西 南 交 通 大 学

本科毕业论文

基于IC卡和GPS数据的公交出行OD

推算平台搭建

Platform establishment for transit OD

estimation base on IC and GPS data

年 级：2019级

学 号：2019112757

姓 名：陈正贤

专 业：交通运输

指导教师：徐占东

二零二三年五月

西南交通大学

本科毕业论文学术诚信声明

本人郑重声明：所呈交的毕业论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，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。

作者签名：

日期： 年 月 日

西南交通大学

本科毕业论文版权使用授权书

本毕业论文作者同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权西南交通大学可以将本毕业论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本毕业论文。

**保密**□，在 年解密后适用本授权书。

本论文属于

**不保密**□。

（请在以上方框内打“🗸”）

作者签名： 指导教师签名：

日期： 年 月 日 日期： 年 月 日

**毕业设计（论文）任务书**

班 级 交运（茅班）2019-01班 学生姓名 陈正贤 学 号 2019112757

发题日期：2022年12月25日 完成日期：2023年5月18日

题 目 基于IC卡和GPS数据的公交出行OD推算平台搭建

1、本设计（论文）的目的、意义 本选题旨在基于成都市交通多源海量数据，包括地面公交运行轨迹数据集、地面公交发车时刻表数据集、地面公交IC卡数据集、地面公交线路站点数据集设计算法推算居民出行OD，并开展出行结构特征的相关分析工作，为网络结构优化、服务品质提升、智能公交系统升级提供决策依据。

2、学生应完成的任务 （1） 搜索国内外相关文献，深入学习并总结分析；（2）搭建数据处理和分析平台；（3）基于SQL语言设计算法开展OD推算；（4）对成都市公交系统的运营现状进行评估。

3、本设计（论文）与本专业的培养目标达成度如何？（如在知识结构、能力结构、素质结构等方面有哪些有效的训练。）

本毕业设计涉及到数据结构、网络分析、交通规划理论等，能够有效锻炼学生的数学建模、数据处理、软件编程、文献阅读、学术写作等多方面的能力。

4、本设计（论文）各部分内容及时间分配：（共 17 周）

第一部分 归纳整理国内外相关文献，细化技术路径 (2 周)

第二部分 构建交通网络时空多源数据分析处理平台 (3周)

第三部分 设计算法推算常规公交出行需求 (4周)

第四部分 优化搭建平台，实现代码填写和功能运行 (4周)

第五部分 梳理和修改，完成论文撰写及答辩准备 (2周)

评阅及答辩 完成毕业论文的修改、评审及答辩 (2周)

备 注

指导教师： 年 月 日

审 批 人： 年 月 日

摘 要

随着城市经济水平和常驻人口的快速发展，城市内交通的压力日益增大，公共交通逐渐为人们出行的主要方式。以成都市为例，市内有7000余个公交站点和近1000条公交线路，是服务公共出行的主力军。然而，已有数据并未记录乘客的出行链，推算公交乘客的出行OD对公交站点的选址优化，及客流量的预测具有重要的意义。

公交出行OD数据作为城市公共交通特征研究的基础数据，在公交客流分析，预测和公交路网规划与优化等工作中扮演了重要角色。其主要研究内容分为上车站点的匹配和下车站点的推算。本文基于IC卡、公交车载GPS数据、线路站点数据等一系列公交基础数据，运用时间段配对和出行链闭合理论分别对上下车站点进行推算。并且基于C#、数据库软件和地图软件搭建一个推算平台集数据导入、查询、OD推算和可视化展示功能于一体。

首先，本文总结了国内外学者基于IC卡和GPS数据推算公交OD的研究方法。上车站点匹配的主要方法为利用公交到站时间匹配乘客刷卡时间。针对下车站点的推算，国内外的学者往往仅对一种出行特征的人群进行推算，其推算结果存在下车站点的丢失，因此本文将对人群进行分类后进行下车站点推算。

其次，本文对数据进行预处理，主要为IC卡数据、GPS数据的清洗和数据融合。并且通过时间段匹配的方法匹配上车站点，再基于出行链闭合的理论推算下车站点。基于乘客上下车站点的数据，本文统计得到交通大区之间的出行OD、出行时段的统计数据和线路站点的登降量数据。。

最后，本文基于C#语言搭建了公交OD推算平台，连接了外部数据库软件Oracle和SQL Server以及地图软件ArcGIS，最终通过调用存储过程实现有关功能。利用ArcGIS软件将推算得到的OD数据和衍生数据如线路登降量数据在C#搭建的窗口中实现数据的可视化展示。

综上，本论文的研究在数据层面实现了公交数据的管理和深度挖掘，在应用层面为成都市公交运营的管理规划提供了数据支持和软件基础。

关键词：公交出行；上下车站点推算；OD矩阵；平台搭建

Abstract

With the rapid development of urban economic level and resident population, the pressure of urban traffic is increasing day by day, and public transport has gradually become the main way for people to travel. Take Chengdu as an example. There are more than 7,000 bus stops and nearly 1,000 bus lines in the city, which is the main force serving public travel. However, the existing data do not record the travel chain of passengers. Calculating the travel OD of bus passengers is of great significance to the optimization of bus site selection and the prediction of passenger flow.

As the basic data for the study of urban public transport characteristics, bus travel OD data plays an important role in the analysis and prediction of bus passenger flow and the planning and optimization of bus network. The main research content is divided into the matching of the boarding station and the calculation of the exiting station. Based on a series of basic bus data, such as IC card, bus GPS data, route station data, etc., this paper uses the theory of time pairing and travel chain closure to calculate the loading and unloading stations respectively. And based on C#, database software and map software to build a calculation platform set data import, query, OD calculation and visual display functions in one.

Firstly, this paper summarizes the research methods of bus OD calculation based on IC card and GPS data. The main method of bus stop matching is to use bus arrival time to match passengers' swipe time. For the calculation of drop-off stops, scholars at home and abroad usually only calculate the crowd with one kind of travel characteristics, and the calculation results usually include the loss of drop-off stops. Therefore, this paper will classify the crowd and then calculate the drop-off stops.

Secondly, this paper preprocesses the data, mainly for IC card data, GPS data cleaning and data fusion. And through the method of time matching to match the station, and then based on the theory of travel chain closure calculate the station. Based on the data of passenger loading and unloading stations, this paper calculates the travel OD between traffic regions, the statistical data of travel time and the data of the boarding and dropping volume of line stations. Then ArcGIS was used to visualize the above data in the map of Chengdu City.

Finally, this paper builds a bus OD calculation platform based on C# language, connects external database software Oracle, SQL Server and map software ArcGIS, and finally realizes related functions by calling stored procedures. Using ArcGIS software, OD data and derived data such as line landing volume data can be visualized in the window built by C#.

In conclusion, the research of this paper realizes the management and in-depth mining of bus data at the application level, and provides data support and software foundation for the management planning of bus operation in Chengdu.

**Key words**: Bus travel; Calculation of loading and unloading stations; OD matrix; Platform building

目 录

[第1章 绪论 1](#_Toc135400146)

[1.1研究背景及意义 1](#_Toc135400147)

[1.2 研究目标 2](#_Toc135400148)

[1.3 国内外研究现状 2](#_Toc135400149)

[1.4 研究内容及技术路径 5](#_Toc135400150)

[1.4.1 研究内容 5](#_Toc135400151)

[1.4.2 章节安排 5](#_Toc135400152)

[1.4.2 技术路径 6](#_Toc135400153)

[1.5 本章小节 7](#_Toc135400154)

[第2章 公交基础数据 8](#_Toc135400155)

[2.1 公交IC和车载GPS技术介绍 8](#_Toc135400156)

[2.1.1 公交IC卡技术介绍 8](#_Toc135400157)

[2.1.2 车载GPS技术介绍 9](#_Toc135400158)

[2.2 基础数据介绍 10](#_Toc135400159)

[2.2.1 公交IC卡数据 10](#_Toc135400160)

[2.2.2 公交车载GPS数据 11](#_Toc135400161)

[2.2.3 快速公交（BRT）数据 11](#_Toc135400162)

[2.2.4 GIS地图数据 12](#_Toc135400163)

[2.2.5 线路站点数据 13](#_Toc135400164)

[2.2.6 联锁转乘站点数据 14](#_Toc135400165)

[2.2.7 站点区域数据 14](#_Toc135400166)

[2.3 基础数据的预处理 15](#_Toc135400167)

[2.3.1 公交IC卡数据 15](#_Toc135400168)

[2.3.2 公交车载GPS数据 15](#_Toc135400169)

[2.3.3 快速公交（BRT）数据 16](#_Toc135400170)

[2.3.4 公交单日站点发车时刻表 16](#_Toc135400171)

[2.4 本章小节 17](#_Toc135400172)

[第3章 OD推算方法 18](#_Toc135400173)

[3.1 上车站点的推算 18](#_Toc135400174)

[3.1.1 基本假设 18](#_Toc135400175)

[3.1.2 上车站点匹配算法 19](#_Toc135400176)

[3.2 下车站点推算 22](#_Toc135400177)

[3.2.1 乘客出行链闭合假设 23](#_Toc135400178)

[3.2.2 往返出行的下车站点推算 25](#_Toc135400179)

[3.2.3 换乘过程的下车站点推算 29](#_Toc135400180)

[3.2.4 一次出行的下车站点推算 30](#_Toc135400181)

[3.3 本章小节 30](#_Toc135400182)

[第4章 公交出行OD推算平台 31](#_Toc135400183)

[4.1 推算平台软件基础 31](#_Toc135400184)

[4.1.1 数据库软件Oracle、SQL Server 31](#_Toc135400185)

[4.1.2 平台搭建软件Visual Studio 33](#_Toc135400186)

[4.1.3 地图软件ArcGIS 34](#_Toc135400187)

[4.2 推算平台结构 35](#_Toc135400188)

[4.2.1 平台主界面 35](#_Toc135400189)

[4.2.2 交通数据界面 35](#_Toc135400190)

[4.2.3 OD推算及查询界面 39](#_Toc135400191)

[4.2.4 出行查询界面 40](#_Toc135400192)

[4.2.5 地图展示界面 42](#_Toc135400193)

[4.3 推算平台主要功能 44](#_Toc135400194)

[4.4 推算平台特点 45](#_Toc135400195)

[4.5 平台搭建意义 46](#_Toc135400196)

[4.6 本章小节 46](#_Toc135400197)

[第5章 实验结果 47](#_Toc135400198)

[5.1 乘客上车站点推算结果 47](#_Toc135400199)

[5.2 出行OD统计结果 47](#_Toc135400200)

[5.3 平台运行结果 48](#_Toc135400201)

[5.3.1 数据处理界面 49](#_Toc135400202)

[5.3.2 GIS展示界面 53](#_Toc135400203)

[5.4 本章小节 54](#_Toc135400204)

[结论与展望 56](#_Toc135400205)

[致谢 58](#_Toc135400206)

[参考文献 59](#_Toc135400207)

**第1章 绪论**

1.1研究背景及意义

在成都这样的大都市中，随着经济水平的飞速发展和城市化进程的不断发展，城市规模的扩大也带来了城市公共交通设施配套不足、区域用地结构单一、职住分离等一系列社会问题，导致了居民通勤时间延长、费用增加等问题。国内外的研究经验表明了，优先完善城市内的公共交通是解决城市交通问题的一条有效途径。特别的，根据成都市的交通规划，成都市内公共交通的战略地位也不断提升。

近年来，智慧交通、数字孪生、智慧城市等新兴概念的提出也推动了目前相对成熟的智能技术如数据采集技术、大数据集成技术等加入交通领域。特别是在目前“互联网+”的时代背景之下，一方面大数据的加入能够为目前的研究提供新的思路和研究方向，进而为城市提升治理交通问题的能力，缓解城市内交通资源紧张的现状；另一方面，大数据技术取代传统的人工调查，问卷调查的方式，不仅提高了数据的可靠性也解放了大量的人力资源。

虽然目前国内大部分城市都实现了公交刷卡和车载GPS定位，但是以往的公交IC卡和GPS数据研究存在以下问题：一是大量的公交IC卡数据并没有切实运用于城市公交管理和规划之中；二是此类数据的利用过于独立，没有能够和其他交通类大数据进行合作；三是IC卡和GPS数据的研究大多注重于上车站点的研究，而忽视了乘客的下车站点以及换乘站点的匹配上，其对应的OD也并不完善。

面对大数据发展的浪潮，把握好交通大数据的机遇，基于IC卡数据和GPS数据研究城市公共交通如城市公交的出行需求，对于城市交通系统的管理规划以及智慧型城市的建设都起到了重要的作用。因此本文的研究内容在现实中，首先，实现城市内公共交通数据的综合利用分析与深入挖掘，获得了实时、充分、准确的公共出行客流的运行信息，实现公交上车站点和下车站点的双重匹配，提高了数据结果了可靠性与准确性。第二，城市内多种交通方式之间的配合也将得到进一步优化，城市路网的性能进一步提高，为提升城市交通服务质量提供了多方位、多层次、多渠道的帮助。第三，为城市公交数据的管理提供了平台支撑，使城市交通出行需求结果得到具象化的展示，为城市交通管理与规划提供了研究思路，实现对社会交通资源的合理利用。

1.2 研究目标

本文基于成都市公交系统的IC卡数据、公交车载GPS数据以及相关的公交线路站点数据，针对不同类人群分别研究，实现对成都市公交乘客的出行OD进行推算。本文主要实现以下的研究目标：

1. 基于GPS和IC卡数据对公交乘客的出行OD进行推算。
2. 搭建公交OD推算的平台，实现数据导入、处理、OD推算、查询和数据可视化功能

1.3 国内外研究现状

在公共交通领域，出行OD（起点和终点）是非常重要的信息。它不仅为公共交通运营提供了基础数据，还可以用于识别服务不足的区域、规划新的线路等。然而，获取准确的OD数据一直是一个挑战。公交车乘客的OD研究内容主要有三个部分——上车站点、下车站点以及换乘站点。对于这三个方向的研究以时间来划分主要有两个阶段，分别是人工调查再利用OD矩阵反推验证阶段和近几年大数据兴起后利用GPS数据和IC卡等多源数据来直接预测上下站情况并推导OD矩阵。

在第一个阶段中，针对上、下车站点的研究，最常见的方法为人工记录和票据记录。在这一阶段，外国学者的开展早于国内且国外的学者的研究成果较国内研究者更加先进而全面。Barry[1]等提出上一段出行的下车点是下一段出行的下车点的假说，并以此来反推OD。

在第二阶段，随着大数据技术、GPS定位的精准化和公交IC卡的大面积普及，研究者们开始将这些数据综合利用，通过多源数据的关联来研究分析公交乘客的上车点、下车点以及换乘车站。但是不同的城市之间所用的公交刷卡系统和车载GPS系统都存在差异，其记录的公交出行数据结构也不同。在国内的一些城市中，IC卡可以同时记录乘客的上车站点和下车站点，因此IC的录入数据在完成数据清洗后就可以得到精准的乘客出行记录。但是目前国内大部分城市中所使用的公交刷卡系统仅会记录卡号、上车时间、公交编号等信息。为了得到完整的出行信息还要借助公交的GPS数据和站点GIS数据。近几年的一些学者在他们的研究使用了比较典型的研究方法，马伯浩[12]、尹安藤[16]、WangYunshan[21]等以第二次的刷卡记录匹配最近的站点，Hussain Etikaf[19]、Masood Jafari[20]、Huang Di[18]、秦政[15]基于海量前人的研究数据包括现有的研究OD和人工调查的OD推算未来的公交OD，赫伟娜等[13][14][23]采用了MAC等数据结合IC卡和GPS数据综合分析推算公交OD。同时OD研究的重点可以集中在上车站点的推算和下车站点的推算这两个部分。

1. 上车站点推算方法

针对上车站点的推算，国内外的学者主要运用的方法出现了一致性。绝大部分学者采取了利用公交的车辆编号，及其对应的运行线路，匹配IC卡刷卡数据中的刷卡时间和公交编号这两个字段。再结合线路站点的GIS数据，得到公交的进出站时间段。最后将乘客刷卡的时间和公交进出站时间段进行匹配。舒国辉[1]（2009）首先把乘客的刷卡数据和对应车辆关联起来，之后对乘客的刷卡时间进行了聚类，最后将公交在站的停靠时间同聚类后的刷卡时间相匹配，以此获取乘客的上车站点。经过分析公共交通系统中乘客的刷卡行为，李海波[3]等（2015）提出了一种新的对刷卡数据时间进行校正的方法，即利用公共交通中的AVL数据，以最高的站点识别率为基准，确定整个系统的可信实际时差，以此来避免由于刷卡设备时间误差而造成的乘客上车站点匹配失败的情况。Yu等[4]（2018）提出了一种将时序控制设备的信息提取和区间匹配的算法，用以解决车载GPS信息和IC卡数据中记录的车辆编号不一致的情况，并以青岛的公交为实例，验证了此算法的可行性和有效性。张彤[5]（2019）为了克服仅依据空间位置匹配上车站点时出现的配站错误问题，提出了在利用GPS数据和IC卡数据进行站点匹配时，除了考虑公交进出站的空间位置，还需要加上公交运行时的方向，判定是上行、下行或者是环行

1. 下车站点推算方法

仅利用IC卡和车载GPS数据并不能像上车站点一样通过时间段匹配得到准确度高的下车站点，目前国内外针对下车站点的推算方法大致可以分为集计模型、非集计模型和机器学习方法。

集计模型是一种统计学方法，可以通过对已有数据进行分析，建立数学模型，预测未来的结果。在公共交通领域，使用集计模型可以通过观察上车和下车时的乘客数量、时间、地点等因素来预测乘客下车站点。柳伍生（2018）[6]将乘客的个人出行特征与乘客的出行距离作为车站吸引力的权数，获得旅客的旅行时间推算概率，并与区间不确定度相结合，获得旅客的旅行时间间隔。胡继华（2014）[7]定义了高频站点集等集合变量，并提出一种融合乘客个人出行特性的车站吸引力加权计算方法。

非集计模型通常针对单日内有多条刷卡数据的乘客，利用出行链理论，通过另一次出行的下车站点以及出行时间段来推算本次的下车站点。胡继华[7]（2014）以广州的一条公交线路为研究对象，基于现有的出行链和出行OD来确定目标乘客的到站概率，并对模型结果进行了检验。李佳怡[8]（2018）通过统计乘客刷卡频数的站点和运行线路下游各站点的吸引权重来分配乘客的下车站点。崔紫薇[9]（2019）基于多源数据，利用融合后出行链对下车站进行推断，将成功匹配的下车站点整理为基础数据集，并基于其中的个人出行特征对于未确定的下车站点进行推算。

机器学习的方法是指利用等人工智能、机器学习和深度学习等技术，并结合多源数据如GPS数据、IC卡数据、移动支付数据等，以提高预测准确率。温晓丽[10]（2021）通过乘客历史出行记录，利用均值Shift聚类算法得出每个乘客的居住地中心点O与工作地中心点D，采用阈值内最近邻算法计算离工作地中心或居住地中心最近的距离，进而补充无法识别乘客的下车站点。杨鑫等[11]（2019）采用结合规则和Bi-LSTM-CRF序列化标注算法的两阶段公交客流推断方法，可以预测所有乘客的下车站点。

基于IC卡和GPS数据推算公交出行OD的方法已经经过了长时间的发展，在国内外针对上车站点推算的研究已经相对成熟、完善、匹配成功率也相当高且不同国籍的学者之间研究方法出现了同质化。而基于IC和GPS数据对下车点的推算方法则比较多元化，不同方法之间的匹配结果也差异较大。整体而言，下车站点推算结果的可信度和可靠度均不如上车站点的匹配结果。特别的是，大部分的研究学者仅对单次出行之间的上下车站点推算进行了研究，而忽视了乘客在一次完整的出行中换乘的可能，没有对换乘站点的推算进行深入研究。未来，随着数据采集和处理技术的不断提升，这个领域还有很大的发展空间和挑战。

1.4 研究内容及技术路径

1.4.1 研究内容

本文采用成都市内数日的公交IC卡数据和公交定位系统中的GPS数据进行乘客的OD推算并基于推算流程搭建一个公交OD推算平台。其主要步骤为数据导入、数据清洗、上车站点推算、下车站点推算、OD统计、平台搭建。

因此本文的研究内容有如下几个部分：

1. 数据处理：首先将收集到的成都市IC卡和GPS数据以及一系列的公交基础数据，通过Oracle进行数据的导入和预处理。再结合SQL Server数据库对预处理后的数据进一步清洗，对错误或无效数据进行剔除，剔除有用的字段，为后续上下车站点的识别提供基础。
2. 公交乘客上下车站点的识别：根据前人的研究方法和IC卡、GPS数据的特性，通过时间范围可以对乘客的上车站点进行匹配。下车站点推算包括了对换乘站点的推算。本文基于乘客出行链闭合的假设，将人群按出行特征进行分类，分别进行下车站点的推算。
3. OD统计分析：本文根据成都划分的行政区，将成都市划分为了16个交通大区和若干个交通小区。基于推算得到的乘客上下车站点信息中包含的线路站点信息和定位信息，本文对乘客的上下车量进行了统计，制成了交通大区之间的出行OD。同时统计了不同公交线路和公交站点的乘客登降量

