人类活动轨迹的分类、模式和应用研究综述[J].地理科学进展,2014(7):938-948.
- [6]贾兴无.基于网约车数据的居民出行需求特征分析及需求预测[J].交通工程,2018,18(5):39-45.
- [7]林基艳,张雅琼,张慧.基于出租车 GPS 轨迹数据挖掘的居民出行特征研究[J].计算机时代,2017(5):37-39.
- [8]刘惠宇,刘鹏,张婷.基于出租车 GPS 数据的城市人群出行模式研究[J].江苏科技信息,2019,36(17):48-51.
- [9]贾冲,冯慧芳,杨振娟.基于出租车 GPS 轨迹和 POI 数据的商业选址推荐[J].计算机与现代化,2020(2):21-25.
- [10]魏海涛,肖天聪,胡宝生,王薇,王晨曦.基于出租车 GPS 数据的居民出行行为和城市空间结构分析--以北京市为例[J].上海城市规划,2020(1):69-76.
- [11]吴华意,黄蕊,游兰,向隆刚.出租车轨迹数据挖掘进展[J].测绘学



报,2019,48(11):1341-1356.

[12]方琪,王山东,于大超,朱鸿博,李贺.基于出租车轨迹的居民出行特征分析[J].地理空间信息,2019,17(5):128-130.

[13]周英迪,张兴国,潘晓芳,吴超红.基于出租车轨迹的居民出行时空分析[J].城市勘测,2020(3):17-23.

[14]孙贵治.基于出租车 GPS 轨迹数据的热点区域出行需求预测[D].北京交通大学,2019.

[15]陈志明.基于大数据的出租车乘客出行特征提取和分析方法研究[D].长安大学,2017.

[16]刘入嘉.基于出租车 GPS 数据的城市出行空间结构研究[D].东南大学,2018.

[17]姜海林.基于出租车数据的城市居民出行时空特征研究[D].武汉大学,2018.

[18]崔羽,顾琼.大数据环境下出租车 GPS 轨迹对居民出行研究.沈阳市规划设计研究院信息中心

[19]周劬,秦昆,陈一祥,李志鑫.基于数据场的出租车轨迹热点区域探测方法[J].地理与地理信息科学,2016,32(6):51-56.

[20]秦昆,王玉龙,赵鹏祥,徐雯婷,徐源泉.行为轨迹时空聚类与分析[J].自然杂志,2018,40(3):177-182.

[21]秦昆,周劬,徐源泉,徐雯婷,罗萍.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(9):1149-1157.

## 二、研究内容、研究目标及拟解决的关键科学问题

### 2.1 研究内容

本项目对网约车数据和 POI 数据的城市居民出行规律进行研究,主要研究内容包括以下三个方面:

#### (1) 基于网约车轨迹数据挖掘城市热点出行区域

网约车轨迹是蕴含着居民出行行为的地理时空大数据,从网约车轨迹数据中挖掘居民出行的热点区域和移动模式,利用基于网格密度的 GScan 聚类算法,得到数据生成地居民网约车出行热点区域的时空分布,进而分析居民的出行行为。

## **(2) 利用网约车 GNSS 数据探寻居民出行时空分布**

从网约车 GNSS 数据中提取时间特征和空间特征，根据需要分别研究出行 OD 点空间分布、出行 OD 点时间趋势、出行强度空间结构、出行时空模式空间结构、不同空间尺度分布特征。从中探寻居民出行时空分布。

## **(3) 基于网约车数据的城市居民出行需求预测**

在城市的不同交通区域，网约车的出行需求时空分布不均，在研究城市人群出行规律的基础之上，进一步预测不同条件下城市居民出行需求。

## **2.2 研究目标**

(1) 使用并完善基于网格密度的 GScan 聚类方法，并基于网约车数据进行实验分析，进一步结合 POI 数据和时空因素对居民的出行行为进行挖掘，并验证方法的可行性。

(2) 根据研究问题的不同，从网约车数据中挖掘居民出行空间分布的各种特征规律，研究区域的活跃度，进一步从时间和空间维度上考察城市居民出行随时间的动态变化。

(3) 基于深度神经网络回归算法，对不同条件下城市居民出行需求进行预测，根据 DNN 算法的相关步骤，构建 DNN 回归预测模型。

## **2.3 拟解决的关键科学问题**

### **(1) 结合网约车数据和 POI 数据挖掘城市热点出行区域**

利用具体的网约车轨迹数据，提取数据中的上车点与下车点数据，并进行可视化显示。之后利用 GScan 算法对数据进行运算，寻找对应于不同时段数据合适的参数（网格大小，网格密度阈值），对不同出行时段的出行热点区域进行挖掘。可以进一步结合 POI 数据对热点区域进行归类划分。

