

文章编号: 1005-8451 (2016) 09-0038-05

数据挖掘典型案例及在高速铁路运营期的应用展望

吴艳华, 李 平, 李 芳

(中国铁道科学研究院 铁路大数据研究与应用创新中心, 北京 100081)

摘 要: 中国高速铁路运营管理缺少经验可供借鉴, 围绕其核心职能, 通过数据挖掘技术和手段为管理者提供科学、客观的管理决策, 是辅助提升高速铁路运营管理水平的有效方法。构建资产运营、运输监管和安全监督等数据分析典型场景, 满足高速铁路数据分析需求, 为管理决策提供数据支撑。

关键词: 高速铁路; 数据分析; 应用

中图分类号: U238 : TP39 **文献标识码:** A

Typical case of data mining and its application during operation periods of high-speed railway

WU Yanhua, LI Ping, LI Fang

(Research and Application Innovation Center for Big Data Technology in Railway, China Academy of Railway Sciences, Beijing 100081, China)

Abstract: There is no existing management experience on the operation of China High-speed Railway, how to focus on its core functions, and how to provide scientific and objective management decisions techniques and tools for managers, it is an effective method to enhance the operation management level of high-speed railway through data mining. Construction of asset management, transportation safety regulation and supervision of the data analysis of typical scenarios to meet the high-speed railway data analysis needs could provide data support for management decisions.

Key words: high-speed railway; data analysis; applications

截至 2015 年底, 中国高速铁路运营里程已达 1.9 万 km, 现有高速铁路公司 30 余家, 多以股份制的形式组成。在国家高速铁路“走出去”战略的要求下, 高速铁路的发展将走向“资产管理精细化、企业运营高效化”的道路。自勘察设计阶段开始, 在工程建设、联调联试和运营维护等高速铁路生命周期的各个阶段, 各业务系统和日常管理中都产生了大量的价值数据, 但由于技术、管理手段、管理模式^[1]的限制, 这些价值数据大部分孤立地存储于各个应用系统中, 无法交互与共享, 导致形成数据孤岛; 还有部分数据分散于管理人员手中或由于应用系统存储空间的限制而丢失。高速铁路公司的数据资产无法发挥其潜在价值, 逐渐变成无用数据被浪费, 难以为高速铁路精细化管理提供科学决策的客观支持^[2]。

收稿日期: 2015-06-15

基金项目: 中国铁路总公司科技研究开发计划课题 (2015X003-F); 中国铁道科学研究院院基金重大项目 (1551DZ8004)。

作者简介: 吴艳华, 副研究员; 李 平, 研究员。

目前, 资产密集型企业 and 肩负运营任务的企业, 都充分认识到数据的价值, 并通过数据挖掘发现数据背后的规律, 支持实际的生产和运营。本文从各企业的典型案例入手, 从高速铁路公司的核心业务职能出发, 深化高速铁路数据挖掘分析需求, 提出高速铁路数据分析典型应用场景, 充分挖掘数据的价值, 为构建“建设领先、运营领先、维护领先”的中国高速铁路科学管理提供思路。

