

数据驱动的公交线路智能识别系统开发

刘瑶秋, 赵苗苗

(扬州亚星客车股份有限公司, 江苏 扬州 225116)

摘要:针对车联网监控系统中公交线路收集和分析困难的现象,开发了一套高效、准确的公交线路智能识别系统。系统由公交信息采集模块、路线智能识别模块和界面交互模块构成,分别负责收集路线数据、执行识别任务 and 提供直观的用户界面。经测试验证,此系统具有较好的准确性和稳定性。

关键词:数据驱动; 公交线路; 智能识别; 车联网

中图分类号: U471.2

文献标志码: A

文章编号: 1006-3331(2024)05-0022-05

Development of a Data-driven Intelligent Recognition System for Bus Routes

LIU Yaoqiu, ZHAO Miaomiao

(Yangzhou Yaxing Motor Coach Co., Ltd., Yangzhou 225116, China)

Abstract: Aiming at the difficulties of collecting and analyzing bus routes in-vehicle networking monitoring system, this paper develops an efficient and accurate intelligent identification system for bus routes. The system consists of a public transport information collection module, a route intelligent recognition module, and a user interaction module, which are respectively responsible for collecting route data, performing recognition operations, and offering an intuitive user interface. Through test verification, the system has better accuracy and reliability.

Key words: data driven; bus route; intelligent recognition; vehicle networking

当前的车联网监控系统难以有效地收集和获取各公交公司的运营路线。传统方法是依赖人工检视公交车辆的历史轨迹数据来推断运营路线,效率低下,无法满足日益增长的大数据分析需求。这种路线识别方式的界面既不直观也不易于操作,用户、平台监管者、数据分析工程师都需要一个更加高效、准确且界面友好的方式来获取和分析公交车辆的运营路线^[1-2]。

因此,研究和设计一种高效、准确且界面友好的公交线路智能识别系统显得尤为重要。这样的系统不仅能提高数据收集和处理的效率,还能通过直观的界面来提升用户体验,从而更好地满足车联网监控系统的应用需求。

1 系统概述

本系统基于模块化架构设计,分三个部分:公交信息采集模块、路线智能识别模块、界面交互模块。系统结构图如图1所示。

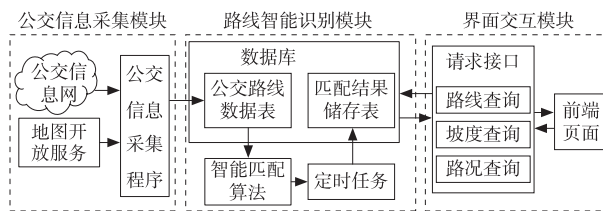


图1 系统结构图

其中公交信息采集模块负责基于公开的公交信息网和地图开放服务,收集并存储公交线路信息;路

收稿日期: 2024-01-31。

基金项目: 扬州市科技计划项目 (YZ2022019)。

第一作者: 刘瑶秋 (1992—), 女, 硕士; 工程师; 主要从事车辆智能网联开发的研究工作。E-mail: liuyaoqiu@asiastarbus.com。

线智能识别模块是系统的核心,由智能匹配算法、定时任务以及数据库组成,负责完成路线识别任务;界面交互模块包含请求接口和前端页面,负责满足路线查询、坡度查询、路况查询需求,并通过界面将查询结果和路线状态以图形的形式展示给用户^[3-5]。

2 公交信息采集模块设计

本系统的基础数据由公交信息采集模块提供,其设计和实现包含的三个关键步骤如下:搜集公交线路名称、获取公交线路的详细信息以及数据入库存储。

1) 搜集公交线路名称。采用 Scrapy 框架,编写爬虫脚本,对公交信息网站的静态页面进行数据搜集。先确定城市定位,然后分析网站的 HTML 源码,进而逐一解析出指定城市内所有公交线路的名称。

