

智能交通管控大数据平台设计开发与应用

梁达

(上海电科智能系统股份有限公司, 上海, 200063)

摘要: 随着交通供需矛盾日增、交管执勤警力不足问题突出, 准确掌握道路交通拥堵, 主动管控道路交通流, 快速智能处理交通事件, 分析挖掘海量交通数据, 为交通管理提供智能支持。本文设计建设的交管平台基于 Hadoop 技术、机器学习, 实现交管大数据的接入、存储计算、语义化, 场景化, 高清可视化, 通过科技赋能, 为交通管理提供有力支撑。

关键词: 交管大数据; 机器学习; 语义分析; 交管业务场景

Development and Application of Big Data Platform in intelligent traffic control

Liang Da

(Shanghai Seari Intelligent System CO., LTD, Shanghai, 200063)

Abstract: With the increasing contradiction between traffic supply and demand, the problem of insufficient police force on the traffic control, accurate grasp of road traffic congestion, active control of road traffic flow, rapid and intelligent handling of traffic incidents, analysis and mining of massive traffic data, to provide intelligent support for traffic management. Based on Hadoop technology and machine learning, the communication platform designed and built in this paper provides strong support for traffic management through the access, storage.

Key words: Traffic Big Data; Machine Learning; Semantic Analysis; Traffic Business Scenario

1 引言

随着城市的快速发展车辆的迅速增加, 交通需求发展大于对道路资源、管理资源所能提供的交通供给, 交通供需矛盾日增, 公安交警执勤警力不足的问题日渐突出。同时物联网的发展, 交通的数据种类和数据指数级膨胀, 形成海量交通信息。传统单一的交通技术管理面对海量数据存储管理分析利用存在不足, 作为公安交通管理工作, 亟需依靠信息技术改革创新加强。公安交管部门提出运用云计算、大数据、物联网、互联网、人工智能等信息技术手段, 以信息主导警务, 以科技引领勤务, 通过技术向智慧交管赋能。

技术赋能的关键, 是根据海量交通大数据和智慧交通管控的新需求, 打破传统技术单一处理能力的限制, 建设智能大数据交管平台, 通过云计算的资源池存储能力, 实现对视频图

像、过车号牌、GPS、交通事件等数据分布式存储；通过大数据算力，实现对交通业务、车技战、图像语义化的计算处理，通过机器学习深度学习算法模型，实现交管业务智能化处理。

2 需求分析

交通管控平台作为一个综合集成管理与指挥应用的大平台，需综合集成图像设备、信号控制设备、检测设备、诱导设备的智能设备类信息，通讯设备、定位设备、单兵设备类的移动设备信息，路况信息、事件信息、管控信息、路面信息类的交通信息、交通运输、气象行业、规划建设信息类的行业信息。基于集成各类系统静态数据，支撑交警的各项业务应用。通过对道路交通研判、事故预警、管控优化的精准分析，警情处置、特勤保障、应急调度指挥的扁平化智能化，实现对城市交通主动、精确、合成、实效的管理。

