中国大学生计算机设计大赛

作品编号：2024039338

作品名称：基于多源数据融合的人群出行模式识别系统

作　　者：黄麒文、饶旭乾、史晓岩、邓方利

版本编号：V1.0

填写日期：2024 年 4 月 30 日

目录

第一章　需求分析

1.1 开发背景

在中国城市化迅速发展的今天，城市交通需求的激增给城市空间及其基础设施建设造成了空前的压力。有效地分析和理解城市居民的出行模式成为了研究城市交通发展状况的关键。面对传统问卷调查方法的低效与高成本，大数据技术在城市交通中的应用开辟了新的研究途径。手机信令数据作为城市大数据的重要组成部分，因其覆盖面广、获取便捷及时效性强等特点，已经成为分析城市交通模式的重要工具。然而，目前许多系统仅依赖手机信令数据的隐含特征来识别交通模式，忽略了例如地铁出行的乘客轨迹经过地铁站点的行为特征，导致有偏差的识别结果。此外，手机信令数据的庞大体量也为研究带来了挑战，如何通过大数据技术降低处理延迟，并实现出行轨迹和流量的实时提取及可视化，也是当前研究的核心问题之一。

针对以上问题，本系统通过整合多源数据来优化人群出行模式识别，并基于 Flink 流计算引擎和 Alink 机器学习平台，开发了一款人群出行模式识别系统。

1.2 项目目标

本系统利用多源数据进行人群出行模式识别，并实现了人群流量的可视化，旨在帮助交通规划部门和城市规划部门提供详细的人群出行方式结构信息和人群热点分布状况，帮助政府合理分配公共资源以及有效监控和管理热点地区人群聚集，以促进城市交通系统的优化和城市管理的提升。本系统还可以为应急管理部门评估紧急情况下的影响范围和人群分布，为紧急响应和救援行动规划提供支持。

1.3 系统目标

利用多源数据实现对人群出行模式识别，并利用大数据可视化大屏界面，将计算获得的人群出行模式通过百度地图 API 热力图和 Echarts 图表进行可视化展示。详细功能如下：

（1）利用手机信令数据进行实时轨迹提取，该功能细分为手机信令数据清洗、轨迹流提取、驻留分析。

（2）根据获取的用户出行轨迹进行出行模式识别并引入公共交通站点数据对识别结果进行纠正，该功能细分为该功能细分为人群出行流量计算、出行特征构建以及出行方式聚类分析。

（3）通过构建可视化界面对人群出行模式分析的结果进行展示，包括人群各出行方式占比、人群各出行方式平均速度、人群各出行方式距离、人群各出行方式平均时间的 Echarts 图表展示和人群驻留情况、人群密度的热力图展示。

第二章　概要设计

2.1 用户需求分析结果

（1）系统可以基于 Flink 流计算引擎对海量手机信令数据进行实时处理，并提取大量实时无界的轨迹数据，并进行人群出行流量分析和人群驻留分析；

（2）系统可以基于 Alink 机器学习平台根据取轨迹流中的出行特征，识别出不同的交通出行方式，并应用公共交通站点对识别的交通出行方式进行校正；

（3）用户能够通过大数据可视化大屏查看人群出行流量和驻留地点的热力图，以及通过 Echarts 图表呈现的各种出行方式的占比、平均速度、平均距离和平均时间等统计数据。

2.2 功能模块划分

本系统主要分为三个主要功能，分别为实时轨迹提取功能、出行模式识别功能、出行模式可视化功能。详细的功能模块如图 1 所示。

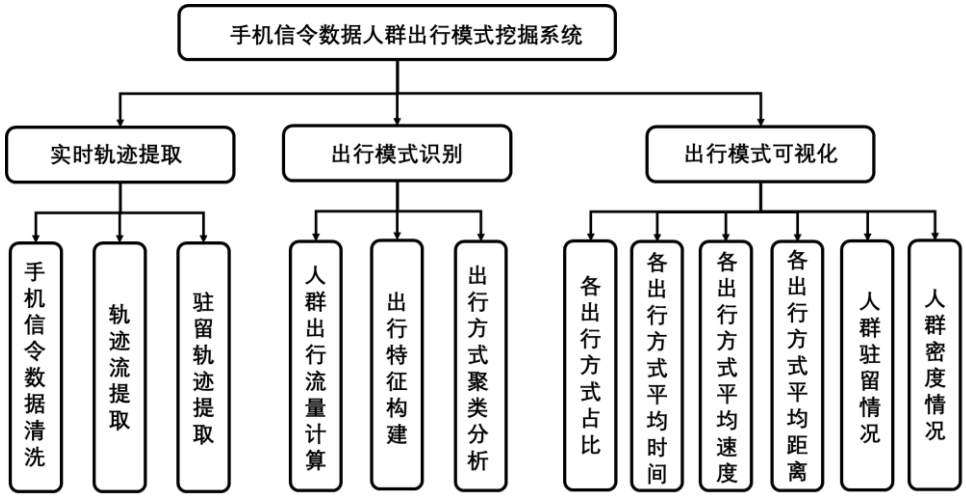


图 1 系统功能结构图

2.2.1 实时轨迹提取功能

实时轨迹提取功能主要是通过原始的手机信令数据进行处理获得手机用户的移动轨迹。该功能细分为手机信令数据清洗、轨迹流提取、驻留分析等功能。手机信令数据清洗包括除去实时产生的原始手机信令数据中出现的不完整数据，同时将原始手机信令数据转换为数据流存入消息队列中。轨迹流提取功能是指从实时产生的清洗后的手机信令数据中提取每个用户的轨迹，这一过程还包括对轨迹中的乒乓数据以及漂移数据进行优化。驻留分析功能是指对一段时间长期处于同一区域的轨迹进行筛选。

2.2.2 出行模式识别功能

出行模式识别功能根据实时获取的用户出行轨迹计算人群出行流量并判断用户的出行方式。该功能分为人群出行流量计算、出行特征构建和出行方式聚类分析。其中人群出行流量计算是指从清洗后的手机信令数据中，获取离手机信令数据最近的基站点位从而计算出人群的出行流量。出行特征构建是指从获取的有效出行轨迹中分析其速度、出行距离、出行时间等特征。出行方式聚类分析功能是指根据出行特征利用机器学习聚类算法对出行模式进行聚类，并依据公共交通站点对识别出为公共交通出行方式的类别做二次校正。

2.2.3 出行模式可视化功能

出行模式可视化功能主要通过构建可视化界面对人群出行模式分析的结果进行可视化展示，包括人群各出行方式占比、人群各出行方式平均速度、人群各出行方式距离、人群各出行方式平均时间的 Echarts 图表展示和人群驻留情况、人群密度的热力图展示。

2.3 系统体系架构

考虑到人群出行模式可视化结果的可访问性，在设计系统时我们采用了 B/S 结构。鉴于数据量庞大，本系统的数据处理主要在客户端进行，而处理后的可视化图表与热力图则通过 Web 端展示。用户只需通过网页即可查看和分析处理后的可视化结果，从而提升了系统的使用便捷性和效率。在体系架构方面，本系统采用分层架构设计，分为数据层、服务层和功能层。详细的系统体系架构如图 2 所示。