2) 搜集公交线路详细信息。基于 1) 条中获取的公交线路名称列表,调用高德地图的开放服务,使用公交导航 API,逐个获取各路线的站点名称和站点的地理位置信息。

3) 数据入库存储。将 1)、2) 条获取的路线名称、站点、位置等详细信息归档到路线数据库中,构建公交线路查询数据表,使数据检索、更新和维护更加高效。

3 路线智能识别模块设计

路线智能识别模块融合智能识别算法、定时任务调度及数据库存储功能,以实现路线的高效匹配。

3.1 智能识别算法

路线智能识别算法流程如图 2 所示。首先读取公交线路数据表,构建智能识别字典;同时处理车辆监控数据,提取车辆轨迹序列。然后进行公交线路的站点匹配,并计算匹配率,以准确地识别出具体的公交线路。

智能识别算法的核心是构建一个精确的查询字典。该字典从数据库中提取所有公交线路及其站点的位置信息。在该字典中,站点名称为键,其位置信息及关联路线信息为值,以便进行匹配查询。某城市公交线路构建的查询字典结构示意图如图 3 所示,每个公交站点对应的值有站点位置(经纬度),站点位置用于识别轨迹是否途经公交站点。同时每个公交站点还对应了经停的相关路线,即经停该站点的所有

公交线路。

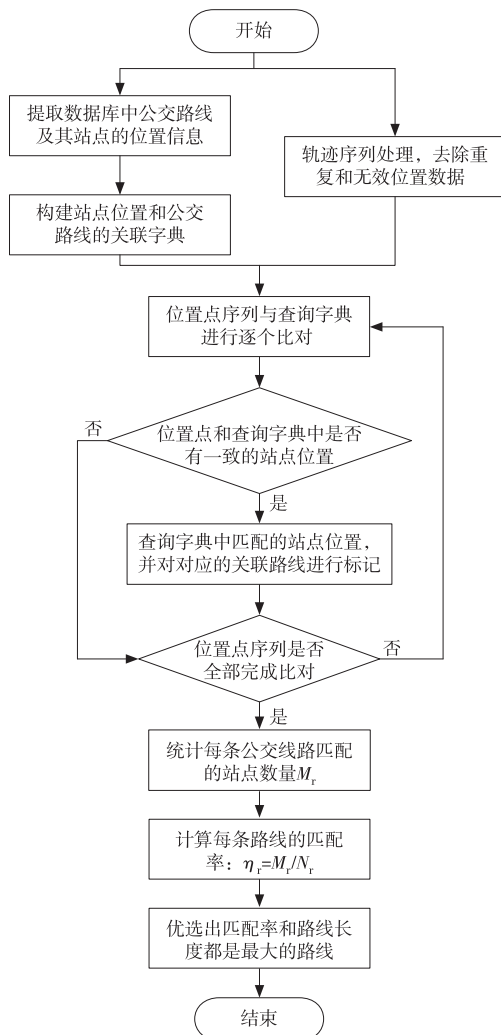


图 2 路线智能识别算法流程图

键	值	
站点名称	站点位置	经停的相关路线
文昌阁	(119.433 723, 32.396 828)	128 路、20 路、23 路、51 路(晚班车)...
何园	(119.449 76, 32.385 651)	旅游观光巴士, 古城巴士
...		

图 3 查询字典结构示意图

完成字典构建后,处理路线轨迹点序列,对监控平台采集(每隔 10 s 采集一个点)的原始轨迹点序列进行标准化处理,过滤掉重复或无效数据。将轨迹点与查询字典进行比对,识别所有经过的或附近(距离在 100 m 范围内的)的公交站点,再看这些与车辆轨迹经过的或附近的站点是否是各公交线路经停的站点,如是公交线路中的站点,则视为该公交线路的匹

配站点,包含匹配站点的公交线路均称为关联公交线路。统计所有关联公交线路和各公交线路的匹配站点数量。计算各关联公交线路的匹配率:

$$\eta_r = M_r / N_r \quad (1)$$

式中: N_r 为公交线路 r 的总站点数量; M_r 为公交线路 r 的匹配站点数量; η_r 为公交线路 r 的匹配率,且 $0 \leq \eta_r \leq 1$ 。

