

# 工业大数据的应用与实践

袁爱进<sup>1</sup>, 岳滨楠<sup>2</sup>, 闫鑫<sup>2</sup>, 黄健<sup>2</sup>

1. 三一集团有限公司, 上海 201299; 2. 上海华兴数字科技有限公司, 上海 201299

## 摘要

随着物联网和信息物理系统时代的来临,更多数据可以被收集和分析,工业大数据也成为行业创新和转型的重要驱动力。与传统互联网大数据不同,工业大数据与工业逻辑紧密结合,详细对比了工业大数据与互联网大数据的区别,结合典型案例,阐述了需求驱动传统制造业变革的过程和工业大数据的发展之路,介绍了工业大数据平台的架构和工作原理。最后以挖掘机业务为例,展示了工业大数据平台的应用效果,并提出共享数据和模型将会使基于数据的应用产生更大的价值。

## 关键词

工业大数据;共享数据;平台;智能服务

中图分类号:TP399

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2017059

## *Application and practice of industrial big data*

YUAN Aijin<sup>1</sup>, YUE Binnan<sup>2</sup>, YAN Xin<sup>2</sup>, HUANG Jian<sup>2</sup>

1. Sany Group Co., Ltd., Shanghai 201299, China

2. Shanghai Rising Digital Co., Ltd., Shanghai 201299, China

## *Abstract*

With the advent of the Internet of things and the era of Cyber Physical System, more data can be collected and analyzed, and industrial big data have become important driving force for industry innovation and transformation. Unlike traditional Internet big data, industrial big data are closely integrated with industrial logic. The difference between industrial big data and internet big data was compared in detail. Combined with the classical cases, the process of demand driven transformation of traditional manufacturing industry and the development path of industrial big data were expounded. The structure and working principles of industrial big data platform were introduced. At last, taking the excavator business as an example, it shows the application effect of the industrial big data platform, and that having a share in data and models will make data based application produce more value was put forward.

## *Key words*

industrial big data, share data, platform, intelligent service

## 1 引言

### 1.1 工业大数据的兴起

随着工业进入信息化时代,工业大数据已成为新一轮产业革命的重要动力。工业大数据以工业系统的数据收集、特征分析为基础,对设备、装备的质量、生产效率、用户体验以及产业链进行更有效的优化,并为未来的制造系统搭建无忧的环境<sup>[1]</sup>。无论是德国提出的“工业4.0”,还是美国定义的“工业互联网”,大数据在其中都发挥着重要作用。美国IBM公司认为,“工业4.0”是大数据驱动的智能制造<sup>[2,3]</sup>,无数据不智能;美国通用电气(General Electric, GE)公司表示,工业互联网最重要的就是数据分析,未来每一家工业企业也必须是一家软件企业,有了软件必然会有数据。

大数据在工业领域的兴起主要由以下因素决定<sup>[4]</sup>。

- 设备自动化过程中,控制器产生了大量的数据,然而这些数据蕴藏的信息和价值并没有被充分挖掘。
- 传感器技术和通信技术的发展使实时数据的获取成本不再高昂。
- 嵌入式系统、低能耗芯片、处理器、云计算等技术的兴起使设备的运算能力大幅提升,具备了实时处理大数据的能力。
- 制造流程和商业活动变得越来越复杂,依靠人的经验和分析已经无法满足复杂的管理和协同优化的需求。

工业领域中,如果设备数据、运维数据、事件数据、舆论数据、服务数据、公司数据、市场数据和上下游产业链数据等能够在统一的平台管理,大量的数据将会使原本孤立的系统相互连接,使设备之间可

以通信和交流,也使生产和服务过程变得更加透明。

### 1.2 工业大数据与互联网大数据的区别

大数据正在改变着人们的生活,过去几年,无论是健康、交通、公共安全,还是生活、购物、旅游、娱乐,都已经逐步建立起了大数据分析系统。无论是国家还是企业,对大数据的投入都数以亿计。大数据的应用也从开始的互联网领域走向了金融、医疗、环境以及工业领域,这其中应用最成功的是互联网。互联网以其开放、自治与共享的理念,与社会各个领域的结合,带动了生产和社会的巨大发展和进步<sup>[5]</sup>。

工业大数据是大数据与智能制造的交叉点,工业大数据是指在工业产品全生命周期的信息化应用中产生的数据,是工业互联网的核心,是工业智能化发展的关键<sup>[6]</sup>。工业大数据基于网络互联和大数据技术,贯穿于工业的设计、工艺、生产、管理、服务等各个环节,使工业系统具备描述、诊断、预测、决策、控制等智能化功能。