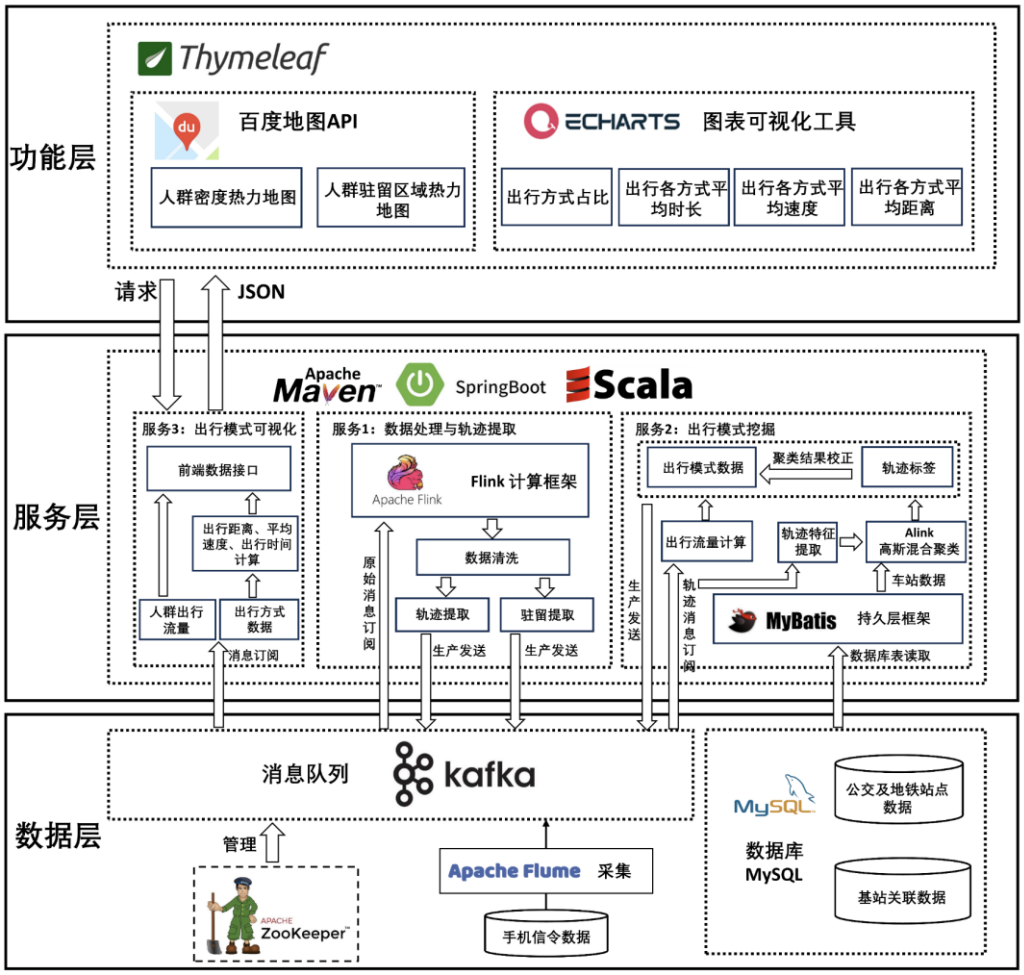


图 2 系统体系架构

数据层主要负责存储流数据和静态数据，其中流数据通过 Kafka 消息队列进行暂存，静态数据则存储在 MySQL 数据库中。

服务层由三个服务组成，分别负责数据的清洗、轨迹流提取、人群出行流量计算及特征工程构建，通过高斯混合聚类分析对出行方式进行识别，并引入公共交通站点数据对识别结果进行校正。

功能层则使用 Thymeleaf 框架和百度地图 API 实现人群出行模式的可视化展示，包括热力图和基于 Echarts 的动态图表，展示人群流量、人群驻留、出行方式的占比、速度、距离和时间等关键信息。

第三章　详细设计

3.1 功能设计

本系统主要分为三个主要功能，分别为实时轨迹提取功能、出行模式识别功能、出行模式可视化功能。功能模块划分在概要设计中功能模块设计已具体介绍，在此便不再赘述，本处主要介绍各个功能的详细设计。

在本系统中，使用 Kafka 消息队列作为实时数据流存储的核心，实现了手机信令数据及其分析生成的实时数据的存储与传输功能。系统设置了五个 Kafka Topic 分别为 timeStream、heatMapData、stayData、routeChain、GaussianMixture，分别用于存放原始手机信令数据、人群流量数据、驻留轨迹数据、轨迹流数据以及经高斯混合模型聚类后的出行方式标签化轨迹数据。各功能模块将通过这 5 个 Topic 传输实时数据流。

3.1.1 实时轨迹提取功能

（1）手机信令数据清洗

本系统中，手机信令数据通过文本文件作为数据源被系统读入，再以一定时间间隔向 Kafka 消息队列中的 timeStream 主题持续发送手机信令数据，以该过程来模拟实时手机信令数据接入系统。在该过程中，首先会创建生产者身份，并读取手机信令数据。利用 Flink 算子对手机信令数据进行预处理，将得到的数据流以生产者的身份向 Kafka 的 timeStream 主题发送预处理后的手机信令数据。具体的程序时序流程如图 3 所示。

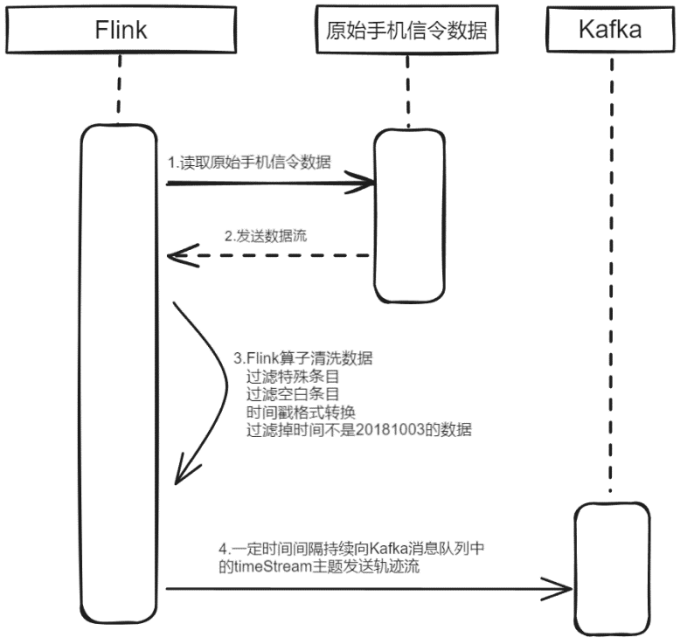


图 3 手机信令数据清洗流程时序图

（2）轨迹流提取

轨迹流提取功能设计为首先将该功能模块注册为消费者，当 Kafka 消息队列的 timeStream 主题中产生清洗后的手机信令数据时，该模块监听 timeStream 主题并实时拉取其中的数据用于轨迹流的提取。对于拉取得到的流数据，利用 Flink 流计算引擎提取轨迹，该计算过程包括对每一个信令数据匹配基站 id 获取数据库基站表中对应的经纬度信息，构

建时间窗口对时间区间内的手机信令数据提取对应轨迹数据，优化乒乓轨迹以及漂移轨迹。最终再注册为生产者向消息队列中的 routeChain 主题发送轨迹流数据。具体的程序时序流程如图 4 所示。

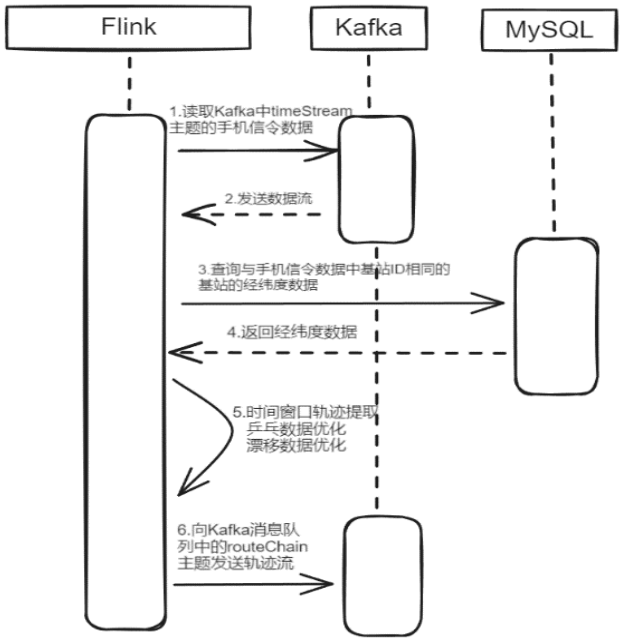


图 4 轨迹流提取过程时序图

（3）驻留分析

由于驻留分析需建立在已获取轨迹流的前提下，但为提高整个可视化系统的底延时性，该过程将与轨迹流提取同时进行，不等待轨迹流提取完成。该功能前置步骤设计为与轨迹流提取过程一致，首先对提取轨迹流，在轨迹流提取的时间窗口中，添加对轨迹长时间驻留的提取器，得到指定驻留时间的轨迹流。最后注册生产者身份，将驻留轨迹流向消息队列的 stayAnalyse 主题发送。