由于存在包含路线(例如某条较长的公交线路覆盖了另一条较短的路线,两条路线有重合部分),完成各路线的匹配率计算后,可能出现多条匹配路线。针对上述情况,从匹配率较高的多条路线中选择路线长度最大的路线:

$$\begin{aligned} \text{优选路线} &= \operatorname{argmax} P(r) \\ r &\in \{R | \max(\eta(r))\} \end{aligned} \quad (2)$$

式中: R 为所有匹配率较高的公交线路集合; $P(r)$ 为路线 r 的路线长度。

3.2 定时任务调度

采用后台定时任务调度工具,实现定时批量识别。任务设定在每日公交运营量最低的凌晨时段,最大程度上减少了对平台资源的占用。

任务首先是批量加载各城市的公交信息数据,随后根据车辆的历史轨迹确定目标城市,并执行识别算法,最终存储识别结果。

3.3 数据库存储

数据库包含两类数据表:公交线路数据表和识别结果数据表。

公交线路数据表按照不同城市存储的公交线路分别记录在独立的表中。表名为城市名称,表设计字段包括:路线 ID、路线名称、首站位置、末站位置、途经站点名称及其地理坐标。

识别结果数据表为定时任务执行完成后的匹配结果。表设计字段包括车辆识别码、匹配日期、匹配路线名称、匹配率、路线长度。考虑到表容量和检索效率,匹配结果数据表会根据容量周期性新建存储,时间段会标在表名中^[6]。

4 界面交互模块设计

界面交互模块包括请求接口和界面设计。请求接口包括识别路线查询、路况查询、坡度查询三类接

口,界面设计主要是用户图形界面设计。

4.1 请求接口设计

识别路线查询接口可查询指定车辆和时间的路线识别结果,接口通过查询识别结果数据表返回。识别路线查询接口提供了基于日期的筛选功能,用户可指定查询的起始日期和结束日期,获取特定时间段内的匹配结果。为便于前端处理大规模数据集,接口提供了分页功能查询。同时接口允许用户通过多种条件进行筛选,包括车辆识别码、车型、订单号等。用户可根据不同的需求进行精确的数据检索^[7-8]。

路况查询是对识别路线进一步的详情展示,路况查询接口返回的是路线长度、行驶轨迹以及途经站点信息,此接口响应 HTTP GET 请求,使用户能够获取目标公交线路的 GPS 坐标和途经站点数据,查询城市和公交线路名称作为必填参数。查询过程包括从数据库读取指定路线的站点信息,调用地图服务获取实时交通数据,以及综合站点和交通态势信息后返回给用户。

坡度查询是对识别路线的路线坡度展示,通过 HTTP GET 请求使用户能够查询城市内指定公交线路的坡度数据。为实现精确的坡度查询,接口设计的计算步骤如下:

- 1) 路线数据读取。读取数据库中的公交线路 GPS 位置序列。
- 2) 距离计算。计算连续 GPS 坐标点之间的水平距离。
- 3) 高程匹配。读取高程数据表,筛选出公交线路的位置范围,再使用 k-d 树数据结构将高程数据与 GPS 坐标匹配。
- 4) 梯度拟合。在指定水平距离间的点集上应用最小二乘法线性回归,计算坡度(梯度)。
- 5) 坡度输出。算法计算出路线各段的坡度后,通过与预定义的阈值比较,识别出显著的坡度区间,并输出最大坡度值、该坡度延续的距离以及对应的 GPS 坐标。

4.2 界面设计

界面采用 Vue.js 框架和 Element UI 库构建。设计的重点在于提供清晰的用户输入界面和高效的数据展示方式。界面包含多种 Element UI 组件,如输入

框、选择器和日期选择器,用于精确设置查询条件。以表格形式展现数据,包括车辆 VIN、车型、订单号等关键信息,其中匹配率字段通过颜色编码突出显示,以增强数据的可读性和直观性。界面底部的分页控制器进一步优化了大数据集的管理。此外,脚本部分通过 Vue 和 JavaScript 实现了与后端服务的交互逻辑,包括数据管理和事件处理,确保用户操作的流畅性和响应速度。图 4 为公交线路识别界面实现效果展示。