与互联网大数据相比,工业大数据具有更强的专业性、关联性、流程性、时序性和解析性等特点,仅仅依靠传统的互联网大数据分析技术已无法满足工业大数据的分析要求<sup>[7]</sup>。两者的区别见表1。

因此工业大数据分析并不仅仅依靠算法工具,而是更加注重逻辑清晰的分析流程和与分析流程匹配的专业技术体系。

互联网大数据可以从数据端出发看问题,但是工业大数据则应该从价值和功能端思考。也就是说,传统装备企业在进行物联网建设时,如果只是强调数据获取的途径、量级,没有考虑到数据的具体分析和利用以及相应的功能与目标,很可能就会造成许多数据采集回来之后没有用,而一些关键数据反而没有采集的情况。

表1 工业大数据与互联网大数据区别

分类	工业大数据	互联网大数据
数据要求	全面样本数据,以覆盖工业过程中各类变化条件,时效性要求高	大量样本数据,时效性要求低
特征提取	注重特征背后的物理意义以及特征之间关联性的机理逻辑	依靠统计学工具挖掘属性之间的关系,不注重属性的具体含义
分析手段	数据建模、分析更加复杂; 需要专业领域的算法,不同行业、不同领域的算法差异很大; 强调跨学科技术的融合	具备成熟的数据挖掘算法; 轻属性含义、重价值挖掘,从看似无关的属性中找出内在价值
应用领域	健康诊断、故障预警、工况识别、市场预测等	图像识别、语音识别、语义分析、偏好推荐等

## 2 需求驱动变革

以上海华兴数字科技有限公司(以下简称华兴公司)为例,它是三一集团有限公司(以下简称三一集团)的全资子公司,为三一集团旗下工程机械装备(如挖掘机、旋挖钻机、装载机、履带起重机等)提供控制系统和信息系统。华兴公司从2007年开始机器物联的实践,由于设备控制器终端、硬件、语言学模型(linguistics model, LM)软件平台、组态软件、控制算法以及通信、传输、采集都是自主研发,因此可

以将任意所需数据实时上传至大数据平台。目前平台上有20多万台三一集团设备的5 000多种参数,低频数据上传周期为5 min,高频数据上传周期为20 ms,另外还有故障数据、维保数据以及相关事件数据等。同时开发出易维讯(E-Vision, EVI)系统的网页版与应用(App)版用于管理这些数据,并通过数据进行故障报警、故障预测、智能服务、辅助研发和信用管理等,很好地支撑了公司的核心竞争力,帮助公司实现了快速增长。