### **(2) 利用网约车 GPS 数据探寻居民出行时空分布**

从网约车轨迹数据中提取出有用的特征信息后，进行处理、统计，可视化展示。研究不同角度问题上，居民出行空间分布规律。

### **(3) 利用 DNN 算法进行居民出行需求预测**

1. 在 DNN 前向传播算法与反向传播算法中，其使用的损失函数是均方差，而激活函数是 Sigmoid 函数由于在均方差和 Sigmoid 激活函数组合的反

向传播算法中，梯度的变化值是在每一层向前递推的过程中乘以  $f(z)$  的导数。因此，在大多数时候，梯度变化值都很小从而导致  $b$  更新到极值的迭代很慢，该算法的收敛速度较慢。

2. 在深度学习的过程中，如果模型基于一个数据集训练时，其正确率非常高，而当出现该模型未训练过的数据集的时候，模型对该数据集很难做出正确的响应，那么该模型就存在过拟合现象。

### 三、拟采取的研究方案及可行性分析

#### 3.1 研究方法

在研究过程中，项目组将全面了解国内外最新发展动态，特别是有关地理大数据、行为轨迹分析、空间人文社会学、路径优化算法等方向的最新成果，对其相关理论与方法进行合理借鉴与充分利用。

项目组将以武汉市为研究区域，利用经有关公司处理过的武汉市的网约车 GPS 采样数据，并收集和整理武汉市及周边 POI（即兴趣点）数据。同时，利用社会调查获取受教育程度、经济收入、工作状况等社会学数据。利用这些数据为本项目的实验研究提供丰富的数据源。

项目组紧密团结、分工协作，同时与相关单位和部门进行紧密合作。本项目与我院“地理多元流相互作用机制和异常探测相关内容”的研究项目相合作，利用同种数据相互帮助，相互促进。