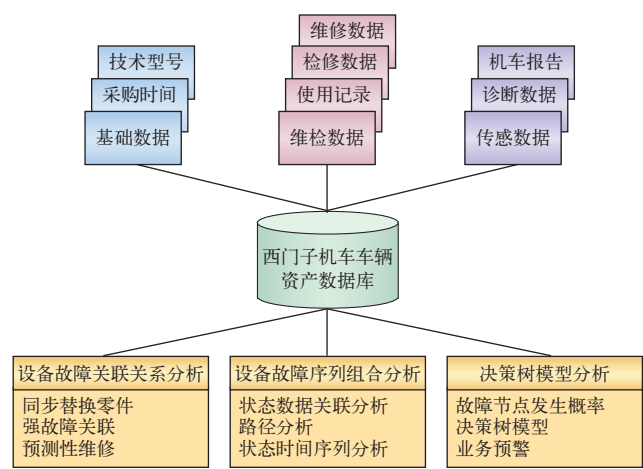
1 国外典型案例分析

在国内外铁路、电网、航空等领域, 相关企业都认识到数据价值带来的经济效益, 纷纷通过数据挖掘为日常管理提供决策。

1.1 欧美铁路机车车辆管理应用

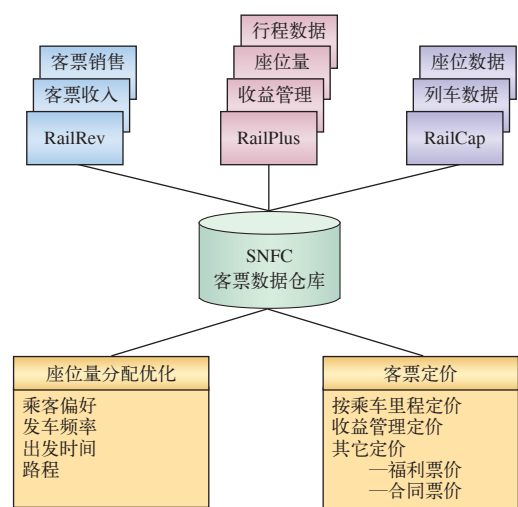
欧洲和美国的多家铁路公司租用西门子公司公司的机车, 并要求在其生命周期内提供日常、返厂等维修服务, 还有些要求其评估设备运营期的关键指标

体系，并要求保证设备的安全性和可靠性。为了能够为各铁路公司提供上述服务，西门子公司建立了机车车辆资产数据库，首先为每辆机车建立基础信息数据库和履历信息数据库，同时根据机车的使用记录、检修数据、维修数据等，并通过部署传感器等采集机车重要部件的实时状态数据，如图 1 所示。通过这 3 类数据,开展机车支行状态和设备故障分析，对机车车辆进行更精细化、科学化的管理。



1.2 法国铁路客票管理应用

20 世纪 90 年代初，法国国营铁路（SNCF）采用 SABRE 方案开发了法国铁路客票系统，实现客票预订、分配和综合决策支持，系统主要包括收益管理（RailRev）、时刻表计划管理（RailPlus）和席位管理（RailCap）3 个子系统^[3]，如图 2 所示。



RailRev 是 SABRE 方案的基础管理系统，主要存储和管理客票收入和销售数据；RailPlus 是为了支

持 SNCF 高速铁路服务的基本时刻表而设计的集成计划与分析系统，主要模块包括工作设置经理、收益率模块、可行性模块、座位量分配模块以及行程安排模块，RailPlus 每天以 RailRev 提供的销售和收入数据为基础，开展基于数据的客流变化预测，并以此预测结果提出运能调整和运行方案设计建议；RailCap 中存储全部列车、线路席位数据，是一个整体计划编制和分析系统，RailCap 系统根据 RailRev 系统的每日更新的需求预测，调节列车座位量，其核心是座位量调节模型，该模型在保证现实中各项操作限制的情况下，寻求期望的、最大化的收益增加方案。这 3 个子系统为决策者使用全局的方法进行决策，提供高级操作研究模型集合。

SNCF 的 SABRE 方案还支持座位量分配优化、客票定价分析等决策分析。

座位量分配优化帮助 SNCF 公司设计出符合乘客偏好的时刻表（发车频率、出发时间、路程），并按照满足需求和减少损失的乘客数量的要求分配座位量。同时提高了顾客满意度，现在第一需求被满足的乘客人数增加了 3%。SNCF 进行了座位量分配对时刻表排班收益率影响的研究。

在客票定价方面，法国铁路在各类系统数据的支持下，同时采用“按乘车里程定价”和“收益管理定价”两种方式。此外，还采用福利票价、合同票价等辅助定价方式。里程定位一般按照统一制定的费率表计算票价；收益管理定价适用于较高等级列车，包括传统干线改进型列车和 TGV 高速列车；其它定价方式包括福利票价和合同票价，福利票价由国家规定，合同票价由法国国营铁路公司与有关部门约定产生。多种定价机制并存,是以一定的数据为支撑，引入收益管理模式，在政府宏观调控下基于动态定价理论实现运价的合理浮动。

2 国内典型案例分析

2.1 国家电网公司资产管理应用

国家电网公司自 2006 年开始实施 SG-186 工程，2008 年开展了统一数据平台的建设，通过平台实现各应用系统中的数据共享和交换，实现丰富的数据展现和分析。数据平台的建设过程历经 3 个主要阶段：

(1) 通过建设 ODS (操作数据存储)、数据仓库和数据集市, 实现简单的查询, 直接上报查询结果数据 ; (2) 挑选一些关键的业务 / 主题域 (如营销和生产) 进行实践, 主要提供联机分析和综合报表等分析结果 ; (3) 收集其余业务 / 主题域的数据, 提供更全面的数据分析和数据展现, 对数据进行更深层次的挖掘。