华兴公司工业大数据之路发展至今(如图1所示),每一项功能的开发都面向用户的实际需要。

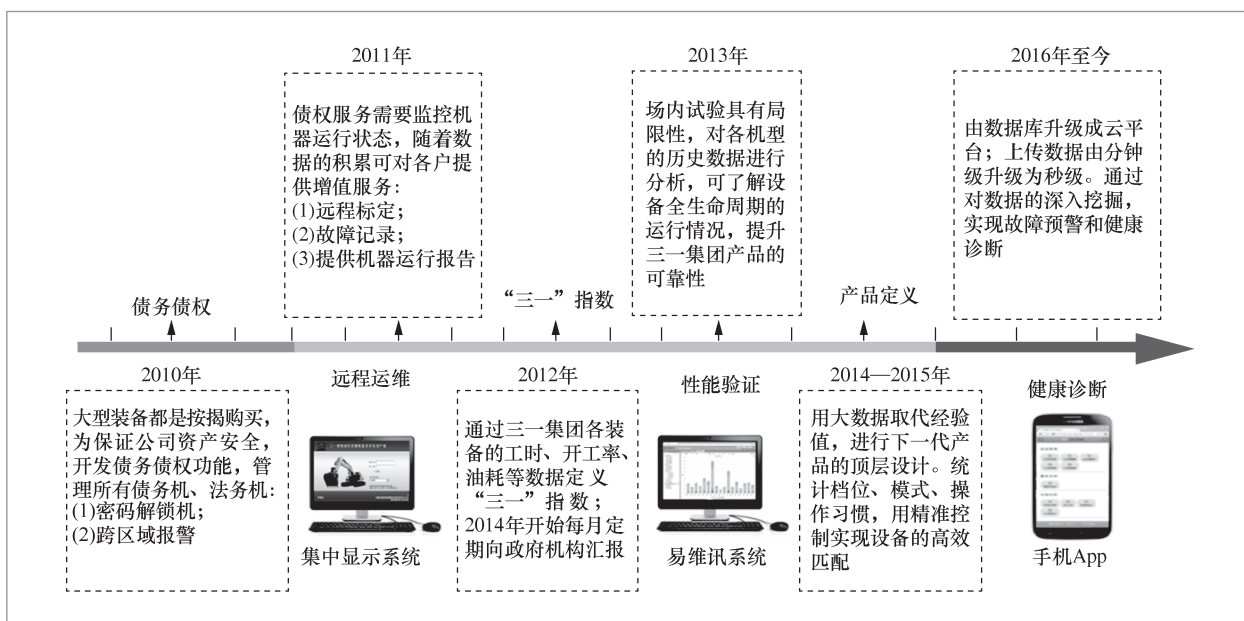


图1 华兴公司工业大数据发展历程

从满足用户的需求到通过数据为公司、用户创造价值，华兴公司的工业大数据经历了如下3个阶段。

#### (1) 集中显示阶段

工程机械的销售特点与其他商品不同，大部分都是贷款购买，按月还款。如果某台车不能按时还款，公司需要及时对其进行法务处理，因此需要知道每台贷款车的地理位置。基于这样的原因，华兴公司2007年开始为三一集团每台在外运行的设备安装全球定位系统(global positioning system, GPS)模块，并通过控制器将数据实时上传，记录位置信息的同时，也将机器运行的信息上传到服务器。

随着数据慢慢积累得越来越多，为了展示数据的统计结果，华兴公司2011年开始筹备开发易维讯系统。从开始面向公司内部提供机型分布、开工率、解锁机统计等定期的报表，到后期面向客户提供工时、油耗、周报、月报等功能，功能越来越全面，不仅为领导决策提供了支持，还提升了用户的设备使用体验。

#### (2) 统计分析阶段

易维讯系统的成功应用使公司重视了数据带来的潜在价值。三一集团的设备遍布全国，设备的运行状态某种程度即可反映国家基础设施建设以及宏观经济的情况。因此将易维讯系统中的数据进行整理，提出了“三一”指数的概念，2014年开始每月定期向政府机构提供该数据。

数据积累之前，公司对产品的定义和验证都基于有限样本进行，有了大量的数据以后，所有已售出的三一集团机器都可作为试验机，验证实际运行过程中其性能是否达到了设计要求。对于新产品的开发，也可以第一时间通过分析前期产品的使用情况，从数据中提炼出下一代产品定义的相关指标。

#### (3) 智能服务阶段

随着云计算、大数据技术的逐步发展，华兴公司从2016年开始升级了服务器，将原本的物理服务器全部转移至云上，通过数据中心操作系统(data center operating system, DCOS)统一管理，并将原来数据存储的最快间隔时间由分钟级提升至毫秒级，根据不同的需求，对数据进行分级存储。

通过这些技术以及一些成熟数据挖掘算法的使用，使数据产生了更大的价值，华兴公司开发出工况识别、操作手识别、故障预测、故障诊断、市场预测、配件预测等功能，为公司的决策和用户操作体验的提升提供了重要支持。

## 3 工业大数据平台

华兴公司的工业大数据平台建立在公司核心产品控制系统和信息系统的基础上，将运行设备(如挖掘机、风机、起重机、旋挖钻机等)互联，同时使企业数据(如生产数据、客户相关数据、产品生命周期数据等)与运行数据互通，通过3个平台(计算平台、业务平台与可视化平台)对数据进行存储、分析与展示，最终面向主机厂、用户、二手手机市场、配件市场等提供全方位的增值服务。平台整体架构如图2所示。

与市场上通用的平台即服务(platform as a service, PaaS)平台相比，华兴公司工业大数据平台是面向工业装备的专业化平台，数据从设备端采集上来，经过网关、存储、计算到客户端的应用展示，全程都是定制化的，满足了不同种类设备的需求。