3.1.2 出行模式识别功能

本系统将该功能模块设计为一个持续等待轨迹流数据的过程。该模块将注册为消费者，当 timeStream 中产生清洗后的手机信令数据后，该模块将对数据流以 5s 为时间窗口计算当前的人群出行流量，并创建生产者身份向 heatMapData 主题发送人群流量数据流。当实时轨迹流在 Kafka 消息队列的 routChain 主题中产生时，该模块会实时对该主题的消息队列进行消费，该数据流分批送入出行特征构建模块以及出行方式聚类分析模块中，识别轨迹流出行模式类别。最后注册为生产者，将识别出出行类别打上标签的轨迹流向 GaussianMixture 主题的消息队列发送，该过程时序图如图 5 所示。

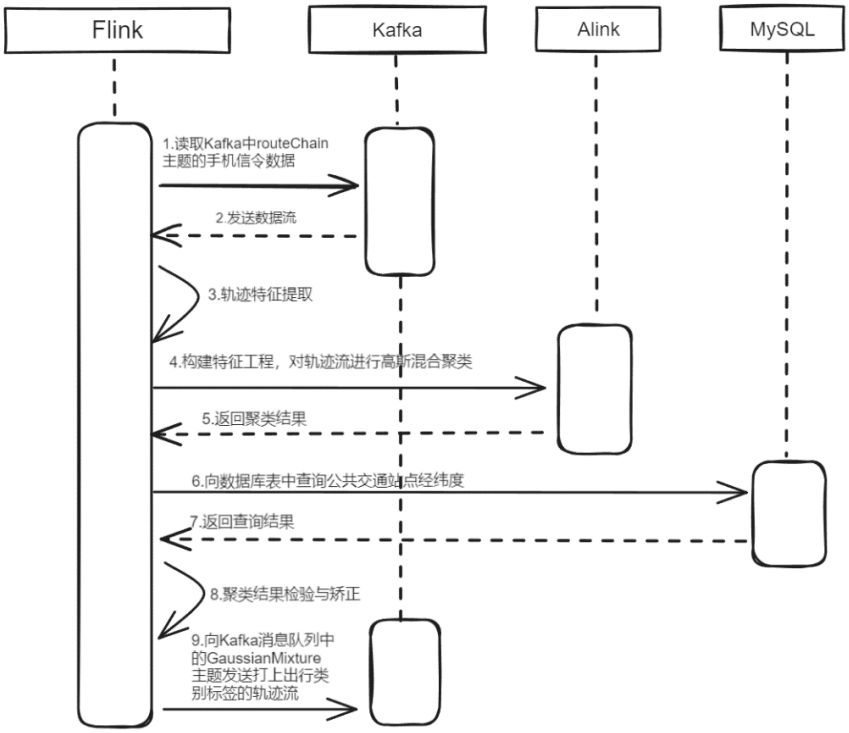


图 5 出行方式聚类分析过程时序图

3.1.3 出行模式可视化功能

该功能模块主要需要构建后端数据接口以及前端可视化界面。对于后端服务，当Kafka中 GaussianMixture、stayAnalyse、heatMapData 主题的消息队列产生数据时，后端服务将会注册为消费者实时对以上主题的数据流进行消费，提取数据流中的出行模式信息，并将其封装为 Json 格式构建成接口以供前端可视化服务调取。前端可视化界面需要以一定的时间间隔向后端所提供的接口发送请求，以获取最新的出现模式数据。通过请求获取该数据后，需要对接收的数据进行解析并调整为可视化组件所要求的数据格式，进行正常的可视化展示。

3.2 系统界面设计

3.2.1 系统界面设计图

依据出行模式可视化功能的设计，本系统需对实时挖掘分析出的出行模式信息进行可视化展示，这包括各出行方式占比、各出行方式平均速度、各出行方式平均距离、各出行方式平均时长、人群密度以及人群驻留情况等信息。本系统界面设计参考主流的大数据可视化大屏，通过动态图表以及热力图对实时数据进行展示。对于人群实时密度情况、人群驻留情况通过百度地图 API 热力地图进行可视化展示，对于各出行方式占比、各出行方式平均速度、各出行方式平均距离则通过 Echart 图表进行可视化，界面设计如图 6 所示。

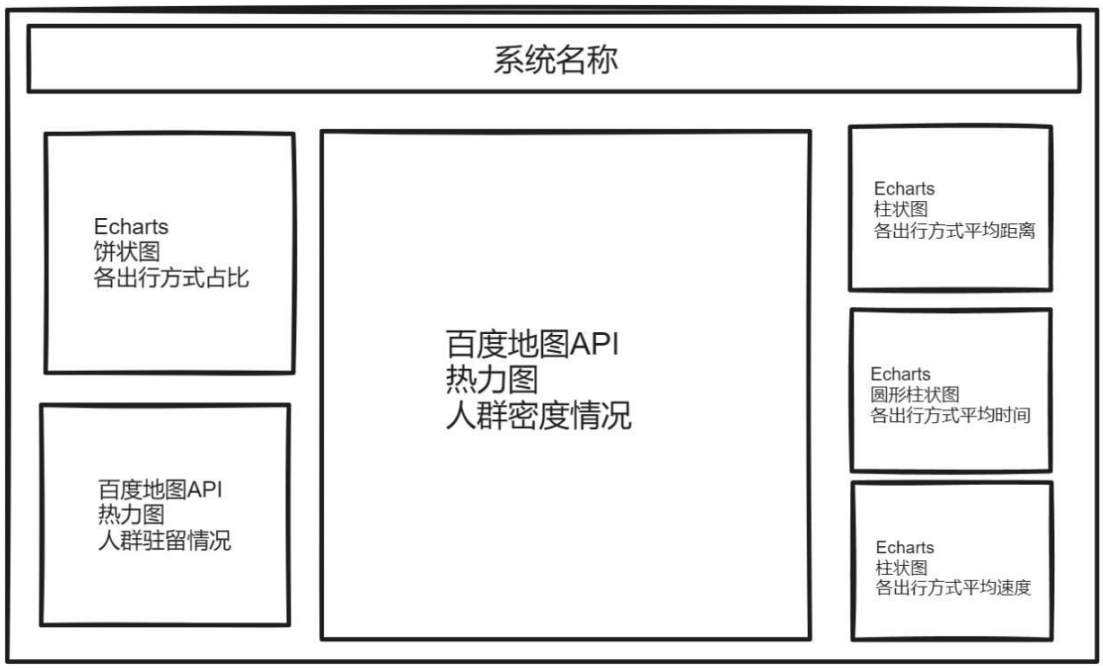


图 6 系统界面设计

3.2.2 界面效果展示图

图 7 系统界面效果图

3.3 数据库设计

3.3.1 基站表设计规范

根据具体的功能需求，在实时轨迹提取功能中，原始手机信令数据缺少经纬度信息，需要根据数据中包含的基站 ID 信息向基站信息中查询并获取基站的经纬度信息。其所对应的数据库表设计如表 1 所示。