#### 3.2 技术路线

##### 3.2.1 总体技术路线

总体技术路线图如下图 1。



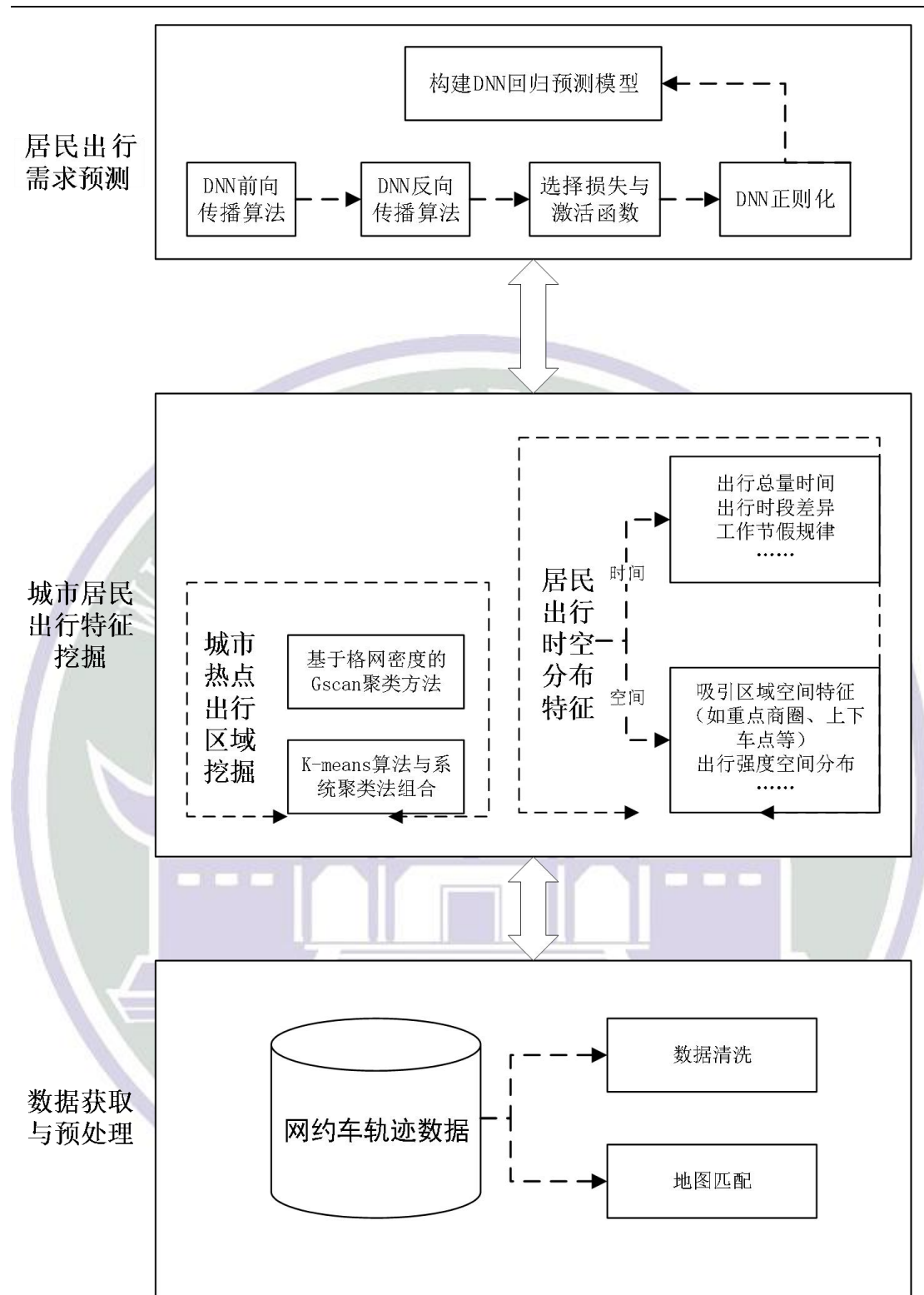


图 1 总体技术路线图

### 3.2.2 各部分技术路线

#### 1) 网约车轨迹数据预处理

网约车的轨迹数据通常由于设备异常或其他人为因素会产生误差,因此为确保计算的精确,应当进行数据的预处理。

网约车的轨迹数据通常由于设备异常或其他人为因素会产生误差,因此为确保计算的精确,应当进行数据的预处理。网约车轨迹预处理主要包括数据格式转换、异常数据清洗、地图匹配等步骤。

### (1) 数据格式转换

将每天的 txt 轨迹数据转换为一个 Shapefile 格式的文件,从而将文本数据空间化。

### (2) 数据清洗

通常数据清洗是综合几种常见的异常错误相对应产生的措施。主要分为以下四点:

#### ① 经纬度越界数据的处理

当数据经纬度不在所研究城市经纬度范围内,说明数据错误,需要删除。

#### ② 重复数据的处理

GPS 设备采集轨迹信息有一定的时间间隙,在网约车原始 GPS 数据中,存在很多相同的记录,当多条记录的“id”字段和“time”字段一致时,此类数据也属于同一辆网约车在同一时间上传的重复数据。对重复数据予以剔除。

#### ③ 异常数据的处理

若 GPS 数据无效,将其删除。若网约车全天载客或全天未载客,删除记录。若网约车的行驶方向取值不在 $[0, 360]$ ,属于异常数据,删除。

#### ④ 记录不全数据的处理

当原始 GPS 数据集中的某条轨迹数据存在数据不全的情况时,比如某天 GPS 记录经度或者纬度为空,那么该条数据的存在,会对后续的出行需求时空分布及热点区域聚类等研究造成影响。网约车行驶过程中的车号、载客状态、时间、经度及纬度等信息中任意至少一个字段的缺失,都会对下文进一步的研究造成影响,所以,对于有缺失的数据,直接剔除,以保证后续研究的数据可靠性。

### (3) 地图匹配

目前地图匹配算法多种多样,其中最常见的是点到点,点到线,线到线这三种。我们预计使用点到线的方法。

以网约车 GPS 点作为圆心，40 米为半径作圆（半径的选取要控制在误差的范围内，GPS 正常误差在 25 米的范围内，经验以 40 米为半径最为合理）。选择这个圆内所有的道路节点，并以这些道路节点所匹配的路段作为待匹配路段，求待匹配节点与所有待匹配路段的距离，选择与此节点距离最短的路段即为最终的结果。

当出现一个待匹配节点对应多条匹配路段或者没有与其对应的匹配路段时，将该网约车轨迹的相邻点所匹配的路段加入到当前点的待匹配路段中。最后，距离点 P 最近的待匹配路段即为最后的修正点。