### 3.1 数据采集

数据采集部分采用LM自主组态软件与不同频次数据的分级处理。

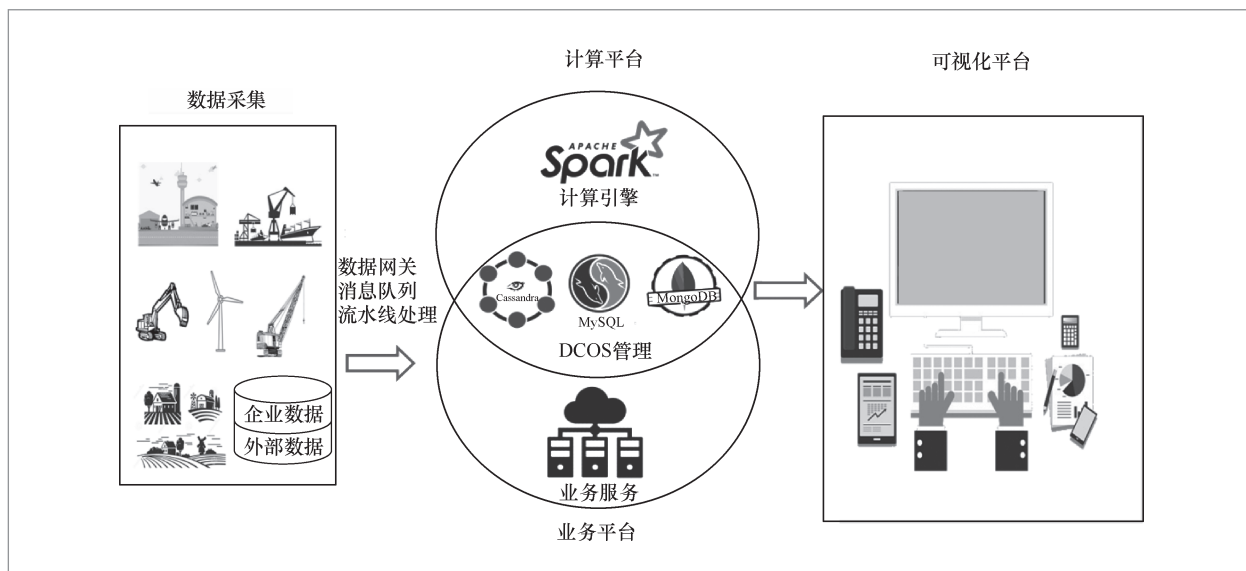


图2 工业大数据平台架构

远程控制终端由控制器和显示屏组成，是数据采集的核心部分，分布在每个工程机械设备上，负责接收来自GPS的卫星定位信息，并计算出工程机械装备的位置、速度、方向等信息。数据通过Wi-Fi/通用分组无线服务（general packet radio service, GPRS）通信网络传送至数据网关，通过消息队列进行流水化处理后，进入计算平台。

设备的核心控制程序由华兴公司自主开发的组态软件LM编写，因此除了GPS数据外，其他机器运行的数据（如通过控制器局域网络（controller area network, CAN）总线采集的发动机相关数据、控制器采集的设备传感器数据以及机器运行中的一些过程数据）都可以实时上传。针对振动等高频数据，在程序中提取高频信息的特征值，转为低频后再上传；针对故障数据，底层软件记录故障前后1 min的高频数据，打包后在非工作时段上传；针对设备运行数据（如主压、转速、排量等），采用秒级上传；针对设备的定位等低频数据（如位置、工时），则每15 min上传一次，

节省云平台的存储量。

### 3.2 计算平台

计算平台负责数据的快速存取和业务计算，由通信网关、消息队列、数据库、计算引擎等构成，软件栈如图3所示。

为了满足10万级台量每秒钟数据的实时接入需求，系统根据通信流量对通信网关进行水平伸缩，数据通过Kafka消息队列进行缓冲，再写入NoSQL数据库Cassandra。数据库包括工程机械装备详细的装备属性、工况信息、报警信息、故障信息、保养信息、维修信息、位置信息、处理后的音频/视频信息等。

计算平台使用高性能的Apache Spark作为计算引擎。原始数据通过Spark进行数据清洗，统计报表、故障预测、节能提示等业务需求则通过Spark SQL与Spark机器学习库（MLlib）进行计算，并将计算结果写入MongoDB，供可视化平台呈现。