表 1　基站表设计规范

数据项名	字段	数据类型	长度	说明
经度	Longitude	Varchar	50	Not null
纬度	Latitude	Varchar	50	Not null
基站 id	Laci	Varchar	50	Not null

3.3.2 交通站点表设计规范

在对轨迹类型初步分类后，需要对其进行再次校正以排除误分类的情况，本系统利用公共交通站点对其进行检验并校正。其所对应的数据库表设计如表 2 所示。

表 2　交通站点表设计规范

数据项名	字段	数据类型	长度	说明
经度	Longitude	Varchar	50	Not null
纬度	Latitude	Varchar	50	Not null
站点类型	Mode	Varchar	50	Not null

3.4 关键技术

3.4.1 分布式消息队列 Kafka

Kafka 是一个分布式消息队列系统，旨在处理大规模实时数据流。它由 LinkedIn 公司开发，以其高扩展性、高吞吐量和低延迟的特性而广受欢迎。Kafka 支持数据以主题形式组织，并可分布在多个分区和服务器上以提高性能。系统中的生产者将消息发送到指定主题的分区，而消费者从这些主题拉取数据。Kafka 的这些特点使其成为实时数据传输、日志聚合及实时分析等场景的理想选择。Kafka 的基本工作原理如图 8 所示。

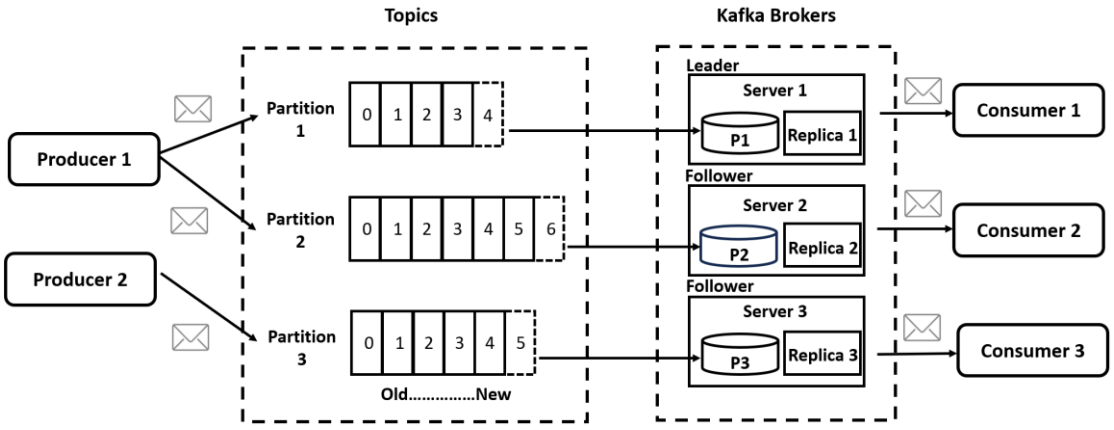


图 8 Kafka 基本工作原理示意图

3.4.2 实时流计算引擎 Flink

Flink 是一种高效的流式处理引擎，专门设计用于实时处理和分析大规模数据流。它的

核心优势在于真正的流数据处理能力，具备高度的可扩展性、容错性和高可用性。Flink 支持精确的时间处理机制，包括事件时间、处理时间和摄取时间，这使其能够处理带有时间戳的数据流，并能有效管理数据的时间顺序和时间窗口。这些特性使 Flink 在处理时间敏感的实时数据分析方面表现出色，特别适用于需要低延迟和精确时间控制的应用场景。通过其独特的时间窗口和水位线技术，Flink 能够处理数据的时间乱序，确保按照事件的时间顺序进行计算和结果输出，从而成为实时数据分析的首选工具。Flink 的时间概念示意图如图 9 所示：

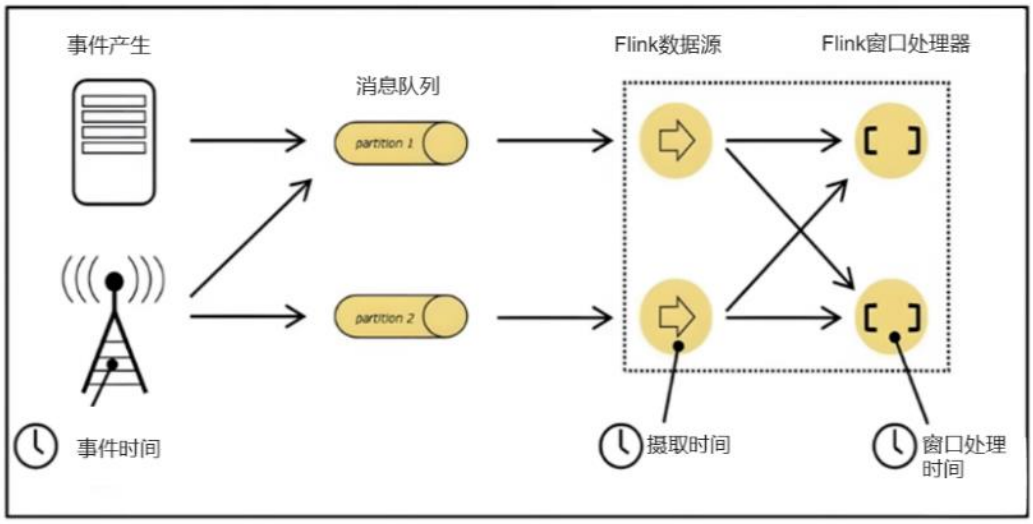


图 9 Flink 时间机制示意图

3.4.3 Alink 机器学习平台

Alink 是由阿里巴巴开发的一个开源机器学习平台，主要用于处理大规模数据。Alink 构建在 Apache Flink 上，继承了 Flink 的分布式计算能力，可以在 Flink 的生态系统中无缝集成，利用 Flink 处理大数据流的优势来执行机器学习任务。本系统利用 Alink 机器学习平台 Alink 机器学习平台中提供的高斯混合聚类方法对所有轨迹流特征进行聚类，并引入公共交通站点数据对聚类结果进行校正。

3.4.4 WebGIS 技术

本系统的 WebGIS 技术集成了多种前沿技术和框架，以实现高效、动态的地理信息系统和数据可视化。系统采用 SpringBoot 作为后端开发框架，利用其快速开发和自动配置的特性，提高了开发效率并简化了微服务架构的部署。前端模板引擎 Thymeleaf 与 SpringBoot 无缝集成，支持动态 HTML 内容的生成，增强了前端页面的灵活性。地理信息可视化方面，系统集成了百度地图 API 进行热力图可视化，使得地理数据的展示更为直观和动态。此外，系统还采用 Echarts 图表，通过动态图表展示后端分析数据，增强了数据可视化的互动性和信息表达的清晰度。

第四章　测试报告

4.1 系统性能需求

4.1.1 实时性

系统需要有实时分析处理流数据的能力。系统需要满足较低的延迟性，对于产生的海量数据，对其进行快速的收集、处理、存储，延迟需要被控制在毫秒范围。同时，对于海量实时产生的数据，利用流式计算方法对无界流数据进行实时计算分析，使系统较快反馈实时分析结果。

4.1.2 可扩展性

系统要求模块化，对数据进行计算、分析、存储的模块需要低耦合度，使各模块间的依赖性降低，各层级间影响较小。这在系统增加新功能时，可以更小的对其他模块产生影响。

4.1.3 易用性

系统的可视化界面要求风格简洁美观整体的主题要求鲜明。对于数据展示要求直观，用户可以较轻松的读取数据变化情况。设置部分交互效果，增加用户在系统中的沉浸感。

4.1.4 可维护性

要求系统在开发过程中使用统一的代码风格且确保有注释，对于接口设置要求接口设计按照统一的规范进行设置。同时拥有较全的文档说明，并确保文档与系统的实际版本保持同步。

4.2 系统开发环境需求

4.2.1 软件需求

本系统开发所需的工具软件如表 3 所示。Intellij Idea 用于编写 Java 后端以及前端代码 Vmware 用于创建 Linux 虚拟机，Xshell 用于远程连接虚拟机系统，Xftp 用 于与远程虚拟机传输文件。