图 4 公交线路识别界面效果展示

5 系统功能测试验证

5.1 路线识别验证

为验证系统的路线识别功能准确性,从系统自动产生的识别结果中随机抽取一辆公交车的运行轨迹数据。基于 3.1 节中介绍的公交查询字典和比对计算,识别出该车辆运行轨迹经过的和附近的公交站点共 97 个。将这 97 个站点和各公交线路经停站点进行比对,统计得到各关联路线的匹配站点数量 M_r ,各关联路线的总站点数量 N_r 和路线长度 $P(r)$,以及由式(1)计算得到的各公交线路匹配率 η_r (表 1)。

表 1 某公交车轨迹与关联路线的匹配计算

关联路线	M_r	N_r	$\eta_r/\%$	$P(r)/\text{km}$
622 路	57	59	96.6	54.5
627 路	47	57	82.5	51.4
230 路	27	33	81.8	29.2
612 路	48	59	81.4	52.8

然后根据式(2),先筛选出匹配率较高的路线(本系统选用 85%作为匹配率较高的阈值),再选出路线长度最大的路线,最终得到最优匹配路线为 622 路,即该公交车的运行路线被识别为 622 路。

为更加直观地展示本案例,绘制了与该公交车轨迹匹配较高的几条路线的匹配示意图(图 5)。从图 5 中可以明显看出,该车轨迹与 622 路线匹配率最高,而 230 路线、627 路线均与 622 路线有相同路段,这也解释了不是该路线但匹配率仍较高的原因。此外,230 路线较短,仅覆盖 622 路线一半左右,但仍有较高匹配率,这也很好地说明了本系统算法需要结合匹配率和路线长度 2 个指标来确定识别路线的原因。

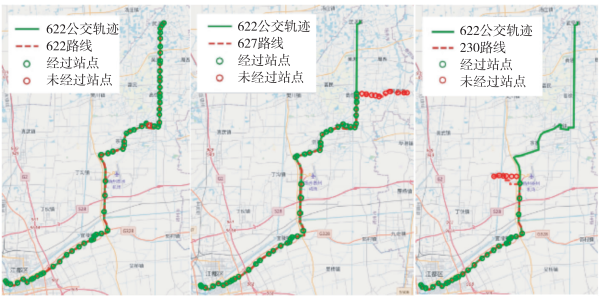


图 5 622 路公交轨迹匹配路线示意图

将系统自动识别的路线和人工分析结果进行对比,部分结果见表 2。从表 2 中可看出,系统智能识别的准确性很高。

表 2 部分路线识别准确性验证

车辆编号	路线	自动 匹配率/%	人工审核 匹配率/%	结果 一致性
1	622 路	97	97	一致
2	627 路	100	100	一致
3	6 路	100	100	一致
4	236 路	100	100	一致
5	67 路	87	87	一致
6	无识别	/	/	一致

