

# 基于大数据应用的智能公交决策分析平台

齐冬浩

华录智达科技股份有限公司

**摘要：**城市化快速发展，城市流动人口激增，城市交通拥堵现象日益严峻。加快推进智能公共交通建设，成为当前城市交通发展的必由之路。随着智能公共交通建设步伐的推进，产生了庞大的公共交通数据信息，而由于现存的数据普遍存在着源头多样化、数据结构不同、数据采集困难以及数据处理复杂等问题，严重影响到智能交通建设步伐。基于此，本文搭建一个智能公交决策分析平台，借助大数据技术，打造公交运营、财务和客流等各个层面的数据分析决策平台，从整体上对公共交通企业运营情况进行评判，为公共交通相关运营管理和决策提供科学的数据支撑。

**关键词：**大数据；智能公交；平台

【DOI】10.12254/j.issn.2096-6539.2022.10.008

公共交通是城市的重要组成部分，也是缓解城市交通拥堵、减少环境污染和能源消耗压力的有效手段。为了提高城市公共交通系统的吸引力，实施科学的规划和调控，公共交通企业的信息化程度逐年提高。目前，信息系统越来越多，大部分系统独立运行，由不同的部门负责，系统数据相互分离。随着公交业务的积累和进一步发展，数据通信隔离和效率低下已成为重要的制约因素。因此，如何对企业内部各类生产数据进行整合，实现数据共享和交换，构建公交企业大数据应用决策平台，最大限度地实现数据价值，改变决策方式，实现公交企业业务升级和创新发展，成为公交企业信息化建设升级的重要目标。在交通运输领域，大数据和人工智能等新技术一直被视为缓解交通压力的技术工具。在公交行业，应用大数据相关技术、工具，对日常产生的各类公交业务数据进行采集、整理、分析，有助于全面了解城市公交的实际运营情况、人群出行规律和交通需求，为公交运营精细化、管理精准提供基于数据分析的综合决策，为公交企业带来高效运行效率和长期的成本节约。

## 一、智能公交系统的基本构成

随着社会经济的快速发展，交通管理部门、公共交通运营商和出行客流对智能公共交通提出了更高的要求，智能公共交通已成为智慧城市的重要组成部分。智能公交系统一般由五部分组成：数据采集层、基础设施层、数据层、中间支撑层和应用层。数据采集层包括各种前端数据采集设备；基础设施层包括服务器、存储设备、网络设备等设备；数据层包括实时运行数据、外部数据；中间支撑层包括GIS平台、Webservice、数据分析中间件等服务组件；应用层可分为公交公司管理与监控应用、政府管理控制与决策应用和公共出行信息服

务。

## 二、基于大数据应用的智能公交决策分析平台

以下建立了以大数据为基础的公交智能决策与分析平台，实现了数据汇聚接入、数据采集、ETL、关联算法、数据存储等功能。根据客流统计和应用支持，对原有的数据信息进行处理，形成了大量的数据和流量统计，并通过应用支持系统进行分析和发掘，实现了多维、直观的客流统计。该体系从总体运营状况、客流各环节的运营状况、客流各环节的运行状况，深入挖掘导致运营变化的具体因素，对运营异常状况进行合理的客流预测，为公交企业的运营管理和决策提供科学的数据支撑。旅客的同时，在运行调度，机务管理，场站管理，安全管理，营运客流管理，考核管理，信息服务，移动办公，财务会计等九大行业的客户服务升级，建立成本，收益，运营，服务、安全客流等多项决策模式，以单车作为计算单位，逐级进行分类，形成单车、线、车队，通过建立一个有效的决策支持体系，可以为乘客做出准确的决策，改变当前的公交运营管理模式，提高公交运营的运营效益。

### （一）数据汇聚与整合

数据汇集、整合来自各种信息化系统、标准接口、物联网设备、数据文件等数据格式参差不齐，数据质量参差不齐，数据收集过程中需要根据数据来源的要求，实现数据格式的标准化。建立数据标准，并对原始资料进行归档和归档。下一步，通过ETL对数据进行提取、加工，并通过HBase进行数据的储存，最后生成数据仓库，为各种统计、分析和决策提供数据支持。