表 3　系统开发所需软件列表

工具名	版本号
Intellij Idea	2023.1.4
Linux	Ubuntu 1604
Xshell	7.0
Xftp	7.0
Vmware	16.0

4.2.2 硬件需求

表 4　系统部署所需的硬件配置列表

名称	配置
CPU	6 核 12 线程及以上
内存	32G 及以上
硬盘	128G 及以上
网卡	1Gbps 及以上

4.3 测试报告

4.3.1 功能性测试

对于功能性测试，主要用于测试系统中的所有的功能模块，确保每个功能能够按照预期执行。本系统处理手机用户人数超过 300 人后对系统进行功能检测，系统的功能模块测试结果如表 5 所示。

表 5　系统各功能模块测试结果表

一级模块	二级模块	主要功能	是否通过
出行模式识别	出行特征构建	提取获取的手机用户轨迹流数据的特征，特征包括速度、距离、时间。	是
	出行聚类分析	对提取出的轨迹流的特征进行高斯混合聚类，聚类出 5 类结果分别为步行、自行车、自驾、地铁、公交车。	是
	人群出行流量计算	在原始手机信令数据清洗后，以 5s 为时间窗口提取人群出行点位，并反映人群出行流量。	是
实时轨迹提取	手机信令数据清洗	对原始的手机信令数据进行清洗，去除数据中的空数据、包含特殊字符的数据，转换时间戳格式。	是
	轨迹流提取	根据时间窗口，提取时间窗口内的用户移动连成的轨迹。	是
	驻留分析	对长期停留于同一区域的手机用户的情况进行分析判断。	是
出行模式可视化	各出行方式占比	饼状图展示城市 5 类出行方式步行、自行车、自驾、地铁、公交车的实时占比情况。	是
	各出行方式平均速度	柱状图展示城市 5 类出行方式步行、自行车、自驾、地铁、公交车的实时平均速度情况。	是
	各出行方式平均时长	三角柱状图展示城市 5 类出行方式步行、自行车、自驾、地铁、公交车的实时平均时长情况。	是
	人群密度情况	以热力地图的形式展示，人群密度高的地方颜色越深，反之则越淡。	是

	各出行方式平均距离	饼状图图展示城市 5 类出行方式步行、自行车、自驾、地铁、公交车的实时平均出行距离情况。	是
	人群驻留情况	以热力地图的形式展示，人群驻留越密集的地方颜色越深，反之则越淡。	是

4.3.2 非功能性测试

非功能性测试主要针对本系统的实时计算性能表现情况，对 Kafka 消息队列进行压力测试。

对于 Kafka 消息队列的测试，本文主要对 Kafka 的吞吐量、数据延迟情况、计算机资源使用情况进行测试。本文使用脚本以生产者的形式对 Kafka 进行生产测试。

写入 10 万级的数据时，其结果如下图 10 所示。从图中可以看到，在 10w 的数据量测试下，每秒能控制在毫秒的延迟以内。

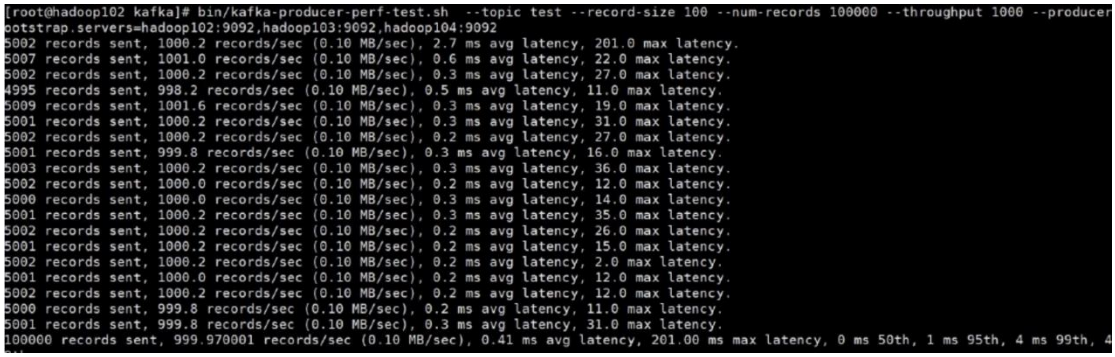


图 10 Kafka 生成者压力测试结果

使用脚本以消费者的身份对 Kafka 中 timeStream 消息主题的手机信令数据流进行消费测试，测试最大吞吐量以及平均消费时间的情况如图 11 所示，可观察到消费的数据水平也达到 10 万级别。

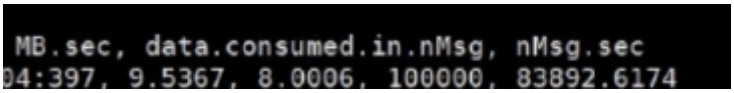


图 11 Kafka 消费者压力测试结果

第五章　安装及使用

5.1 系统开发与运行环境

本系统运行所需的软件、组件以及数据库如表 6 所示。由于部分组件的不满足向下对版本兼容，因此可能出现版本问题。表中的版本为实际开发测试过程中具体需要应用的版本。

表 6　系统运行所需软件列表

工具名	版本号
Zookeeper	apache-zookeeper-3.6.4
Flume	apache-Flume-1.7.0
Kafka	kafka-2.12
Flink	flink-1.9.1
Scala	scala-2.14.3
Java	Jdk-8
SpringBoot	SpringBoot-2.2.5
Alink	alink-2.11
Mybatis	Mybatis-3.4.6
MySQL	MySQL-5.5
Thymeleaf	Thymeleaf-3.0.11

在开发过程中，Kafka、Flmue、Zookeeper、MySQL 将部署于 Ubuntu 系统中，Scala、Flink、Java 需要单独安装于 windows 环境下，其他工具将整合于 SpringBoot 框架下。

本系统的开发设备配置如表 7 所示：

表 7　系统开发设备配置列表

部件名	参数
CPU	Intel i9-13980HX，24 核 32 线程
GPU	NVIDIA GeForce RTX 4080 m
系统	Windows11
内存	64G
固态硬盘	2T

5.2 使用流程

5.2.1 服务启动

首先，进入 zookeeper 安装目录，运行 zkServer.sh 将 zookeeper 服务进行启动。确保 zookeeper 的实例成功启动后，找到安装目录 bin 目录下 kafka-server-start.sh 服务启动文件以及 kafka 配置文件 server.properties，启动 kafka 服务，如图 12 和图 13。