1.4.2 章节安排

针对上述研究目标，本文将从数据的预处理、公交OD推算平台的搭建、上车站点推算算法这三个方面进行研究，并且将文章划为5个章节，各章节的研究内容安排如下：

第一章 绪论。阐述本文研究的背景意义以及国内外学者在此主题下的研究现状，并以此提出了本文的研究目标，进而介绍论文中的主要研究内容和组织框架。

第二章 公交基础数据。介绍了公交IC卡和GPS定位系统以及本文中涉及到的其他数据，进而介绍了上述数据的预处理过程。

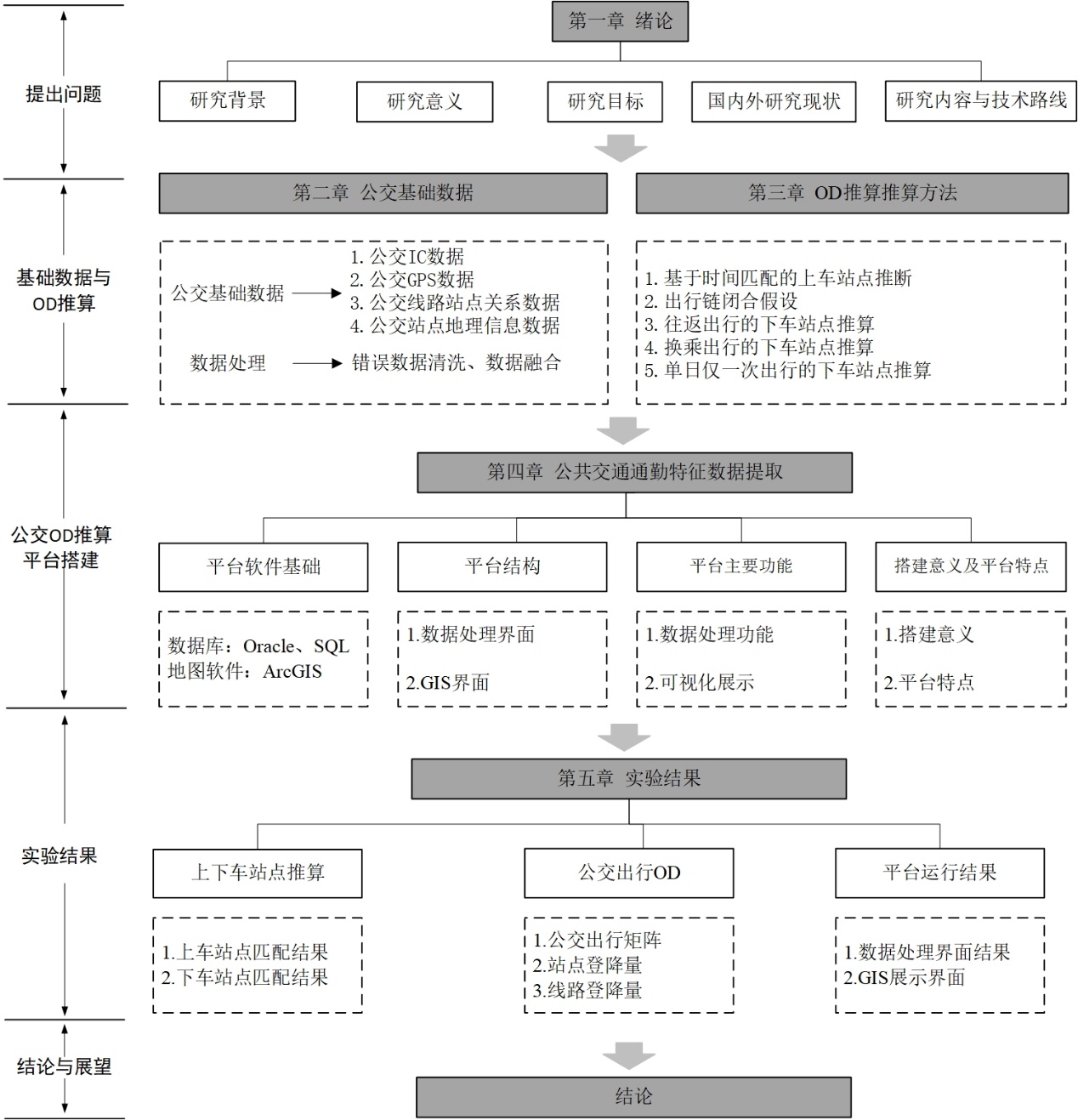
第三章 公交出行OD推算平台。介绍了推算的软件基础、平台结构、主要功能、平台特点优势以及平台的搭建意义。

第四章 OD推算算法。介绍了上车站点的推算算法。提出了出行链闭合假设作为下车站点推算的基础，并且分类介绍了针对不同出行特征乘客的推算方法。

第五章 推算结果。介绍了OD推算的结果，包括上下车站点的信息、OD矩阵、出行时段统计数据以及登降量数据，并且展示了推算平台的运行结果。

1.4.2 技术路径

本文的整个技术路径展示如下：



**图 1-1 文章技术路线**

1.5 本章小节

本章首先介绍了本文的研究背景及意义并且简单阐述了本文的研究目标。其后从推算公交OD的发展过程、公交上车站点的匹配和公交下车站点的匹配三方面对国内外的相关研究文献进行了梳理，总结了以往研究的成果方法和存在的一些不足并且介绍了本文的研究方向和研究的方法。最后展示了本文的研究内容、本文的章节以及推算公交OD的技术路径。

**第2章 公交基础数据**

公交基础数据包括IC卡、GPS、线路站点等数据，是推算公交出行OD的基础。本章主要介绍公交IC卡和车载GPS数据的采集GPS数据、IC及其数据结构以及一系列本文研究中将运用到的公交基础数据，并且对此类数据中的异常数据进行清洗处理。

2.1 公交IC和车载GPS技术介绍

2.1.1 公交IC卡技术介绍

上世纪90年代初，我国部分城市引入了公交IC卡技术。在初期阶段，只有国内的一线大城市采用了这一技术，并针对各自特点开发了相应的公交IC卡管理系统。随着技术的进步和其他城市的跟进，截至2013年，全国已有440多个城市建立了不同规模的公交IC卡计费系统。为规范建设行为、节省投资、数据共享和保障数据安全，住建部制定并出台了公交IC卡统一密钥安全体系技术标准，目前全国160多个城市、4个直辖市、大多数省会城市和地级市均按照该标准建设，公交IC卡已经成为我国各大城市公共交通系统中标识乘客和实施计费的主要载体。该卡由芯片和天线组成，组装在PVC卡片中。相比传统的带磁卡片，IC卡具有明显的优势：可以存储更多数据信息，信息稳定性高且抗干扰性好；加密技术成熟，保障信息安全，使用寿命长；射频识别技术无需电源模块，并非必须与刷卡机密切接触

公交IC卡技术被广泛应用于社会生活的各个领域，特别是公共交通领域。从问世开始，公交IC卡已在全球范围内得到应用。通过在卡片中存储钱包账户信息和建设对应的计费系统，每次刷卡时可自动扣费，并为IC卡用户提供优惠政策，显著提升了用户体验和选择公共交通出行的意愿，为公共交通的发展做出了重要贡献。

公交IC卡技术的普及和推广，不仅方便了乘客的出行，还有助于提高城市公共交通系统的效率和管理水平。通过对乘客刷卡信息的收集和分析，可以更好地了解乘客出行习惯和需求，为运营商提供更准确的乘客服务。此外，公交IC卡还可以与其他应用场景相结合，如停车场、门禁系统等，使得卡片的功能更加多样化，提高了其使用价值。

然而，公交IC卡也存在一些问题。例如，不同城市的IC卡标准和系统不兼容，导致用户在跨城市出行时需要更换或兼容不同的卡片，给用户带来不便；此外，IC卡的信息安全性也面临着挑战，存在被盗刷、破解等风险。

因此，未来需要进一步完善和规范公交IC卡技术和管理体系，建立统一标准和互联互通机制，保障乘客信息安全，提高系统稳定性和用户体验。同时，还应加强对公众对公交IC卡的宣传和普及，提高其认知度和使用率，推动公共交通发展和城市可持续发展

2.1.2 车载GPS技术介绍

全球定位系统（GPS）最初是军事领域的发明，但现在已广泛应用于社会生产和生活中，如手机定位、汽车导航、测量绘图、资源调查和紧急救援等。将GPS技术应用于城市公共交通，是该技术在应用领域的一次重要拓展。早期公交车报站系统主要依赖司机人工操作，不仅繁琐易错，还存在安全隐患。引入GPS技术后，车载系统可以自动报站，减轻司机工作负荷，提高报站准确度，保障行车安全。在站台设置显示列车到站倒计时的PIS屏幕，可以为乘客提供信息导引；同时利用车辆GPS定位，可掌握各路段实时交通信息，对途中车辆进行动态导引，缓解拥堵情况，提高道路通行能力。

通过GPS技术，系统能够动态监测车辆的行驶速度、发车间隔等指标，并对司机超速行为进行提示和管控。根据车辆实时运行情况，可进行实时调度，避免运力的浪费。当车辆发生故障时，车辆GPS技术可及时报告故障，并记录车辆故障发生位置，有利于救援工作的开展。同时，GPS系统还能记录车辆行驶里程数、发车次数、发车间隔等数据，为后续与行车相关的数据查询、统计和分析提供可能。从车辆行驶速度和发车间隔数据出发，得出的车辆运营指标，对公交公司车辆调度、维修等方面提供良好指导。同时GPS技术还可以为公交公司提供更多精准的数据支持。例如，通过车辆GPS定位系统，后台可以实时监测车辆位置、发车时间、行车路线等信息，进一步提升运营效率和维护质量。同时，通过数据分析，可以深入了解乘客出行需求、时段特点，从而制定更加科学的运营策略，提高服务水平。

此外，结合智能手机等移动终端设备，可以利用GPS技术对公交线路进行实时查询和导航，在一定程度上提高乘客的出行体验和便捷度。而当GPS技术与其他先进技术（如人工智能、云计算等）相结合时，将大大拓展其应用场景和潜力，进一步提升公交行业的创新和发展。

2.2 基础数据介绍

本文利用的主要数据包括公交IC卡数据、公交车载GPS数据、快速公交（BRT）刷卡数据。这些数据主要记录的是乘客的刷卡时间与刷卡类型、乘客的进站时间、公交车的位置信息、地铁上的进出站、进出站时间以及卡的型号。这些信息主要是用于匹配乘客的公交上车站点与上车时间并且经过算法推算匹配其换乘站点或者判断是否为无需换乘的通勤。同时此文还运用到线路站点数据、联锁转乘站点数据、站点区域数据等公交基础数据。这些数据主要记载的是成都市公交线路的编号、站点信息、公交自身信息以及站点之间相互的距离等信息。同时这些信息构成了数据库平台搭建所需要的基础数据。

2.2.1 公交IC卡数据

公交IC卡数据是指乘客在乘坐公交时使用IC卡在系统中记录下来的用户行为和信息。目前国内很多城市都在大规模地使用IC卡，成都市已经实现了公交刷卡系统的全覆盖，自动售票系统已经实现了在公共交通系统中的广泛应用。例如成都市在2021年1月13日的IC卡记录就达到了2530763条。

本文采用的公交IC卡数据包括但不限于：单日出行序号、IC卡类型、卡号、IC卡线路名称、刷卡时间、公交编号、消费类型。

**表 2-1 IC卡数据介绍**

|  |  |  |  |
| --- | --- | --- | --- |
| 列数 | 含义 | 内容 | 示例 |
| 1 | RID | 单日出行序号 | 1、2、3 |
| 2 | ICTYPE | IC卡类型 | BUS |
| 3 | CARDNO | 卡号 | 6100000000028080 |
| 4 | IC\_LINENAME | IC卡线路名称 | 651 |
| 5 | CONSUMEDATE | 使用时间 | 2021-01-13 08:23:09.000 |
| 6 | BUSNO | 公交编号 | 169009 |
| 7 | CONSUMETYPE | 消费类型 | 1、2、3 |

2.2.2 公交车载GPS数据

公交车载GPS数据是指公交车通过车上的全球定位系统（GPS）上传的位置和状态信息。公交车载GPS数据以单辆公交为载体，能够以20-30s为间隔上传数据，反应车辆在公交固定运行线路上的开行状态。

本文采用的公交GPS数据包括但不限于：线路编号、公交编号、运行时间、当前经度、当前纬度、站点编号、是否到站。

**表 2-2 公交车载GPS数据介绍**

|  |  |  |  |
| --- | --- | --- | --- |
| 列数 | 含义 | 内容 | 示例 |
| 1 | LINENUM | 线路编号 | 90002 |
| 2 | BUSNO | 公交编号 | 066620 |
| 3 | ACTDATETIME | 运行时间 | 2021-01-13 07:30:10 |
| 4 | LONGITUDE | 当前经度 | 104.122151 |
| 5 | LATITUDE | 当前纬度 | 30.66887 |
| 6 | STATIONNUM | 站点编号 | 10193 |
| 7 | ISARRLFT | 是否到站 | 1、2、null |

2.2.3 快速公交（BRT）数据

成都的特色公交——快速公交（BRT，Bus Rapid Transit）是一种高效、快速、舒适的公交系统，通常采用专用车道、快速登车和离线以及其他技术来优化运营。

快速公交数据是指在快速公交系统中收集的有关乘客出行情况、运营信息以及车辆位置等数据。快速公交数据包括乘客流量数据、班次信息、车辆位置数据、站点到达时间等。这些数据可以提供实时运营状态，帮助运营商改进服务质量和提升效率。

本文采用的快速公交BRT的数据包括但不限于卡号、消费日期、消费类型、快速公交站点编号。

**表 2-3 快速公交BRT数据介绍**

|  |  |  |  |
| --- | --- | --- | --- |
| 列数 | 含义 | 内容 | 示例 |
| 1 | CARDNO | 卡号 | 6100100021485559 |
| 2 | CONSUMEDATE | 消费时间 | 2021-01-13 |
| 3 | CONSUMETYPE | 消费类型 | 1000 |
| 4 | STATIONNUM | 快速公交站点编号 | 91005 |

2.2.4 GIS地图数据

地理信息系统（GIS）是一种基于计算机技术的地理空间数据处理和分析工具，可以帮助人们更好地理解和管理地球表面的事物。GIS地图数据是GIS中最重要的部分。GIS地图数据可以从多个来源获取，包括卫星数据、GPS数据、航空摄影、传感器网络等。通过对多源数据的融合，GIS可以提供更为精确的地图信息。GIS地图数据包括矢量数据和栅格数据。矢量数据由点、线、面等几何对象组成，描述地理对象的位置和属性。栅格数据则是将地理区域划分为一个个像素，在每个像素位置上记录相应的属性值。GIS地图数据可以包含各种地理信息，如道路、河流、山脉、建筑物、城市轮廓等。此外，还可以包括人口统计、社会经济等非地理信息，以便支持分析和决策。

GIS地图数据广泛应用于城市规划、交通规划、环境保护、土地利用、资源管理等领域，也被用于导航、地图制作、游戏等普通用户的日常生活中。

总之，GIS地图数据是一种记录地球表面地理信息的重要数据形式，它通过整合多源数据，描述各种地理对象和属性，并提供支持分析和决策的功能。GIS地图数据应用范围广泛，是现代社会中不可或缺的数据资源之一。本文中将利用成都市的地图配合最终生成的OD数据实现可视化，使结果展示更加具象化。

本文运用到的GIS数据包括成都市的地图数据、公交线路数据和公交站点数据。在此基础上，根据推算得到的公交OD数据将添加交通大区之间的OD数据等。

2.2.5 线路站点数据

线路站点数据主要是记录的是一条公交运行线路中会涉及到的公交车辆所属的信息，以及其中的站点信息。这些信息主要是用于匹配公交运行中的上车站点和下车站点。因此线路站点数据包括但是不限号于：公交公司名称、车队名称、线路名称、区段名称、运行方向、站点编号、站点名称、线路到站顺次、站点经度、站点纬度、站点类型、站点类型名称、线路编号、距离线路下一站的距离、线路目前总运行距离、区域ID、是否丢失GPS、线路类型、大区编号。

**表 2-4 线路站点数据**

|  |  |  |  |
| --- | --- | --- | --- |
| 列数 | 含义 | 内容 | 示例 |
| 1 | ORG | 公交公司名 | 郫县公司 |
| 2 | ORGNAME | 车队名称 | 1202车队 |
| 3 | LINENAME | 线路名称 | P214 |
| 4 | SEGMENTNAME | 区段名称 | P214s |
| 5 | DIRECTION | 运行方向 | 上行 |
| 6 | STATIONNUM | 站点编号 | 51601 |
| 7 | STATIONNAME | 站点名称 | 古城公交站 |
| 8 | SEQ | 线路到站顺次 | 1 |
| 9 | LONGITUDE | 站点经度 | 103.92851 |
| 10 | LATITUDE | 站点纬度 | 30.89721 |
| 11 | STATIONTYPEID | 站点类型 | 8 |
| 12 | STATIONTYPENAME | 站点类型名称 | 副站上客站 |
| 13 | LINENUM | 线路编号 | 1200214 |
| 14 | DISTANCE | 距离前一站的距离 | 618.38288364 |
| 15 | milepost | 目前总运行距离 | 1844.63754404405 |
| 16 | zoneid | 区域ID | 304 |
| 17 | LOSTGPS | 是否丢失GPS | 否 |
| 18 | LINETYPE | 线路类型 | BUS |
| 19 | largezoneid | 大区编号 | 10 |

2.2.6 联锁转乘站点数据

假设所有的乘客为了到达目的地都按照最短路径来换乘公交，那么为了匹配出乘客可能的换乘站点则需要获取某线路上的某站点和其附近最近的两个站点之间的距离，和这三个站点本身的信息。因此，联锁转乘站点数据包括但是不限于：线路1的名称、线路1的运行方向、线路1中的到站顺次、站点1的编号、线路2的名称、线路2的运行方向、线路2中的到站顺次、站点2的编号、站点1的横向坐标、站点1的纵向坐标、线路1到站的距离、线路2到站的距离。

**表 2-5 联锁转乘站点数据**

|  |  |  |  |
| --- | --- | --- | --- |
| 列数 | 含义 | 内容 | 示例 |
| 1 | LINENAME1 | 线路1的名称 | 304 |
| 2 | direction1 | 线路1的运行方向 | 下行 |
| 3 | SEQ1 | 线路1的到站顺次 | 14 |
| 4 | stationnum1 | 站点1的编号 | 42006 |
| 5 | LINENAME2 | 线路2的名称 | 1076 |
| 6 | direction2 | 线路2的运行方向 | 环行 |
| 7 | SEQ2 | 线路2的到站顺次 | 1 |
| 8 | stationnum2 | 站点2的编号 | 31774 |
| 9 | DX | 站点1的横向坐标 | 18.870001 |
| 10 | DY | 站点1的纵向坐标 | 16.174747 |
| 11 | MILEPOST1 | 线路1的到站坐标 | 3114.1628 |
| 12 | MILEPOST2 | 线路2的到站距离 | 5881.6764 |

2.2.7 站点区域数据

站点区域数据指的是将整个成都市划分为一个一个的大区,如金牛区、郫都区、青羊区等，并且在大区之下又划分为多个小区，并且将小区之中的线路进行记录。此类的数据将用于按照区域分类时匹配各个线路对应的站点编号，是在上下车站点推算完成后形成最终OD矩阵的基础，同时是辅助画出蛛网图和热力图，实现OD数据可视化的重要基础。站点区域数据包括但是不限于：线路类型、站点编号、小区编号、大区编号。

**表 2-6 站点区域数据**

|  |  |  |  |
| --- | --- | --- | --- |
| 列数 | 含义 | 内容 | 示例 |
| 1 | LINETYPE | 线路类型 | BUS |
| 2 | STATIONNUM | 站点编号 | 60357 |
| 3 | smallzoneid | 小区编号 | 574 |
| 4 | largezongid | 大区编号 | 8 |

2.3 基础数据的预处理

需要预处理的数据主要包括公交IC卡数据、车载GPS数据、BRT数据、地铁刷卡数据。由于各种情况原始数据会出现数据错误或者是数据丢失的情况，特别的，公交IC卡数据、公交车载GPS数据等原始数据源本身体量大，会拖缓系统的处理速度，因此需要利用Oracle和SQL Server来对基础数据进行预处理。数据预处理主要是剔除数据中的错误数据如定位丢失、卡号丢失、公交编号丢失等情况。此外，为了生成OD表还需要通过匹配公交车载GPS数据和线路站点数据来生成一张公交单日进出站发车时刻表。

2.3.1 公交IC卡数据

公交IC卡数据在乘客刷卡之后会存储于对应车辆的车载读卡器中，需要通过数据采集盒将IC卡记录导出至计算机，再到Oracle和SQL SERVER 软件之中进行处理。对于IC卡数据主要进行剔除操作，即保留卡号、时间、公交编号等信息，并对出现下列情况的IC卡记录进行删除：

1. 刷卡时间不在研究范围内的数据；
2. 出现完全一样的重复数据；
3. 卡号和时间属性相同但是匹配的公交编号不同的数据；
4. 单日出行序号、IC卡类型、卡号、IC卡线路名称、使用时间、公交编号或消费类型字段为空的数据

2.3.2 公交车载GPS数据

公交车载GPS数据记录了公交车辆的运行位置信息、车辆信息等，这些信息可用于形成单日的公交进出站时间表、匹配乘客的上下车站点、匹配不同条件下的换乘站点等。是OD推算的重要数据源，但是GPS上传位置信息的频率为20-30秒左右，加上成都市本身大量的公交线路和公交车辆，因此公交车载GPS数据的体量巨大。为了提高整个数据源的有效性、减轻计算机的负担、优化软件的处理速度，需要对大量的公交车载GPS数据进行剔除操作，仅保留公交编号、线路编号、是否进出站等重要信心，并且按照下列条件删除无用的数据：

1. ISARRLFT数据如果为NULL；
2. 经纬度数据超出了成都市的边界或超出了此条公交线路的合理范围；
3. 同一公交编号在同一时间匹配了不同的公交站点或者不同的经纬度信息；
4. 线路编号、公交编号、运行时间、经纬度信息、站点编号丢失的数据。

2.3.3 快速公交（BRT）数据

快速公交（BRT）数据是最终生成OD表的重要一环，此数据本身的字段比较简单，但还是会出现无效数据的情况。因此，在保留卡号、刷卡时间、站点等重要信息的情况下，对出现下列情况的数据进行清洗：

1. 同一卡号在同一时间匹配了不同站点的数据
2. 卡号、消费时间、消费类型或快速公交站点编号为空的数据
3. 同一卡号、同一站点在短时间内出现两次的数据

2.3.4 公交单日站点发车时刻表

在生成OD表的过程中，成都市内公交在各自线路上各个站点的发车时刻表是重要的一环。公交单日站点发车时刻表主要包含了公交编号、线路编号、站点编号、线路运行方向、站点在运行线路中的到站顺次、发车时间等信息。此类信息分散在公交车载GPS数据、线路站点数据中，无法直接获取。首先需要提取公交车载GPS数据中ISARRLFT字段值为2的数据，此数据代表了公交车离开某一站点。再将筛选完成的数据按照公交编号为主字段，运行时间为次字段进行排序。此后，将上述数据与线路站点数据的公交编号、运行线路编号进行匹配。

完成之后，对原始的公交单日站点发车时刻数据表进行数据清洗工作，如数据出现了如下情况则进行清洗：

1. 公交车发车时间在短时间（5min以内）匹配了不同站点；
2. 公交车在不同的发车时间内匹配了相同站点，且次站点非终点站或始发站；
3. 公交车的发车站点不属于其对应的公交运行线路；
4. 任意字段为空的数据；

2.4 本章小节

本章首先介绍了地面公交IC刷卡数据、公交车载GPS数据、线路站点数据、快速公交的刷卡数据、GIS地图数据、联锁转乘站点数据和站点区域数据的特征和结构。其次，对出现重复记录、缺失数据、异常数据等情况的数据进行了预处理，并通过线路站点数据和GPS数据融合处理生成了公交单日进出站发车时刻表，为后文中进行上车站点、下车站点的匹配和出行OD的推算提供数据基础。

**第3章 OD推算方法**

基于IC卡和公交车载GPS数据，搭配现有的公交线路站点等数据推算而来的公交出行OD数据可以用来进行交通规划和设计，并提供更好的公共交通服务，这对城市交通调度和未来的公交线路的优化都具有重要意义[33]。对比于传统的通过问卷调查的形式来获取上下车的站点并形成OD数据。通过IC卡数据和车载GPS数据推算公交OD能够大大降低数据搜集的成本并且扩大数据样本。但是IC卡数据中仅有乘客的消费时间和对应的公交编号，并没有直接记录下上车站点的信息，同样的GPS数据中也不会记载有关乘客上下车的信息。因此，推算公交的OD必须依据IC卡数据和公交车载GPS数据进行上下车站点的匹配[32][31]。同时，鉴于运行时间处于夜间23点至6点之间的公交线路稀少且出行人群在一天中的占比少，人群的出行特征也难以分析，因此本文中只对出行时间位于白天6点至夜间23点的乘客进行出行OD的推算。

