

文章编号: 1005-8451 (2015) 11-0042-05

基于大数据分析的铁路自动售检票 监控系统研究

王 成^{1, 2}, 史天运¹

(1. 中国铁道科学研究院 电子计算技术研究所, 北京 100081)

2. 中国铁道科学研究院, 北京 100081)

摘要: 本文提出铁路自动售检票监控系统总体框架由基础层、管理层、分析层和应用层组成。利用多维数据分析、分布式文件系统存储和MapReduce计算、复杂事件处理、数据挖掘等技术, 实现对铁路自动售检票系统的故障预警和故障率分析、设备利用率分析、业务优化分析以及OD热点分析、异常旅客识别、设备易用性分析等以旅客行为分析为基础的增值业务, 为铁路自动售检票系统的未来发展提供一种新思路。

关键词: 信息技术; 监控; 大数据分析; 铁路自动售检票系统; 复杂事件处理

中图分类号: U293.22 : TP39 **文献标识码:** A

Railway Automatic Ticketing and Gate Monitoring System based on big data analysis

WANG Cheng^{1,2}, SHI Tianyun¹

(1. Institute of Computing Technologies, China Academy of Railway Sciences, Beijing 100081, China;

2. China Academy of Railway Sciences, Beijing 100081, China)

Abstract: This article proposed the general frame of Railway Automatic Ticketing and Gate Monitoring System(RATGS). The System was consisted of 4 layers, which were the infrastructure layer, the management layer, the analysis layer and application layer. The System was introduced technologies such as multidimensional data analysis, the distributed file system storage MapReduce, Complex Event Processing(CEP), data mining and etc., to implement the value added services based on passenger behavior analysis, such as fault early warning, analysis of failure rate, the utilization rate analysis of equipments, business optimization analysis, OD hotspot analysis, abnormal passenger recognition, usability analysis of equipment. All of these pointed out a new method for the future development of RATGS.

Key words: information technology; monitoring; big data analysis; Railway Automatic Ticketing and Gate System(RATGS); Complex Event Processing(CEP)

中国高速铁路(简称: 高铁)自动售检票系统的软硬件复杂度较高。自动售票机发售的是符合ISO7816标准尺寸的磁介质车票, 支持现金和银行卡支付, 同时设纸币和硬币找零; 自动售票软件系统支持旅客自助获取车次、发到站、乘车日期、票种、席别, 为旅客提供流程化购票服务, 并为管理人员和操作员提供交互式管理界面和维护向导。自动检票系统通过接口服务器连接客票系统, 获取车次目录、列车停靠站、电子票等信息, 通过专用接口连接旅服系统, 获取进出站开停检时间、列车早晚点等信息,

收稿日期: 2014-04-10

作者简介: 王 成, 副研究员; 史天运, 研究员。

自动检票机支持读取磁介质车票、二代身份证电子票、中铁银通卡储值票过闸^[1~3]。

随着自动售检票设备装机量的增加, 自动售检票系统在高铁客运作业中发挥的作用也越来越明显, 系统的可靠、稳定运行也越来越重要。通过建设铁路自动售检票监控系统和铁路自动售检票监控系统大数据分析平台, 采用大数据分析技术, 实现对系统交易数据、状态数据、用户操作数据、日志数据、视频数据的分析, 进一步从这些海量数据中找出故障规律, 挖掘出故障与系统环境、用户行为间的关系, 从而提高系统的监控水平, 保障自动售检票系统的正常运行^[1]。

1 铁路自动售检票监控系统总体框架

基于大数据分析的自动售检票监控系统可自下而上分成的4层：(1) 基础层。包括：以计算虚拟化、存储虚拟化和网络虚拟化为基础的云资源池，通过采集代理收集数据库中的交易数据、状态数据，操作系统主机运行环境数据，日志文件以及视频数据，形成自动售检票系统的多维大数据源；(2) 管理层。在多维数据源基础上，提供基于海量并行处理架构(MPP)和MapReduce的离线数据的处理和基于复杂事件处理模型的流数据管理；(3) 分析层。建立大数据分析平台，并在其上执行针对监控数据的聚类分析、关联规则发现、分类预测等数据挖掘和机器学习算法，建立故障树和事件树分析模型、统计分析模型、辅助决策模型等；(4) 应用层。提供铁路自动售检票监控管理平台，集成运行环境监控、关键业务监控、设备监控、统计分析、辅助决策等多项应用，并且支持多样化应用间的无缝集成。大数据自动售检票监控系统总体框架图如图1所示^[4-5]。