平台运维采用Mesosphere公司的数据中心操作系统（data centre operation

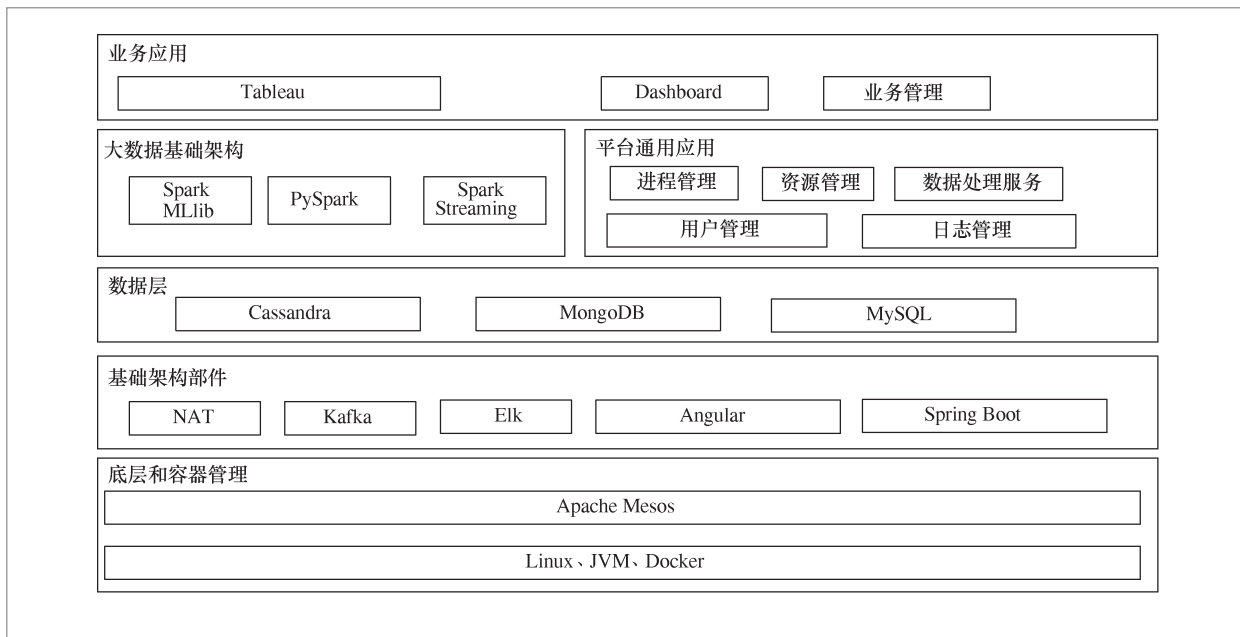


图3 计算平台软件栈

system, DC/OS) 进行管理。传统数据中心采用服务器和虚拟机的模型, 具有资源碎片化、管理和运维复杂、部署和更新过程冗长、资源利用率低等缺点。DC/OS将物理主机、虚拟主机聚合成资源池, 易于扩展, 支持多租户, 资源利用率高。DC/OS在统一的平台上对微服务容器、NoSQL数据库集群、Spark集群进行管理和调度, 生产环境高度可迁移, 避免了云服务商锁定, 更适用于现代化的运维管理。

### 3.3 业务平台

业务平台目前以挖掘机为主要对象, 同时也接入了风机、旋控钻机等其他三一集团装备。结合自身对业务理解的优势, 借助云计算和数据挖掘技术, 从故障维护、订单预测以及挖掘机工况识别等方面为不同客户提供更多智能服务, 如图4所示。

以挖掘机业务为例, 业务平台通过对挖掘机运行数据、App事件数据、厂内试验数据、服务订单数据、营销售后数据的联通,

精准打造挖掘机画像和用户画像, 如图5所示。其中, 挖掘机画像由单机画像与机群画像组成, 标签内容包括运行信息、统计信息、保养信息、故障信息、残值信息、健康信息等; 用户画像由代理商、操作手、服务工程师、挖掘机老板和研发人员组成, 标签内容包括人口属性、工作属性、性格属性、购物意向等。

工业大数据的价值主要体现在诊断和预测, 见表2。利用全方位联通的数据以及挖掘机和用户的画像, 面向不同客户选取与之匹配的标签进行计算。不仅对挖掘机, 也可对操作手、代理商、服务工程师等进行全方位的诊断, 同时诊断结果可用在保养、销售、二手机、库存、意向客户等方面的预测。

### 3.4 可视化平台

可视化平台从初始端的数据源到末端的客户端, 基于全息数据可视化构架以及分布式控制系统, 实现施工现场、企业控制中心、装配线智能车间监控以及公司会议设备等视频源的无缝衔接, 形成端到