为实现上述业务的需求，需充分利用人工智能、大数据、云计算即 ABC 技术，建设智能交通管控平台，为交通管理提供赋能。

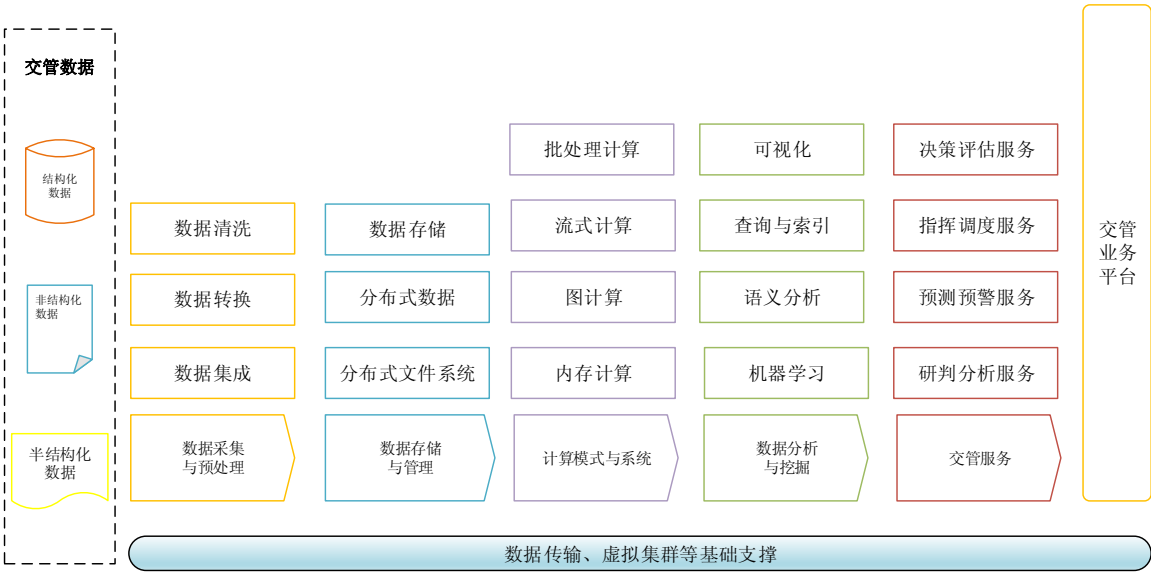


图 1 交管平台的技术需求

- (1) 支持云计算：利用计算虚拟化、存储虚拟化、网络虚拟化、安全管理等技术通过大量服务器虚拟化，形成虚拟资源池，使用户能够按需获取计算资源、存储资源、网络资源的技术服务即基础即设施服务。具有多租户、动态可扩展、按需部署、灵活性高可靠性高、性价比高、可扩展性的特点，为业务应用提供统一的基础支撑；
- (2) 支持大数据存储：采用分布式存储技术，提供一个高可扩展和高吞吐率的海量数据存储服务；
- (3) 支持大数据实时查询：针对大规模数据存储管理和查询，需支持分布式 NoSQL 列式数据库，实现 TB、PB 级别的海量数据实时或准实时的高速查询访问。平台同时还支持实现面向结构化的数据快速查询；

- (4) 支持批处理计算: 针对海量数据批量处理要求, 具备并行执行大规模数据处理任务, 基于磁盘或内存上用于大规模数据集的并行运算, 支持 MapReduce、Spark;
- (5) 支持流式计算: 支持来自不同数据源、连续到达在时间分布和数量无限的动态数据集合体的流数据, 通过实时分解成一系列短小的批处理任务完成实时计算, 支持 spark streaming 或 storm;
- (6) 支持大数据搜索: 应支持基于 Elasticsearch、Solr 全文搜索引擎实现数据搜索;
- (7) 支持大数据分析挖掘: 利用分布式并行编程模型和计算矿建, 结合机器学习的监督学习, 半监督学习, 无监督学习和强化学习和数据挖掘算法, 实现海量数据的处理分析。支持回归分析、聚类分析、分类分析、关联分析、空间聚类分析、时空分析、神经网络、深度学习等算法模型;
- (8) 支持大数据展示: 对分析结果进行可视化呈现, 帮助用户更好理解数据、分析数据;

3 总体设计

智能交管大数据平台分基础服务层、平台服务层、数据服务层、交管服务层和用户端, 总体框架如下图:

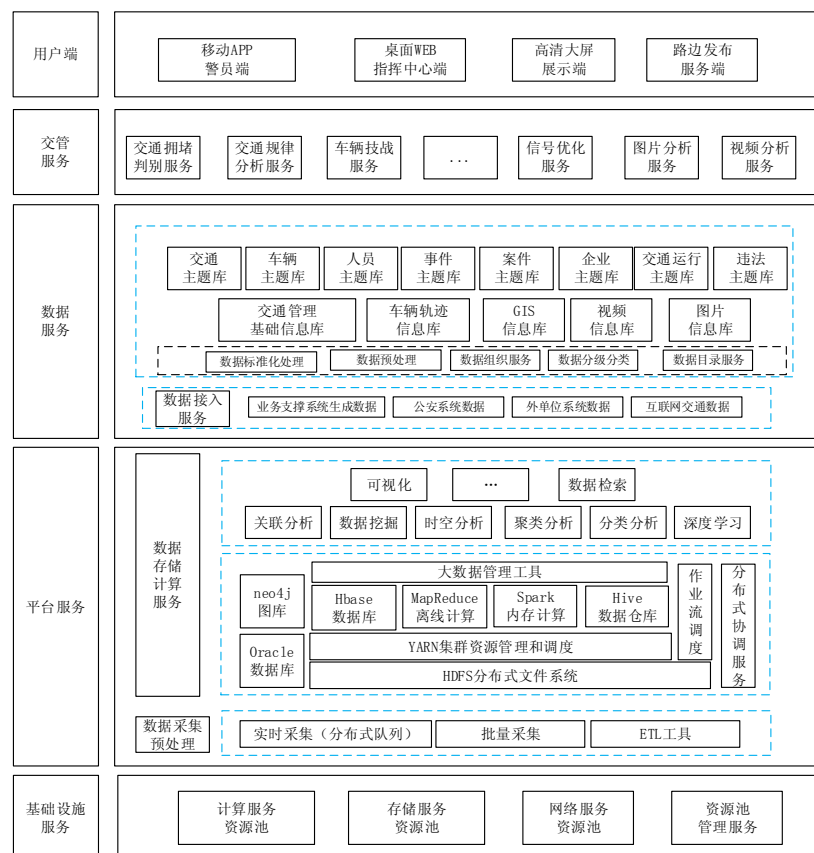


图2 智能交管平台总体架构如图

- (1) 基础服务层：基于云技术服务将物理资源进行虚拟化，所有资源整合后在逻辑上以单一整体资源池的形式呈现，资源将可进行动态扩展和配置，对外提供弹性计算服务；
- (2) 平台服务层：平台服务提供对异构交管数据采集汇聚接入，利用 HDFS 分布式文件系统、分布式面向列库 Hbase 及业务关系数据进行存储、应用 MapReduce、Spark 计算框架，Elasticsearch、solr 全文检索引擎，实现大数据高效计算及检索；运用机器学习和 GPU 资源对数据的分析挖掘，作为实现交管数据的关联、融合分析、图像识别、语言识别等智能分析的基础支持；
- (3) 数据服务层：提供各类资源数据接入、处理、管理及服务，建立道路、车辆、人员、时间、案件、交通违法、交通规律等交管主题库，完成数据资源整合、共享、分析及展示；
- (4) 业务服务层：基于大数据提供各类算法构建的交通路况判别模型、信号优化模型、交通规律分析模型、车辆技战模型、视频检测分析模型、语音分析模型等业务应用模型，为应用层提供业务支撑；
- (5) 用户服务层：通过桌面 WEB 端、移动 APP 端、大屏高清展示端多种方式为警员、指挥中心、道路参与者提供交通管理及服务；

4 功能设计

交管大数据平台从道路感知设备、警员移动设备、交通及相关信息、道路地理信息等收集多源异构数据，进行清洗、关联、融合、管理等预处理，经过交通管控算法模型计算分析，为事件、勤务、信控、诱导、应急、交通规律分析等业务提供智能识别、主动处置、量化评估闭环式的业务应用服务。

从功能角度出发，交通管控平台涉及众多的智能交通业务平台，对应不同层次、维度、粒度的各类型数据接入，设计建立统一的交管数据标准资源管理，实现各交通系统的数据融合共享建立统一的交管数据中台；面对交管数据多源异构海量的特性，设计建立数据快速存储、计算能力；为适应前端业务需求快速响应，提供交通规律分析模型、信号优化、视频分析模型、语音分析模型、仿真模型等，作为业务中台支持；为满足多用户应用需求，构建移动 APP 端、桌面 WEB 端、大屏高清端等多样化、场景化的前端应用支撑。

4.1 数据接入及管理

交管数据接入及管理主要分为数据接入、数据处理、数据管理服务三部分。

(1) 交管数据接入分为信息采集接入、数据汇聚接入、Kafka 实时消息处理接入三种方式。

- 信息采集接入数据主要为卡口、电警、信号、诱导等视频图像、交通流、违法、警员警车 GPS 位置等感知设备和交警移动端设备的信息，基于与设备通信协议建立通信接口的方式采集设备的原始数据；