同时, 国家电网公司为实现资产全生命周期中综合绩效的管理, 建立了决策层、管理层和执行层三级分析指标体系, 实现资产的精细化管理和分析, 如图 3 所示。

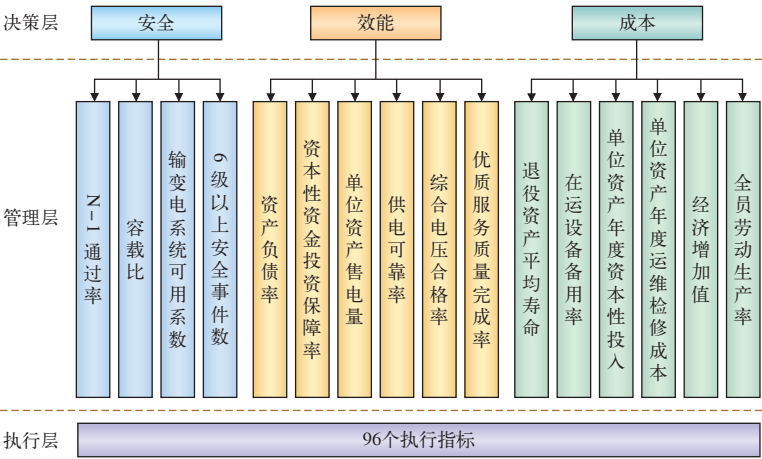


图3 国家电网资产全生命周期管理综合绩效指标体系

2.2 中国国际航空公司MSAS应用

中国国际航空股份有限公司 (简称 : 国航), 推行一体化运营, 具有强大的运行控制能力, 计算机飞行计划系统从飞机性能数据库、全球导航数据库、全球机场数据库和高空气象数据库中提取信息, 优选航路, 制定飞行计划, 把握飞机性能, 确认飞行资格, 严格把关放行。同时, 为了提高航空运营服务水平, 国航提出了面向客户运营的市场销售分析系统 (MSAS)。

MSAS 通过对各业务系统中分散的客户数据进行整合, 逐步建立国航的客户知识体系和客户数据标准, 以支持国航的市场和销售体系, 从客户的角度进行市场研究和规划、营销创新以及个性化服务, 并提升竞争力。

MSAS 所需的数据源包括市场、离港、收入、订座、促销、航班计划、常旅客、客户沟通、网站访问等数据。基于这些数据, 从业务视角出发开展

客户行为、上座率趋势、价格体系调控等应用分析, 支持管理层的定价与收益管理决策, 如图 4 所示。

以上座率趋势监控分析为例, 其目的就是通过提前跟踪某一航班和不同子舱位的订座趋势, 能够及早采取适当的措施和营销策略, 从而提高航班收益。系统中主要实现: (1) 对上座率进行同比和趋势分析; (2) 对不同航线、不同航班、不同舱位、不同渠道 (销售点 / 旅行社) 进行比较分析; (3) 与当前可销售座位数进行比较。

通过多维分析对比, 提前监控舱位的上客速度, 并进行同比和趋势分析, 及时监控节日产品的销售动态, 结合历史数据预测节假日需求; 通过上座率趋势监控分析, 能够及时总结不同航线、航班的订座变化规律, 调整不同舱位的布局, 及时制定促销运价和策划促销活动。

3 高速铁路数据分析需求

高速铁路从建设期转入运营期后, 公司职能也相应地从“以工程建设为核心”逐渐转变为以“资产运营”为核心, 同时兼顾“运输监管”和“安全监督”, 因此高速铁路的数据分析, 也是为资产运营、运输管理和安全

