工程与管理科学
Engineering and Management Science
第 7 卷◆第 7 期◆版本 1.0◆2025 年
文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

大数据实时计算的车辆套牌分析研究

肖潇　段旻（通讯作者）
重庆财经职业学院　重庆永川　402160

DOI:10.12238/ems.v7i7.14267

[摘　要] 通过对城市道路卡口监测数据的分析，可以快速有效地发现套牌车辆，本文展示了通过使用大数据工具 Kafka 组件作为大数据实时计算数据源，Spark Streaming 组件作为流式计算引擎，Redis 组件作为快速分析的存储工具，搭建大数据平台进行套牌车辆分析的方法。引入多台硬件计算资源协同处理大规模数据下的套牌车检测，显著提高了计算性能，进一步提高了套牌车辆被检测和识别的实时性和准确性。

[关键词] 大数据；套牌车辆；Kafka；Spark Streaming；Redis

[中图分类号] TP3　　[文献标识码] A

Big Data in Real Time Computing Vehicle Set Analysis Research.Computer Engineering and Applications

Xiao Xiao　Duan Min (corresponding author)
Chongqing Finance and Economics Vocational College Chongqing Yongchuan　402160

[Abstract] By analyzing the monitoring data from road bayonet in the city, we can find deck vehicles efficiently.This paper shows the methods of analysing deck vehicles by using different components to set up a big data platform.We use the Kafka components as a real-time computing data source and the Spark Streaming components as a stream computing engine, as well as the Redis components as a store for rapid analysis respectively.Furthermore, we have introduced many sets of hardware computing resources to test data of deck vehicles under mass data processings, which significantly improves the computing performance and possibility of the deck vehicles being detected and identified.

[Key words] Big date　Deck vehicles　Kafka　Spark Streaming　Redis

1 引言

套牌车（俗称克隆车）是指未通过正规渠道办理合法手续，通过伪造或非法套用他人车辆号牌上路行驶的车辆。这是一种侵犯真实车主权益的交通违法行为。2024 年，全国机动车保有量达 4.53 亿辆（其中汽车 3.53 亿辆），机动车驾驶人达 5.42 亿人（其中汽车驾驶人 5.06 亿人）。同年，新注册登记机动车 3583 万辆，新领证驾驶人 2226 万人。随着机动车保有量的快速增长，套牌车辆的数量也相应攀升。由于套牌车隐蔽性强、识别难度高，部分车辆甚至长期使用未被查处，严重扰乱了道路交通安全秩序。此类车辆常伴随肆意闯红灯、超速行驶、违规变道、乱停乱放等危险驾驶行为，给公众出行安全带来极大隐患。同时，套牌行为也极大地增加了交通肇事逃逸案件的侦破难度[1]。

文献[1]、[2]提出基于 MapReduce 的交通流大数据套牌车分析计算框架，虽然该框架有解决交通流大数据下的性能瓶颈的优点[1][2],但是基于 HDFS 分布式文件系统上的计算框架 MapReduce 不能随即读取，故它不适应实时大数据应用的需求。

文献[3]中大数据云计算技术在缉查布控系统的应用中，采用 Kafka 作为实时数据接入组件，Spark Streaming 作为实时计算引擎，HBase 作数据存储[3][5]，但是由于 HBase 仍然采用的是磁盘 IO，存取速度满足不了实时应用的需求，同时 HBase 只能做简单的 Key value 查询，无法进行复杂的 sql 统计。鉴于上述原因，本文将 Hbase 换为读写速度更快，性能更高的内存数据库 Redis 来提高数据存取速度，从而使得数据分析更实时。

2 技术方案

通过分析道路卡口摄像机拍摄的车牌号、车身颜色、车标和位置等信息，从相同车牌在外观（车型、颜色）上是否一致以及相邻两条记录出现的时间和位置是否合理两个方面判定和分析各监控点的车辆信息，实现套牌车的识别。

2.1 实现原理

基于城市道路卡口设备的车辆特征识别与套牌分析系统，其技术实现路径可优化为以下三个步骤：

（1）在城市道路关键节点部署高清卡口设备，通过图像识别算法实时提取过往车辆的三维特征数据（含车牌号码、

车身色值、车标类型），同步采集车辆通过时刻、设备编号等时空信息。所有结构化数据经边缘计算节点预处理后，通过专网实时回传至大数据平台，构建动态更新的车辆特征数据库，为后续分析提供基础数据支撑。

（2）当新数据流入时，首先与历史库进行全量特征匹配。若发现相同车牌但特征组合（颜色+车标）不一致，直接触发套牌预警；若特征组合完全一致，则启动时空矛盾校验，计算当前卡口与前序卡口的时间差，结合 GIS 系统获取的路径距离，通过公式"时间差×路段限速＜实际距离"验证车辆移动合理性，违反物理规律的判定为套牌嫌疑。