3.1 上车站点的推算

国内外的学者针对根据IC卡数据和GPS数据对乘客的上车站点进行推算的研究方法相对一致，本文同样采取基于时间段匹配的算法。此算法的研究相对成熟且匹配度高，适用性广。其基本原理是通过公交进出站的时间段和乘客的刷卡时间进行匹配，从而得到乘客的上车站点。

3.1.1 基本假设

公交车的行驶时间段中可以分为停站时间和在途时间，而乘客的刷卡时间一定是在公交的停站时间内的。选取公交车载GPS数据中距离站点最近的一次定位数据判定为到站。此条数据中的GPS数据上传时间判定为到站时间TGPS，同时考虑到GPS的上传间隔为20至30秒，所以乘客的刷卡时间TIC应该满足公式3-1：

TGPS-30s<TIC<TGPS+30s 公式 (3-1)

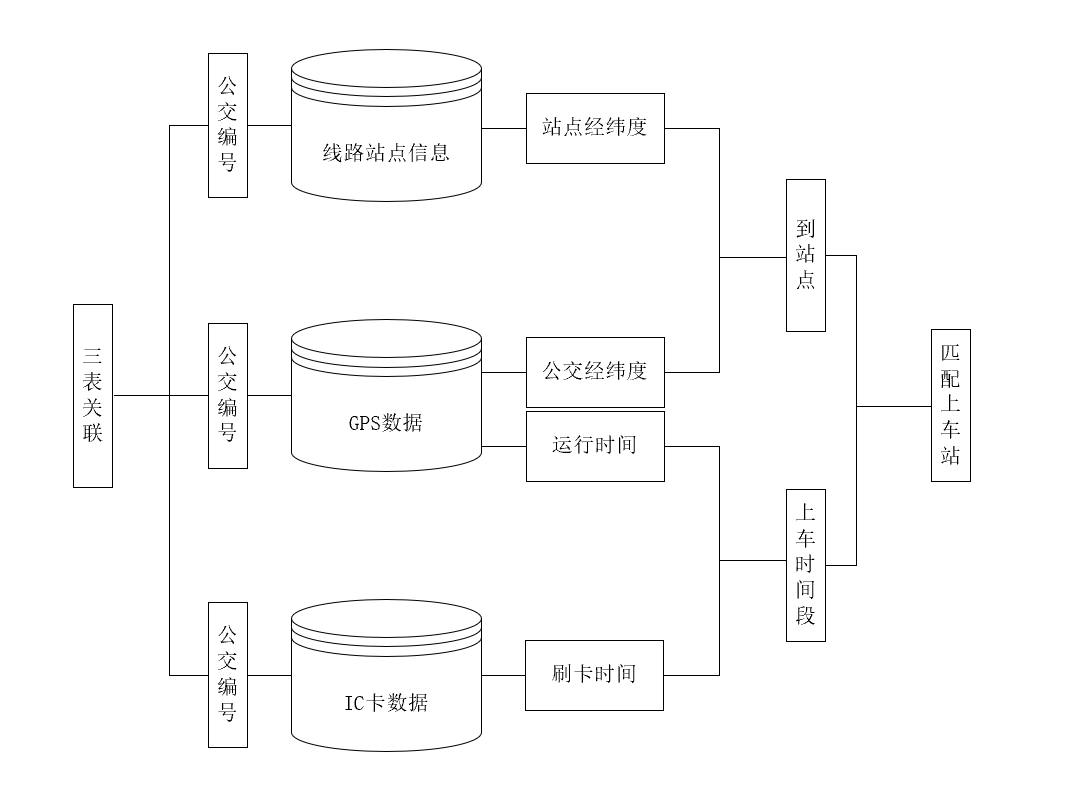
而现实情况中，公交车辆的停站时间和GPS上传的时间内有一定的时间差异，并且车辆的停站时间不一定为1分钟[29]，乘客的数量、公交站的拥挤程度都有可能加长停站时间。此类误差将导致乘客上车站点的匹配丢失，或者错误匹配。因此为了提高上车站点的匹配准确性，本文以线路站点数据中站台的经纬度为圆心设定一个半径为5米的圆形范围，一旦公交的定位经纬度第一次碰到了圆形的边界，则判定公交进入了停站状态即到站，第二次碰到了圆形边界时，则判定公交进入了在途状态即离站[27]。将此两次GPS数据的上传时间设定为Tarr和Tlft。因此可以将上一公式优化为公式3-2

Tarr<TIC<Tlft  公式 (3-2)

优化之后的公式可以提高上车站点的匹配率，并且通过判定公交的进出站时间可以简化数据处理的时间成本和算法的复杂程度。

3.1.2 上车站点匹配算法

首先根据IC卡数据中的公交编号锁定对应的公交车辆和运行线路，并确定对应公交线路中的站点信息，在此基础上，将车辆进出站时间和乘客的刷卡时间相结合，最终得到乘客刷卡时最有可能的上车站点。其算法的匹配流程如图4-1所示。上车站点的匹配主要是通过公交编号联锁匹配公交运行线路站点、基于站点和公交运行经纬度匹配进出站点及时间、公交进出站时间段和IC卡刷卡时间匹配这三个步骤[26]。



**图 3-1 山车站点匹配流程图**

1.步骤一：通过公交编号联锁匹配公交运行线路站点

在匹配上车站点之前，首先需要将乘客乘坐的公交编号和公交运行线路匹配出来。公交车载GPS数据中包含了公交的编号，同时IC卡数据中也有公交编号的信息，将两份数据库的公交编号（BUSNO）进行匹配可以确定乘客所乘坐的公交以及此公交的运行线路（LINENUM）。将已经锁定的公交线路放入线路站点数据库中查找可以得到该运行线路上所有的站点信息，其中最重要的是每个站点的经纬度。

按照上述步骤可匹配到2021年1月13日IC卡号为61005\*\*\*\*的乘客所乘坐公交的编号为049846，对应的公交运行线路编号为180010，该线路的部分站点信息如表4-1所示：

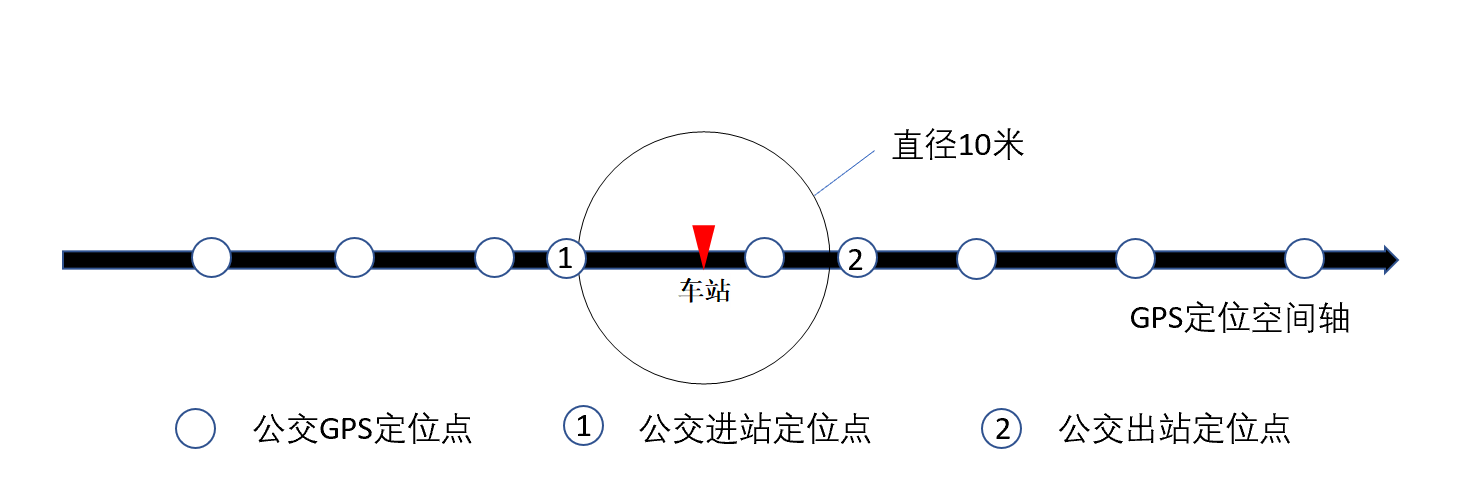
**表 3-1 线路180010部分站点信息示意表**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 站点到站顺次 | 站点名称 | 站点编号 | 站点经度 | 站点纬度 |
| 1 | 红砂村公交站 | 41292 | 104.157805 | 30.581911 | |
| 2 | 花博路口 | 40192 | 104.1524 | 30.58517 | |
| 3 | 华西第二医院锦江院区 | 40190 | 104.148331 | 30.587628 | |

2.步骤二：基于站点和公交运行经纬度匹配进出站点及时间

判断一辆公交车已经进站或者离站是利用时间段匹配乘客的上车站点的重要前提。结合步骤一中已经得到了运行线路中的所有站点的经纬度，利用公交车载GPS数据中记载的经纬度进行关联，并按照空间上距离最近的原则将离车辆最近的站点确定为公交的到离站点[28]。

在现实情况中，因为公交站短时公交车辆拥堵的情况、站台候车乘客数量过多、司机停车失误等多种情况，公交的实际进站经纬度信息和线路站站点的经纬度信息存在一定偏差，公交的确切进站时间同样会受到影响。这种偏差可能缩短公交的进出站时间段并最终导致乘客的IC卡刷卡时间没有合适的匹配结果，出行OD也会相应受影响。因此，为了提高乘客上车站点匹配结果的准确度和避免出行OD因为没有匹配对象而丢失，本文为每一个公交站设置了一个距离范围。其原理是以线路站点信息中的公交经纬度为中心向外拓展，形成一个直径为5米的圆形，当公交车的经纬度接触到圆形的边界时就视为公交进出站。将GPS数据按照时间顺序排列，当公交第一次进入这个圆形范围时就视为公交进站，并以此时的时间节点记作进站时间，此后当公交第一次到达圆形范围外视为公交出站，并以此时的时间节点记为出站时间，两次时间节点的差即为公交在站停留时间。整个进出站确定的过程如示意图4-2所示。



**图 3-2 公交进出站判断示意图**

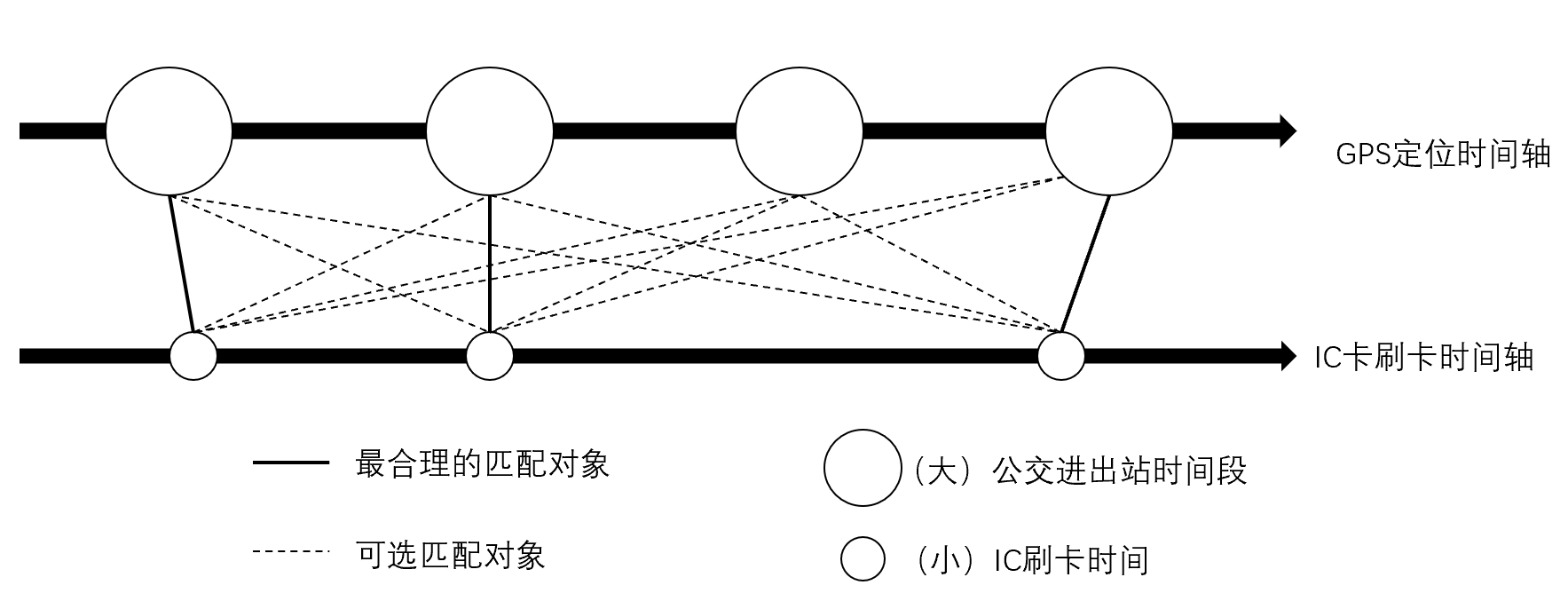
在SQL Server数据库中进行上车站点推算的具体方法是通过提取线路站点信息表和GPS数据表中经纬度数据，通过经纬度之差计算公交与站点之间的距离[29]，当计算的距离第一次小于10米时，将公交车载GPS数据中ISARRLFT字段的值更新为1，即标记为公交进站定位点；当公交与车站距离第一次大于5米时，将公交车载GPS数据中ISARRLFT字段的值更新为2，即标记为公交进站定位点。其匹配后的公交车载GPS数据案例如表3-2所示。

**表 3-2 进出站匹成功的GPS数据**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 线路编号 | 运行时间 | 公交编号 | 站点编号 | 是否到站 |
| 56 | 2021-01-13 18:18:43.0 | 049846 | 20198 | 1 | |
| 56 | 2021-01-13 18:20:41.0 | 049846 | 20198 | 2 | |

3.步骤三：公交进出站时间段和IC卡刷卡时间匹配

通过步骤二的结果，公交车载GPS数据不仅匹配好了是否进出站的信息，还记录了对应了时间段。因此将进站的时间节点记为T1，出站的时间节点记为T2。对应的IC卡中的刷卡时间（CONSUMEDATE）记为T。将每一组的T1,T2和刷卡时间T进行对比，选取时间差距最小的一组，并且把此组进出站时间段匹配的公交站点判定为乘客的上车站点[30]。此步骤的匹配流程简易图可由图3-3展示：



**图 3-3 公交进出站时间段和IC刷卡时间匹配**

匹配好的上车站点数据表包括了单日出行次数、IC卡号、刷卡时间、公交编号、站点经纬度、站点编号、公交运行线路编号、公交运行方向、站点的到站顺次。匹配好上车站点后的部分数据如表4-3所示：

**表 4-3 成功匹配上车站点的数据（部分）**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 卡号 | 刷卡时间 | 公交编号 | 站点编号 | 方向 |
| 6100\*\*\* | 2021-01-13 6:25:19.0 | 028519 | 40829 | 下行 | |
| 6100\*\*\* | 2021-01-13 6:25:21.0 | 017320 | 31231 | 下行 | |
| 6100\*\*\* | 2021-01-13 6:25:21.0 | 019921 | 32429 | 下行 | |
| 6100\*\*\* | 2021-01-13 6:25:21.0 | 038576 | 31577 | 上行 | |
| 6100\*\*\* | 2021-01-13 6:26:07.0 | 161151 | 15000653 | 下行 | |
| 6100\*\*\* | 2021-01-13 6:26:10.0 | 032052 | 123 | 下行 | |
| 6100\*\*\* | 2021-01-13 6:26:12.0 | 169558 | 1500650 | 下行 | |
| 6100\*\*\* | 2021-01-13 6:26:15.0 | 072026 | 70007 | 上行 | |

3.2 下车站点推算

目前成都市的IC卡数据中并不记载乘客的下车时间或者下车站点等信息，无法直接通过现有数据表之间的字段匹配方法来得到下车站点的信息。目前针对下车站点推算的主流方法有如下两种：基于刷卡的换乘分析和基于出行链规律的分析。鉴于公交乘客出行规律的多样性和差异性，单单依靠一种方法进行下车站点的推算，会造成大量的IC卡数据无法匹配下车站点，算法的匹配度会下降，最终的OD的可靠性同样会大大降低。因此，本文利用两种算法针对不同的人群进行下车站点的推算，以此提高最终的下车站点的匹配度。

3.2.1 乘客出行链闭合假设

城市内的公交运行线路和公交站点在常态下都是固定的，且城市内的居民的出行需求具有一定的规律，如通勤类人群、通学类人群，这两个条件决定了乘客利用公交出行时具有一定的规律性。因此本文可以根据乘客表现出的出行规律提出对应的出行链假设[31][29]。

本文提出的出行链闭合假设的内容如下：

1. 乘客单日内的首次出行从A地点出发之后，无论后期经过几次换乘，无论最后一次出行和第一次出行之间的时间间隔为多久，最后一次出行的目的地均为A地点。即当乘客的单日出行次数超过一次时，其单日内的首次出行和最后一次出行互为往返关系；
2. 乘客的公交为连续的，即乘客在数次公交出行之间不会和其他的交通工具之间换乘操作，仅在不同的公交运行线路之间换乘；
3. 乘客如需换乘则采用最少换乘次数或者最短距离的换乘方式，即乘客第一次出行的下车站点和第二次出行的上车站点为同一车站，或者是第一次出行的下车站点和第二次出行的上车站点之间距离短，一般认定为步行时间在20分钟之内或者是直线距离在300米以内。此条件可以配合联锁换乘表来推算下车站点。

为了实现出行链闭合理论的可行性，根据相关文献中对于往返出行人群的特征判别研究，本文从时间纬度和空间纬度一共提取了3个描述往返出行行为的特征点：单日出行次数、出行时间段、一周内的总计出行次数。即单日出行次数需要大于或等于两次，出行的时间段在白天，一周内的出行总次数需要在10次及以上。由这三个特征作为出行链闭合理论可行的前置要求。

同时本文认为，在一天之内出行数（RID）大于9次的乘客不是常规乘客，将不划入考虑范围内，因此在数据融合处理的过程中删掉了一天之内乘坐公交次数大于9次的乘客。此类数据每日的数据量一般在几万条，加入计算则会影响最终的出行需求推算的结果。

例如在剔除了单日出行数大于9次的乘客之后，2021年1月13日的单日IC卡数据量为1190862条。同时，经过数据处理后，可以得到单日出行中不同出行的次数的乘客总人数的统计数据，在融合2021年1月13日和14的数据后其具体数据如表4-4所示。

**表 3-4 乘客出行次数统计表**

|  |  |  |
| --- | --- | --- |
| 出行次数 | IC卡数量 | 数据占比 |
| 1 | 448597 | 37% |
| 2 | 441980 | 37% |
| 3 | 130723 | 10% |
| 4 | 102248 | 9% |
| 5 | 35124 | 3% |
| 6 | 18225 | 2% |
| 7 | 7165 | 1% |
| 8 | 3784 | 1% |
| 9 | 1463 | 1% |

可以看出来，成都市的公交乘客中单日出行次数在一次和两次的为大多数，其具体数据基本持平。随着出行次数的增加，其占据的比例越小，而从整体数据来看，成都市公交乘客单日内多次出行的占据多数，占据总数的63%，而在多次出行的数据中，出行两次的占比最大，可以推测为成都市工作日中大部分乘客的出行规律符合往返类人群的出行规律。因此可以采用基于出行链闭合原理的乘客出行规律进行下车站点的推算。

同时将IC卡数据按照时间线进行处理排序，同样剔除单日出行次数大于9次的人群，按照一小时为间隔统计了出行人次。将2021年1月13日和14日的出行人次进行数据处理后，其具体结果由图4-4展示展示。

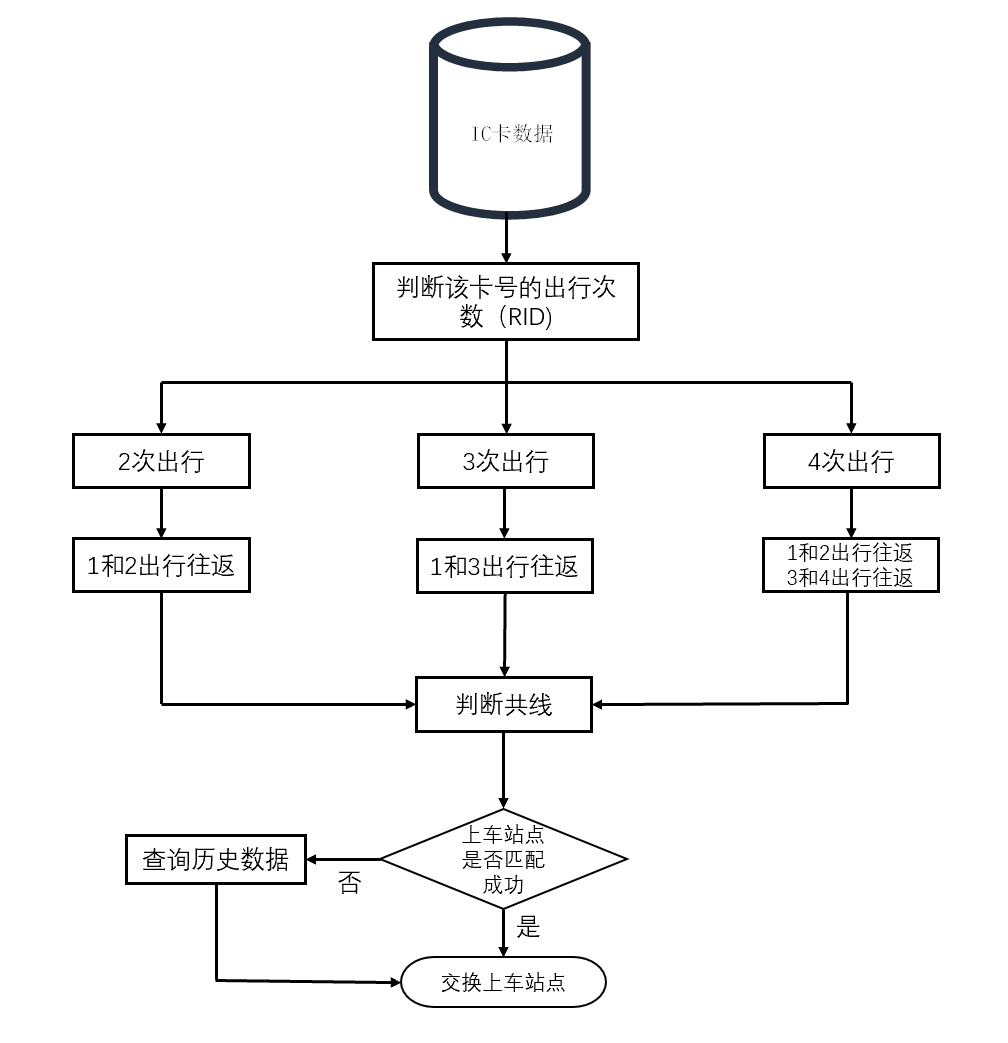
**图3-4 2021/1/13-14日出行人次-时间图**

2021年1月13日和14日皆为工作日，根据图3-4所示，两者的结果基本一致，图形曲线高度重合，只是数值上不同。同时，两者都呈现了出行双高峰，其时间段都在7-9点和17-20点之间，且早高峰的出行人数都在30万以上，晚高峰都略低于30万，且8-20点之间的统计数据占据了全天的85%以上，也符合往返类出行条件中的出行时间段要素。同时可以得出成都市人群的出行呈现明显的规律性和固定性，由此判定单日出行人群的主体满足一周内出行次数能达到10次及以上。