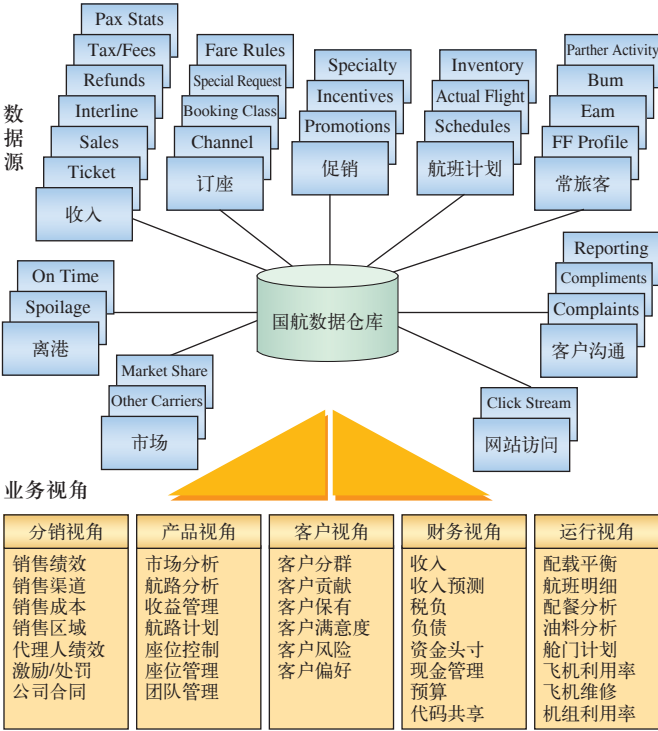


图4 MSAS中数据来源

监督等提供决策支持服务。

3.1 资产运营数据分析需求

高速铁路公司负责全线资产的管理和维护工作，包括运营期间全部基础设施设备的档案管理、设备设施的更新改造、维修保养、处置报废和备品备料的采购、储存等资产全生命周期的管理，同时拓展车站和沿线土地、商业的利用和开发。

为保障线路的安全、高效运营，高速铁路基础设施状态、运输生产、设备维修、节约成本、放置等，都是资产运营管理过程中最为关注的问题。

3.2 运输监管数据分析需求

高速铁路公司在现行的铁路体制和业务模式下，负责旅客运输专线业务的运营监管和市场拓展，包括运输业务的监管和辅助决策、市场规划和营销拓展等，为服务质量的提升、运输效率的提高和市场竞争力的增强等提供监管技术手段和决策依据。

为提高高速铁路运营服务的能力和质量，开展线路开行方案相关性分析及优化、实行客户关系分析和精准营销，分析客票销售策略、与航空等其它运输方式的关联等，分析运输收支相关因素，在为旅客提供高质量服务的同时，保障高速铁路运营收入。

3.3 安全监督数据分析需求

高速铁路公司负责监督高速铁路的运输安全，包括日常运输的安全监督、设备设施安全状况，特别是源头性安全的检查、识别和管理、设备设施故障和事故的辅助处理等。

安全风险发展为安全事故是有其发展过程的，如何通过日常安全风险数据的监控，分析安全事件的发展规律，及时发现其变化关键因素，控制进一步演化，都是数据分析工作需要开展的相关研究。通过即有业务系统数据，为管理提供 3 个方面管理的决策支持，就成为高速铁路数据分析的主要需求。

4 高速铁路数据分析典型场景

根据高速铁路公司的数据分析需求，可将其数据分析划分为 3 大类典型应用场景：资产运营、运输监管和安全监督，如图 5 所示。

4.1 资产运营数据分析场景

高速铁路运营期的资产运营管理核心职能，决

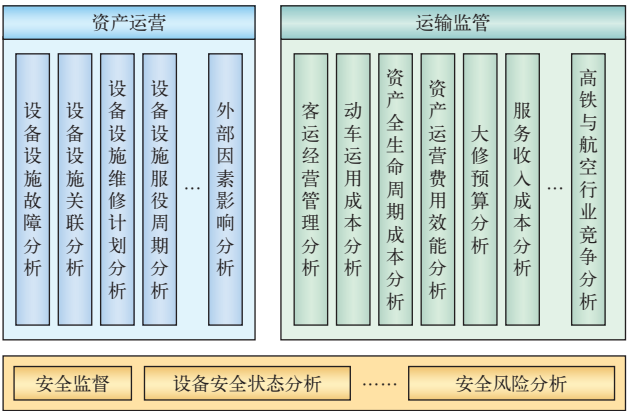


图5 高速铁路运营期数据分析典型应用场景

定了与设备、基础设施和资产相关的数据分析，主要关注于设备和资产的状态，如设备和基础设施状态的好坏，设备间的相互影响、影响的程度，设备周边的气候、自然因素对设备产生影响，设备设施的自身变化趋势和规律等，从而为设备设施的维修保养提供参考。

4.2 运输监管数据分析场景

高速铁路公司对运营情况的监管职能，决定了其在客运经营、运输收入支出、运输效能管理，以及与其它运输行业之间的竞争力的关注。开展客运开行方案分析、客流预测分析，通过数据反映列车开行和运营情况，并以此为依据优化列车开行方案，及时满足旅客出行需求，提升高速铁路客运服务的满意度。同时，开展运输收入支出和效能分析，提高高速铁路运营收入；开展铁路与航空、公路运输的优势和需求分析，深入发掘高铁客运开行与客票销售策略，提供特殊情况下（天气、大型活动）的高铁客运服务保障能力。

4.3 安全监督数据分析场景

高速铁路的安全监督职能，是保证高速铁路运营安全的重要环节。开展高速铁路设备安全状态和安全风险分析，发掘影响安全的主要因素，并对各因素的影响程度进行重点分析，从而为优化重点安全因素的检查和监督提供支持。