- 数据汇聚接入数据主要为违法、事件、警情、勤务、道路基础、地图、天气等信息基于外部系统，通过第三方 ETL 实现对外部系统数据库数据的抽取；
- Kafka 实时消息处理系统的任务是将由信息采集系统、信息汇聚系统所采集的实时数据分发给上层数据处理及应用。

(2) 数据处理：实现数据标准化处理、清洗、脱敏等处理。

- 数据标准化：根据数据字典的数据标准化定义，进行协议类型转换，将原始协议数据格式统一转换为交通信息资源库标准化数据处理单元格式；
- 数据清洗：根据业务规则定义，对数据进行去重处理、一致性检测、有效性检查、完整性检查、业务规则检测等处理；
- 数据脱敏：敏感信息进行数据变形处理，实现敏感隐私数据的保护；

(3) 数据管理：实现元数据管理、数据质量管理、数据共享交互等管理服务；

- 元数据管理：对整个交管信息资源库的元数据信息进行统一管理，包括数据目录管理、数据字典管理、数据版本管理、数据编码管理、GIS 编码管理；
- 数据质量管理：通过模型对交管数据质量管理，对数据质量实时监控，质量报警，质量分析；
- 数据共享交互服务：与交通管理内部系统、与大公安业务系统、与外部单位系统制定共享数据的相关统一接口规范，实现各系统间隔的数据接入传输与共享交换服务。

4.2 数据存储、计算

平台实现大数据基于 Hadoop 架构，存储与计算主要组件 HDFS、Hbase、Spark 以及一些计算框架和组件。

(1) Hadoop 是一个由 Apache 基金会所开发的分布式系统基础架构。充分利用集群的威力进行高速运算和存储。Hadoop 的框架最核心的设计就是：HDFS 和 MapReduce。HDFS 分布式文件系统为海量的数据提供了存储，具有高容错性，可部署在低廉的硬件上，高吞吐量适合访问有着超大数据集的应用程序。MapReduce 则为海量的数据提供计算，是面向大数据并行处理的计算模型、框架和平台[4]。

(2) HBase (Hadoop Database)，是一个高可靠性、高性能、面向列、可伸缩的分布式存储系统，利用 HBase 技术可在廉价 PC Server 上搭建起大规模结构化存储集群[4]，Pig 和 Hive 为 HBase 提供了高层语言支持，支撑基于 HBase 数据统计处理，Sqoop 为 HBase 提供了方便的 RDBMS 数据导入支持抽取传统数据库数据[4]。

(3) Apache Spark 是一种通用并行计算框架，与 Hadoop 相似的开源集群计算环境，但是两者之间还存在一些不同之处，Spark 启用了内存分布数据集，Job 中间产生结果保存在内存中不再读写 HDFS，除了能够提供交互式查询外，它还可以优化迭代工作负载，能更好适用于数据挖掘、机器学习的模型算法[4]。

4.3 数据挖掘分析

(1) 数据分析：交管大数据分析挖掘技术方法有数据检索、聚类分析、分类分析、关联分析、地理编码等技术应用[1]。

- 聚类分析：聚类分析指将物理或抽象对象的集合分组为由类似的对象组成的多个类的分析过程[4]。在交管业务方面，通过聚类分析从海量交通数据中提取特征，实现道路交通事件影响等级标定、交通状态评价等级标定等交通对象分组应用，为监督学习算法提供了必要的先验信息；
- 分类：对海量数据，根据内在数据特征划分成若干类别或组别。在交管业务方面，通过 K-NN、支持向量机（SVM）、神经网络等有监督学习算法，实现交通流数据修复，交通状态判别及预测等业务处理；
- 关联分析：是从大量数据中发现项集之间有趣的关联和相关联系，在交管业务方面通过关联分析，实现交通拥堵成因、交通道路事故黑点、交通对象对环境影响等业务分析；
- 地理编码：将交管业务中道路、设施设备、行政管辖等交通对象地址信息建立空间坐标关系的过程，在交管业务方面，通过对交通对象的地理编码，制定出便于交通信息交互，精确表达交通信息地理编码。