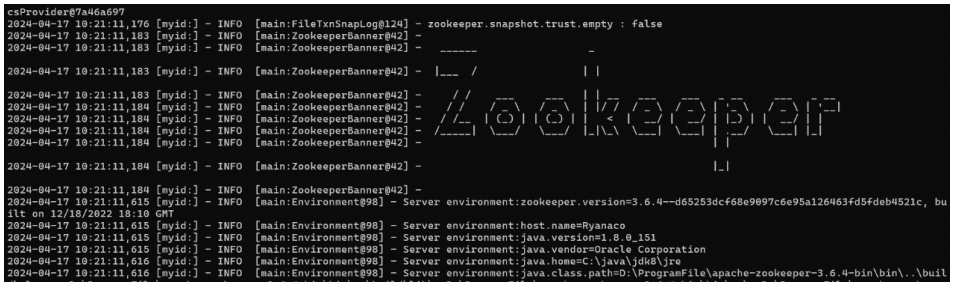


图 12 启动 zookeeper 服务

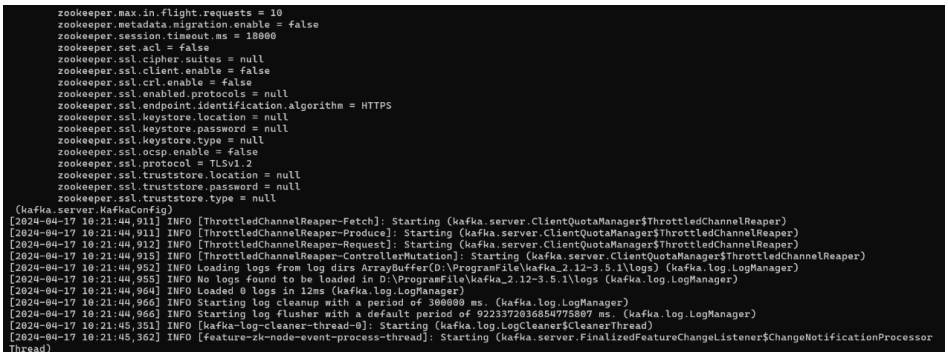


图 13 启动 kafka 服务

5.2.2 实时轨迹提取

在 zookeeper 与 kafka 成功启动后，启动实时轨迹提取服务，模拟向系统输入原始手机信令。模拟输入系统原始手机信令数据如图 14 所示。

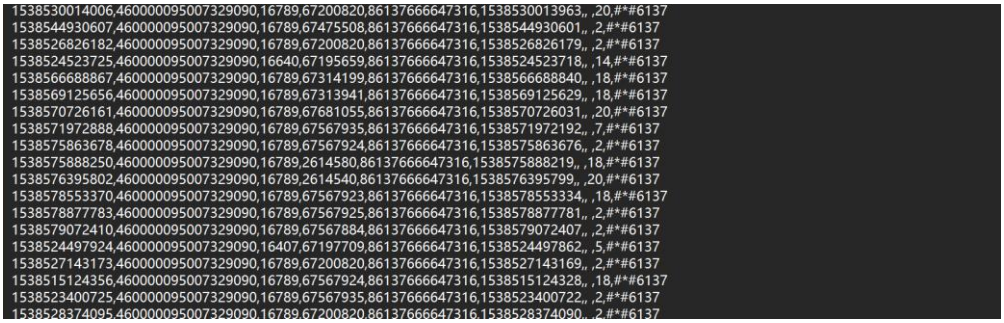


图 14 模拟输入的原始手机信数据

通过手机信令数据清洗功能，原始手机信令数据经系统清洗后的数据流如图 15 所示。

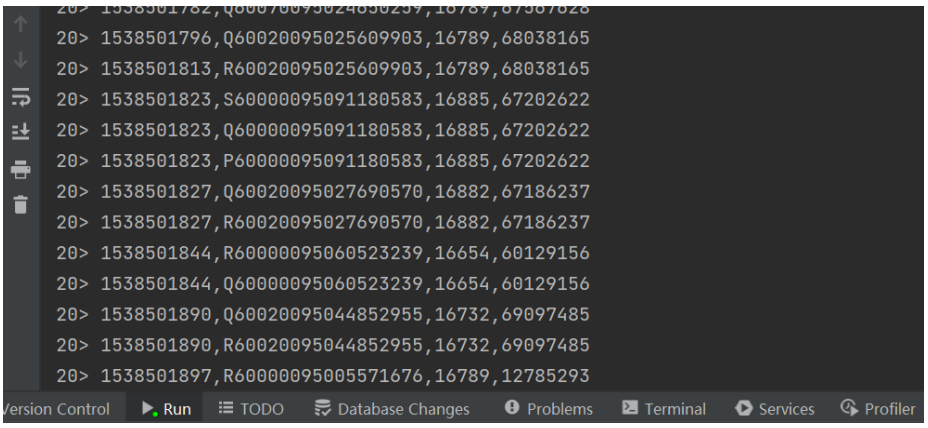


图 15 清洗后手机信令数据展示图

通过轨迹提取功能，清洗后的手机信令数据以 30 分钟为时间窗口获得的轨迹流数据如

图 16 所示。

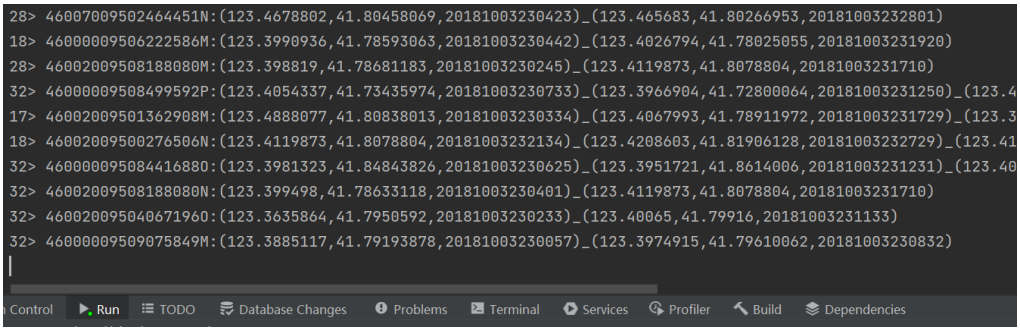


图 16 轨迹流数据展示图

通过驻留分析功能，对提取的轨迹数据进行驻留判断，得到的驻留轨迹数据流如图 17 所示。

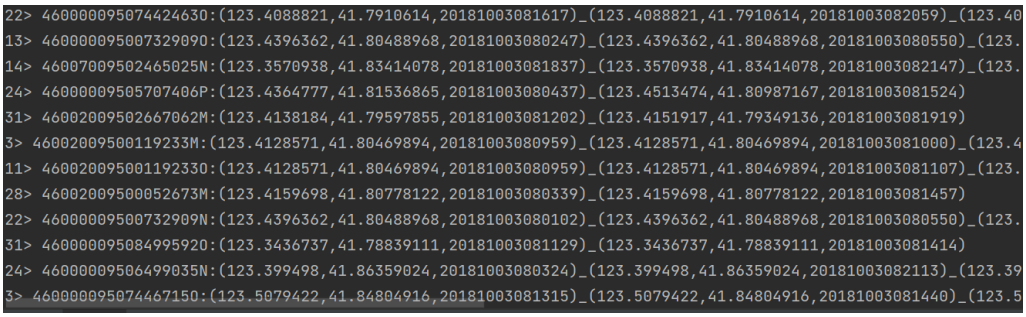


图 17 驻留轨迹数据展示图

5.2.3 出行模式识别

启动出行模式识别服务，该服务将通过 kafka 消息队列获取清洗后的手机信令数据与轨迹流数据，计算人群出行密度并识别出行方式。

人群出行密度计算，将通过 5s 的时间窗口获取实时的人群经纬度点位，人群的实时密度情况如图 18 所示。

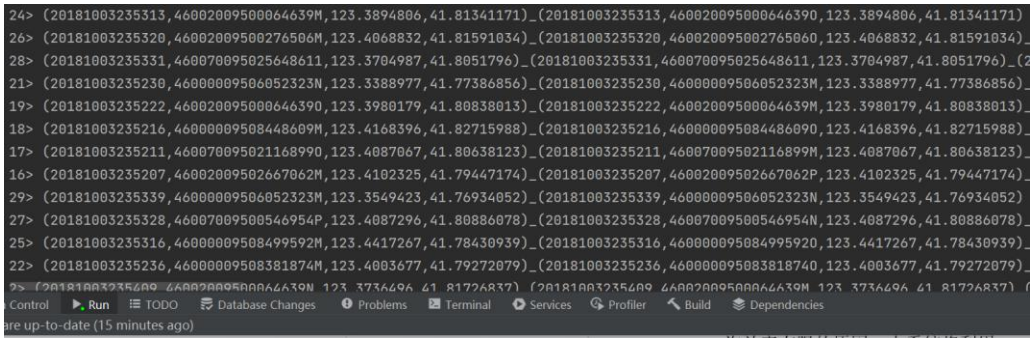


图 18 人群出行流量数据展示图

通过出行方式识别功能，对实时获取的轨迹流数据进行聚类分析，识别出 5 种出行方式，并利用公共交通站点信息对分析结果进行校正。

聚类结果如图 19 所示，其中第一行为 5 种出行方式的代号，下方为每一个被打上对应出行类别标签的轨迹所对应的平均速度。

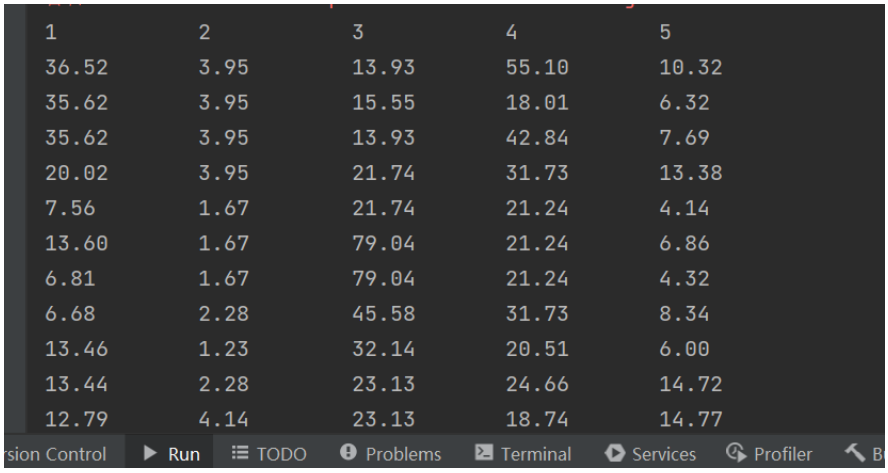


图 19 聚类结果展示图

校正结果如图 20 所示，以图中的情况为例，对于该批次的聚类结果，有 6 条轨迹被校正为了地铁出行方式。

图 20 聚类结果校正情况展示图

5.2.4 出行模式可视化