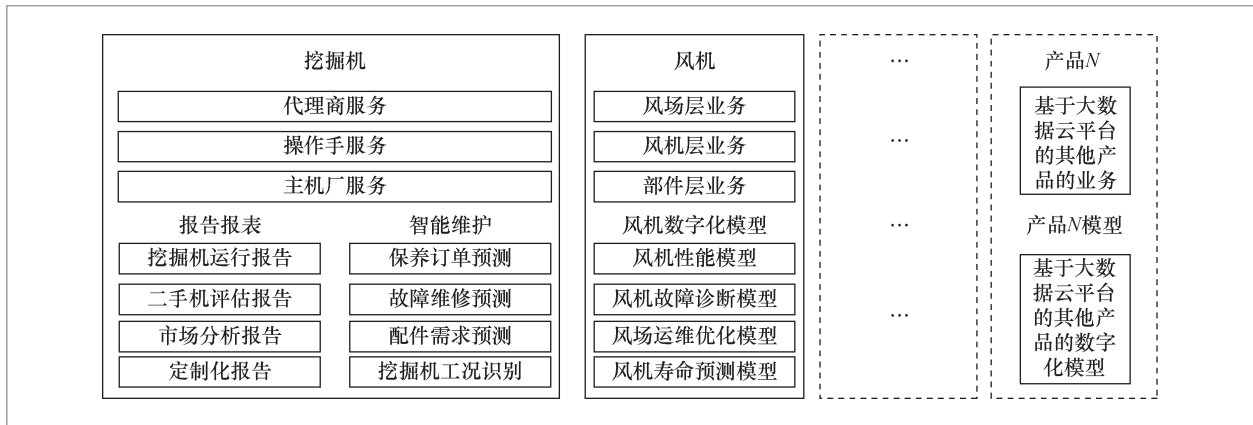


图 4 业务平台模块

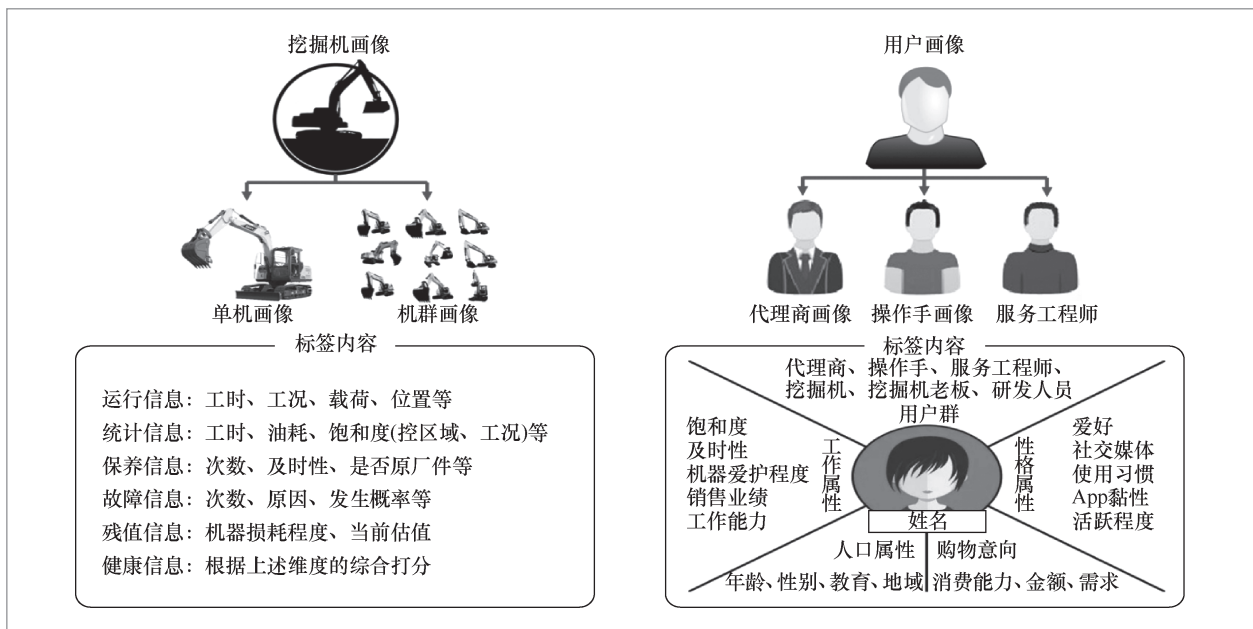


图 5 精准画像

表 2 健康诊断与需求预测功能

模块	功能	维度
健康诊断	挖掘机诊断	效能、易损件残值、核心件残值、状态信息
	操作手诊断	工作饱和度、机器爱护度、使用黏性
	代理商诊断	销售业绩、服务业绩
	服务工程师诊断	工作饱和度、工作能力、工作态度
需求预测	保养订单预测	精准预测保养订单数量, 合理安排工单
	故障维护预测	预测机器可能故障, 提醒客户提前采取措施
	整机销量预测	预测不同区域不同型号机器销量, 为主机厂提供决策支持
	配件销售预测	预测不同区域配件需求状态, 供代理商参考
	核心件库存预测	预测核心件消耗状态, 降低库存储备
	二手机价值预测	评估二手机残值, 为市场提供官方指导价格
	客户流失预测	根据用户画像和使用黏性, 预测客户流失, 并采取合理措施
	意向客户预测	根据用户画像和使用黏性, 预测客户购买意向

端的可视化和信息化管理。硬件设备可按需扩展,不会造成重复建设;软件部分模块化,数据源直接读取计算平台的结果数据,能够将各种数据、视频、音频信号以任意大小、任意位置方式呈现在任意的显示设备上,兼容主流浏览器,提供强大灵活的定制化显示。可视化平台拓扑如图6所示。