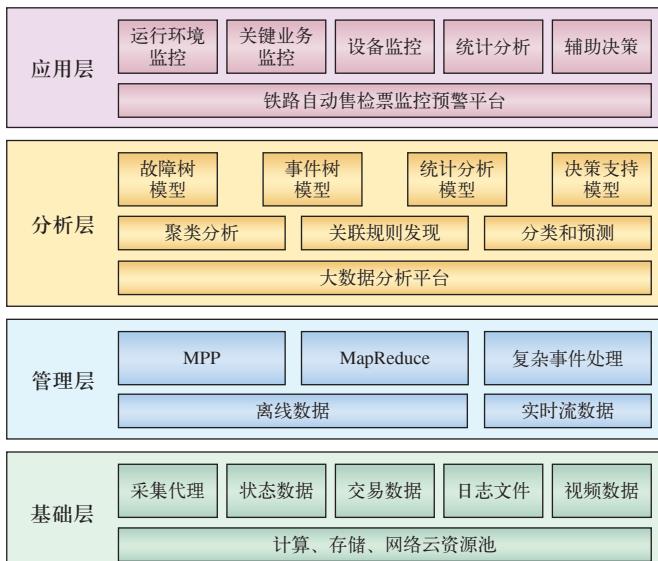


图1 基于大数据分析的自动售检票监控系统框架图

2 大数据分析平台的搭建

传统数据平台通过ETL(Extract, Transform, load)等集成技术将多种结构化数据源导入到基于传统关系数据库的数据仓库中，可采用多种手段分析并展现数据，如图2所示。

大数据分析平台设计原则要满足4V(即：数据

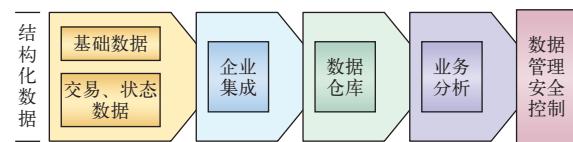


图2 传统数据分析平台

量大，数据类型繁多，价值密度低，处理速度快。)需求，完成海量数据解析，需要部署分布式并行处理架构。面向实时处理和批处理需求，需要采用不同的技术，对于实时数据，采用key-value数据存储(NoSQL数据库)可提供高效分析；对于批处理数据，通过MapReduce技术根据特定业务对数据进行过滤，过滤完成后可直接在本地分析，也可以加载到数据仓库中，关联其他业务数据一起执行分析业务，如图3所示。

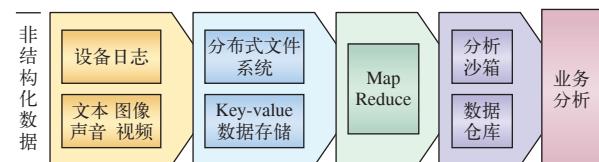


图3 非结构化数据分析平台

在很多情况下，需要将大数据和传统的结构化数据联合才能开展分析工作，因此，大数据分析平台不仅要满足快速应用和快速数据解析需求，还要和原有的结构化数据做关联分析。采用Hadoop的障碍之一就是缺少和现有BI生态系统之间的集成。为解决铁路自动售检票监控系统和原有的客票监控系统、自动售票监控系统之间的融合，将结构化的大数据分析结果和原有数据仓库数据合并，及时高质量的完成针对不同数据源、不同处理需求的数据分析的能力。整合原有架构和大数据分析能力的铁路自动售检票集成大数据分析平台如图4所示。

对于系统采集的基础数据、交易和状态数据、设备日志、多媒体数据，可存储在关系数据库、文件、分布式系统(如NoSQL和HDFS)中。

中间的组织层不仅要组织所有的数据类型和域，缩小传统数据采集、数据处理框架与大数据采集和处理框架之间的差异，还要具有处理高速、高频数据的能力、具有处理大量数据的能力，弥补不同数据结构间的差距，实现Hadoop/MapReduce与传统数据仓库和交易数据库之间的双向业务集成。