在原始手机信令数据模拟输入系统后，手机信令数据经过清洗、轨迹提取、出行模式分析等过程后，通过访问前端 url 地址，大数据可视化界面将会实时更新后台服务中分析出的实时出行模式，包括人群密度、人群驻留、出行方式占比、各出行方式平均距离、各出行方式平均时间、各出行方式平均速度。系统运行时整体可视化界面如图 21 所示。

图 21 系统界面效果图

以系统所处理的手机信令数据的时间戳为上午 7 点和下午 6 点的人群出行模式分析结果为例。当接收到手机信令数据对应的时间戳在上午 7 点时，从可视化平台中观测到出行模式情况如图 22 所示。当接收到手机信令数据对应的时间戳在下午 6 点时，观测到的市区居民出行模式情况如图 23 所示。从上午 7 点与下午 6 点时刻实时的人群出行量密度（a）、人群驻留热力图（b）可以发现人群出行热力情况。

(a) (b)

图 22 上午 7 点人群密度以及人群驻留热力图

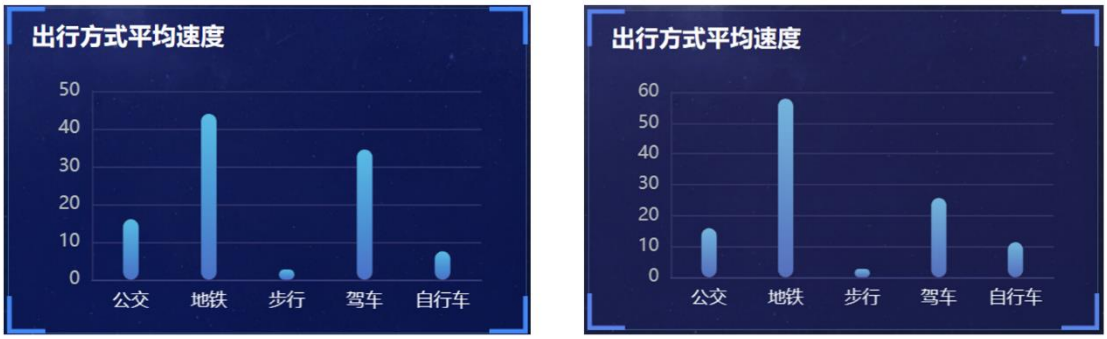
(a) (b)

图 23 下午 6 点人群密度以及人群驻留热力图

对比图 22（a）与图 23（a）可以观察出，上午 7 点市中心的人群密度相较于工作时段较为分散，而晚高峰下午 6 点人群较多集中于市中心的区域。同时，上午 7 点人群驻留区域主要为社区及周边区域如图 22（b）区域所示，而晚高峰时人群主要驻留在购物中心以及写字楼区域如图 23（b）所示。

同时，针对出行方式分析识别的结果，可以反映城市居民不同时间段出行方式选择的规律与交通状况的变化。

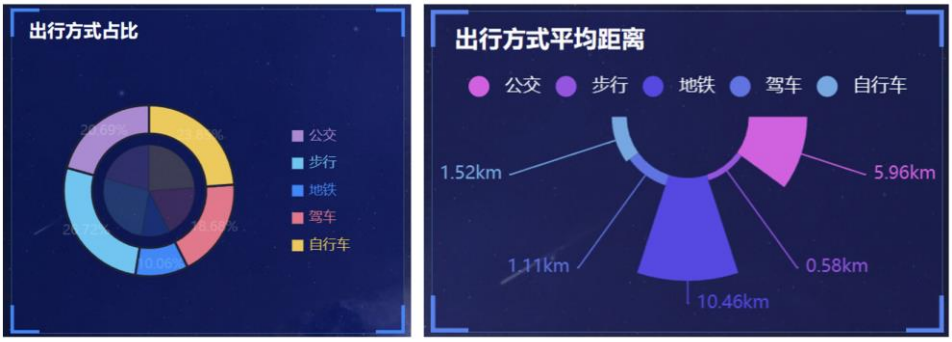
通过各出行方式平均速度可以体现当下城市交通是否拥堵。上午 7 点时通过可视化界面观察的各出行方式平均速度结果如图 24（a）所示，晚高峰下午 6 点时的各出行方式平均速度如图 24（b）所示。可以观察到，相较于晚高峰时刻，上午 7 点时公交、自行车、驾车出行的平均速度高于晚高峰时刻的速度，这反映了晚高峰时有一定的交通拥堵情况出现。



(a) (b)

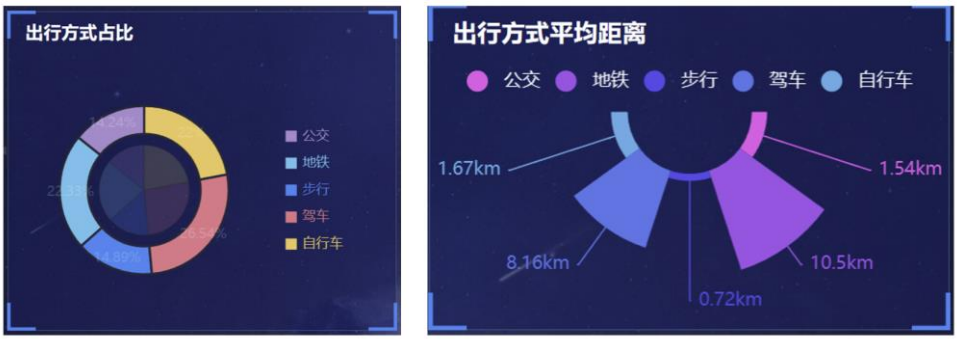
图 24 上午 7 点与下午 6 点人群出行平均速度

通过不同时间段各出行方式的比例以及平均出行时间可反映人群在不同时间段对不同出行方式的使用倾向。上午 7 点的各出行方式比例、平均时间如图 25 所示。下午 6 点时的情况如图 26 所示。可以观察到，早晨时居民使用公共交通以及步行为出行方式的情况较多，且公交和地铁的平均出行距离较远。而晚高峰时更多居民使用驾车、地铁出行，且驾车和地铁的平均距离较远。