（3）对首次出现的车辆记录建立初始档案；对重复出现的车辆，根据比对结果分类存储——确认正常的记录转入历史库形成轨迹链，触发预警的记录则进入待核查库并关联时空轨迹。系统支持按车牌号查询实时位置、按时间段回溯历史轨迹，为执法部门提供可视化轨迹分析工具。

2.2 大数据平台

由于现如今数据时效性随时间衰减，要求事件触发后即时处理而非批量积压。为应对海量实时流数据，需构建低延迟、高扩展、强容错的流式处理架构，确保每个事件在产生瞬间即完成分析决策，最大限度释放数据即时价值。

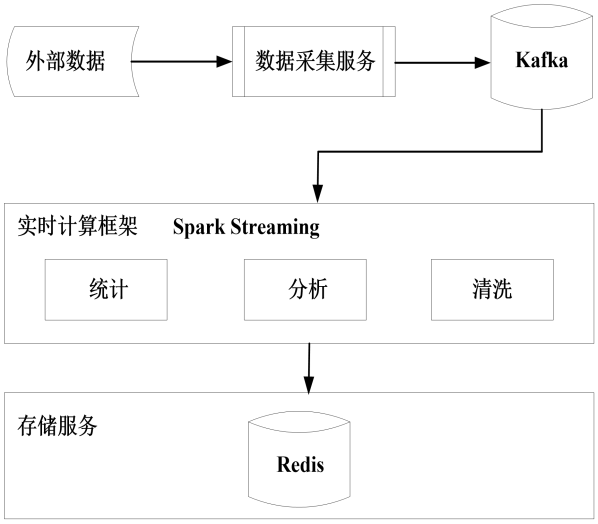

图 1　实时计算框架组件架构和数据流图

因此为了满足套牌车辆实时识别与报警的需求，结合当前比较流行的实时大数据处理技术，本文使用以下产品作为首选方案：

（1）利用 Kafka 作为实时计算的数据源[6]。Kafka 是一款分布式高吞吐的发布订阅消息系统[7]，即使是非常普通的硬件 Kafka 也支持每秒数十万的消息。

（2）利用 Spark Streaming 作为计算引擎。Spark Streaming 是构建在 Spark 上的实时计算框架[8]，扩展了 Spark 处理大规模流式数据的能力，能运行在超过 100 个节点上，并达到秒级延迟，具有高效和容错的特性。Spark Streaming 能对从数据源 Kafka 接入的实时数据做快速高效的处理。并且还提供了一套高级的 API，通过简单的调用和组合就能实现强大的业务逻辑功能[9]。

（3）利用 Redis 作为中间结果（或状态）存储引擎[10]。Redis 是一个开源的使用 ANSI C 语言编写、支持网络、可基于内存亦可持久化的日志型、Key-Value 数据库，并提供多种语言的 API[11][12]。Redis 能读的速度是 110000 次/s，写的速度是 81000 次/s，性能极高。

大数据平台实时计算框架组件架构和数据流图如图 1 所示。

2.3 大数据平台实现原理

本文使用的大数据技术解决车辆套牌分析问题的实现原理是：

（1）数据采集服务进行外部实时数据（车辆卡口数据）的接入和初步清洗后，以生产者的角色将数据写入 Kafka 组件。接入方式有多种，包括远程过程调用（如 Web Service），传统关系型数据库连接，文件读取等。

（2）Kafka 会将生产者写入的数据有序地存放于消息队列中供消费者读取。Kafka 以 Topic 来进行消息管理，每个 Topic 包含多个 Partition，每一个 Partition 对应数条过车数据，在本文中，设置 Topic 为“过车数据”，按照过车数据中的车牌号码的进行处理，Topic 和 Partition 每个服务器上的消息存储位置。前端数据采集服务作为生产者按照 Partition 规则不断生产过车文本数据，并发布到每一台服务器上（Broker），消费者订阅 Topic，并从 Broker 拉取数据，从而消费这些已发布的消息[3]。

（3）流式计算框架（Spark Streaming）以消费者的角色从 Kafka 中获取数据，进行统计、分析和数据清洗等工作，在套牌车辆分析时主要实现嫌疑车辆比对、区间测速、套牌车分析等实时业务分析。从 Kafka 消息队列中获取的实时过车文本数据流。然后将流式计算分解成一系列短小的批处理作业，批处理引擎是 Spark，即 Spark Streaming

在接收到实时数据流后，将数据流以时间片（秒级）为单位分成一段一段的数据（Discretized Stream），然后以类似批处理的方式处理每个时间片数据，每一段数据都转换成 Spark 中的 RDD（Resilient Distributed Dataset，弹性分布式数据集），然后将 Spark Streaming 中对 Stream 的 Transformation 操作变为针对 Spark 中对 RDD 的 Transformation 操作[3]，分析之后将处理的结果保存于外部存储系统 Redis 供二次查询或分析。