鉴于成都市公交绝大多数的乘客满足此假设的三个前置要求，所以本文认为此类乘客的出行可以判定为符合出行链闭合理论中的描述，故采用出行链闭合理论作为下车站点推算的前置假设，以便简化下车站点的推算研究。

3.2.2 往返出行的下车站点推算

根据出行链闭合假设，首先可以进行乘客单日内首次出行和末次出行的下车站点的推算。利用往返出行的规律，第一次出行匹配得到的上车站点可以判定为最后一次出行的下车站点。如出现多次出行的情况，则进行分类判别，其具体的推算流程可如下图展示：



**图 3-5 2往返类人群下车站点推算图**

对于往返类出行人群可以根据单日内出行次数再划分为简单出行类乘客和多次出行类乘客。其中的简单往返类乘客主要由直达通勤和直达式出游人群构成。直达通勤类乘客指的是从居住地到达工作地点之间只需要乘坐一次公交，不需要进行转乘就可以直接到达目的地的人群。通常此类通勤人群的出行特征为IC卡刷卡纪律集中在7-9点和17-20点之间的早晚上班高峰期，自然形成了一个往返链。直达式出游人群的出游时间分散于各个时间段，不具有规律性，但其主要特征是单日出行次数仅有两次。而多次出行类乘客的出行次数在2次以上，但是一般不会超过7次，且最后一次出行和第一次出行依旧构成往返关系，其匹配的运行线路会有多个重复站点。此类乘客的出行无法通过简单的出发地直达目的地的规律来分析，只能利用往返规律来分析。

针对这些乘客的下车站点推算需要进行如下情况的分析讨论。

（1）单日出行为2次：

情况一：第一次出行的上车站点已经成功匹配，并且第二次的上车站点也成功匹配，同时两次匹配的上车站点连带的公交运行路线编号（LINENUM）也相同，则可以判定为乘客是采取单一线路进行通勤，直接将两次出行的上车站点进行交换，便可以得到该乘客单日内两次出行的下车站点，即另一次出行的上车站点。

情况二：第一次出行的上车站点已经匹配成功，并且第二次的上车站点也匹配成功，但是两次匹配的上车站点连带的公交运行线路编号不同。此时则需要判断两条线路中是否共线，并且包含了另一次出行中的上车站点。如果确定包含了另一次出行中的上车站点则同样认为是情况1的类似情形，同样将两次出行的上车站点进行交换，便可以得到该乘客单日内两次出行的上车站点。

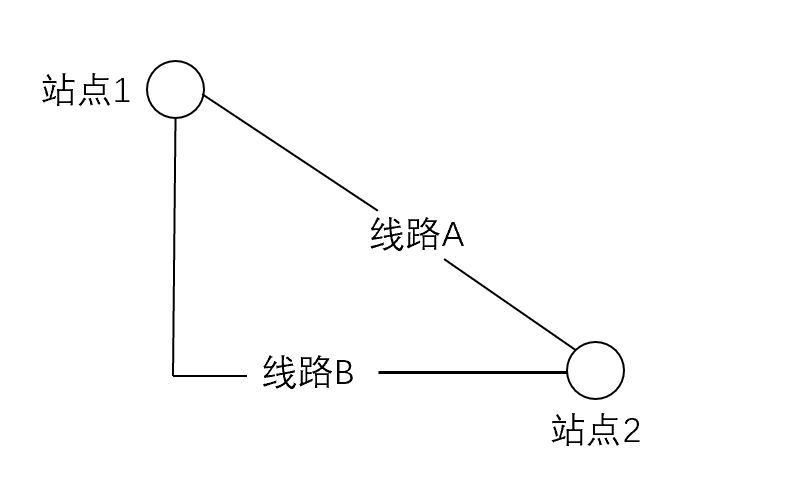
情况三：两次出行的起点均未匹配成功，则匹配历史出行中的上车站点，本文所拥有的数据为2021年1月13日至16日成都市内的公交IC卡记录，例如13日有两次出行都未能匹配的则去后面三日的数据中寻找相同卡号的IC卡记录，并定位其出行的上车站点，如其他日期中的上车站点及运行线路相同，则在得到后交换两次的上车站点作为另一次出行的下车站点。

情况四：两次出行的起点均未匹配成功，则匹配历史出行中的上车站点，本文所拥有的数据为2021年1月13日至16日成都市内的公交IC卡记录，例如13日有两次出行都未能匹配的则去后面三日的数据中寻找相同卡号的IC卡记录，并定位其出行的上车站点，如其他日期中的上车站点及运行线路不同。此时则需要判断两条线路中是否共线，并且包含了另一次出行中的上车站点。如果确定包含了另一次出行中的上车站点则同样认为是情况1的类似情形，同样将两次出行的上车站点进行交换，便可以得到该乘客单日内两次出行的上车站点。

（2）单日出行为3次，认为第一次出行和第三次出行是互为往返关系。并且按照单日出行为2次的情况进行分类分析。

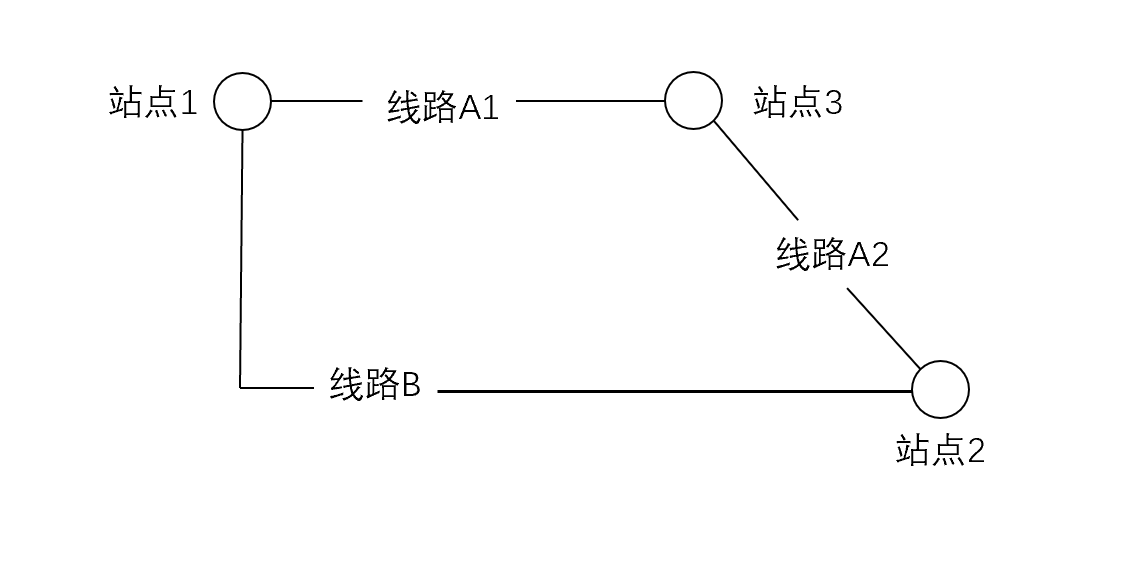
（3）单日出行次数为4次，认为第一次出行和第四次出行是互为往返关系，第二次出行和第三次出行是互为往返关系，并按照如下分类处理：

情况一：寻找线路A和线路B共有的站点，如果存在相同的站点，则认为两者共线，其示意图如3-6所示；



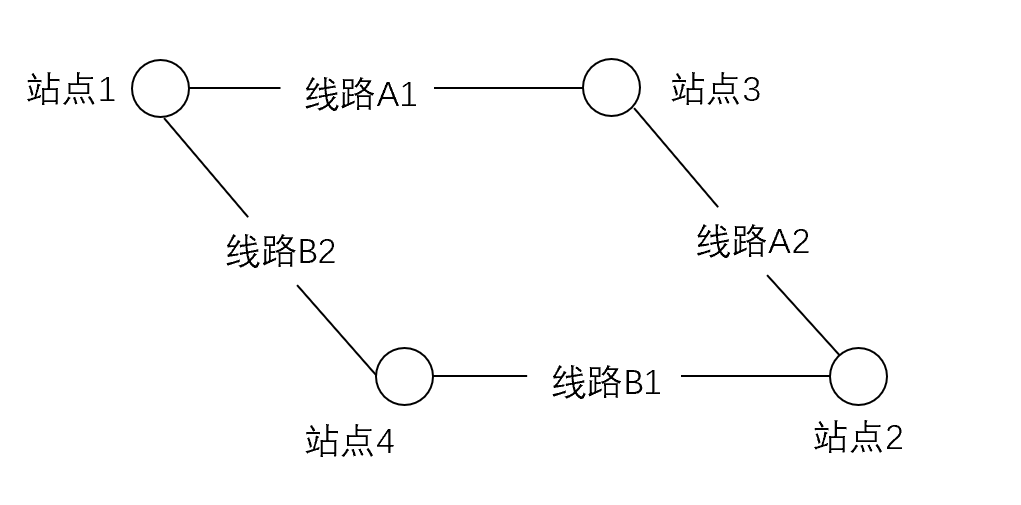
**图 3-6 共线线路类别1**

情况二：寻找线路B和线路A1的共有站点，然后找线路B和线路A2共有的站点，如果存在相同的站点，则认为两者共线，其示意图如3-7所示；



**图 3-7 共线线路类别2**

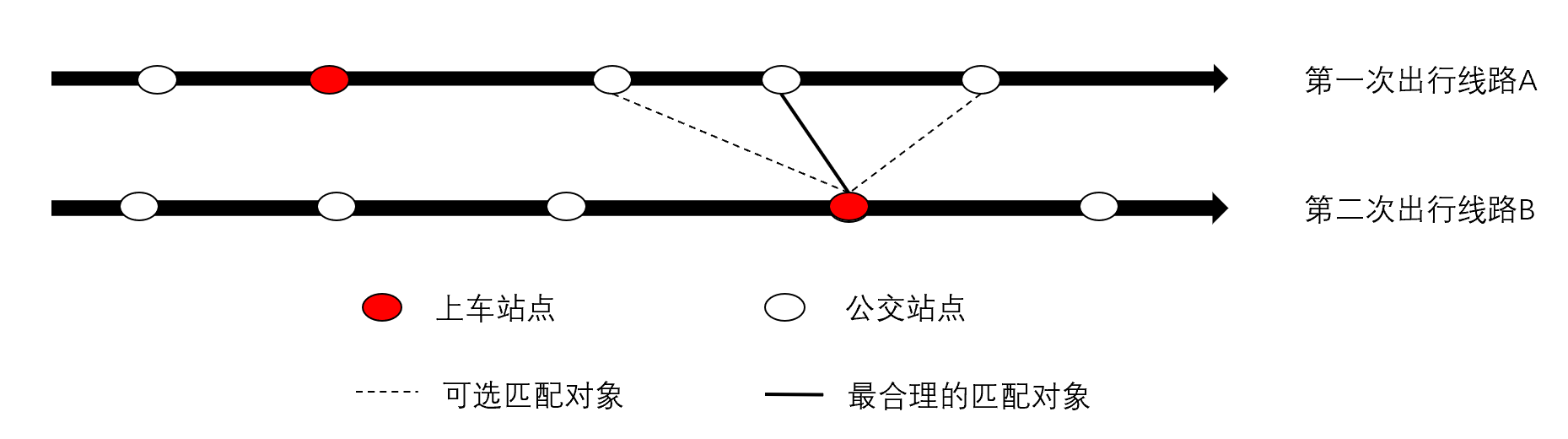
情况三：找线路B1和线路A1共有的站点，然后找线路B2和线路A2共有的站点，如果存在相同站点，则认为两者共线，其示意图如3-8所示。



**图 3-8 共线线路类别3**

3.2.3 换乘过程的下车站点推算

单日出行大于2次的乘客便可以判定为其单日出行中出现了换乘的情况，则换乘的站点可以视为中间站点，此类情况一般出现在出行次数为奇数次的乘客中，且中间一次的出行记录判定为换乘出行。例如单日出行为三次的乘客，其中第二次出行则视为换乘出行，且无法和第一次或最后一次出行形成往返关系。针对此类的乘客，沿用出行链闭合假设中的第二条内容即假设所有乘客的出行方式采用最短行程，如需要进行换乘则换乘出行的下车站点和下一次出行的上车站点有两种情况；一是换乘出行的下车站点和下一次出行的上车站点为同一站点；二是换乘出行的下车站点和下一次出行的上车站点的直线距离在300米以内。其具体的推算流程可以如图4-9展示：



**图 3-9 换乘站点匹配**

本文以出行次数为3次为示例，其余奇数次出行与三次出行的情况类似。如出行次数为3次，则第二次出行判定为换乘出行。其具体推算流程如下展示：

步骤一：如第二次出行的上车站点未匹配成功，则判定此次出行为无效数据，进行剔除处理；

步骤二：如第二次出行的上车站点已经匹配成功，则通过上车站点中一并得到的运行线路编号（LINENUM）和站点编号（STATIONNUM）信息进行分情况分析。

情况一：第三次的上车站点在第二次出行中运行线路的下游，则判定乘客是在同站进行换乘，即第三次出行的上车站点为第二次出行的下车站点；

情况二：第三次的上车站点不在第二次出行中运行线路的下游，则将第三次上车的编号和联锁转乘站点数据表进行字段匹配，得到最有可能的转乘车站Szc。转乘车站和第三次出行的上车站点之间符合直线距离小于300米。如Szc不在第二次上车站点匹配的运行线路中则视为无效匹配，归为出行断裂的乘客；如Szc在第二次上车站点匹配的运行线路中则将Szc判定为第二次出行的下车站点。

3.2.4 一次出行的下车站点推算

特别的，单日出行次数只有一次的乘客无法通过其他的出行记录来判定其下车站点，但是单日出行次数为一次的乘客占据了整体数据的37%，因此依旧是研究重点。

针对此类人群，主流的推算方法有两种，一是扩大数据考虑时间，二是通过下游站点的现有OD进行分配。

扩大数据考虑时间指的是将下车站点推算的时间跨度从单日提升到一周。鉴于乘客的出行需求的多样性，此类乘客的出行规律可能要在多日内体现即乘客的出行可能在一周的时间跨度中呈现往返的规律性。例如每两天才能完成一次往返且两天中都仅有一次出行记录。再参照往返出行的下车站点推算的情况便可以实现对下车站点的推算。因此，通过扩大数据考虑时间可以实现部分仅一次出行乘客的下车站点推算。

通过下游站点的现有OD进行匹配站点是在无法通过长期数据推算下车站点时。借助其他日期中已经生成的OD数据作为参考，以历史下车人数来确定乘客在此站的下车概率，其中获取了最多下站人数的车站即为最有可能的下车站点。最终将乘客的下车站点按照概率进行分配。

3.3 本章小节

本章以上车站点推算和下车站点推算两个部分为主要内容，首先介绍了上车站点的推算的基本原理，并展示了其具体的推算过程。之后介绍了下车站点的推算方法，提出了出行链闭合的假设作为推算的前提并展示了成都市主体人群的出行特征。其后将人群分为三类后，分别展示了其对应的下车站点推算方法及流程图。

**第4章 公交出行OD推算平台**

在推算得到公交车出行OD矩阵的基础上，本文将会以Visual Studio软件中的C#语言为基础，连接多款软件搭建一个集交通数据导入、已有数据的查询、OD推算等功能于一体的公交出行OD推算平台。特别的，该平台将连接ArcGIS软件，实现OD数据的可视化。该系统能够高效地存储和处理交通数据，具有良好的实用性和可扩展性[34]。本章将介绍平台的软件基础、平台结构、平台主要功能、平台的特点优势以及平台搭建的意义。

4.1 推算平台软件基础

公交出行OD推算平台需要实现OD的导入、数据的查询、OD的推算、OD的查询、数据的可视化等功能。因此在基于Visual Studio平台中C#语言搭建推算平台窗口的基础上，需要连接多款外部软件以实现对应的功能。首先，数据的处理和OD的推算需要数据库软件如Oracle和SQL Server。其次，实现OD数据的可视化需要地图软件ArcGIS。

4.1.1 数据库软件Oracle、SQL Server

数据库是处理各类大数据的重要平台和工具，利用数据库可以对数据进行存储、管理、查询、计算等功能，是本文对基础数据预处理的重要软件基础，同时也是最终公交出行OD推算平台搭建的重要一环。常见的数据库软件有Oracle、SQL Server、MySQL等。本文中使用的数据库软件为Oracle和SQL Server。

其中Oracle是一种商业关系型数据库管理系统，可以使用多种语言进行数据处理，其主要的查询和编程语言是SQL，适用于大型企业级应用程序。Oracle数据的特点为

1. 可拓展性：Oracle数据库可以自主拓展空间以满足大容量的数据存储和处理；
2. 高可用性：Oracle数据库具有高可用性和容错性，会进行备份和恢复，可以大大提高数据处理的可行性；
3. 可移植性：Oracle数据库支持在多平台和操作系统上使用，并且兼容多种编程语言和应用程序。

在本文中Oracle数据库将主要负责基础数据的导入和清洗工作。本文原始数据中IC卡和公交车载GPS数据均为DMP文件，需要调用CMD命令台导入Oracle，并在其中进行一些简单处理，最终生成ICDATA、GPS\_BUS和BRTDATA这三张表。

SQL Server是本文中主要利用的数据库平台，是目前主流的数据库之一，其被广泛用于大型企业、政府机构和组织，以存储、管理和处理大量结构化的数据。对比于其他的数据库软件，其具有如下特点：

1. 安全性：SQL Server提供了强大的安全功能，包括基于角色的访问控制、加密、审核等。此外，它还提供了细粒度的安全策略，使管理员能够对不同用户或用户组设置不同的权限；
2. 高可用性：SQL Server可以提供高可用性解决方案，包括数据库镜像、AlwaysOn可用性组、故障转移群集等。这些功能保证了系统的连续性和可靠性，即使在硬件或软件故障的情况下也能保持数据的可访问性和完整性；
3. 强大的编程接口：SQL Server支持多种编程语言，包括T-SQL、Java、C++等，并提供了多种API和工具来与其他应用程序进行集成；
4. 可定制性：SQL Server提供了可定制的存储过程、触发器、UDF以及CLR函数，使开发人员可以为自己的应用程序创建定制的存储过程和函数；
5. 易管理性：SQL Server提供了易于使用的管理工具，帮助管理员监控和管理数据库实例、备份和恢复数据、维护数据库对象、优化查询等；

本文中，SQL Server数据库会将在Oracle中已经简单处理的原始数据导入，并且通过运行预先编写的存储过程生成对应日期的IC表和GPS表。此后SQL Server将完成匹配公交上车站点和推算公交下车站点的功能，并且完成各类数据的统计存储，如不同时段的公交登降量、交通大区的出入量等等。

4.1.2 平台搭建软件Visual Studio

推算平台主要的功能为数据的导入、查询以及最终的OD推算展示提供便利。通过与数据库软件和地图软件的连接，推算平台可以实现上述功能的一体化，集成化，结果的可视化。为公交路网的优化调整和交通的运营调度提供可靠的数据支撑，有效缓解城市内的交通拥堵问题，实现公共交通资源的合理利用，保障了城市内交通的协调、高效。

对于推算平台的搭建，本文使用的是Visual Studio中的C#。C#在Visual Studio中是一种强大的面向对象的编程语言，能够帮助开发人员创建各种类型的应用程序，并提供丰富的可视化工具和集成开发环境。对比于其他软件，C#具有如下优势。

1. 可靠性：C#是一种类型安全的语言，它能够检测到编程中的类型错误并防止其在运行时发生。这使得C#应用程序更加可靠，减少了由于类型错误导致的崩溃和故障；
2. 性能高：与其他语言相比，C#具有更好的性能和效率。它可以通过使用.NET Framework进行优化，使开发人员能够高效地开发应用程序；
3. 跨平台：C#不仅可以用于Windows平台，还可以用于其他平台，例如Linux和Mac OS X。这意味着开发人员可以将同一代码库用于不同的操作系统，并且不需要为每个平台创建单独的代码库；
4. 安全性：C#提供了强大的安全功能，包括内存管理、类型安全、异常处理和代码访问权限等。这些功能可以帮助开发人员编写更安全的应用程序，并保护用户的数据；
5. 可维护性高：C#是一种易于维护的语言，因为它具有清晰的语法和结构。此外，使用Visual Studio等开发工具可以使代码更易于阅读和理解；

本文中，Oracle，SQL Server，ArcGIS三款软件都将和VS进行连接，并且调用命令台和两款数据库软件中的存储过程实现对数据的导入，清洗等工作。并为操作人员提供查询不同条件下的IC卡和GPS数据。最终与地图软件连接实现OD的可视化展示。

4.1.3 地图软件ArcGIS

GIS是地理信息系统的缩写，它是一种将空间数据和属性数据结合起来进行存储、管理、分析和展示的技术系统。GIS主要由硬件、软件、数据和人员等组成，其目的是为了更好地理解和利用地球表面上的空间数据。GIS能够对地球表面上的各种空间信息进行处理和分析，并且可以通过图形化的方式来展现这些信息。例如，GIS可以用于制作地图，分析社会经济数据，分析气象数据，进行环境保护监测等。GIS还可以用于城市规划、土地利用规划、灾害防治等领域，帮助政府、企业和个人做出更加明智的决策。

推算平台最终的OD展示功能，本文采用的是ArcGIS。ArcGIS是一款由Esri公司开发的地理信息系统（GIS）软件套件，它为用户提供了丰富的地理数据管理、地图制作、空间分析和可视化等功能。ArcGIS可以在桌面、服务器、云和移动设备上运行，并且支持多种平台和操作系统。对比于其他地图软件ArcGIS具有以下特点：

1. 数据管理：ArcGIS可以处理各种类型的地理数据，包括矢量数据、栅格数据、地形数据和卫星影像等。
2. 地图制作：ArcGIS提供了强大的地图制作工具，用户可以创建漂亮、专业的地图，添加各种注记和符号等。
3. 空间分析：ArcGIS支持各种空间分析功能，包括缓冲区分析、网络分析、3D分析和地形分析等。
4. 可视化：ArcGIS可以将地理数据可视化为各种图层、标注和符号，帮助用户更好地理解和展示地理信息。
5. 扩展性：ArcGIS支持多种编程语言和开发平台，用户可以通过自定义开发和集成扩展其功能。

本文中ArcGIS将通过ArcEngine作为中间接口与SQL Server数据库连接，并基于推算完成的OD矩阵及上下车站点数据，在成都市的地图上呈现公交出行的OD图和公交线路的OD图，使得OD数据可视化，并且根据需要，自由切换交通大区和小区。