#### 1. 财务数据采集

收集和校验公交企业的多维度数据，制定数据规范，通过数据清洗、匹配，整合收入（现金收入、信用卡收入、网络收入）、安全检测数据（车辆检测、车辆保险费、安全生产费等）；意外损失，固定资产补偿（车辆固定资产，驾驶员工资，管理人员工资），维护数据，能源消耗；材料消耗，工时，补贴数据等。

#### 2. 客流数据采集

公交乘客信息的收集主要包括公交调度系统、E卡系统、公交现金收入系统；车辆技术、车辆安全、公交收入结算平台等信息采集，主要有公交线路站点数据、公交到离站数据、路单数据；现金交易数据、公交刷卡结算数据、公交分段计费线OD数据等，主要是为了求取各个车站的上下车流量。在此基础上，以E通在各个车站的出勤率为基础，对每个车站的出票量进行了分类。公交车到出站的资料主要来自汽车的实时位置，其中包含了车牌号，经度，纬度，时间，线路，站点、上

车人数、下车人数等。

### 3. 基础数据采集

基础数据主要来自公交线路、车辆和站点的智能化调度；为确保与原有基本信息相符，司机、运营资料等均使用后台处理软件进行数据的实时处理，从而防止因人为输入而造成的差错；既能确保资料的可靠性，也能有效地实现数据的同步。

### 4. 数据预处理

公交公司的业务数据是庞大的，为了提升数据的处理速度和数据的精确度，公交的营收、客流的刷卡、公交的每日到站数据等，都是非常必要的。

#### (1) 财务数据预处理

以营收数据为例子，平台建立统一规范的数据规范，将现金收入、刷卡收入、网络收入等基本收入数据进行归类收集。在进行入库的过程中，对入库的基本资料进行二次分析，根据车辆、线路、分公司等不同时段的不同时段（每日、每月）进行统计和分析，并将二次分析的结果保存到系统中。

#### (2) 客流数据预处理

公交客流数据预处理是根据公交运营数据、刷卡收入等数据，应用大数据技术和工具自动清洗、分析、匹配数据、公交上下车匹配和客流统计，融合形成可分析使用的客流基本数据等，整合后的主要数据有：交通卡消费 OD 明细数据、线路车辆各站点投币上下客流明细数据、线路班次（含各站点客流信息）明细数据等。在交通卡的 OD 详细资料包括线路名称，车辆，POS 机号，卡号，交易时间；以交通卡消费详细资料为依据，以消费时间、下次乘坐时间、车辆编号和乘坐时间等为依据；通过 GPS 等相关的信息对乘客的上下车位置进行定位，其识别精度必须在 80% 以内。

### (二) 运营数据分析模型

#### 1. 营收分析

对刷卡收入，互联网收入，现金收入。深入剖析移动支付收入、定点专线收入构成的发展态势，可从集团、分公司、线路等多个维度全面掌握收入渠道变动情况，为制订提高非支付比率的相关政策提供可量化的数据基础。