3. 套牌分析算法

3.1 算法描述

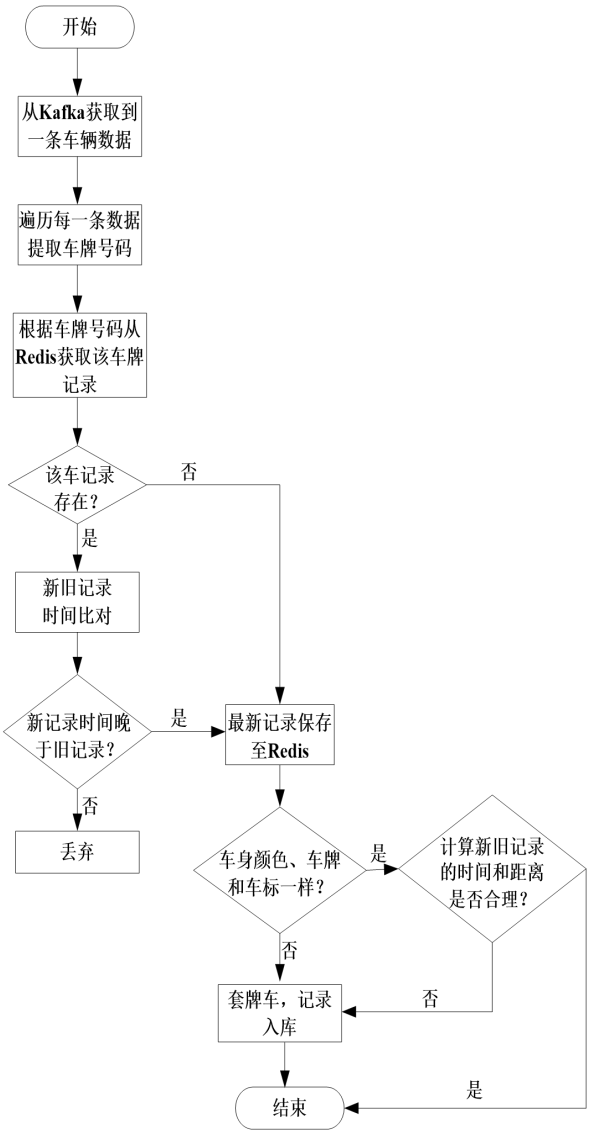

图 2　套牌车辆算法流程图

本文采用的套牌车辆分析算法为：

（1）在内存（或内存数据库中）维护一个状态区（内存缓冲区），用来保存车辆的最新记录；

（2）当有更新的记录流入，判断状态区是否存在该车对应的记录，若不存在，则将记录添加到状态去；若存在，则在状态去中保留两者中时间较新的记录；

（3）将新旧车辆记录做属性比对（车身颜色、车牌颜色、车标），若有不同，则判定其是套牌车；若都相同，则判定新旧记录的时间和距离是否合理，若不合理，则判定其是套牌车，否则不是套牌车。

3.2 算法处理流程

本文采用的套牌车辆分析算法流程如图 2 所示。

4. 实验分析

由于大数据套牌车辆分析要求处理海量的数据，处理的效率要足够高、数据要足够准确，因此，本文根据上述算法，使用广东省某市的真实交通流数据信息进行实验。

4.1 实验环境

为保证实验的顺利完成，本文搭建的实验平台，硬件条件如下表 1 所示，软件条件如下表 2 所，测试条件如下表 3 所示。

表 1　硬件环境需求表

3 台物理机	规格参数
CPU	2 颗，共 12 核 24 线程（Intel（R） Xeon（R） CPU E5-2620 0 @ 2.00GHz）
内存	20G，共 20G * 3=60G
硬盘	2T（SAS）+4T（SATA）
网络	1000Mb/s

表 2　软件环境

组件名称	版本	说明
Kafka	2.11-0.10.1.1	在三台机器上安装好集群，并创建 Topic
生产者程序	1.0	向 Kafka 生产定量数据，能够线性并发，以便进行压力测试
Spark Streaming	1.0	能够从 Kafka 消费数据，并实现套牌分析算法
Redis	3.2.8	存储状态信息、中间结果信息等

表 3　测试条件

参数名称	参数值
Spark 各节点内存	20G
测试总数据量	25000000 条
平均每秒产生数据量	约 20000 条
批次间隔	2s

4.2 实验样本数据

本文实验数据样本如下表 4 所示。

表 4　实验数据样本

car_num	car_date	car_color	car_num_color	car_type	dev_chn_name

粤 B**6C8	2017-04-01 00：00：01	18	9	0	西南二车道 前排
粤 Y**991	2017-04-01 00：00：01	17	9	0	东向二车道 前排
赣 K**361	2017-04-01 00：00：02	15	5	0	东北一车道 前排
粤 V**168	2017-04-01 00：00：04	15	9	1	东北一车道 后排
粤 V**667	2017-04-01 00：00：05	18	9	0	东向二车道 后排