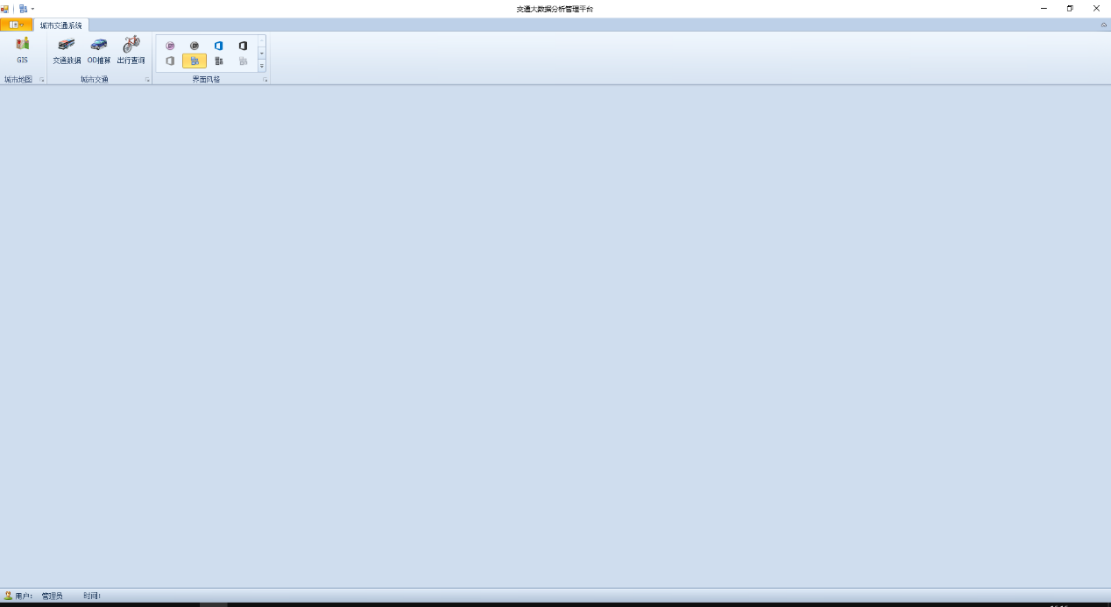
4.2 推算平台结构

公交推算平台主要实现数据的导入查询、OD推算、OD查询和可视化展示四个主要功能，因此本文搭建的公交出行OD推算平台在主界面以外主要包含了交通数据界面、OD推算界面、OD查询界面和地图展示这四个窗口，以及若干的子窗口。

4.2.1 平台主界面

公交出行OD推算平台的主界面主要包含交通数据界面、OD推算界面、OD查询界面和地图展示界面的互动控件，以及一个界面风格切换栏。主界面主要用于打开另外四个主要的窗口以及为用户提供不同的界面风格。

主界面的内容和细节如下图展示：



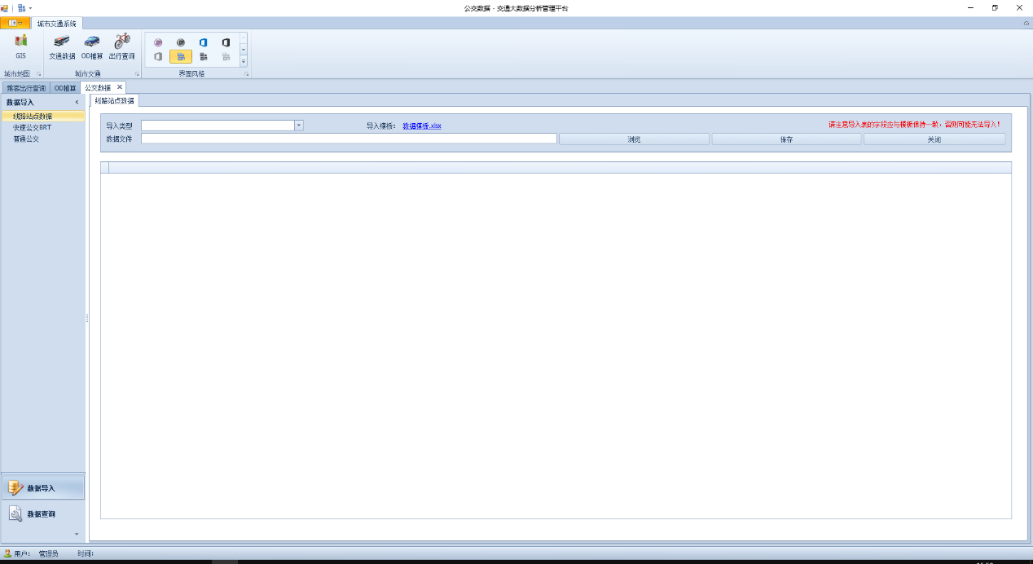
**图 4-1 推算平台主界面**

4.2.2 交通数据界面

交通数据界面为用户提供了导入基础数据如IC卡数据、车载GPS数据、线路站点数据和查询已经被导入数据的功能。交通数据界面提供两个主要的互动控件：数据导入和数据查询，

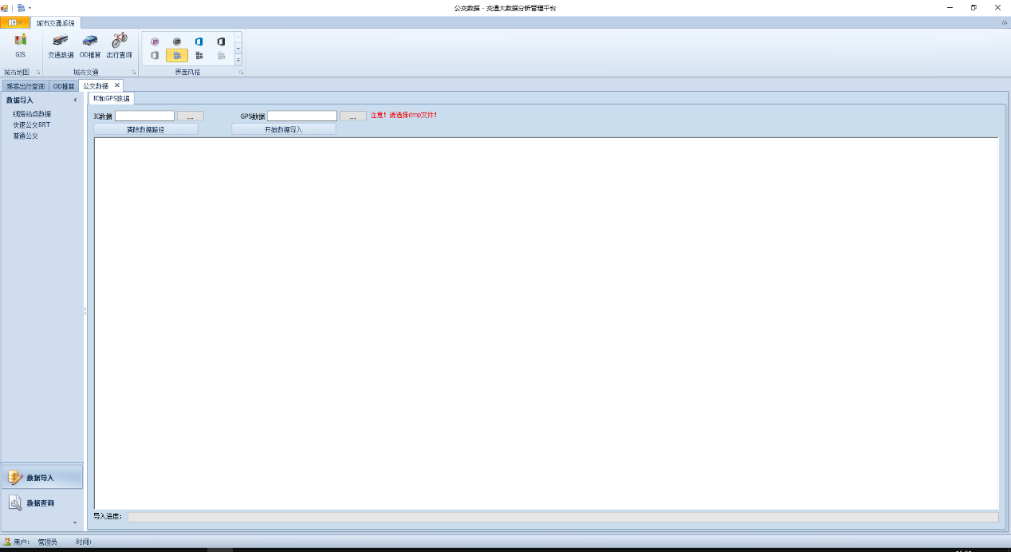
数据导入的子界面对应了原始数据导入和清洗的功能，其中主要包含三个部分。分别是线路站点数据、快速公交BRT和普通公交。上述三个子界面的结构相似，用于导入不同类型的原始数据。其导入流程为操作人员选择指定的文件目录后与导入按钮互动，平台将与数据库软件连接自主实现对原始数据的导入和数据的清洗。同时界面中的控件会以文字和进度条这两种方式显示目前的导入和清洗进度。

线路站点数据导入界面的内容和细节如下图展示：



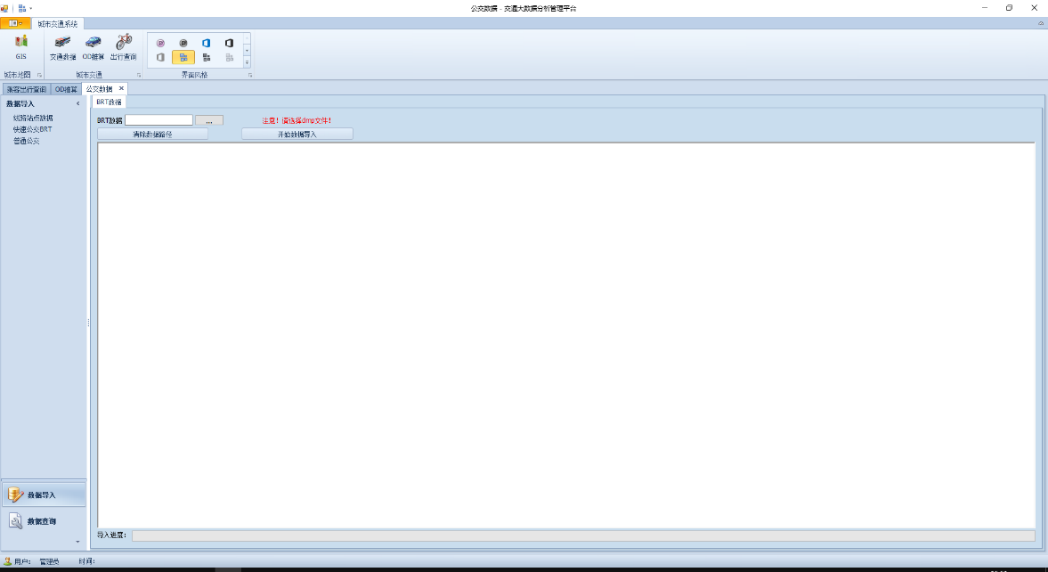
**图 4-2 线路站点导入界面**

其中，普通公交数据导入界面的内容和细节如下图展示：



**图 4-3 普通公交IC卡和GPS数据导入界面**

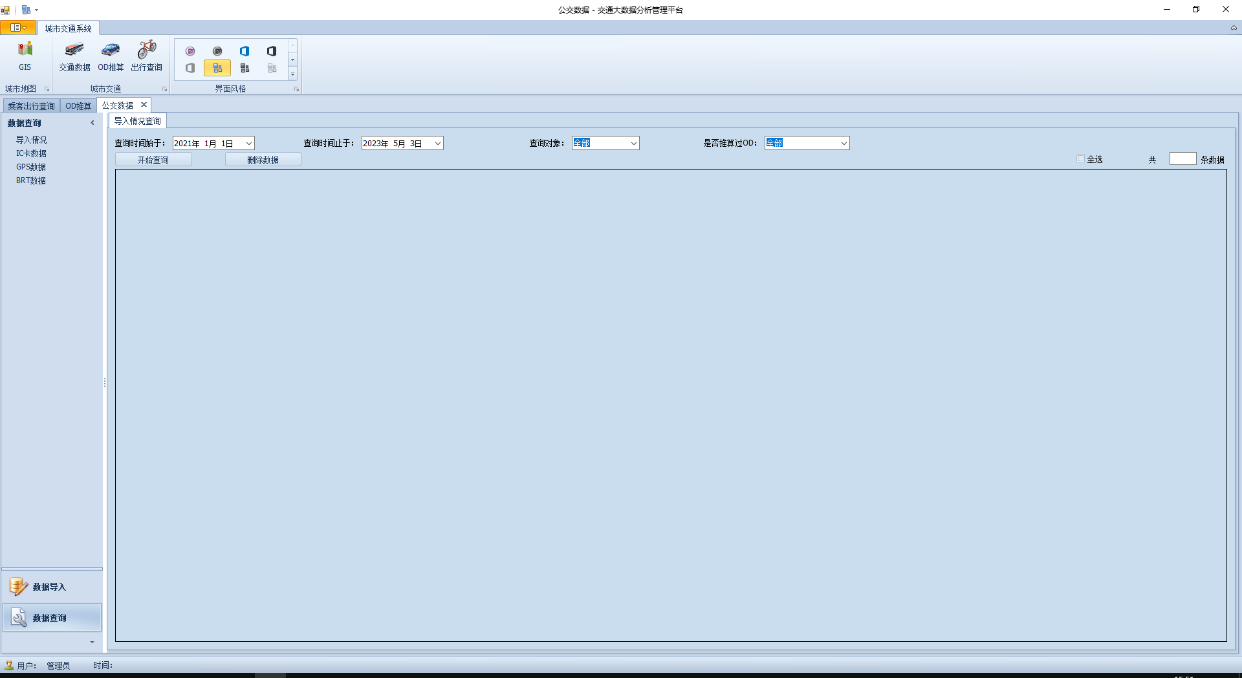
快速公交BRT数据导入界面的内容和细节如下图展示：



**图 4-4 线路站点导入界面**

数据查询的子界面对应了数据查询的功能。其中包含了四个部分。分别是导入情况查询、IC卡数据查询、GPS数据查询和BRT数据查询。其中导入情况查询窗口可以对目前已经导入的数据进行整体性的查询，其主要呈现的是目前已经导入的数据时间，以及对应日期的IC卡、GPS数据和BRT数据的导入量，以及此数据是否进行OD的推算，未导入或者未推算OD的数据将显示为null。同时，此界面为用户提供了查询的日期范围、是否推算过OD以及查询对象的查询条件进行选择。

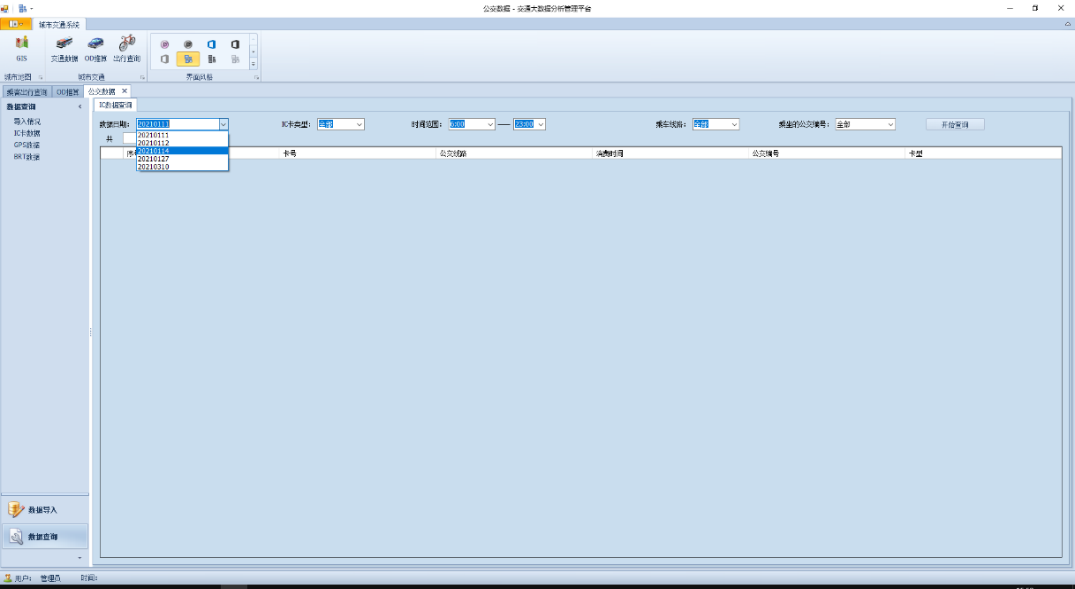
导入情况查询的界面内容和细节如下图展示：



**图 4-5导入情况查询界面**

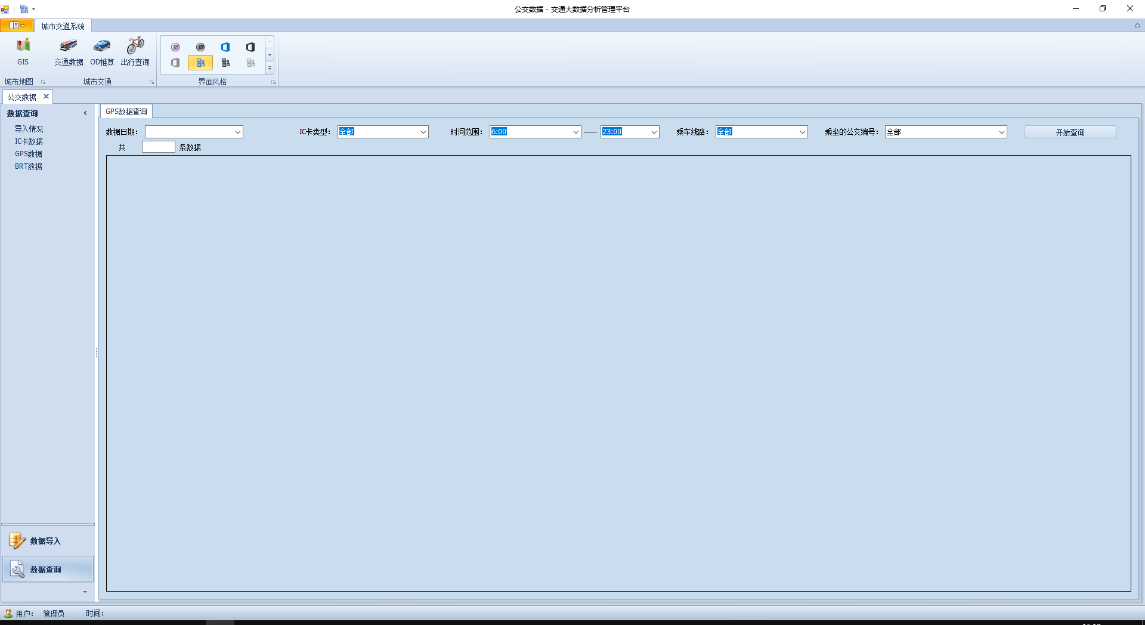
IC卡数据查询、GPS数据查询和BRT数据查询界面的结构相同，其查询结果为对应数据的具体内容，例如查询IC卡数据将会得到刷卡的时间、IC卡的卡号、公交编号、公交线路编号、消费类型等数据。GPS数据查询的主要结果为公交到离站、公交运行线路编号和公交运行时间。同时，IC卡数据查询界面为用户提供刷卡时间范围、乘车线路、公交车牌号和IC卡类型这四种参数进行选择。其中IC卡类型包括普通卡、军人卡、老年卡和爱心卡。特别的，乘车线路和公交编号支持用户自主输入和通过下拉框选择。GPS数据查询界面为用户提供了公交运行时间段、线路编号、公交车牌号和线路站点这四种参数进行选择。BRT数据查询界面为用户提供了公交站点的参数选择。上述的所有参数选择在默认状态下均为全部。

IC卡数据具体情况查询的界面内容和细节如下图展示：



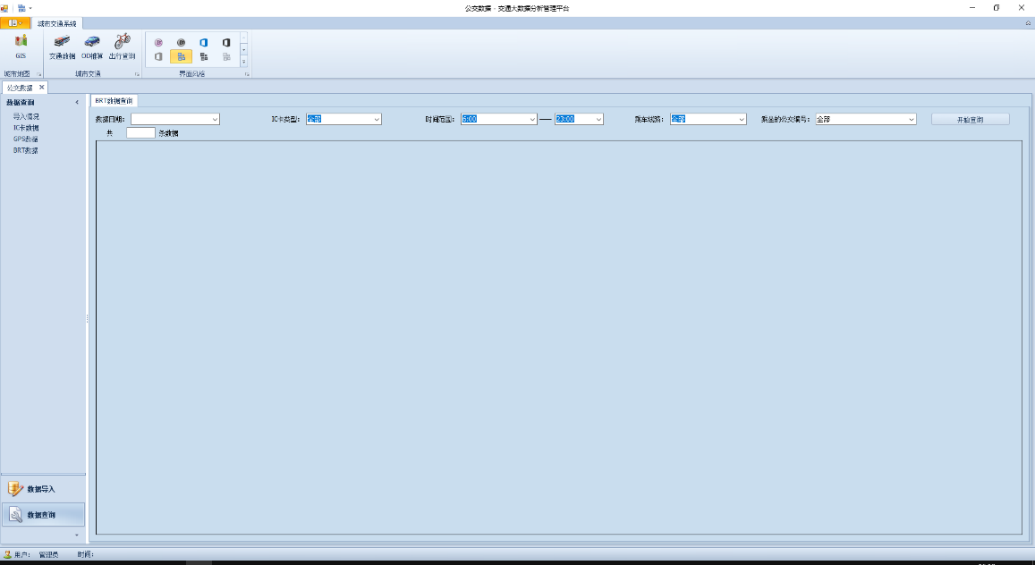
**图 4-6 IC卡数据查询界面**

GPS数据具体情况查询的界面内容和细节如下图展示：



**图 4-7 GPS数据查询界面**

BRT数据具体情况查询的界面内容和细节如下图展示：



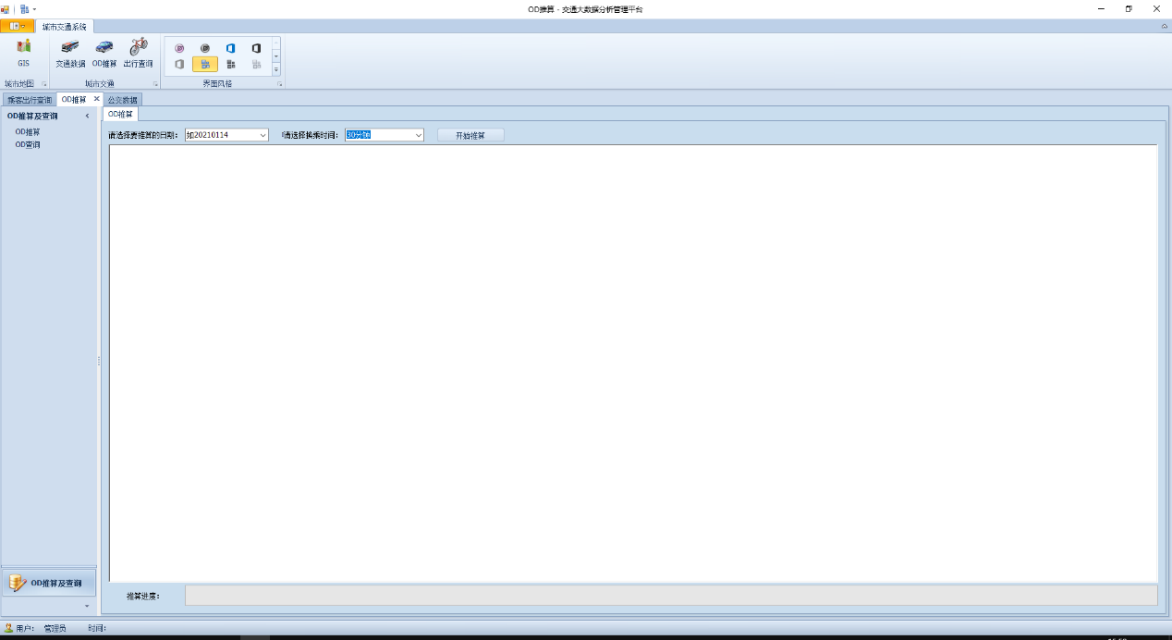
**图 4-8 BRT数据查询界面**

4.2.3 OD推算及查询界面

OD推算及查询界面为用户提供了推算公交出行OD和查询OD矩阵的功能，其中包含了两个子窗口，分别是OD推算和OD查询。其中OD推算子窗口用于对未进行过上下车站点推算的日期进行OD推算，而OD查询子窗口则可以对数据库中已经推算过OD的日期进行查询，得到交通大区的公交出行OD矩阵