5 结束语

随着信息技术在铁路的成熟应用，越来越多的数据积累于各应用系统中，数据的价值也越来越被

(下转 P45)

表3 2006~2015年铁路科研突变词列表

序号	关键词	突变强度	起始年	结束年	2006—2015
1	青藏铁路	6.552 2	2006	2010	
2	多年冻土	3.986 5	2006	2008	
3	遗传算法	3.670 2	2006	2008	
4	移动通信	3.028 9	2006	2008	
5	钢管混凝土	2.58 3	2006	2009	
6	铁路桥梁	2.572 8	2006	2008	
7	动力学	2.485 3	2006	2007	
8	客运专线	3.582 9	2007	2009	
9	无砟轨道	2.804 1	2007	2008	
10	GSM-R	3.364 4	2010	2012	
11	高速铁路	17.221 2	2013	2015	
12	高速动车组	4.724 9	2013	2015	

2006 ~ 2010 年,我国铁路的潜在研究重点主要集中于青藏铁路、多年冻土、遗传算法、移动通信、钢管混凝土、铁路桥梁、动力学等 7 个方面,也是我国铁路科研前沿专业较多的阶段;2007 年,客运专线、无砟轨道成为新的研究前沿,得到研究者迅速和广泛的关注;2013 ~ 2015 年,高速铁路、高速动车组成为新的研究前沿,特别是在高速铁路,吸引了众多的科研工作者投入其中。

通过近 10 年我国铁路领域热点和前沿问题的数据分析可以看出,铁路科研工作与时俱进,与现代社会技术进步、铁路建设规划及相关政策密切相关。铁路科研人员的工作与我国铁路发展的实际需要和发展过程中遇到的新问题紧密关联,不断地将研究成果应用到铁路建设之中。

3 结束语

本文利用 CiteSpaceIII 科学知识图谱工具,对 CNKI 数据库中 2006 ~ 2015 年铁路行业 3 种 EI 期刊的发表文章进行了分析研究,对近 10 年来我国铁

路领域的研究热点与研究前沿进行了定量分析,分析结果表明,我国铁路研究近 10 年来的热点领域主要为高速铁路、有限元、高速列车、城市轨道交通、数值模拟、客运专线、模型试验、盾构隧道、无砟轨道、青藏铁路、接触网等方面,未来的发展前沿将集中在高速铁路、高速动车组等专业。

由于本研究仅以 2006 ~ 2015 年的《铁道学报》、《中国铁道科学》、《西南交通大学学报》这 3 种期刊的所有科研论文为对象进行分析,样本虽具有一定代表性,但不能覆盖所有的铁路科研内容,得出的结论可能有一定偏颇。

通过 CiteSpace 这种关键词词频共现和突变率统计分析的新量化手段,对近 10 年我国铁路的研究发展情况进行了总结与梳理,希望能为我国铁路科研人员更好地掌握研究动向提供参考。

参考文献:

[1] 肖 明.知识图谱工具实用指南[M].北京:中国铁道出版社,2014.

[2] 代峰刚,蔡焕杰.基于 CiteSpace 的农业节水研究进展与发展趋势[J].水资源与水工程学报,2015,26(1):212-218.

[3] 赵蓉英,许丽敏.文献计量学发展演进与研究前沿的知识图谱探析[J].中国图书馆学报,2010,36(5):60-68.

[4] 马海群,吕 红.基于中文社会科学引文索引的中国情报学知识图谱分析[J].情报学报,2012,31(5):470-478.

[5] 朱 亮,赵瑞雪,寇远涛,等.基于 CiteSpace II 的“共引分析”领域知识图谱分析[J].数字图书馆论坛,2012(12):58-63.

责任编辑 付 思

(上接 P41)

管理决策者所认识。同时,大数据技术的兴起和逐步成熟,也让管理人员能够运用各类技术手段实现对数据的挖掘。高速铁路运营管理是中国铁路管理面临的新挑战,基于高速铁路资产运营、运输监管和安全监督的核心职能,在保证高速铁路日常运输管理的同时,通过数据为管理决策提供科学的依据是提升高速铁路管理能力的有效途径。

参考文献:

[1] 王 东.高速铁路委托运营管理模式研究[J].铁道运输与经济,2014(6):12-14.

[2] 王英杰,侯日根,吴艳华,等.高速铁路资产管理系统动态数据接入研究[J].铁路计算机应用,2016,25(6):1-4.

[3] 韩保花,贾光智.法国铁路客票定价机制分析与借鉴[J].中国铁路,2013(2):90-92.

责任编辑 陈 蓉