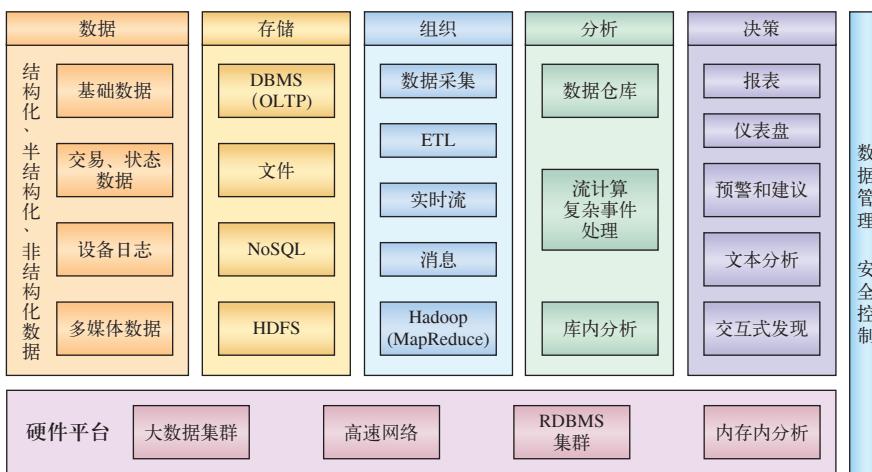


图4 大数据集成分析平台

分析层将大数据分析结果加载到数据仓库或交易数据库中执行进一步的分析和展现。复杂事件处理可针对实时的流数据执行有针对性的实时业务分析。

决策层通过采用库内统计分析、数据可视化、高级分析（数据挖掘等）等方法提供报表、仪表盘、预警、决策建议等信息。

3 铁路自动售检票监控系统大数据应用

根据自动售检票监控系统的业务目标和数据结构，采用数据仓库、分布式计算、MapReduce、复杂事件处理、数据挖掘分析理论和方法等，建立故障树和事件树分析模型、统计分析模型、辅助决策模型等，利用大数据分析的成果，结合日志数据分析和故障分析的通用方法，可开展铁路自动售检票系统故障快速定位及故障率分析、设备故障预警及预处理、设备利用率分析、业务优化分析、旅客行为分析等。

3.1 故障快速定位及故障率分析^[6]

针对自动售检票系统维护作业中复杂问题定位困难、过分依赖维护人员经验人工处理的问题，通过对系统响应时间数据做聚类分析，可由稠密簇找出自动售检票业务正常性能指标的边界值，对应一组性能指标项形成的序列，可用作故障树模型的参数输入。系统建设初期，可根据训练数据和经验值建设初步故障树模型，系统运行一段时间，积累了足够的数据后，再由聚类分析和关联分析得到的边界值、阈值作为参数重新调整故障树的分类规则，从而形成一个动态调整的故障树模型。

3.2 故障预警和预处理

对某些故障的预处理，需要有丰富的维护经验。根据系统当前运行状态，来制定相对应的预处理措施，是实现自动化故障处理目标的主要步骤。利用事件树分析模型找出系统事件之间的关联关系，推断出事件发生引起的后续结果，设计故障预测算法，设计自动化的故障预处理流程，对于无法自动完成的任务，向维护人员发出预警通知。预处理的故障内容包括清理日志、整理磁盘、重启程序、部件复位等。

图5为事件树模型，图6为故障预处理流程的示例。

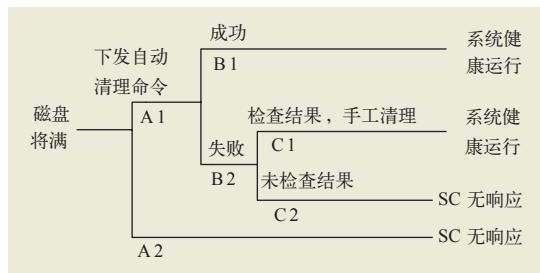


图5 事件树示例

图5描述的是服务器磁盘空间“磁盘将满”这一事件的事件树。按照规定处理流程，维护人员在发现监控系统检测到“磁盘将满”时：(1) 下发自动清理日志命令，启动服务器日志清理脚本开始清理日志；(2) 检查命令执行结果，查看磁盘空间使用率；(3) 如果脚本执行不成功，则手工清理日志，整理磁盘空间。图6描述的是典型的故障预处理流程。

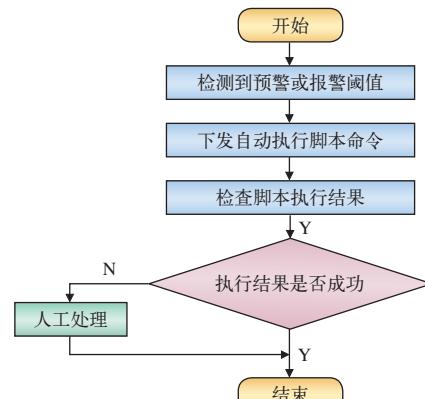


图6 故障预处理流程示例

3.3 设备利用率分析^[6]