OD推算子窗口为用户提供推算日期和换乘条件这两个参数进行选择，确认选择后，后台将自动连接SQL Server数据库并调用存储过程自主实现对目标日期的OD推算，并且更新其他相关联的表。特别的，在推算完成之后将会形成两张数据表分别是交通大区之间出行的数据矩阵和同一交通大区内部的出行量，并且这两张表将会被导出成CSV文件，被保存在ArcGIS的文件路径下。

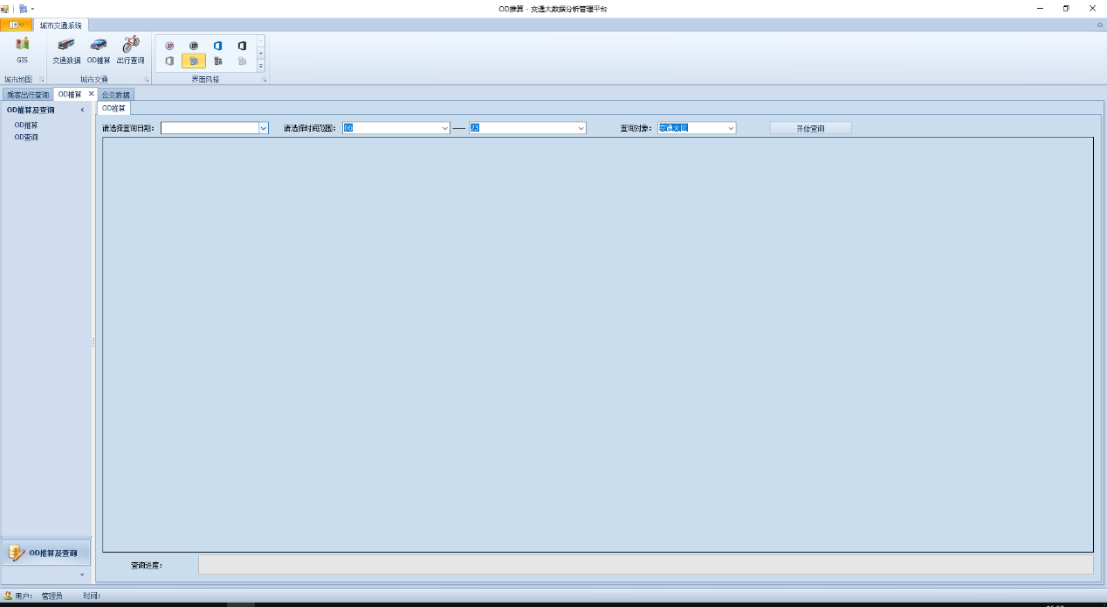
OD推算子窗口的界面内容和细节如下图展示：



**图 4-9 OD推算界面**

鉴于本文仅仅研究公交的出行，因此OD查询子窗口仅为用户提供交通小区和交通大区作为查询对象。因为原始数据中并不包含新津县地区的IC卡和GPS数据，因此在以交通大区为查询对象时将不包含新津县。OD查询子窗口的查询结果为具体的OD矩阵，当查询对象为交通大区时将展示对应大区的名字如晋江区、武侯区。当查询对象为交通小区时，将展示交通小区对应的编号。

OD查询子窗口的具体内容和细节如下图展示：



**图 4-10 OD矩阵查询界面**

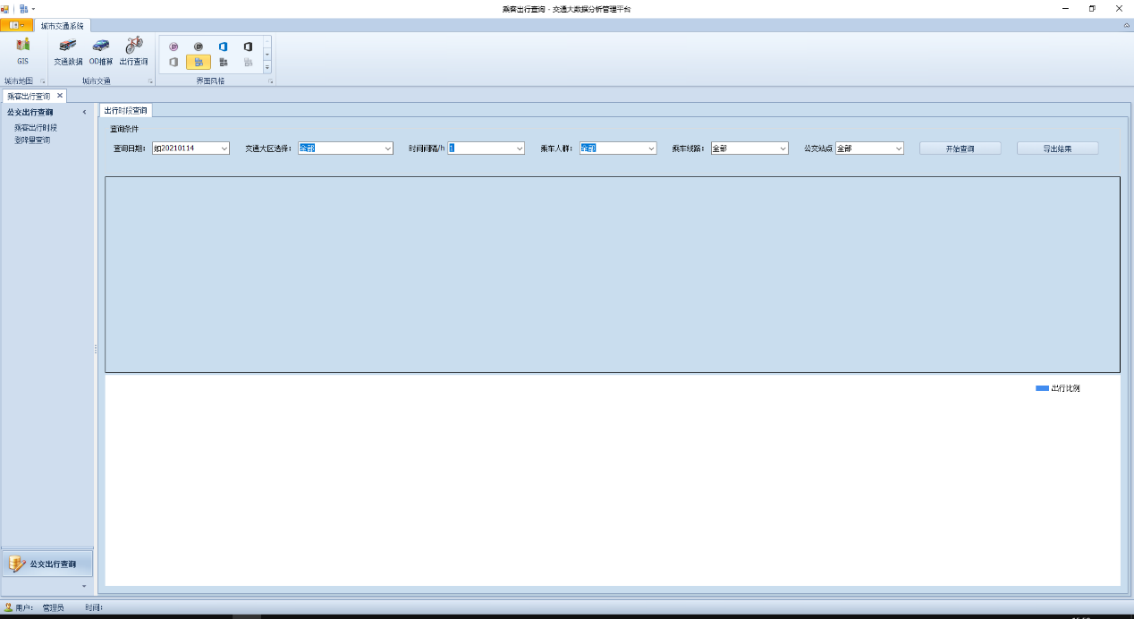
4.2.4 出行查询界面

出行查询界面为用户提供了查询基于OD矩阵和乘客下车站点数据派生得到的其他公交出行数据如乘客出行时段的统计数据和公交登降量数据。出行查询界面包含了两个子窗口，分别是出行时段查询窗口和公交登降量查询窗口。

出行时段查询窗口用于查询不同时间段中乘客出行量的统计数据，并且为用户提供了交通大区、时间间隔、乘客类型、乘车线路、公交站点这五个参数进行选择。其中，交通大区、乘客类型、乘车线路和公交站点的默认值为全部，而时间间隔的默认值为一个小时。乘客类型和IC卡数据查询界面中的相同，交通大区中同样不包含新津县，除默认选项全部外，其余选项分别是锦江区、武侯区、青羊区、金牛区、成华区、高新南区、高新西区、双流区、温江区、郫都区、新都区、龙泉驿区、青白江区、天府新区和东部新区。

其查询结果以表格结构和折线图两种形式进行展示，折线图中会对每一个节点的值进行标注，使查询结果更加直观化。

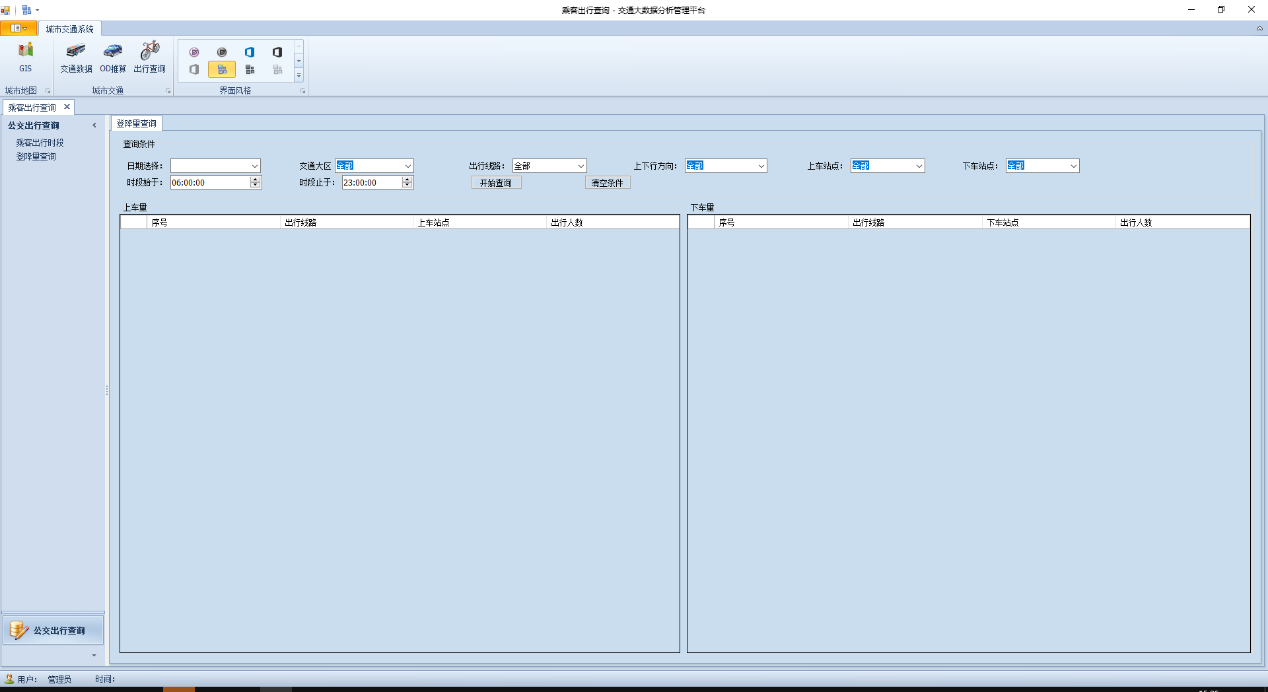
出行时段查询窗口的具体内容和细节展示如下：



**图 4-11 出行时段查询界面**

公交登降量查询窗口用于查询公交站点中乘客上车量和下车量，同时会展示此站点对应的公交线路的编号。窗口为用户提供了交通大区、出行线路、上下行方向、上车站点、下车站点、出行时段的起止时间这6个参数进行选择。其中交通大区、出行线路、上下行方向、上车站点、下车站点的默认值均为全部，出行时段的默认范围是早上6点至夜里23点。鉴于此窗口中可以选择的参数较多，界面涉及中添加了清空条件按钮以便一次性将上述6个参数恢复至默认值。查询的结果将会以上车量和下车量分开展示。

公交登降量查询窗口的具体内容和细节如下图展示：

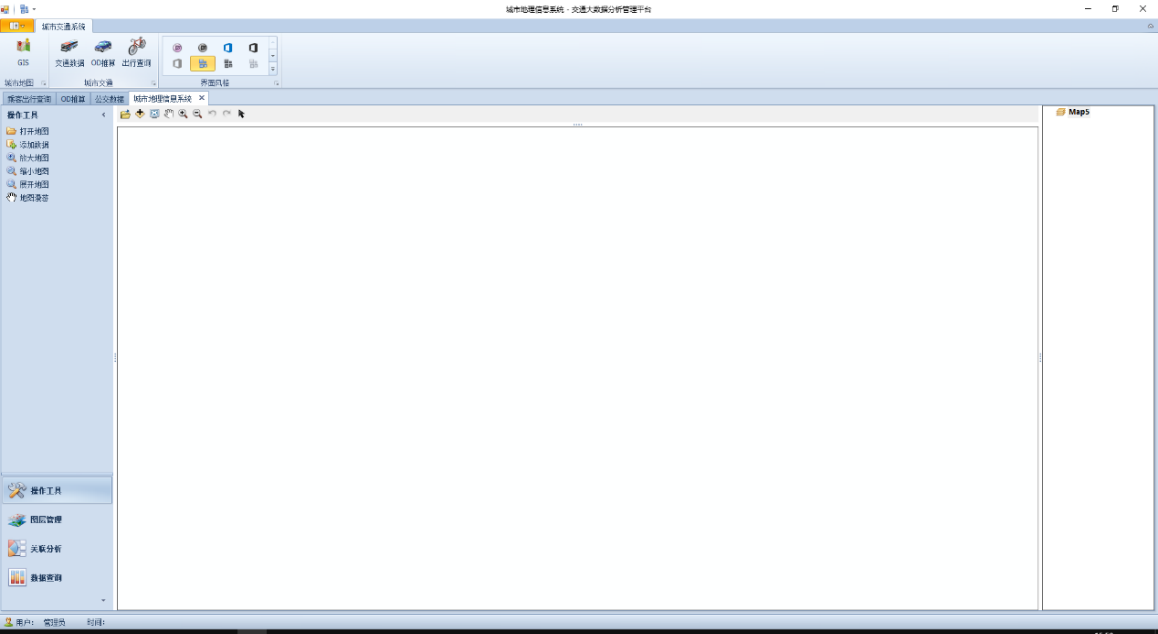


**图 4-12 线路登降量查询界面**

4.2.5 地图展示界面

地图展示界面将基于已经推算得到的OD矩阵连接ArcGIS展示，使用户可以更直观地视察成都市的公交出行状态。此界面包含一个主要窗口用于展示ArcGIS中的地图文件中的地图，以及一个窗口用于展示对应地图中的相关图层。同时窗口中包含了若干控件为用户提供多种方式查看地图，包括地图的缩放、定位、铺满等。特别的，地图展示界面存在一个图层管理功能，用于实现对目标图层的数据符号化管理，其中主要的两个功能为分级色彩符号化和分类符号化。

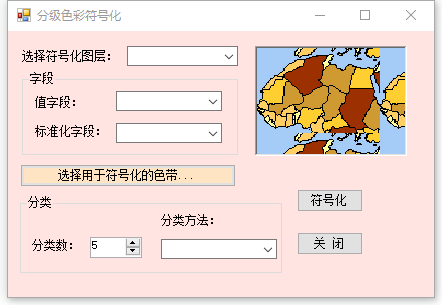
地图展示界面的主体内容和细节如下图所示：



**图 4-13 GIS界面**

分级色彩符号化的功能用于实现大区之间的公交出行的可视化展示，用户通过选择指定日期OD图层中的出行量字段进行符号化，其展示的主要形式为两个交通大区中心点的连线，OD出行量越大的两地之间连线的色彩会更深。其分级方式包括等间隔分级、分位数分级、自然裂点分级和几何间隔分级一共四种。同时此功能让用户自主选择分级数并且提供多种符号化色带进行选择。

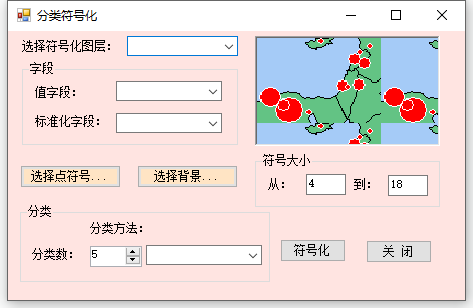
分级色彩符号化的界面内容和细节如下图所示：



**图 4-14 分级色彩符号化界面**

分类符号化的功能用于实现交通大区内部的公交出行的可视化展示，用户通过选择指定日期的内部出行OD图层中的出行量字段进行符号化，其展示的主要形式为在交通大区的中心点形成一个图案，内部出行量越大的交通大区的图案标识越大。其分级方式同样为等间隔分类、分位数分类、自然裂点分类和几何间隔分类。同时此功能让用户自主选择分级并且提供了多种类型的点符号和符号背景进行选择。

分类符号化的界面内容和细节如下图展示：



**图 4-15 分类符号化界面**

4.3 推算平台主要功能

推算平台搭建的目的是实现OD推算的一体化、集成化以及结果展示的可视化。其主要功能包含数据导入、数据查询、推算OD、OD查询和GIS展示。这五项功能的具体内容展示如下：

1. 数据导入

推算平台将通过C#语言完成基于Oracle和SQL Server数据库的基础数据导入和数据预处理过程，并将数据存储于SQL Server的新建表之中。数据导入包括普通公交的IC卡和公交车载GPS数据、快速公交BRT的数据以及线路站点数据。

2. 数据查询

对于导入的数据，推算平台通过C#编写SQL语言，并通过SQL Server查询导入的数据情况，其中包括已导入的数据日期、是否完成所有数据的导入、已经导入数据如IC卡数据的数据量、是否完成了OD的推算。所有未完成的部分，在数据查询中将显示为空。

3. 推算OD

如已完成了某日全部的数据导入，则可以进行OD推算。推算平台将利用C#语言根据用户选择的换乘条件，调用SQL Server不同的存储过程并进行上下车站点的推算。

4. OD查询

对于已经完成OD推算的日期，可以进行OD推算。通过选择交通小区或者交通大区调用SQL Server中不同的存储过程得到每个区块的出行量和吸入量，并且可以人工设定查询的时间段，最终以表格的形式进行展示。同时，此板块还可以查询不同线路的登降量。

5. GIS展示

对于已经完成OD推算的日期，可以通过ArcGIS软件实现基于成都市地图的OD展示。其主要展示形式为蛛网图。主要的展示内容为交通大区之间的出行量和吸入量。特别的，可以根据上下车站点数据得到成都市内单日的线路流量界面图。

4.4 推算平台特点

本文设计的公交出行OD的推算平台实现了公交出行OD推算的集成化、一体化，提供了数据导入、清洗、推算、查询和可视化等多种功能，连接了多款外部软件 。公交出行OD推算平台的主要特点如下：

1. 数据处理量大

成都市单日的IC卡数据可以达到240万条，公交车载GPS数据一般在3500万至3700万之间，单日的总体数据可以达到15G。数据的导入、清洗、预处理都需要大量的时间成本和运行存储，同时对计算机的算力也有一定要求。

1. 可查询性强

推算平台在数据导入情况、OD矩阵和出行统计的查询界面都提供了大量可供选择的参数。用户可以从多个纬度对数据进行查询，筛选得到目标数据，提高了数据的有效性和平台的可查询性。

1. 软件兼容性强

推算平台的基础软件Visual Studio和数据库软件SQL Server，Oracle以及地图软件ArcGIS都使用了目前主流的版本且均具备一定的兼容性。外部软件的改版不会影响整个平台的正常运行。

1. 数据展示直观

推算平台中的数据展示形式除了表格，还存在折线图和ArcGIS的地图展示，使用户更直观、更快速地了解数据，获取数据，同时避免了数据展示的单调性。GIS中可视化的展示使城市交通出行需求结果得到具象化的展示，为城市交通管理与规划提供了研究思路，实现对社会交通资源的合理利用。

1. 集成性强

推算平台将Oracle和SQL Server中的存储过程都集成为了C#语言，转入后台调用，便捷了用户的操作，整个平台的操作流程更加流程明了。

4.5 平台搭建意义

公交大数据的数据量大、字段复杂、整体价值密度低、处理繁琐。表格化的数据难以实现对数据的高效利用。通过平台自主连接数据库软件和地图软件完成数据的导入、清洗、处理、查询和可视化，首先能够实现城市内公共交通数据的综合利用分析与深入挖掘，获得了实时、充分、准确的公共出行客流的运行信息，实现公交上车站点和下车站点的双重匹配，提高了数据结果了可靠性与准确性。第二，城市内多种交通方式之间的配合也将得到进一步优化，城市路网的性能进一步提高，为提升城市交通服务质量提供了多方位、多层次、多渠道的帮助。第三，使城市交通出行需求结果得到具象化的展示[35]，为城市交通管理与规划提供了研究思路，实现对社会交通资源的合理利用。

4.6 本章小节

本章首先介绍了公交出行OD推算平台的软件基础。其次详细介绍了平台的结构，以多款外部软件为基础的推算平台将数据导入、清洗、OD推算、数据查询和可视化展示等功能集于一体，并且在平台的各子窗口中实现了对应的功能，并且展示了部分重要窗口的界面内容和细节。其后介绍了平台的主要功能和平台的特点优势。最后阐述了搭建公交出行OD推算平台的意义。

**第5章 实验结果**

前三章中，本文已经介绍了基于IC卡和公交车载GPS数据针对不同出行特征人群的上车站点和下车站点的推算，即得到了出行的起讫点。在获取了上下车的站点数据之后，即可完成OD矩阵的推算。本章将展示上下车站点的推算结果已经出行OD的推算结果，以及一并统计得到的交通大区的登降量、主要客流时段分布、刷卡的时空数据等。

5.1 乘客上车站点推算结果

乘客的上车站点推算结果中的信息包括但不限于IC卡号、刷卡时间，公交编号，线路编号，上下车站点即其具体经纬度，下车时间。其中下车时间为公交达到对应站点的进站时间。这些信息中重要的信息为线路编号、上车站点、刷卡时间、下车站点和下车时间。其具体的示例如下表展示。

**表 5-1 乘客上下车推算结果（部分）**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 线路编号 | 山车站点 | 上车时间 | 下车站点 | 下车时间 |
| 3 | 20442 | 2021-01-27 16:12:52 | 10221 | 2021-01-27 16:30:44 |
| 4 | 10221 | 2021-01-27 16:45:10 | 20148 | 2021-01-27 16:46:53 |
| 4 | 20148 | 2021-01-27 16:55:47 | 30116 | 2021-01-27 17:05:48 |
| 86 | 20460 | 2021-01-27 17:11:12 | 30208 | 2021-01-27 17:21:15 |

其中，成都市单日的IC卡刷卡记录约为250万条，其中上车站点能成功匹配约240万条，平均匹配率为97%；往返类出行的下车站点能成功匹配约120万条，平均匹配率为50%，经过扩样后整体的下车站点匹配率可以达到86%。例如2021年1月27日的单日IC刷卡记录为230万余条，剔除夜间23点至凌晨6点的IC乘客后共有2480647条数据，上车站点成功匹配了2441085条，其匹配率为98.40%；下车站点成功匹配了1889544条，其匹配率为76.17%.