5.2 路况详情验证

将系统显示的路线信息与实际路线信息进行对比,如图 6 和图 7 所示,结果表明两者的路线长度、地图轨迹和站点信息均保持一致。

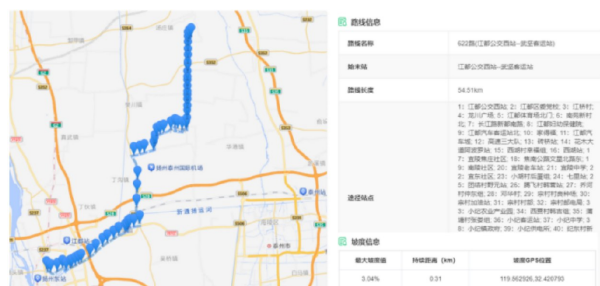


图6 系统展示的路况详情页面

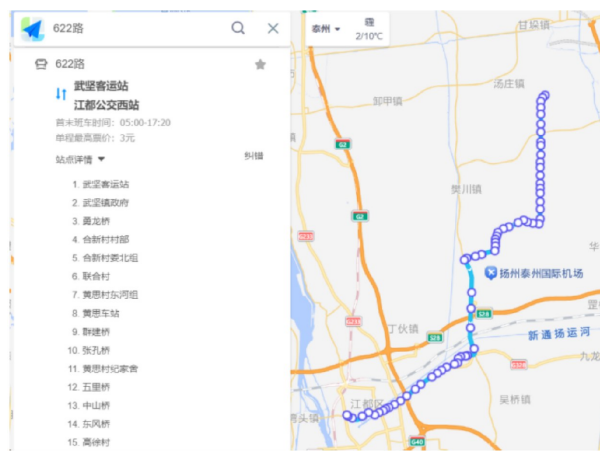


图7 实际路况详情(截自高德地图)

5.3 坡度查询验证

使用专用设备采集某路段实际坡度信息,与系统估算结果进行对比。由图8可以看出,估算坡度的趋势与实测结果基本一致,3%以上的坡度均能较为准确地识别^[9-11]。需要说明的是,本系统的坡度计算是基于云数据,准确度虽然达不到实地测量结果,但可以有效识别陡坡,且具有快速计算、低成本和良好的可扩展性等优势。

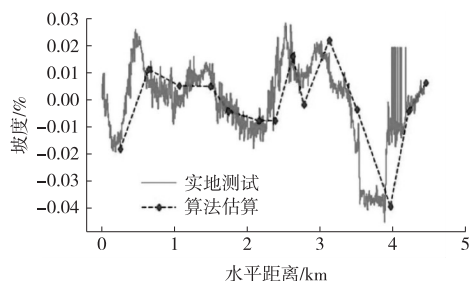


图8 坡度实测与系统估算对比

6 结束语

本文提出的公交路线智能识别系统能快速准确地收集和分析公交运营路线^[12]。通过模块化设计,系统提高了数据处理效率(相比于人工查看地图的传统方法),通过直观的用户界面,用户体验得到了显著提升。应用表明,系统在路线识别精确度、路况详情展示及坡度计算等方面均展现出良好的性能,符合系统的应用需求。

参考文献:

- [1] 邹渊,孙文景,张旭东,等. 智能网联汽车多域电子电气架构技术发展研究[J]. 汽车工程,2023,45(6):895-909.
- [2] 倪俊,姜旭,熊周兵,等. 无人车大数据与云控制技术综述[J]. 北京理工大学学报,2021,41(1):1-8.
- [3] 张鹏飞. 基于出租车历史轨迹挖掘的公交线路优化研究[D]. 长沙:湖南师范大学,2021.
- [4] 陈辉,蒋圭峰,姜桂圆,等. 基于海量公交轨迹数据挖掘的地图匹配算法[J]. 计算机应用,2018,38(7):1923-1928.
- [5] 刘瑶秋,季金强,赵苗苗,等. 整车 CAN 总线数据测试系统的研究[J]. 客车技术与研究,2020,42(6):36-39.
- [6] 王震坡,袁昌贵,李晓宇. 新能源汽车动力电池安全管理技术挑战与发展趋势分析[J]. 汽车工程,2020,42(12):1606-1620.
- [7] 暨南大学. 一种基于便携式 GPS 的道路坡度估算方法:202011006738.0[P]. 2020-11-27.
- [8] 杨峰,杜翠凤,王新宇. 基于车联网的交通预警系统设计[J]. 物联网技术,2024,14(1):22-25.
- [9] 王泽天,高岭,高全力. 基于改进马尔科夫链的移动轨迹预测方法[J]. 西安工程大学学报,2020,34(2):97-102.
- [10] 李尊,孔盼,史艳龙. 重型商用车质量和道路坡度估算方法[J]. 重型汽车,2020(3):19-20.
- [11] 郝胜强,罗培培,席军强. 基于稳态卡尔曼滤波的车辆质量与道路坡度估计[J]. 汽车工程,2018,40(9):1062-1067.
- [12] 周桂良. 车联网环境下私家车拼车优先导向的出行效能提升研究[D]. 镇江:江苏大学,2019.