在得到匹配点的匹配路段后，首先判断该点处于道路线段的范围内还是范围外。假设网约车 GPS 点在道路线段范围内，可以直接求该点在该路段的投影点，并求出投影点的经纬度坐标，该投影点即为修正后的点。

当出现网约车 GPS 点在道路线段范围外时，匹配路段上距离该点最近的线段点即为投影点，该线段点的经纬度坐标即为投影点的经纬度坐标，该投影点即为修正后的点。

## 2) 结合网约车数据和 POI 数据挖掘城市热点出行区域

方法 1：利用基于网格密度的 GScan 聚类方法。首先，对空间数据对象所占的区域进行网格划分，将研究区域范围划分成互不重叠的若干个大小为  $k \times k$  的正方形网格单元；其次，遍历原有数据集，将原始空间数据通过映射函数映射到所属的网格单元；接着，通过计算所属网格单元中的数据密度，识别网格集合中的热点网格单元，剔除其中的普通网格单元；最后，根据热点网格单元中的数据点的分布计算热点网格单元的位置，并通过计算网格单元间的距离来合并热点网格单元从而得到热点区域。其中还可对不同时间段进行分类计算，得到指向更加明确的出行热点区域分布。

方法 2：通过 K-means 算法与系统聚类法组合的方式对大规模的数据进行聚类分析，由 K-means 算法进行降维处理，提高系统聚类法的时间效率，然后利用系统聚类法来对 K-means 算法 K 值的选取进行指导。

(1)首先根据先验知识，确定 K 值的大小，然后利用 K-means 算法对 GNSS 数据进行第一次聚类；

(2)比较各聚类后的类别阈值与设置阈值之间的关系，剔除掉不合理的聚类，



将剩余的聚类类别个数和聚类中心作为第二次 K-means 算法的初始值；

(3)设置第二次 K-means 算法聚类的 K 值和初始聚类中心为上一步得到的值并开始聚类，得到聚类类别个数和聚类中心；

(4)对第二次 K-means 算法聚类得到的聚类中心进行系统聚类，分析聚类结果，并将本次系统聚类类别个数和第二次 K-means 算法聚类得到的聚类中心分别作为第三次 K-means 算法聚类的 K 值和初始聚类中心，开始第三次 K-means 算法聚类，得到最终聚类结果。

在该 K-means 算法与系统聚类法组合方式聚类过程中，前两次通过 K-means 算法聚类，对网约车 GNSS 数据进行了降维处理，为之后的系统聚类做预处理。其后的系统聚类是在前两次 K-means 算法聚类降维的基础上，进行聚类，为其后第三次 K-means 算法聚类 K 值的选取提供指导。最后通过第三次 K-means 算法聚类进行分析。

### 方法 3：基于时空数据场的城市热点区域提取

利用数据场理论进行网约车轨迹的聚集性分析首先需要对势函数进行时空扩展。引入归一化的时差系数作为权重参与数据场势值计算，可描述两点间势值随时间和空间的衰减效应，从而实现对数据场的时空扩展。

计算出各上(下)车车点的势值，再将势值相等的点连接成线构成等势线，由等势线的分布可知，在场势的作用范围内，若两个上(下)车点间的势值相差越小，则它们被划分为同一类的可能性越大，可利用数据势场中等势线的自然嵌套结构对数据对象进行分类。因此，利用时空数据场可以对上(下)车点数据的聚集区域进行探测，从而实现热点区域的提取和分析。

### 3) 基于网约车 GNSS 数据的居民出行空间分布

出行 OD 点指网约车乘客的上车点和下车点，蕴含着丰富的信息，可以据此了解某地区人流活动最密集的地区的分布。具体研究方向如下：

①分别对工作日和休息日出行点数据进行核度分析，基于城市尺度，选择合适的半径观察出行点分布规律。

②分别对工作日和休息日每个小时的上车点和下车点数量进行统计，可得到工作日和休息日 OD 点时间变化图，研究出行 OD 点时间趋势。

③将研究区域划分为多个时空单元，然后按日期将所有时空单元分段，根据

网约车上车点与下车点的和描述各段时空单元的出行强度,最后从中提炼出城市的出行强度空间结构。

④城市中不同地区的一天中各个时段会产生独特的出行发生与吸引特征和出行强度特征,这两大特征共同组成了城市出行时空模式。对城市各个空间单元的出行发生吸引时空特征和出行强度时空特征进行聚类,再经过分类与归并的过程就是出行时空模式识别。