5.2 出行OD统计结果

基于推算得到的公交乘客上下车站点编号、刷卡时间段以及公交运行线路等信息，可以统计得到交通大区之间的公交出行量。交通大区公交出行的OD是指从成都市的某交通大区通过公交达到另一个交通大区的出行量。通过上下车站点信息中带有的站点位置信息，可以得到站点所处的交通大区编号，通过数据统计便可以得到交通大区之间的公交出行OD矩阵。鉴于原始数据的缺失，本文的IC卡数据并不包含15号交通大区新津县区，因此在交通大区间OD的统计矩阵中将不包含新津县的数据。

例如2021年1月14日成都市交通大区之间的公交出行矩阵如下表所示：

**表 5-2 2021年1月14日交通大区公交出行OD**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 编号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 16 |
| 1 | 105033 | 13881 | 11889 | 7705 | 23848 | 12448 | 0 | 159 | 0 | 0 | 0 | 8475 | 0 | 556 | 0 |
| 2 | 15048 | 85155 | 21315 | 7015 | 4888 | 16269 | 0 | 2748 | 18 | 0 | 0 | 106 | 0 | 145 | 0 |
| 3 | 14811 | 20495 | 86867 | 32582 | 10390 | 5329 | 254 | 21 | 3654 | 154 | 23 | 76 | 0 | 0 | 0 |
| 4 | 7074 | 8105 | 34660 | 104196 | 21712 | 2294 | 1738 | 15 | 487 | 6174 | 5713 | 268 | 0 | 0 | 0 |
| 5 | 22003 | 4662 | 13242 | 23419 | 136313 | 1709 | 0 | 0 | 0 | 31 | 3574 | 4719 | 112 | 0 | 0 |
| 6 | 13090 | 18546 | 5327 | 1839 | 1799 | 80878 | 0 | 3699 | 0 | 0 | 0 | 541 | 0 | 6367 | 0 |
| 7 | 0 | 0 | 377 | 2784 | 0 | 0 | 5671 | 0 | 654 | 4848 | 0 | 0 | 0 | 0 | 0 |
| 8 | 185 | 3127 | 34 | 25 | 0 | 4411 | 0 | 3467 | 168 | 0 | 0 | 0 | 0 | 1178 | 0 |
| 9 | 0 | 150 | 4852 | 612 | 0 | 0 | 570 | 160 | 7229 | 279 | 0 | 0 | 0 | 0 | 0 |
| 10 | 1 | 0 | 153 | 7692 | 37 | 0 | 4325 | 0 | 201 | 28970 | 85 | 0 | 0 | 0 | 0 |
| 11 | 0 | 0 | 193 | 7548 | 5035 | 0 | 0 | 0 | 0 | 126 | 4490 | 24 | 1341 | 0 | 0 |
| 12 | 7906 | 276 | 62 | 288 | 5641 | 661 | 0 | 0 | 0 | 0 | 21 | 17691 | 11 | 349 | 0 |
| 13 | 0 | 0 | 0 | 0 | 281 | 0 | 0 | 0 | 0 | 0 | 950 | 9 | 4558 | 0 | 0 |
| 14 | 555 | 267 | 0 | 0 | 10 | 6200 | 0 | 793 | 0 | 0 | 0 | 315 | 0 | 27298 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2558 |

其中黄色的单元格中标红的数字为交通大区内部的出行量，及该乘客的上车站点和下车站点均位于同一交通大区内。

上表中的交通大区编号对应的具体交通大区名称如下表所示：

**表 5-3 交通大区编号及名称**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 大区编号 | 大区名称 | 大区编号 | 大区名称 | 大区编号 | 大区名称 |
| 1 | 锦江区 | 6 | 高新南区 | 11 | 新都区 |
| 2 | 武侯区 | 7 | 高新西区 | 12 | 龙泉驿区 |
| 3 | 青羊区 | 8 | 双流区 | 13 | 青白江区 |
| 4 | 金牛区 | 9 | 温江区 | 14 | 天府新区 |
| 5 | 成华区 | 10 | 郫都区 | 16 | 东部新区 |

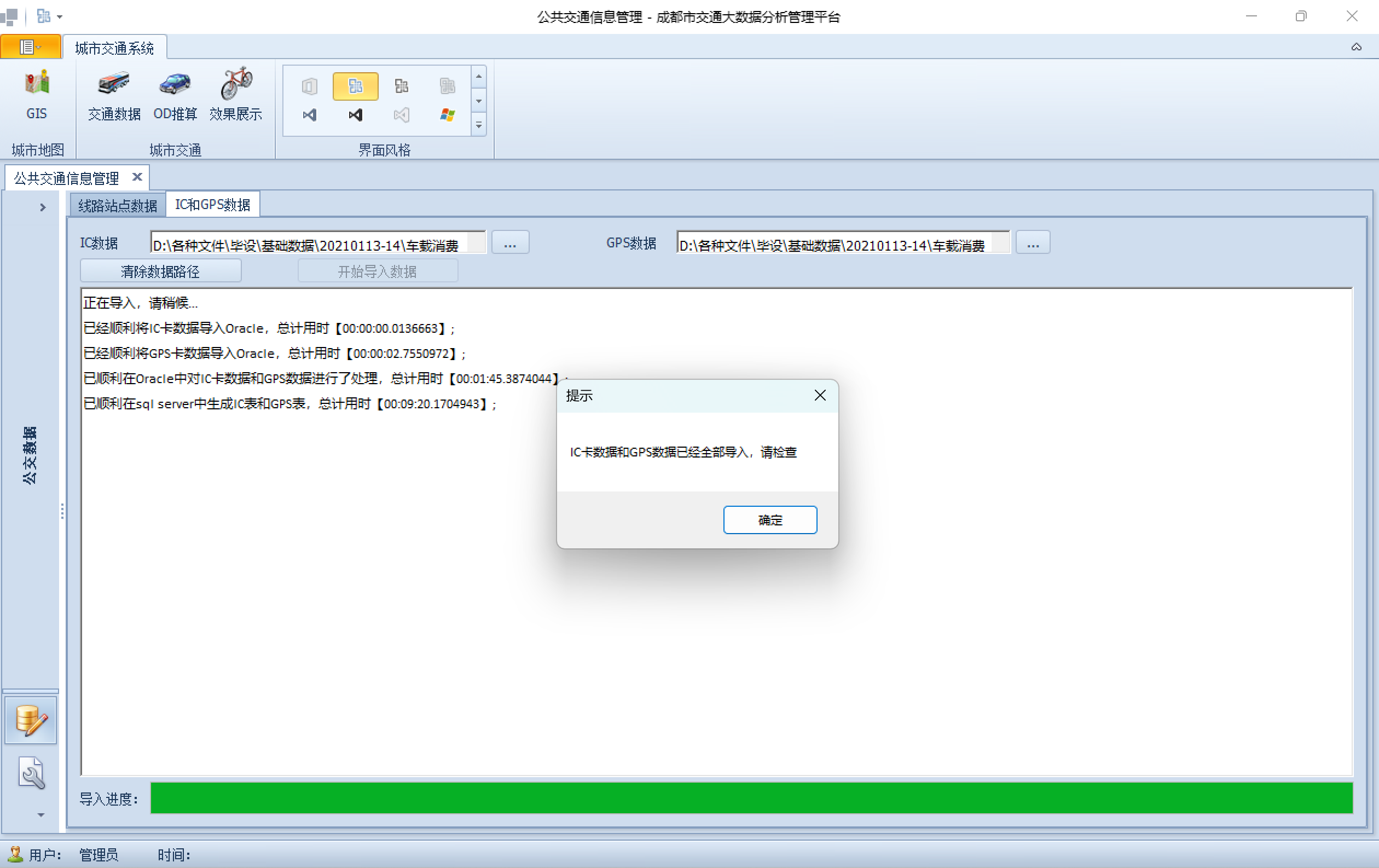
5.3 平台运行结果

基于数据库中预先编写完成的存储过程，利用C#语言进行连接和调用，公交出行OD推算平台实现了公交数据的导入、清洗、处理、处理、查询、OD推算、OD查询以及基于ArcGIS的可视化展示。

5.3.1 数据处理界面

1. IC卡数据和GPS数据导入界面

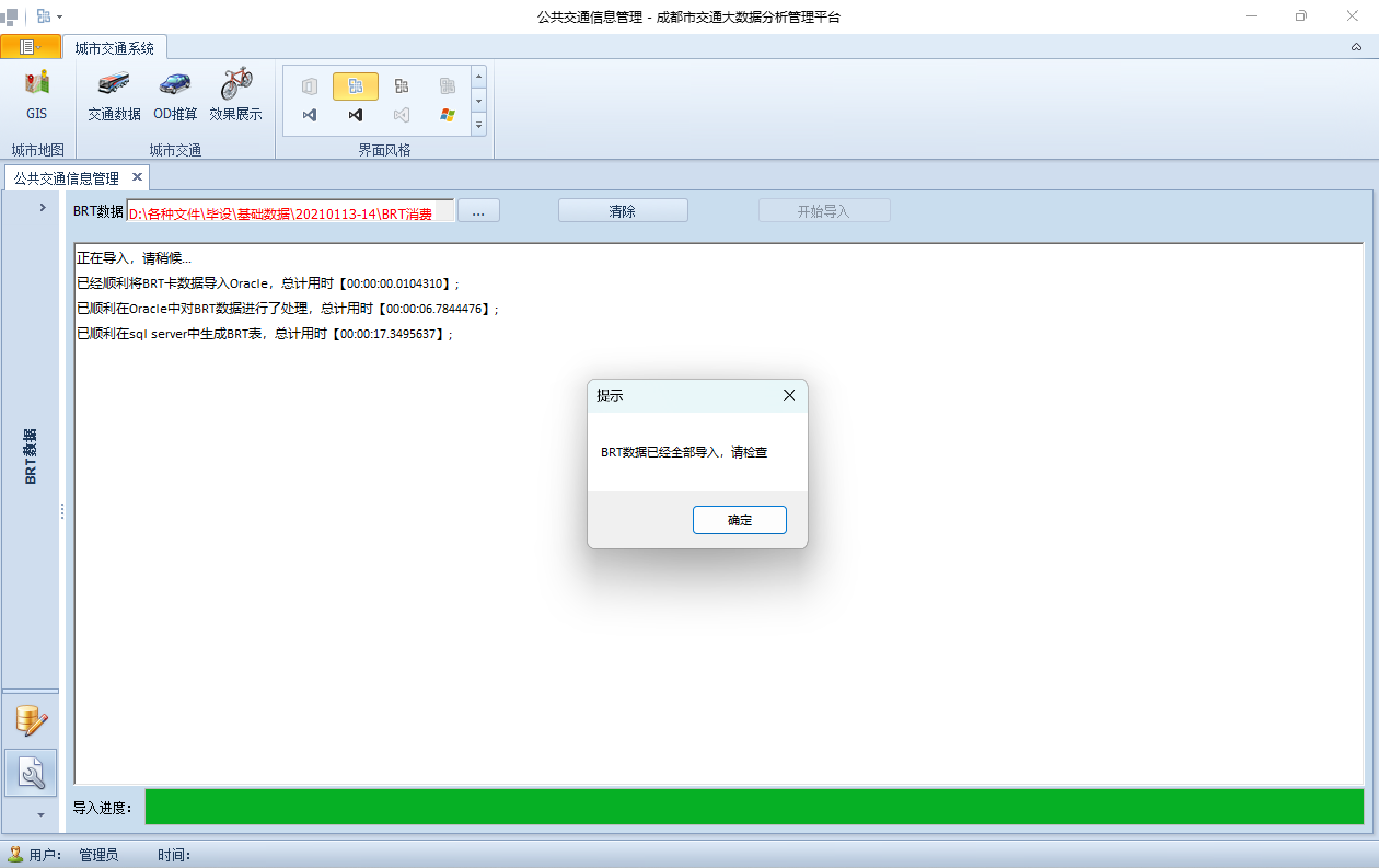
此界面对应导入公交IC卡数据的功能，其成功运行的结果如下图展示：



**图 5-1 IC卡和GPS数据导入界面运行结果**

1. 快速公交BRT数据导入界面

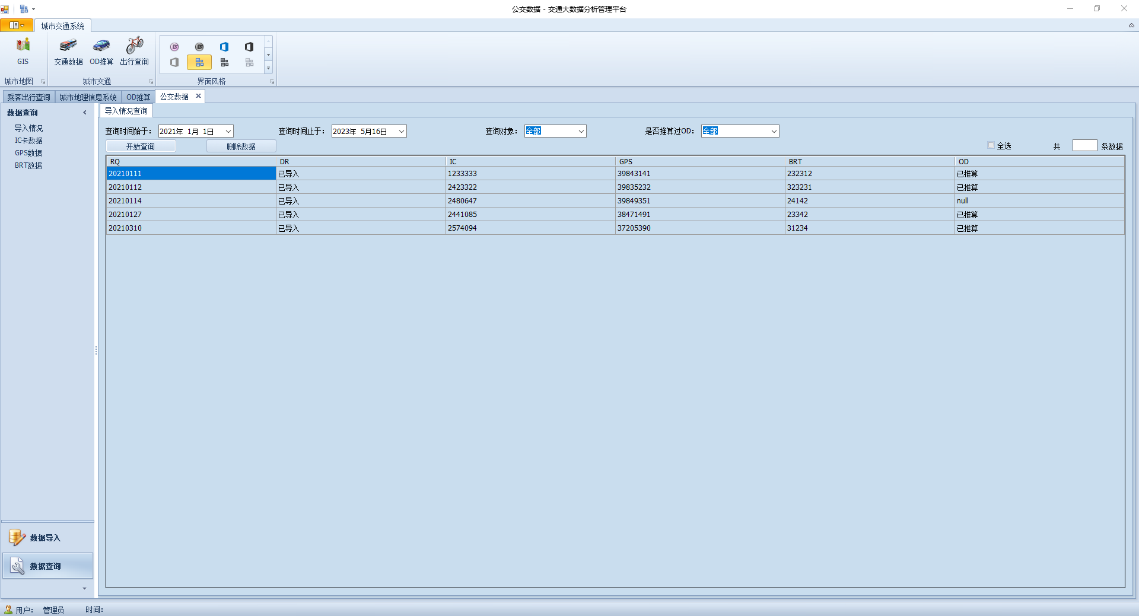
此界面对应导入快速公交BRT数据的功能，其成功运行的结果如下图展示：



**图 5-2 BRT数据导入界面运行结果**

1. 导入情况查询界面

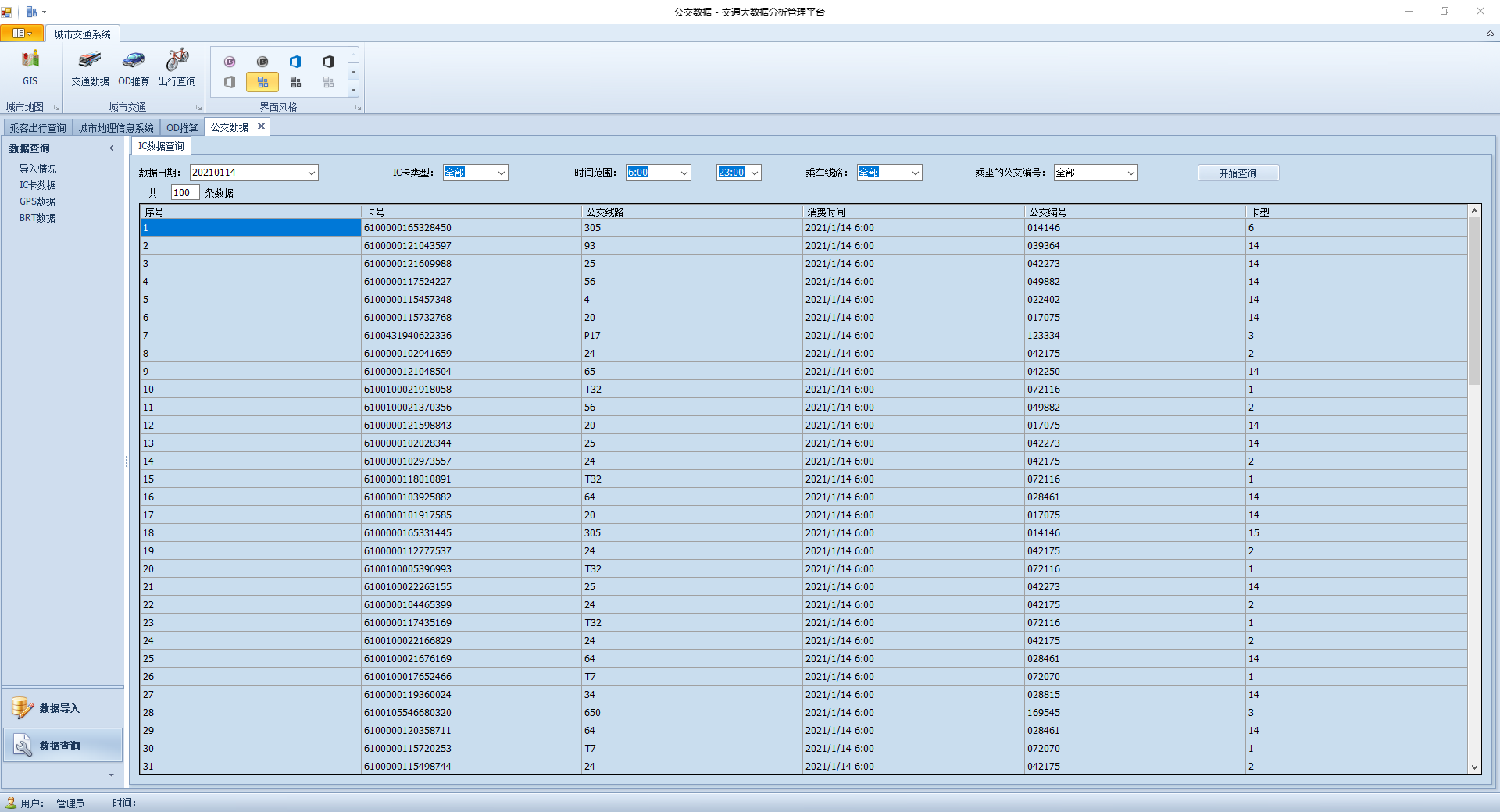
此界面对应查询导入数据整体情况的功能，其成功运行的结果如下图展示：



**图 5-3 导入情况查询界面运行结果**

1. IC卡、GPS、BRT数据查询界面

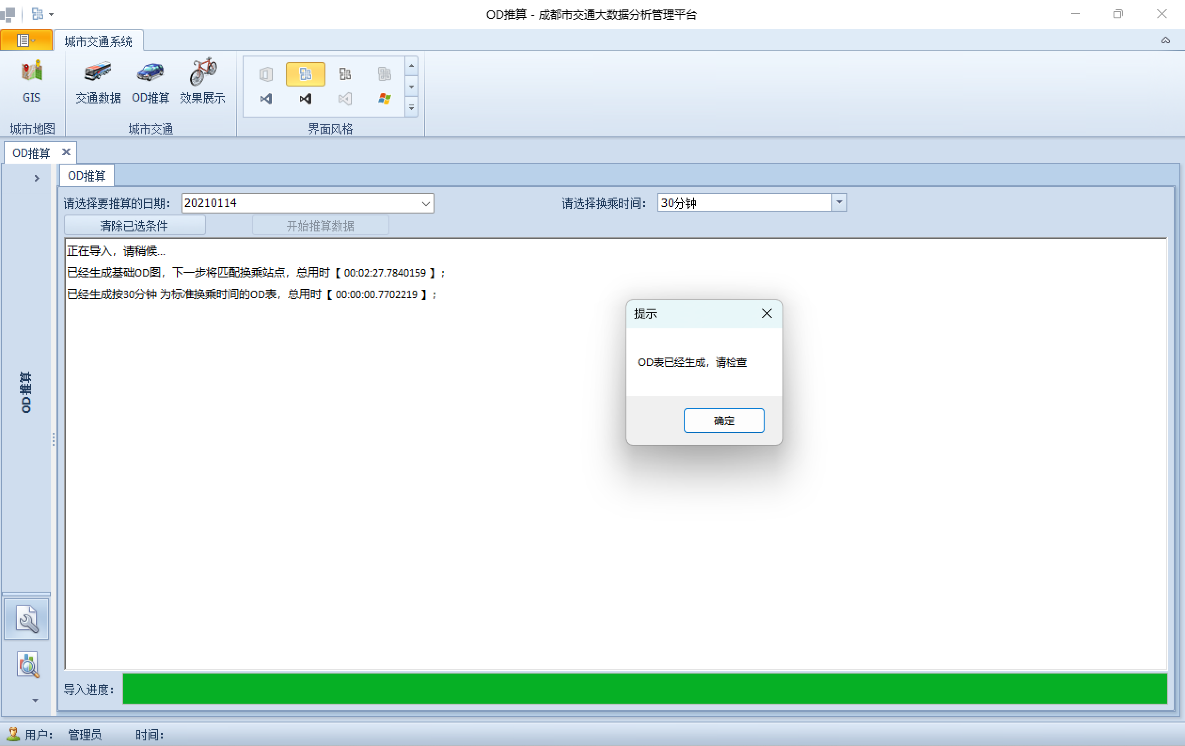
此界面对应查询指定日期IC卡数据、GPS数据、BRT数据的功能，其功能和界面均类似，本文以IC卡数据查询界面为例，其成功运行的结果如下图展示：