4.3 实验性能分析

在本文搭建的实验平台下，采用 2 秒的批次间隔进行数据获取与分析，实验效果图如下图 3 所示。

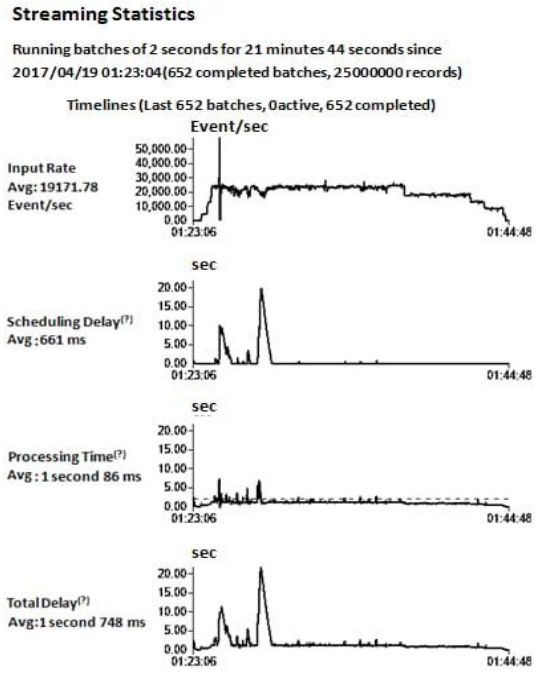

图 3　实验效果

由上图得出，本文测试实验 2500w 数据消费情况，平均延迟仅为 1.748s/批次，已经基本稳定。具体实验性能评测如下表 5 所示。

表 5　性能评测

性能指标	值
平均等待时间	661ms
每个批次处理时间	1.086s
平均延迟	1.748s

总结：一般应用下（具有状态管理功能），在能容忍 5s 延迟的情况下，预估应用能处理 1.2w/s 的数据。

不难看得出，本实验采用的平台完成了使用传统单台机器无法完成的任务，达到了实时处理海量数据的目的。

5 结束语

针对城市交通卡口产生的海量异构数据，传统批处理模式面临实时性不足与算力瓶颈[1]。本文构建的流式计算架构通过实时特征解析引擎，融合车牌号码、车身色值等多维度信息开展动态比对，结合时空轨迹校验模型，可实现毫秒级套牌嫌疑车辆识别。该方案突破传统算法的时效性限制，显著提升复杂场景下的分析精度，且具备弹性扩展能力，能够快速适配智慧交通实战场景的部署需求。

[参考文献]

[1]朱萍，马韵洁.基于大数据技术快速分析套牌撤方法[J].电脑知识与技术，2015，11（34）：20-23

[2]王涛，王顺，沈益民.交通流大数据中的套牌车并行检测算法[J].湖北工程学院学报，2014，34（6）：29-32

[3]徐晓东，孔晨晨，席正祺.大数据云计算技术在全国机动车缉查布控系统中的应用[J].智能交通，2015，1：87-91

[4]卢晓春，周欣，蒋欣荣，潘薇，王峰.基于网格化监控的套牌车检测系统[J].计算机应用，2009，29（10）：2847-2848

[5]周建宁，徐晓东，蔡岗.流式计算在交通管理中应用研究[J].智能交通，2016，1：70-75

[6]孙元浩.如何构建安全的 Kafka 集群[J].电信网技术，2015，8：10-14

[7]杨冬晖.一种分布式消息队列的可靠性研究[J].电脑知识与技术，2015，11（21）：75-79

[8]李祥池.基于 ELK 和 Spark Streaming 的日志分析系统设计与实现[J].电子科学技术，2015，02（06）：674-678

[9]陈丽，王锐.基于 Spark Streaming 流技术的机动车缉查布控系统设计[J].顺德职业技术学院学报.2016，4：10-15

[10]吴霖，刘振宇，李佳.Redis 在订阅推送系统中的应用[J].电脑知识与技术，2015，11（7）：292-294

[11]柳皓亮，王丽，周阳辰.Redis 集群性能测试分析[J].微型机与应用，2016，35（10）：70-71

[12]刘俊龙，刘光明，张黛，喻杰.基于 Redis 的海量互联网小文件实时存储与索引策略研究[J].计算机研究与发展.2015，5：148-154

[13]闫密巧，王占宏，王志宇.基于 Redis 的海量轨迹数据存储模型研究[J].微型电脑应用.2017，4：9-11