城市的出行发生吸引时空特征指某个地区出行发生与吸引随时间变化的模式。可以进一步根据出发性质与吸引性质的强弱。通过一段时间内下客与上客的差值来判断一个地区在某一段时间是属于发生区还是吸引区。出行强度根据出行发生和出行吸引的数量之和来体现。对于每个时间单元,可以得到研究区域两个出行特征的空间画像;而对于每个空间单元,可以得到全时间区间的出行特征变化模式。

简单而言,在每个空间单元上,上客数与下客数之和在时间上的积累形成了出行强度时空特征,上客数与下客数之差则形成了出行发生吸引时空特征,它们共同组成出行时空模式,空间单元上的出行时空模式在空间上的积累形成城市的出行时空模式。

结合网约车 GNSS 数据和空间规划数据,分析不同规划分区、不同行政区、重点交通枢纽、重点商圈的网约车上下客空间分布规律。分别对上车位置、下车位置的空间分析按照每天、工作日与休息日、不同时间段的时间粒度进行统计分析,结合真实数据具体分析各个时间段,空间尺度上居民出行的分布规律。

#### 4) 出行需求时间分布特征提取

(1) 出行总量时间统计。网约车分空车状态和载客状态,将车辆按车牌排序,以空车,载客状态变化 POI 点为上下车点。提取网约车上下车特征点,并进一步筛选,剔除冗余数据。

(2) 出行时间段差异统计。选用单峰阈值法设定单一阈值,提取热点区域。将城市分为不同行政区域,统计不同时间段内热点区域分布,或以较小的时间间隔统计网约车行驶数据的动态变化,获得城市动态热区分布。以此得到时间分布规律。

(3) 工作日与节假日出行统计。工作日与周末，节假日的出行差异是人群出行规律的重要反映，可统计工作日与节假日出行量或基于网约车轨迹数据分析城市不同时段拥堵情况来相关分析人群出行规律。

(4) 出行时长统计。不同人群在不同时间有着不同的出行需求，单次出行时间也是人群出行规律的重要反应，通过统计时段内各特征点上下车时间间隔总量的期望来得到时间段内出行时长。

(5) 网约车运营效率。网约车订单之间的时间间隔，即相邻下上车特征点时间间隔统计，是反映网约车运营情况的一个参考标准，也可以反映城市人群出行密度。

(6) 网约车类型时间分布。统计不同时段内人群对网约车类型的选择，例如滴滴打车、代驾、拼车等类型约车方式的使用情况。

(7) 群体差异性特征与网约车使用频率相联关系。通过统计群体特征（如年龄，性别，职业，收入情况等）与网约车使用频率的统计来分析得到二次特征。

## 5) 居民出行需求预测

### (1) DNN 前向传播算法

输入：DNN 总层数  $I$  输入值向量  $x$ ，所有隐藏层和输出层对应的权重系数矩阵  $W$  偏倚向量  $b$ ；

输出：输出层输出  $a_I$ 。

步骤 1：初始化  $a_1=x$ 。

步骤 2：通过相关公式计算输出值  $a_i$

步骤 3：输出  $a_i$ ；

### (2) DNN 反向传播算法

输入：DNN 的总层数以及训练样本  $(X_1, Y_1), (X_2, Y_2) \dots$ 。每一层的神经元个数，最大迭次数  $M$ ，阈值  $\epsilon$  以及迭代步长  $\alpha$ ；

输出：各隐藏层与输出层对应的权重系数矩阵  $W$ ，偏倚向量  $b$ 。

步骤 1：初始化各隐藏层与输出层对应的权重系数矩阵  $W$ ，偏倚向量  $b$  为随机值；

步骤 2：从 1 到  $M$  开始迭代：

步骤 2-1：for  $i$  in  $(1,m)$

(1)将 DNN 的输入  $a_i$  设置为  $x_i$



(2)for  $i$  in(2,L), 通过相关公式计算  $a_i$

(3)通过损失函数计算输出层

(4)for  $i$  in(L-1,2),通过公式计算所需参数

步骤 2-2: for  $i$  in(2,L)计算 1 层的  $W_1$  和  $b_1$

步骤 2-3: : 当所求的权重系数矩阵和偏倚向量满足小于阈值或迭代次数达到最大迭代次数条件的时候, 转到步骤 3;