为便于评估铁路自动售检票设备在实际铁路客运组织作业中发挥的作用,需要对车站内设备的布局情况、设备的利用情况进行综合分析。铁路自动售检票系统的交易等数据、状态数据等存放在 Sybase ASE 关系数据库中,通过 ETL 等工具将这些数据导入数据仓库,并执行分析处理。

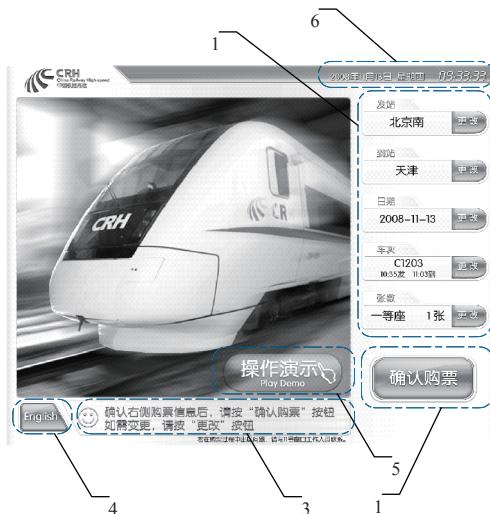
通过数据传输中间件 (DBCS) 定期将存储在 Sybase ASE 中的自动售票存根表、自动检票存根表转储至 Sybase IQ 数据仓库中,建立区域、机器、型号、时间段等不同的维度表,分析设备的售、检票量在不同区域、不同机器、不同时间段的分布,得出设备利用率情况,为设备在车站的布局和工作时间调度提供建议,最大限度发挥自动化设备的作用。

通过 DBCS 将 Sybase ASE 中的自动售检票设备状态表转储至 Sybase IQ,建立服务时间维度表,分析设备的正常服务时间、暂停服务时间、维修时间分别占的比例,得出设备忙闲率,进一步分析暂停服务的原因,为设备运营维护、巡检保养制度的制定提供建议。

3.4 业务优化分析

自动售票系统购票界面主要考虑城际铁路和高铁线路车次密集,到站集中的特点,随着多数的既有线也已经开通使用,如何既能满足原高铁线购票便利要求,又能很好的兼顾既有线线路多、到站多,车次分散的特点,是自动售票系统购票界面的优化方向。

购票首界面共有 6 个部分,各部分的位置标注图如图 7 中所示。旅客通过售票界面选择自己要购买的车票信息并通过自动售票机进行交易。通过详细记录旅客每一步操作过程。挑选全路高铁线和既有线典型车站的设备,将设备日志转储于大数据分析平台的 HDFS 中,执行 MapReduce 分析,得出选票过程中发站、到站、乘车日期、车次、票种、席别、张数、确认购票等不同功能按钮的点选次数和点选顺序,得出设备易用性指标,为选票流程优化提供决策建议。同时,通过分析日志中每笔购票取车次、取票、读取证件、支付、制票各环节分别所用的时间,找出影响售票效率的瓶颈。对于旅客选择造成的低效问题,考虑优化界面提示,加快旅客操作速度;



1-选择车票信息按钮区; 2-确认购票按钮; 3-购票信息文字提示;
4-语言选择按钮; 5-操作演示按钮; 6-日期时间。

图7 自动售票系统购票首界面

对于系统或设备响应低效问题,分析问题原因,改善接口方式,提出优化思路。

3.5 旅客行为分析

铁路服务的自助化为捕捉用户行为轨迹提供了基础条件,从旅客操作轨迹中发现潜在的出行计划,为旅客推出个性化的铁路产品推荐信息,实现客运精准营销,将是大数据在铁路应用的研究热点。另一方面,通过对旅客的行为分析,帮助铁路实时改进和创新铁路产品,并根据旅客的实时反馈信息,改进服务,预测面临的经营风险。针对旅客过去和现在的实时数据有针对性的确定风险类别,及时的采取措施,减少损失。

行为分析的数据源来自自动售检票设备的操作日志,铁路车站的监控视频等,这些数据量庞大,规律发掘困难,必须利用大数据处理分析平台来进行数据的存储、计算以及分析结果可视化展现。通过 CEP 技术对铁路自动售检票系统的运行数据进行实时或者准实时处理,使用当前或过去的数据对未来的事件作出预测。对于高速的流数据,通过建立分布式的流计算平台,提高数据处理的实效性。行为分析包括几个方面。