可视化平台主要面向管理层和客户提供不同的服务。面向管理层,通过数据可视化平台对全球范围挖掘机的主要关注指标进行透明化展示,便于公司管理层统一管理,发现运营短板;面向客户,通过数据可视化平台与实时运行的挖掘机数据对接展示公司的产品竞争力、高效的售后服务能力与运维能力,进一步提高企业形象、品牌价值和客户满意度。

服务内容包括决策、显示和互动。

(1) 决策

平台可与公司现有视频设备兼容,若办公室已安装视频设备,不但可参加公司内部会议,还可将云平台信息以双流形式传送给其他参会会场,所有外接信号均可上屏显示,协助快速精准决策。

(2) 显示

在各个园区的监控中心可通过大屏实

时获取云平台所有展示信息,并将信号源以双流形式进行传输;可查看不同工厂的监控视频源,也可查看各施工现场以及主机内部摄像头的视频源。

(3) 互动

可视化信息可在多个地点的大屏幕上任意调度显示,iPad实现各个装备之间数据的联动;采用浏览器/服务器(B/S)架构,兼容主流浏览器,随时随地访问可视化资源,可接入不同的子系统,任意分割画面播放,自定义模版功能,支持多屏显示功能,每个屏幕显示不同的数据和画面,显示不同三一集团装备的实时数据画面等,通过触摸屏一键式切换大屏进行互动。

## 4 工业大数据应用

华兴公司在工业大数据上的应用主要体现在集中显示、统计分析和智能服务方面。

### 4.1 集中显示

数据集中显示的载体就是易维讯系统,该系统是华兴公司为工程机械领域打造的

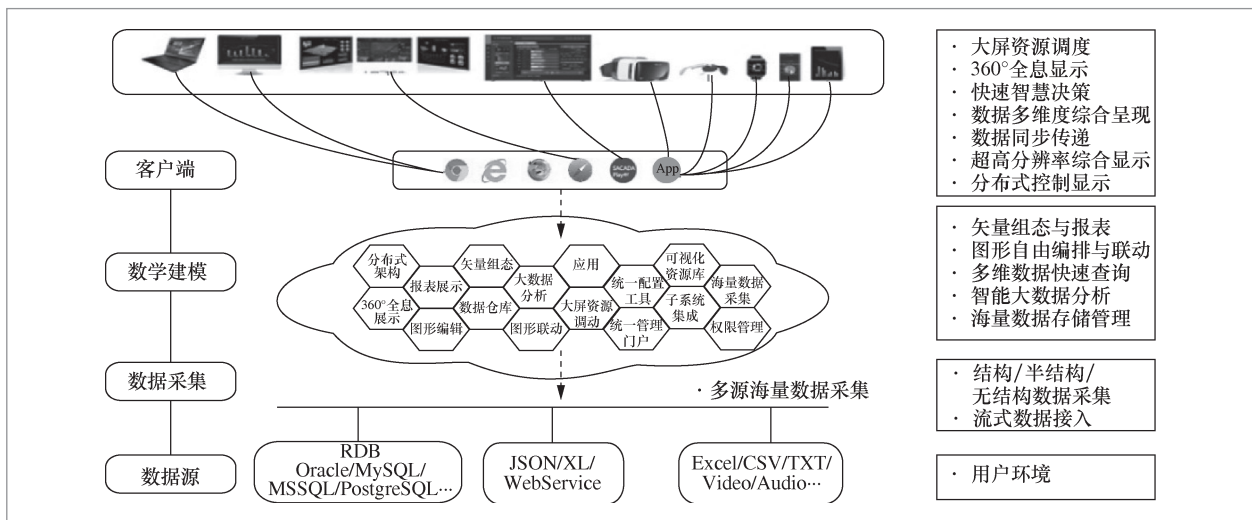


图 6 可视化平台拓扑

信息控制大数据平台。运行至今，平台上共有设备20万余台，每日实时监控其运行信息（如位置、工时、转速、主压、油耗等），是三一重机股份有限公司旗下挖掘机、旋挖钻机 and 装载机等设备全生命周期的“眼睛”。

易维讯系统从2013年12月31日上线至今，共12次改版，其中正式版升级5次。最初版本提供4大类42项子功能，目前系统面向代理商、操作手、挖掘机老板和研发人员4类主要用户提供99项功能，如图7

所示。除网页版外，目前已开发出iOS版和Android版手机App，使用户可随时随地掌握机器的运行状态，极大提升了用户的操作体验以及对设备的满意度。

用户可从易维讯系统中了解市场信息、设备运行信息、保养信息等，如图8、图9所示。挖掘机老板可以查看设备实时运行状态，了解操作手的工作情况，如图10所示。对未按时还贷款的客户进行远程锁机，以保障设备安全，如图11所示。对偏远