步骤 3: 输出  $W$  和  $b$  的值。

(3) 选择合适的 DNN 损失函数与激活函数

(4) DNN 正则化

(5) 构建 DNN 回归预测模型

步骤 1: 首先载入出行需求数据集, 然后通过设定训练集和测试集的比例, 将出行需求数据集划分为训练集和测试集, 训练集用来学习需求特征间的关系, 训练神经网络回归模型, 预测集用来在后续步骤中, 使用训练好的模型对测试数据进行预测;

步骤 2: 定义特征列。根据出行需求特征, 将其转化为定义的特征列 FeatureColumn, 该特征列即为在测试及预测阶段的模型输入;

步骤 3: 模型构建。因为基于网约车 GPS 数据出行需求的预测值为连续变量, 因此, 需要构建深度神经网络回归模型, 然后设置模型的输入特征、各隐藏层层数及神经元数量等, 选择优化器及激活函数, 设置学习率及正则化参数, 最后指定模型保存目录;

步骤 4: 定义传入特征的数据格式, 该步骤包括定义训练阶段使用的数据格式 train\_input\_fn 和测试阶段使用的数据格式 test\_input\_fn ;

步骤 5: 模型训练及测试。在定义传入模型后开始模型训练。训练阶段和测试阶段所用的数据不一样, 训练阶段的数据为相同星期属性历史时刻的出行特征及出行需求, 测试阶段所用的数据为预测日期的出行特征;

步骤 6: 在完成了深度神经网络出行需求回归预测模型的构建之后, 对出行需求进行预测。

### 3.3 可行性分析

#### (1) 投资必要性

在当今的网络时代，人们的工作、生活、消费、旅行等行为正在不断通过泛在传感器的使用而被完整和系统地记录下来并形成数据，改变了过去由于科技的局限，海量的个体信息获取成本巨大、时效性低，而无法有效获取大量的人类行为数据进行研究的状况。这为研究人类行为模式提供了前所未有的机遇和挑战。人类行为模式相关研究不仅可以推动已有理论方法的变革，同时在城市规划、交通设计、商业选址、信息服务等应用领域也可以产生积极而深刻的影响。武汉市是常住人口超过 1000 万的特大城市，城市人群流量巨大，城市交通压力大，人群出行规律研究具有迫切性。所以，基于数据的人类行为模式研究具有必要性。

## **(2) 技术可行性**

网约车数据的特点：

1) 覆盖范围广。网约车轨迹数据在时间和空间尺度上的覆盖范围比其他交通轨迹数据更广。网约车运营时间可达全天 24 h，而且网约车行驶在城市交通路网中，不受线路制约。

2) 采样密度高。网约车数据能完整地记录车辆的行驶路径，具有时空序列性和连续性。

3) 位置精度高。网约车轨迹数据是通过车载 GNSS 采集获得的网约车位置数据，其精度较高，一般为 5~20 m。同时由于车辆始终行驶在城市路网中，因此可以通过地图匹配等方法进一步提高数据的位置精度。

4) 数据规模大。由于城市网约车数量多、运营时间长、采样密度高，因此网约车轨迹数据集往往规模庞大。

5) 蕴含信息丰富。网约车轨迹数据中记录的最直接信息是网约车在每个时间点的位置坐标，这些连续坐标点刻画了网约车在一段时间内的移动轨迹。移动轨迹体现网约车司机的运营特征，包括寻客策略、路径选择偏好、载客范围或欺诈绕路现象等。

所以基于网约车数据研究城市人群行为模式是可行的。

项目组将充分借鉴和利用已有的经过充分检验的研究方法和理论，并在此基础上结合研究目标和实际需求发展创新，对已有的方法和理论做改善或在实践的基础上开发全新的方法和理论。所以，本项目在技术上是可行的。

## **(3) 经济和社会的可行性**

网约车轨迹数据是城市交通状态、城市结构规划及人口流动规律的数据映射,挖掘出的网约车轨迹数据和研究出的城市人群出行规律的成果在智能交通、资源与环境保护、城市规划、社会感知等方面可以得到充分应用,为改善城市交通服务、优化城市规划、提高市民生活质量做出不小的贡献。

本项目的所有研究行为,遵守法律道德准则,旨在顺应学术研究潮流和引导学术研究方向,提高人民生活,推动经济社会持续健康和谐稳定的发展。

#### (4) 研究实力的可行性

项目指导老师秦昆,武汉大学遥感信息工程学院教授,博士,博士生导师。兰州大学自然地理学学士、南京师范大学地图学与遥感硕士、武汉大学摄影测量与遥感博士、计算机科学与技术博士后、伊利诺伊大学香槟分校访问学者。研究方向为:时空大数据分析、智能空间信息处理、空间人文社会科学。主持完成国家自然科学基金项目 2 项,主持或参与其他国家级科研项目和横向课题多项,出版教材 2 部,发表论文 100 多篇。