#### (1) 分公司营收结构占比

根据数据中心的数据，各个分支机构的收入在总收入中的比重，如图 1 所示。

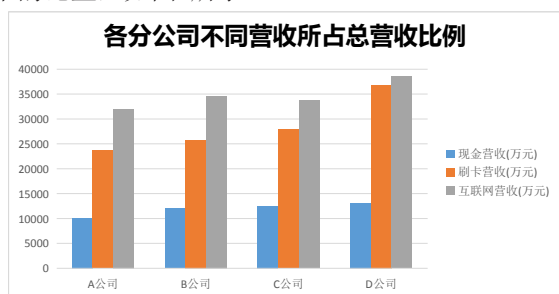


图 1 各分公司不同营收所占总营收比例

#### (2) 分公司营收趋势分析

对各个企业每月或几个月份的收入变动进行统计，用不同的颜色表示各个分支机构，可以直接反映出该企业在一个区域的总体收入走势。

#### (3) 分公司千公里效率分析

对各个分部在不同时期的千公里进行了性能评估。千公里效率以千公里营收和千公里客运量两个指标来衡量，千公里营收即每一千公里的营收总额，千公里客运量即每一千公里客运量。平均票价是由人工设定的。数据中心的数据是按照设定的时间来计算的，数据中心已经将所有客流和收入数据进行了二次分析，对于部门的千米运输能力的分析，数据的处理和数据的精确度都得到了极大的提高。

### 2. 能耗分析

能源消耗的主要是通过对总公司、分公司、线路等不同能源类别的汽车消耗量进行剖析，寻找造成高油耗的根源。以纯电、柴油、LNG 和 CNG 能源为分析变量，分析了各能源的消费构成，并分析了百公里能源的能耗、能耗趋势，从中找出了能耗较高的汽车，并对其产生的具体影响因素进行了分析；提供了可计量的基础资料，以减少能源消耗。

### 3. 公交医生

模拟医师对病情的诊治流程，对线路、车辆的能量消耗等问题进行分析和判定。综合分析营业收入、班次、运营时间、根据人流、停机数据，并定时向外输出诊断报表，以分析人员管理问题、执行问题、线路设置或操作问题，以找到原因，对症下药。

#### (1) 线路诊断

分析了异常线路的原因，其中包含收入趋势分析、班次总数趋势、日均班次走势。平均运营时间，线路砍掉的趋势，运营和非运营的比较。

1) 班次总数 = 各月份线路班次总数。

2) 日均班次（统计期内每日平均车辆运行班次） = 计算期内总班次 ÷ 计算期周期。

#### (2) 车辆诊断

对异常车辆的原因进行了研究，主要有车辆能量消耗趋势、充电量和行驶里程趋势、载客人数，平均运行时间，每月平均运行次数，运行速率。

### (三) 客流与运力调优分析模型

#### 1. 客流 OD 分析

依据客流的出行 OD，与线上的线路 OD、站点 OD、交通小区相关联，显示了整个网络的流量。针对客流来自何处、前往何处、哪些时段、哪些时段客流最密集等问题，可作为线路、站间区间车规划和调整依据。

#### 2. 客流登降量分析

结合线路上的登降量和站点的登降量，构建了客流的时间动态模型，并结合图表和表格对站点断面客流、线路全天客流、线路登降量和线路进行了分析；站点的登降情况排序。通过可视化的方式，可以看到哪些线

路,哪些站点,哪些时段的乘客最多。(1)线路登降量分析。主要体现方向、站点名称和上车客流量、下车客流量、车内客流总量等相关资料。如果站点被撤销,则可以利用功能模型来评价这种变化对线路的作用。

(2)站点登降量分析。通过对一条经过的路线的乘客数量进行统计,可以对车站的客流情况进行统计,从而得出车站的撤销对某一条路线的作用。当列车停靠站减少时,可以评价列车客流的变化。(3)全网线路登降量。对各时间段内各网站的站点登降进行了统计分析。筛选出在不同时段全网站点登降量前10的站点。

### 3. 运力预警分析

通过对客运量和支付数据的分析,实现客流预警、早晚高峰预警和客流分析;对旅客的漏票分析,支付方式的比例和详细的支出分析。(1)客流的统计。通过对各班次的客流、区间和单列的车辆满负荷进行统计,了解一次列车的乘客和乘客数量。(2)平均交通流量的计算。平均班次满负荷是一种反映列车运营中旅客的满负荷水平的一项重要数据,可以用来衡量各单位的汽车使用情况。由此可以得到一日间的平均客运量和日间的平均满载率的变动情况。(3)对早晚高峰期进行的预报。分析了各线路和站点在早晚高峰时间的满负荷情况。