(a) (b)

图 25 上午 7 点人群出行方式占比和出行平均距离



(a) (b)

图 26 下午 6 点人群出行方式占比和出行平均距离

第六章　项目总结

本项目构建了一个基于多源数据的人群出行模式识别系统，该系统通过对手机信令数据进行实时计算与分析，挖掘出其背后人群出行模式，并引入公共交通站点信息对挖掘出的人群出行模式进行校正，并实时可视化展示。本项目的主要工作如下：

（1）采用 Flink 大数据流计算引擎以及 Kafka 分布式消息队列构建实了大数据实时计算分析平台。将手机信令数据流以及计算分析产生的数据流以高效的传输模式在系统多个模块之间传输流转，使系统满足较好的流数据计算分析性能。

（2）设计了手机信令数据预处理、轨迹提取的算法，且利用 Flink 流计算引擎来实现该算法。利用计算清洗后的手机信令数据，实时计算分析了人群出行流量。同时，在轨迹数据获取的基础上，提出利用 Alink 机器学习平台中的高斯混合聚类结合实际公共交通站点信息对手机用户出行轨迹的特征进行聚类，识别出了步行、自行车、自驾、地铁、公交车等 5 类出行方式结果。

（3）实现了人群出行模式可视化界面，通过百度地图 API 热力地图展示了城市实时变化的人群密度情况以及人群驻留情况。同时通过 Echarts 组件展示了手机信令数据识别出的人群出行方式实时变化情况。

经过六个月的不懈努力，由我们团队独立设计与研发的基于多源数据融合的人群出行模式识别系统于 2024 年 4 月 30 日圆满完成。这一开发周期充满了学习、构建功能、深思和编码等环节，期间遇到了许多挑战。虽然这些困难一度让我们陷入停顿，但我们逐渐找到了解决问题的方法，这个过程非常宝贵，也极大地丰富了我们的经验。每当问题得到解决，我们都感到非常欣慰，这六个月的经历让我们受益匪浅，也为我们未来的职业生涯打下了坚实的基础。

项目初期，我们在选题方向上感到迷茫，缺乏明确的目标，时间在不知不觉中流逝，让人感到焦虑。然而，凭借团队中每个成员坚定的信念和不懈的努力，我们不断地自我反省，将专业知识与日常生活、国民经济相结合，最终决定将我们的研究重心放在人群出行模式的实时识别上。确定了研究方向后，我们开始深入挖掘相关数据，发现了手机信令数据在交通模式识别的应用，不断完善我们的系统架构。通过集体讨论，我们深入探讨了系统的结构和功能，每个问题都经过反复的优化和修正。这个过程不仅加深了我们对作为团队合作的理解，也加强了我们对系统功能的可行性和可靠性的认识，最终形成了一个统一且完整的体系。

通过不断地克服困难和挑战，我们团队坚信：有信心、有目标、有毅力，就能解决任何问题。这也验证了那句老话：世上无难事，只怕有心人。在此，我要感谢中国大学生计算机设计大赛组委会提供的这个平台，它不仅让我们有机会展示我们的技术，也是一个锻炼和学习的宝贵机会；感谢我们的指导老师崔海福，她的专业指导和无私支持为我们提供了良好的开发环境。同样，感谢所有支持我们的老师和同学们。在这次比赛的准备过程中，我们学到的不仅仅是关于系统开发的技术知识，更多的是如何在困难面前不屈不挠，与团队成员一起奋斗，共同迈向成功。

参考文献

[1] 中华人民共和国自然资源部. TD/T1073-2023. 国土空间规划城市时空大数据应用基本规定[S].北京：中华人民共和国自然资源部, 2023.

[2] 聂霖杰. 基于移动数据的出行模式识别方法研究[D]. 吉林：吉林大学,2020.

[3] Truong T P, Hensher D A. Measurement of travel time values and opportunity cost from a discrete-choice model[J]. The Economic Journal, 1985: 438-451.

[4] 南敬林.京沪通道旅客行为时间价值研究[J].铁道经济研究,2002(03):36-38.

[5] 马天奕. 通勤者早晚高峰出行时间价值差异分析[D].重庆：西南交通大学,2019.

[6] Cui H, Wu L, Hu S, Lu R. Measuring the Service Capacity of Public Facilities Based on a Dynamic Voronoi Diagram[J]. Remote Sens, 2021,13, 1027.

[7] Mao F, Minhe J I, Liu T. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data[J]. Frontiers of Earth Science, 2016, 10(2): 205-221.

[8] Xiao G, Juan Z, Gao J. Travel mode detection based on neural networks and particle swarm optimization[J]. Information, 2015, 6(3):522-535.

[9] Gong H , Chen C , Bialostozky E , et al. A GPS/GIS method for travel mode detection in New York City[J]. Computers Environment & Urban Systems, 2012, 36(2):131-139.

[10] Bolla R, Davoli, F. Road traffic estimation from location tracking data in the mobile cellular network[C]. In Proceedings from wireless communications and networking conference, 2000, Vo1l.3, pp. 1107-1112.

[11] White, J., Wells I. Extracting origin destination information from mobile phone data[C]. In Proceedings from 11th international conference on road transport information and control, 2002,486:30- 34.

[12] Asakura Y，Hato E. Tracking survey for individual travel behaviour using mobile communication instruments[J]. Transportation Research Part C: Emerging Technologies, 2004,12(3-4): 273-291.

[13] 冉斌. 手机数据在交通调查和交通规划中的应用[J]. 城市交通, 2013 (1): 72-81.

[14] 李振邦, 冉斌, 孟华. 一种基于 C4.5 决策树的手机用户出行方式识别方法[P].中国, A, 105101092A, 2015.

[15] 宋少飞,李玮峰,杨东援.基于移动通信数据的居民居住地识别方法研究 [J].综合运输,2015,37(12):72-76.

[16] 包婷,章志刚,金澈清.基于手机大数据的城市人口流动分析系统[J].华东师范大学学报(自然科学版),2015(05):162-171.

[17] 赖见辉．基于移动通信定位数据的交通信息提取及分析方法研究[D].北京：北京工业大 学,2014.

[18] 陈欢,薛美根.大数据环境下上海市综合交通特征分析[J].城市交通,2016,14(01):24-29+23.[19] 唐小勇，周涛，刘晏霖.基于手机信令的重庆主城职住关系评价[J].城乡规划（城市地理学术

[20] Kalatian A , Shafahi Y , Figueira M , et al. Travel Mode Detection Exploiting Cellular Network Data[J]. Matec Web of Conferences, 2016, 81:03008.

[21] Xu D , Song G , Gao P , et al. Transportation Modes Identification from Mobile Phone Data Using Probabilistic Models[J]. Springer Berlin Heidelberg, 2011.

[22] 卢梦丽, 邬思强, 陈长超, 冯时超, 李伟. 基于手机信令数据的交通时空大数据分析挖掘系统[J]. 计算机科学与应用, 2020, 10(8): 1471-1479.

[23] 杨彬彬. 基于手机信令数据的城市轨道交通客流特征研究[D]. 南京：东南大学, 2015.