项目组负责人陈宇,2019 级遥感信息工程学院地理信息系统方向,学号 2019302130136。共青团员,武汉大学优秀学生丙等奖学金获得者。对编程语言具有浓厚兴趣,视野广阔,善于接纳、学习新事物。具备一定的团队管理和协作能力,能够发现并发挥组员的优势,接受组员提出的建议。求知欲高,勤奋细心,乐于总结、钻研新的知识。本人希望通过在项目中担当负责人,发挥团队的最大优势,提高自身的自主创新和团队协作能力。

项目组成员王宇帆,2019 级遥感信息工程学院地理信息系统方向,学号 2019302130328。共青团员,武汉大学优秀学生乙等奖学金获得者,曾荣获优秀共青团干部、院级社会活动积极分子等荣誉称号。现任遥感院 1906 班班长。本人参加过 2020 年度“有爱无碍——公交车无障碍出行”省级大创项目,并在其中发挥积极作用,有一定的创新创业经验;且有较全面的程序语言能力,掌握 C、C++、java、matlab 等编程语言,并善于实践运用等操作;且有较强的数理基础,在高中时有两年物理竞赛的学习经历,并获得获 34 届,35 届全国中学生物理竞赛湖北赛区二等奖。希望在项目中发扬刻苦钻研的精神,锻炼自主创新能力和团队协作能力,在团队中多多奉献自己,提高各方面综合素质。

项目组成员周佐翼,2019 级遥感信息工程学院空间信息与数字技术方向,



学号 2019302130313。共青团员，担任班级文艺委员、院合唱团宣传委员，曾获丙等奖学金、院级优秀学生干部等奖项和荣誉称号。经过一年半的学习，综合编程能力显著提高，能够熟练使用 C、C#、C++、java、matlab 等编程语言，对 javascript 也有粗浅涉猎。对编程、空间信息、社会地理信息科学等领域充满浓厚兴趣，愿意投入精力和时间学习相关知识，提高水平。兴趣广泛，常思进取，关心时事，重视团体。遥感学院科研团队之精神深入人心，之成果硕果无数，令学生憧憬且向往。希望通过该项目提升自我，培养开拓精神和创新协作能力，更愿为院校团体献智献力，争取荣誉。

项目组成员周祎博，2019 级遥感信息工程学院地理信息工程方向，学号 2019302130289。对基本的几门编程语言均熟练掌握并勤于运用，同时热切希望能够进一步锻炼编程能力并参与遥感科学领域的研究，愿意花费时间，提高自己创新能力，团队合作能力。本人平日勤奋刻苦，热爱学习，团结同学，热爱集体，能够接受新事物，乐于学习新知识。希望能在项目中学到更多的东西，也希望能够将自己所学的知识转换为对社会更有用处的东西。

项目组成员刘子睿，2019 级遥感信息工程学院遥感仪器方向，学号 2019302130322。有一定的软件编程基础。对新兴事物有浓厚兴趣，有意愿参与到此类项目的研究，提高自己的思维创新能力。愿意付出时间精力，有浓厚学习欲望。希望能在过程中锻炼自己，提高自己综合素质和专业能力。

## 四、项目的特色与创新之处

### 4.1 项目创新点

#### (1) 数据来源创新

固定设备数据只能粗粒度地描述人群或车辆在不同固定位置之间的移动，覆盖范围有限；公交车轨迹数据也只能记录运营时间内公交车的固定路线轨迹。网约车不受线路和时间的约束，是最灵活、覆盖范围最广的轨迹数据，且精度较高、较少涉及隐私问题。本项目以网约车数据为数据源，从得到的数据信息初步了解到网约车用户用车时间和频率的群体差异性，并由此入手，分析这些差异性的来源，用户群体特征。

## （2）技术手段创新

由于城市网约车数量多、运营时间长、采样密度高，因此网约车轨迹数据集往往规模庞大，而且蕴含信息丰富。要从如此大规模的信息数据中提取研究所需的数据，必须结合实际需求和研究目的，根据一定的科学方法对数据进行筛选和过滤，并提取有用信息。现有的数据挖掘算法多样而效果不一，项目组将深入研究各种方法，结合它们各自的优缺点，开发出一个适用性更强、效果更可靠的信息提取算法。