3.5.1 起止点的热点分析

通过收集旅客选择的发到站信息,分析在查询热点区间可售票额能否满足旅客购买需求,并分别

(下转 P51)

票检票业务，在京津（北京-天津）线、福夏线（福州-厦门）、沪宁（上海-南京）线、长吉（长春-吉林）线等多条线路开通了中铁银通卡检票业务。铁路自动检票系统提高了车站检票工作效率，简化了车站的客运组织工作，节约了大量的人力、物力，产生了巨大的经济效益。随着中国铁路事业的进一步发展，自动检票系统势必会得到更加深入和广泛的应用。

（上接 P45）

针对有车无票、无车等不同情况，采取相应的票额分配方案、列车开行方案等，在运输淡季实现客运效益的最大化。

3.5.2 异常旅客识别及告警

铁路旅客常携带大件行李，为便于旅客通行，铁路自动检票机采用了半高拍打式闸门，但拍打式闸门对尾随的控制无法达到三杆式闸门的精度。电子票检票业务开通以来，互联网购票的旅客可直接持二代身份证检票。但通过分析电子票检票日志，发现有一小部分的旅客经常在没有购票的情况下，刷二代证尾随过闸。利用 CEP 平台，捕获“刷证-检票失败-尾随”这一关联事件，识别出尾随旅客，并通知闸机发出警告，维护铁路客运公平秩序。

3.5.3 设备易用性分析

自动售检票设备均为开放式自助设备，旅客操作的易用性决定了设备使用的效率。通过对站内设备监控视频的分析，观察不同类型旅客在操作自动售票机触摸选票界面，操作识钞器、读卡器，使用找零口、出票口等部件时的连贯性，提出面向不同类别人群的设备结构改进和旅客操作引导优化方案，如开发针对轮椅旅客和儿童的自动购票设备等。同样，通过观察旅客在使用右手持票规则的自动检票机时动作的流畅度、持大件行李旅客通过闸机时是否存在障碍等，可针对性改善自动检票机结构设计，或增加更明显的操作提示，通过这些措施实现设备利用率最大化。

3.5.4 设备安全监控

除了车站设有设备视频监控外，自动售票机内安装有视频录像机，可详细记录旅客的购票行为，通过对这些视频数据的分析，可识别旅客异常购票、

参考文献：

- [1] 张家峰, 蒋秋华. 郑武高速铁路自动检票系统方案设计 [J]. 铁路计算机应用, 2012, 21 (18) : 22-23.
- [2] 王成, 蒋秋华, 张家峰, 朱建军, 张曦. 基于二代身份证的互联网电子票自动检票系统的研究与实现 [J]. 中南大学学报: 自然科学版, 2013, 44 (51) : 347, 349.

责任编辑 徐侃春

检票行为，以及检测对设备的人为破坏，当发生安全事故时通过告警系统及时通知管理人员采取措施，最大限度保障设备安全。

4 结束语

通过充分利用现有自动售检票业务数据，采用故障分析和预警手段，提高系统稳定性，减少维护时间和成本，是目前自动售检票系统下一步的主要发展方向。基于本文提出的大数据分析平台和大数据分析方法，构建面向全路自动售检票系统的信息挖掘和增值服务系统，将实现系统功能的自检自愈，故障的预警预处理；为设备制造、部件选型以及软件优化提供参考；为铁路车站内设备布局、工作时间调度、机型配比提供建议；识别特殊旅客，识别风险，检测异常等。

参考文献：

- [1] 王成, 张家峰, 蒋秋华. 铁路自动售票监控系统研究与实现 [J]. 中国铁路, 2013 (6) : 51-55.
- [2] 潘浪涛, 王成, 李超. 软件故障树分析在自动售票系统服务中的应用 [J]. 铁路计算机应用, 2011, 20 (9) : 24-30.
- [3] 王珊, 王会举, 覃雄派, 周烜. 架构大数据: 挑战、现状与展望 [J]. 计算机学报, 2011, 34 (10) : 1741-1752.
- [4] 宋亚奇, 周国亮, 朱永利. 智能电网大数据处理技术现状与挑战 [J]. 电网技术, 2013, 37 (4) : 927-935.
- [5] 汤玲丽. 复杂事件处理引擎关键技术研究 [D]. 哈尔滨: 哈尔滨工程大学, 2012 : 5-15.
- [6] Jiawei Han, Micheline Kamber, Jian Pei. Data Mining Concepts and Techniques (3rd Ed)[M]. The Morgan Kaufmann Series in Data Management Systems, 2011.

责任编辑 徐侃春