### 4. 全网客流热力图

对各时段的流量分布进行了研究,并对各时段的流量进行了比较。

### (四) 可视化应用

通过对各种业务指标的集成,实现多维度的可视性显示。动态、实时和有效的数据集中显示可以使公司对公共交通的运行状况有所认识,从而更好的实现对公共交通的决策和管理。

#### 1. 综合展示

对营收、能耗、材料等数据进行全面的统计,包括营收占比、经营状况、工作效率等主要指标。费用和各種经济指数的综合说明。

#### (1) 营运情况

营运状况由营运里程、平均工作时间、平均营运时间、班次总数等四项指标来度量;累计的班次数量,与上个月的数据相对比,得出一个月的数据,反映了一个月的短期趋势。

#### (2) 经济指标

用百公里的运行里程来估算各种消耗,包括:材料耗、轮胎耗、CNG、液氮、电、燃油等。每百公里的材料耗指数:月份各公司材料消耗总金额:营运里程 $\times 100$ 。

#### (3) 营运成本支出

经营费用开支,包含车辆检查费用,车辆保险费用,安全生产费用、交通事故损失,折旧,材料耗和时间成本、胎耗、能源消耗、司机工资。将前一个月的各个部门的成本统计资料汇总到各个部门的汇总,得出各

个部门的成本和费用。

#### 2. 运营里程

通过对运营班次、营业收入、客运量等数据的统计,对班次运营效率、企业营收、客运量趋势进行了统计和分析。班次作业效能包含三项衡量标准,即:计划作业率、砍班率、脱班率。这三个指数可以在一定程度上反应排班的合理性,在发生错班率高的情况下,分析其成因,避免问题,合理安排下个月的排班。另外,通过对每日客流情况的分析,绘制出客流流向的趋势,便于商家更好地反映出各月份客流的动态;由此,对增长下降的成因进行了剖析。

#### 3. 成本能耗

费用包含能源消耗(电力,柴油,液化天然气,CNG),胎耗,材料和工作时间;司机工资、保费等,统计、同月分析等方法,可直接掌握企业的成本构成与构成,为成本调整与管理奠定基础。

#### 4. 维修材料

维护材料的内容包括物料损耗和时间成本。工时成本主要包含意外工时、维修工时、维护工时等。物料消耗量结构的计算采用车辆数、材料金额,对营运车辆投入、维修车辆、事故车辆等进行统计、保养车辆数,根据车辆算出保养率和维修率。

综上所述,现阶段的企业信息系统构建和经营决策的研究仍停留在“以数据+产业体验”为主的方法上,对产业的决策和决策进行了大量的研究。在大数据的今天,企业要善于获取数据、分析数据、运用数据,运用数据推动各种决策分析,是当前我国企业信息化工作的一个重要课题。该系统以Hadoop为基础,利用flink CDC、Spark SQL、HDFS进行数据采集、转换和装载,利用Spark On Yarn、flink进行多源OD数据的处理,并在此基础上完成公交基础信息、实时车辆定位数据、线路运营数据等功能;将财务收入、成本能耗、刷卡结算等数据进行汇总,提取出有意义的信息,对其进行特征分析、规律分析、预测分析,通过数据进行决策,通过数据管理和决策,为企业的决策和决策制定出各种决策和分析的方法,为公司数据创造更多价值。

### 参考文献

[1] 郭群勇,胡振华,张红.基于多源轨迹数据的城市交通状态精细划分与识别[J].交通运输系统工程与信息.2020,(1).83-90.

[2] 陈兴强,唐可,陈岳川,等.新能源公交车辆大数据决策支持平台的研究[J].城市公共交通.2017,(3).25-29.

[3] 游建泳,李文锋,潘梅芳,张鹏,姚辉,程远,阙灿金,林小伟.基于大数据应用的智能公交决策分析平台[J].中国交通信息化,2021(07):134-138.

[4] 贺冬.基于视觉物联网的智能公交信息采集系统设计[J].科学技术创新,2021(17):82-84.