**图 5-4 IC卡数据查询界面运行结果**

1. OD推算界面

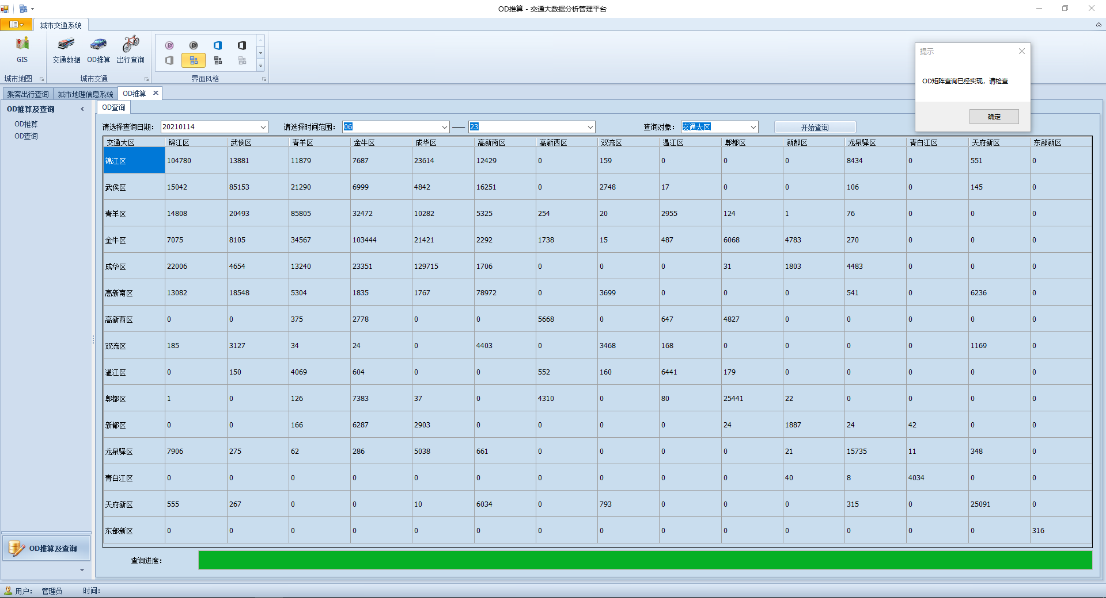
此界面对应推算指定日期公交OD的功能，其成功运行的结果如下图展示：



**图 5-5 OD推算界面运行结果**

1. OD查询界面

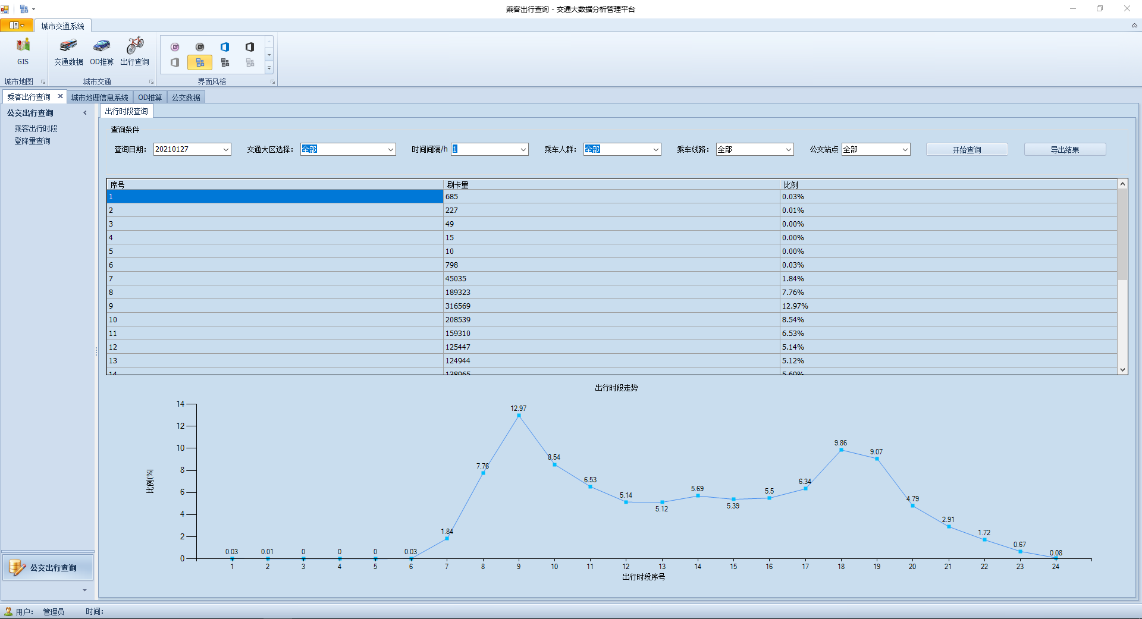
此界面对应查询指定日期OD矩阵的功能，其成功运行的结果如下图展示：



**图 5-6 OD查询界面运行结果**

1. 出行时段查询界面

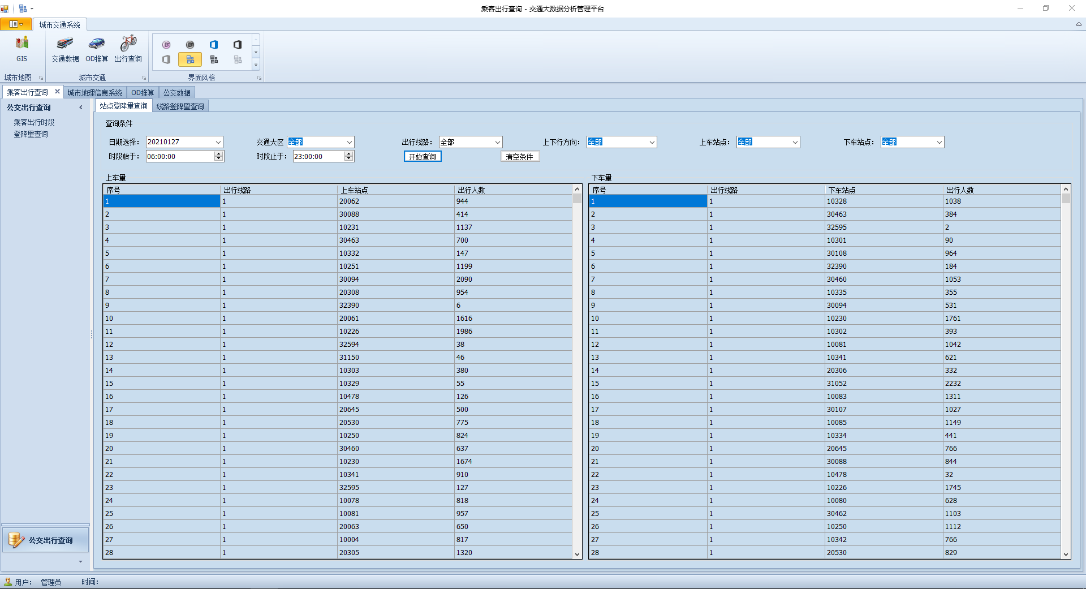
此界面对应查询指定日期出行时段统计数据的功能，其成功运行的结果如下图展示：



**图 5-7 出行时段查询界面运行结果**

1. 站点登降量查询界面

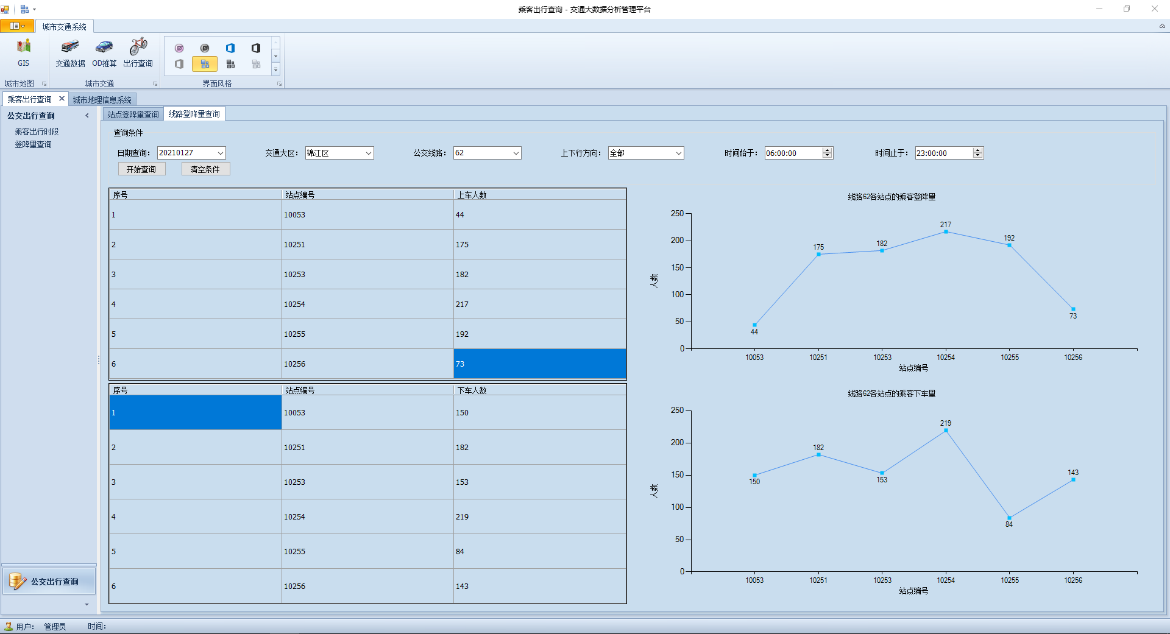
此界面对应查询指定日期公交站点登降量统计数据的功能，其成功运行的结果如下图展示：



**图 5-8 站点登降量查询界面运行结果**

1. 线路登降量查询界面

此界面对应查询指定日期公交线路登降量统计数据的功能，其成功运行的结果如下图展示：



**图 5-9 线路登降量查询界面运行结果**

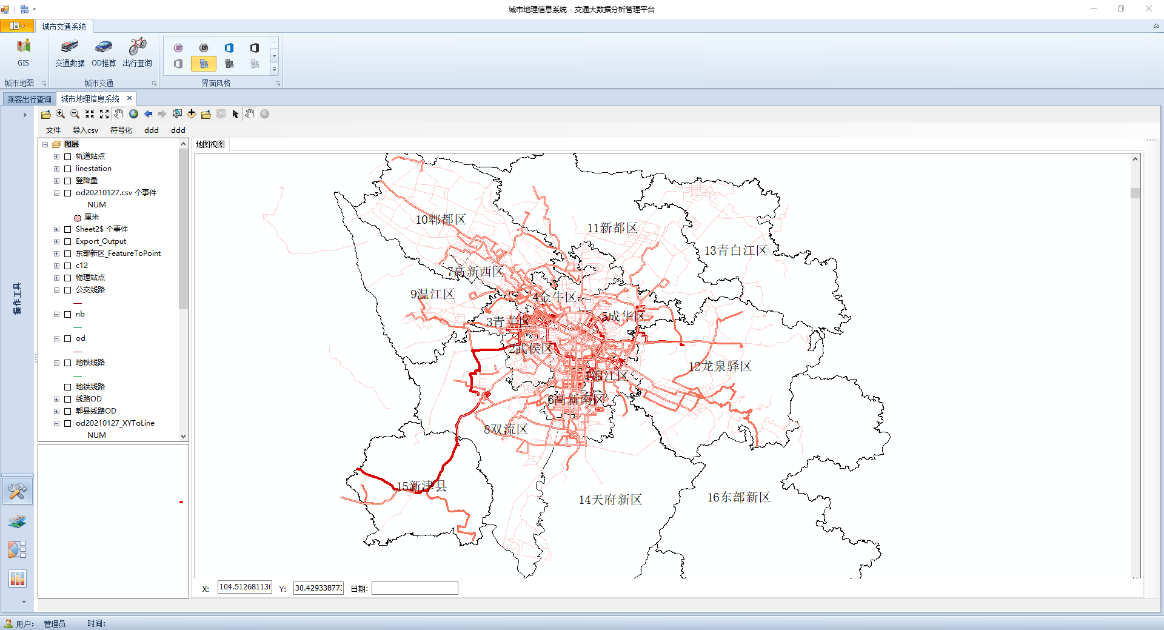
5.3.2 GIS展示界面

GIS展示界面的主要功能为实现OD数据的可视化，并且便捷操作人员的操作，实现GIS软件和平台的联合使用。其展示的主要内容为成都交通大区之间的出行OD和成都市公交线路单日内的乘客登降量。

1.交通大区之间的出行OD

基于交通大区之间的OD矩阵和成都市的地图，将成都市划分为16个交通大区，以线段连接交通大区的地理中心点表示交通大区之间的相互出行，并且以色带的带宽以及色差区分不同量级的公交出行量。

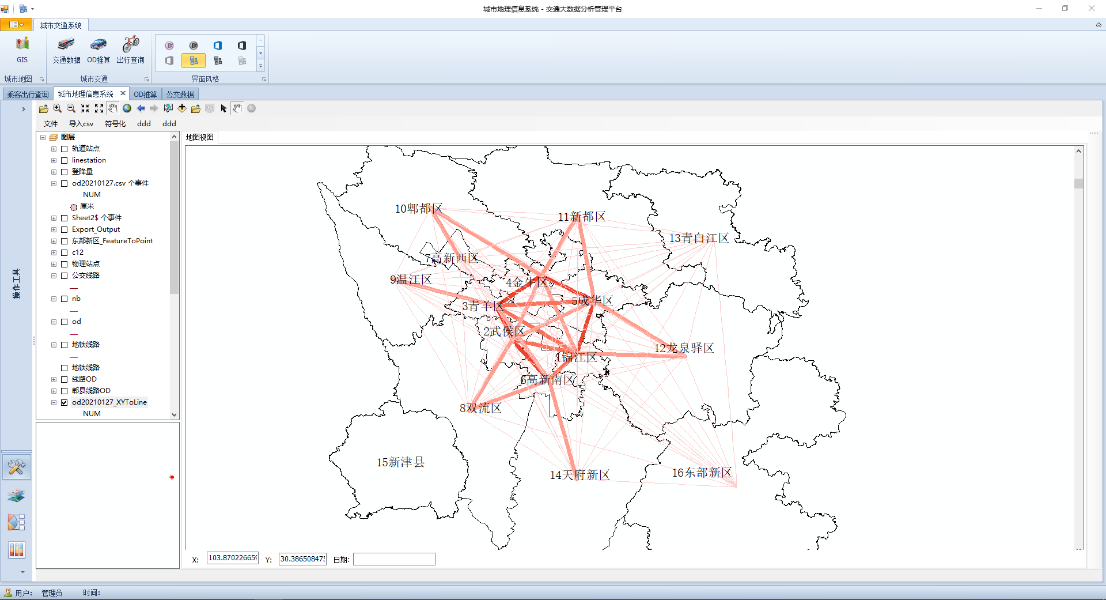
其成功运行的结果如下图展示：

****

**图 5-10 交通大区OD展示界面运行结果**

2.公交线路单日的登降量

在成都内的公交线路地图的基础上，利用基于乘客上下车站点数据统计得到的线路单日内的登降量数据可以显示成都市内公交线路的出行强度。以色带的宽度和颜色的深浅来区别不同量级的线路登降量。



**图 5-11 公交线路单日登降量展示界面运行结果**

5.4 本章小节

本章展示了基于IC卡、公交车载GPS数据以及一系列公交运营数据推算得到了公交乘客上下车数据的结果以及一系列基于上下车数据衍生得到的OD类数据，包括交通大区之间的公交出行数据、交通大区内部的公交出行数据、出行时段统计数据、公交线路登降量和公交站点登降量数据。特别的，以图片的形式详细展示了本文搭建的公交出行OD推算平台中数据推算界面和GIS展示界面的运行成果和功能。

**结论与展望**

1、研究结论

本文利用成都市内2021年内数日的公交IC卡数据、公交车载GPS数据、公交线路站点数据、交通小区编号等公交公司基础数据对持卡出行类乘客的上下车站点进行了匹配和推算，并且将一系列的操作流程集成到了基于C#语言搭建的推算平台中。

主要工作和结论如下：

1. 上车站点的匹配

乘客上车站点的匹配是推算乘客公交出行OD的第一步，同时上车站点还是推算乘客下车站点的基础。针对上车的匹配，本文采用了时序匹配法。结合研究中掌握的数据，将IC卡刷卡记录数据、GPS数据和线路站点数据进行数据融合，以公交的到离站数据为匹配基础，综合考虑乘客上车所需的时间和GPS数据上传的间隔性，最后通过时间范围匹配各条IC卡记录对应的上车站点。其最终的匹配率可以达到97%以上。

1. 下车站点的推算

下车站点的推算是出行OD研究的重点环节。本文基于乘客出行链闭合的假设，将乘客的出行分为往返出行、换乘出行、单日仅以一次的出行，分别进行下车站点的推算研究。根据出行次数的统计结果，成都市内符合出行链闭合假设的人群约占总体人群的83%。除去本文研究范围以外的人群，最终的推算率为75%左右。

1. 推算平台的搭建

在推算得到公交车出行OD矩阵的基础上，本文基于C#语言，连接多款外部软件，包括数据库软件和地图软件，搭建集数据导入、数据清洗、数据处理、数据查询、公交OD推算、OD查询和OD可视化展示于一体的平台

在集成方面，文本主要将数据库中导入、清洗、处理数据和OD推导的SQL语言编写成为存储过程，并且赋予可被定义的参数，再通过C#语言进行数据库连接和存储过程的调用，并通过窗口中的控件对参数进行赋值，实现数据处理和OD推算等功能。特别的，本文基于C#对ArcGIS进行二次开发，将ArcGIS中的一些基础操作植入C#窗口中，不仅实现了数据的可视化，同时实现了OD推导和查询的一体化。

2、研究展望

针对公交OD的推算和平台的搭建，本文的研究中仍存在如下不足：

（1）受原始数据质量的限制，本文的研究数据并不包含新津县的IC卡和GPS数据，因此OD推算的结果中并不包含此交通大区，OD推算未能真正覆盖成都市。

（2）针对下车站点的研究，本文主要基于出行链闭合理论进行推算。但是此理论无法实现推算单日出行仅有一次的乘客，需要通过扩样进行分析，造成此类乘客下车站点的丢失，匹配率为76%。未来可以融合多种下车站点推算方法，进一步提高站点的匹配率并兼顾单日出行仅一次的乘客。

（3）在统计出行OD时，本文未能将乘客的换乘和连续出行连接起来。因此最终的结果仅记录了单次出行的上下车站点数据，未能真实体现乘客的出行目的地和整个行途。未来将通过IC卡号连接乘客的数次出行，推算出公交乘客完整的出行路径。

（4）推算平台的控件和整体布局简单且较单一，整体风格不够专业，可以采用外部软件的控件并重新布局以提高平台的使用感受。鉴于本文研究中使用的设备运行内存不够，部分平台的界面的运行功能未能展示出来，同时，C#语言中类似功能的代码重复率高，降低了平台的运行速度。未来可以将部分功能进一步集成，提高推算平台的性能和运行速度。

**致谢**

四年时光如白驹过隙，我即将在西南交通大学结束我的本科学习。学海无涯，本科的结束虽然不是学习生涯的终点，写到此处却已经是感慨万分了。本科生活不仅丰富了我的知识更是培养了我的科研素养。

值此论文完成之际，首先感谢我的毕业设计指导老师徐占东老师！

无论是论文本身的撰写还是平台的搭建，徐占东老师都给予了我专业的指导。他们严谨的学术作风、缜密的逻辑思维、对细节的把控以及专业的科研素养都让我受益终生。每次和他们交谈都能为我指引前进的方向，如醍醐灌顶一般。他对于科研学术的敬业态度让我无比敬佩，对我的鼓励和帮助也加深了我对学术研究的兴趣。

然后我要感谢我未来研究生阶段的导师唐优华老师！没有唐老师的介绍，我就不能认识徐老师，也就没有这一篇毕业论文。因此我十分感谢唐老师给与的机会，并且期待在研究生阶段能跟随唐老师获得更多的学习机会。

其次，我要感谢实验室各位师兄的帮助和支持。无论是师兄们在论文写作上提供的建议还是日常生活中一起进行的活动，都帮助我度过了写毕业论文这段“痛苦并快乐”的时光。

感谢本科的三位舍友：高浦文、曹铭洋、董俊强。和你们一起生活，一起外出学习，一起探店聚餐的四年时光，使我的本科生活变得丰富多彩。特别是大家对于学习的认真态度，在无形中带动了我。

其次，我要感谢远在家乡的各位亲如家人的朋友：李翔宇、夏张扬、杨成、管彤、俞子龙、戴袁飞等。每每在我感到迷茫和不受理解的时候，他们对于我的关心、理解总能支撑我打破阻碍继续走下去。

最后，我要感谢我的父母。是他们一直以来对我的关切、支持和鼓励给予了我不断向前的动力和信心。

**参考文献**

1. Barry J J, Newhouser R, Rahbee A, et al.Origin and Destination Estimation in New York City with Automated Fare System Data[J].Transportation Research Record:Journal of the Transportation Research Board,2002,1817(1): 183-187
2. 舒国辉. 基于数据挖掘的公交客流规律研究[D].北京交通大学,2009.
3. 李海波,陈学武,陈峥嵘.基于公交IC卡和AVL数据的客流OD推导方法[J]. 交通信息与安全,2015, (6): 33-39, 95.
4. Yu X, Shao F, Sun R, et al. Boarding Stations Inferring Based on Bus GPS and IC Data[C]// Methods and Applications for Modeling and Simulation of Complex Systems. 2018.
5. 张彤.一种利用线路走向信息推算公交上车站点的方法[C]//第十四届中国智能交通年会论文集(2). 2019.
6. 柳伍生,周向栋,匡凯.基于IC卡数据的公交下车站点区间不确定性客流推导方法[J].铁道科学与工程学报,2018, 15(11): 2988-2994.
7. 胡继华,邓俊,黄泽 . 结合出行链的公交 IC 卡乘客下车站点判断概率模型 . 交通运输系统工程与信息,2014, 14(2):62-67, 86.
8. 李佳怡,张锦,张静文等.城市公交乘客下车站点推算方法和有效性评价.武汉大学学报·信息科学版,2018, 43(8): 1172-1177.
9. 崔紫薇,王成,陈德蕾等.基于历史出行记录扩充的公交乘客下车站点推算方法[J].南京大学学报:自然科学版,2020,56(2): 227-235.
10. 温晓丽,何子登,余红玲等.基于Mean Shift的公交下车站点判定方法[J].软件导刊,2021,20(4): 145-149.
11. 杨鑫.基于IC 卡数据的公交客流智能推断方法研究[D]. 北京邮电大学, 2019.
12. 马伯浩,布少聪,夏非凡等.基于多维数据的常规公交时空客流分析[J].中国高新科技,2021,No.94(10): 125-127.
13. 郝伟娜,卢文跃,董红召.基于MAC地址扫描的公交客流动态获取方法[J].浙江工业大学学报,2018,46(02): 199-203.
14. 谢振东,冷梦甜,吴金成.基于一卡通数据的公交站点识别方法分析与研究[J].广东工业大学学报,2019,36(01): 23-28.
15. 秦政.基于公交IC卡和GPS数据的乘客上下车站点研究[J].西部交通科技,2017,No.121(08): 115-119.
16. 尹安藤. 基于公交GPS和IC卡数据的公交OD推算[D].哈尔滨工业大学,2017.
17. Kim Kyoungtae,Lee Inmook,Min Jaehong. Bus OD Matrix Estimation Using Smart Card Data[J]. Journal of the Korean Society for Railway,2018,21(11).
18. Huang Di,Yu Jun,Shen Shiyu,Li Zhekang,Zhao Luyun,Gong Cheng. A Method for Bus OD Matrix Estimation Using Multisource Data[J]. Journal of Advanced Transportation,2020,2020.
19. Hussain Etikaf,Bhaskar Ashish,Chung Edward. Transit OD matrix estimation using smartcard data: Recent developments and future research challenges[J]. Transportation Research Part C,2021,125.
20. Masood Jafari Kang,Shervin Ataeian,S. M. Mahdi Amiripour. A procedure for public transit OD matrix generation using smart card transaction data[J]. Public Transport,2020(prepublish).
21. Wang Yunshan,Zhang Wenbo,Tang Tianli,Wang Dazhong,Liu Zhiyuan. Bus OD matrix reconstruction based on clustering Wi-Fi probe data[J]. Transportmetrica B: Transport Dynamics,2022,10(1).
22. Kusakabe T, Asakura Y. Behavioural Data Mining of Transit Smart Card Data：A Data Fusion Approach[J]. Transportation Research Part C Emerging Technologies, 2014, 46: 179-191.
23. 陈君, 杨东援. 基于APTS 数据的公交卡乘客通勤OD 分布估计方法[J]. 交通运输系统工程与信息, 2013, 13(04): 47-53.
24. 陈娴,龚小林,邬岚.基于IC卡和稀疏GPS数据的中小城市公交乘客上下车站点推算方法[J].交通运输研究,2022,8(01): 79-88.
25. 张志熙,陈玲娟.基于IC卡数据的公交乘客下车站点推算模型[J].昆明理工大学学报(自然科学版),2021,46(01): 142-149.
26. 邓红星,赵志恒,王玮琦.基于公交IC卡和GPS数据的公交OD量推算研究[J].重庆理工大学学报(自然科学),2019,33(06): 220-226.
27. 刘健国. 基于公交IC卡数据的线路客流OD推算方法优化及评价研究[D].西南交通大学,2021.
28. 李佳悦. 基于多源数据的公交乘客OD推算及公交线网优化[D].东南大学,2021.
29. 刘涛. 基于IC卡和GPS数据的公交线路OD推算及客流特征研究[D].东南大学,2020.
30. 王怡峥. 基于公交大数据的OD挖掘与预测方法研究[D].河北工业大学,2020.
31. 钱易. 基于多日公交IC卡数据与AVL数据的公交OD推算方法研究[D].西南交通大学,2017.
32. 戴维. 基于IC卡数据和公交车GPS信息的公交OD矩阵推算[D].华中科技大学,2009.
33. 宋子杭. 基于公交GPS和IC卡数据的公交线网优化方法[D].北京交通大学,2017.
34. 李泰增,程歆玥,李爱迪等.基于分布式架构的多源交通大数据可视化系统设计与实现[J].城市建设理论研究(电子版),2023,No.436(10): 119-121.
35. 赵向宇.基于GIS的公交大数据可视化分析应用[J].交通世界,2019,No.520(34): 18-19+49.