(2) 数据语义分析

基于多源多维异构数据采集过程获得的源数据，通过语义化处理，形成高价值的数据信息，为交管业务应用、情报关系挖掘分析等提供智能应用。

道路视频、电警、卡口采集的视频非结构化图像数据，基于深度学习算法，对采集图像进行结构化处理，在道路环境下对车牌、车头、车尾、非机动车、行人、人脸等道路交通信息进行特征提取，对车牌号码、车辆颜色、车标、车型、人脸等车辆特征信息检测，通过图像跟踪和行为分析技术实现道路交通流信息检测、交通事件检测、违法图像二次分析等图像语义分析处理，为交警在交通流管理、道路事件预警、非现场执法等各业务管理提供智能感知应用。

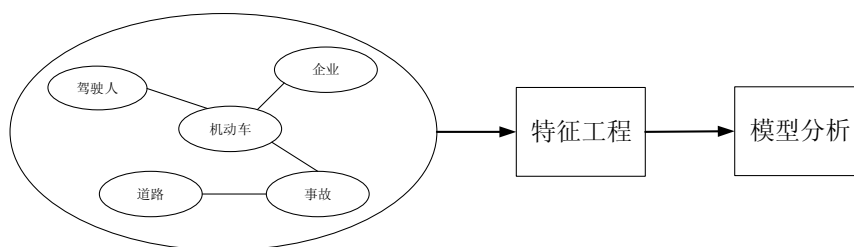


图3 语义化分析图

语义分析更进一阶认知应用，在交管主题库的基础上，依据智能交通系统体系框架和交通信息资源核心元数据，基于交通事件、交通地理、交通工具、交通指标、交通设

施、交通角色、交通信息、交通违法和交通相关信息领域，围绕车辆、驾驶人、事故、违法、道路、设施设备等要素，构建交管场景应用级的交管语义网即交管知识图谱。通过图计算，为交警在交通事故违法关系分析、交通事件处置、应急处理等业务，提供关系要素间的挖掘分析专家推理更深度的业务认知应用[2]。

5 交警业务综合场景智能应用与数据可视化

围绕交通实时指挥管控需求，平台利用场景化技术在警情事件处置、勤务保障线路、勤务管理、应急指挥、诱导、信控等方面，基于业务与交通对象关联，细分每个业务环节需求，建立分析模型，通过实时智能计算，为交警指挥调度、情报分析提供智能化、可视化管理支持。

5.1 交警事件处置场景智能化应用

道路事件处置场景，平台实现事件自动定位与派警、自动跟踪警员位置、主动收集现场信息，为处置各方建立沟通渠道，联动专业系统，实现施救车辆保障、上下游信号联动、周边警力联动等功能，为处置各方提供处置过程的关键信息，主动推送周边交通信息，实现信息查询的指令式交互，为实时指挥主动提供联动决策环境。



图 4 事件处置

事件处置完成后，通过交通关系图谱还原交通事件真相，应用图数据库技术，围绕机动车、驾驶人、事故、道路、企业等多方要素及其关联关系，形成交通要素知识图谱，挖掘交通要素间的潜在关系，分析交通事故责任调查，隐患及异常分析情报信息。

5.2 大数据研判展示

交管大数据分析，主要围绕车辆技战、警情分析、交通态势分析、违法分析、事故分析等信息情报分析，并通过高分显示与指挥调度结合起来，满足交通指挥、情报研判、决策的需要。

下图对警情从大队接警、警情道路分布、警员分布及警情趋势微观宏观两个方面对警情信息分析展示。



图 5 警情分析可视化

6 结束语

平台在大数据、人工智能、云计算技术基础上，打破传统交管平台竖井集成方式，在数据方面实现交管行业数据、行业相关数据、互联网数据的融合；在业务方面实现从功能化向场景化转变，从流程处置化向智能决策化转变，通过后台业务算法模型实时迭代计算，实现业务管理决策的主动辅助，通过多手段、多模式的情报推送，实现交通指挥的扁平、实时、高效协同，为智能交通管理实现技术赋能。

参考文献

- [1] 何承, 朱杨勇 城市交通大数据技: 上海科学技术出版社
- [2] 张滔, 凌萍 智慧交通大数据平台设计开发及应用: 第九届中国智能交通年会大会论文集
- [3] 赛文交通网 https://www.sohu.com/a/207048231_389742 智慧交管平台
- [4] 百度百科 <https://baike.baidu.com/>