## （3）方法运用创新

结合城市网约车、POI 大数据分析，深入分析网约车数据体现的时空规律；结合社会调查数据，从多角度分析众多因素与人群行为模式的关系，以此体现出城市网约车用户大体出行规律。对近年网约车用户 POI 数据及城市交通出行数据进行对比分析，深入分析数据差异，群体差异，判断网约车用户出行规律是否能较准确的代表城市人群。又或者导致明显差异的原因及群体特征。通过对网约车用户进行时间分类，建立模型分析长时间来网约车用户的变化规律，以此判断网约车出行方式在人们出行方式中的发展规律及未来发展方向。得出采用网约车用户群体的潜在规律。

随着传感网、物联网、射频识别技术与智能嵌入技术的发展，信息社会正逐步转变为泛在信息社会，人与物、物与物之间交流所产生的泛在信息关联分析将带来大量潜在新型应用。网约车轨迹数据本身语义信息有限，与其他泛在信息如 POI 数据和社会调查数据等进行联合分析能扩展应用领域范围，进行多领域多学科的综合应用研究，有效解决城市问题。

## （4）服务理念特色

本项目进一步深入挖掘人群行为轨迹数据中蕴含着的反映行为规律与社会交往活动的丰富知识，响应了科学发展需求。以轨迹数据为数据源，同时结合环境、交通、社会经济等数据，从空间交互的角度，研究人与城市之间的相互影响和交互作用，是地理信息科学、管理科学和人文社会科学共同面临的挑战性课题，迫切需要在理论方法方面有所创新，才能促进其研究水平不断提升。同时，本项目旨在利用网约车轨迹大数据反映城市居民的移动行为和出行偏好，为城市规划优化、改善城市环境提供依据和支撑。让健康的人生活在健康的环境里，让生活

在城市中的人都有安全感是本项目的一个理想目标。

## 五、研究计划及预期结果

### 5.1 年度研究计划

(1) 2021 年 3 月——2021 年 12 月

1) 有关调研工作与资料的收集和整理，研究方案和技术路线的细化，拟定详细的研究计划和研究人员的详细分工；

2) 对获得的网约车轨迹数据进行预处理和地图匹配，并完成出行特征提取；

(2) 2021 年 12 月——2022 年 3 月

1) 根据预处理后的数据研究居民出行的时空分布规律；

2) 完成基于网约车数据的热点区域挖掘；

3) 项目中期报告；

(3) 2022 年 3 月——2022 年 9 月

1) 根据前期获得的数据结论完成出行需求预测；

2) 利用建立的实验平台，对研究成果进行全面实验分析；

3) 成果整理，完成项目结题报告，准备项目验收；

4) 发表 1 篇高水平学术论文。

### 5.2 预期研究成果

1) 根据武汉网约车数据得出武汉居民出行的时间与空间分布特征；

2) 挖掘武汉市的热点区域，得出不同时段，不同空间尺度上的居民出行规律；

3) 完成出行需求预测，为居民出行，网约车司机提供出行，接单的策略依据；

4) 在国内核心期刊上发表 至少 1 篇学术论文；

## 六、详细预算

### 6.1 小宗研究设备费用

硬盘：由于涉及大型数据库的管理与操作工作，需要进行大容量硬盘的购置。分别为图像用户比对盘与个人轨道储存盘。以二手 2tb 西数黑盘为例，购置两块



价格共计 764 元。

## 6.2 文印出版费用

- 1、拟在国内核心期刊上发表本项目主要研究成果，所需版面费约 2000 元；
- 2、文印费：包括所有成果展示中主要形式的各类文案，均以一式五份的方式进行印制。其中加入一份汇编成册的文献整理集册。本部分消费累计为 236 元左右。按本科生院要求进行。

表 6-1 开支明细

开支科目	预算经费 (元)	备注
小宗研究 设备费用	764	硬盘：购置两块的价格共计 764 元。
出版费用	2000	拟在国内核心期刊上发表本项目主要研究成果，所需版面费约 2000 元。
文印费用	236	包括所有成果展示中主要形式的各类文案，均以一式五份的方式进行印制。另有一份汇编成册的文献整理集册。

总计费用：3000 元

导师意见：

签 名：

年 月 日

院（系）意见：

签名盖章：

年 月 日

学校意见：

年 月 日