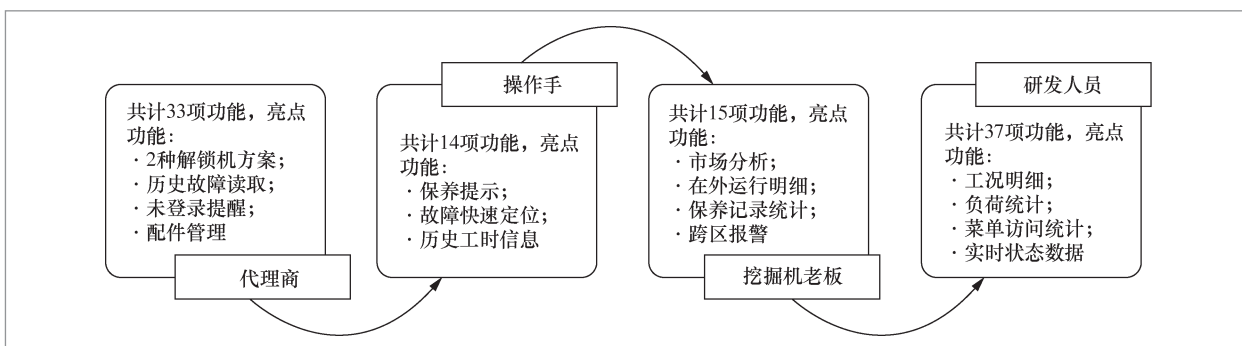


图7 易维讯系统功能分类

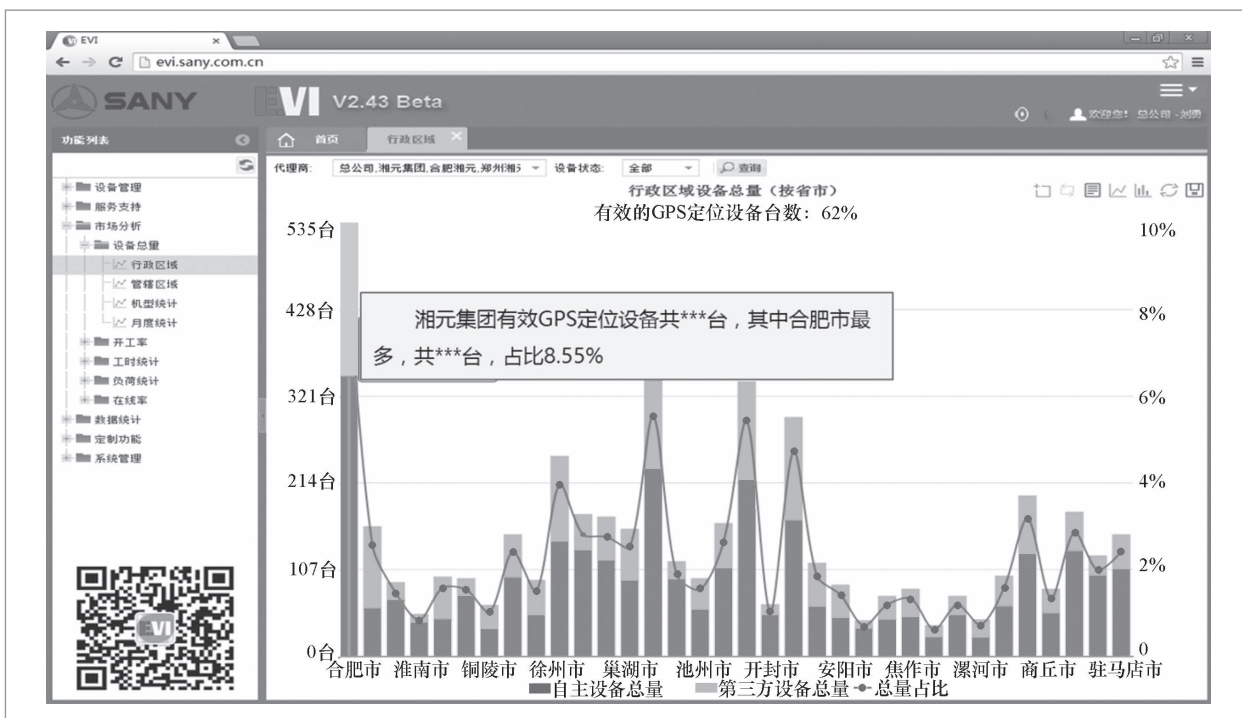


图8 市场分析——行政区域台量统计

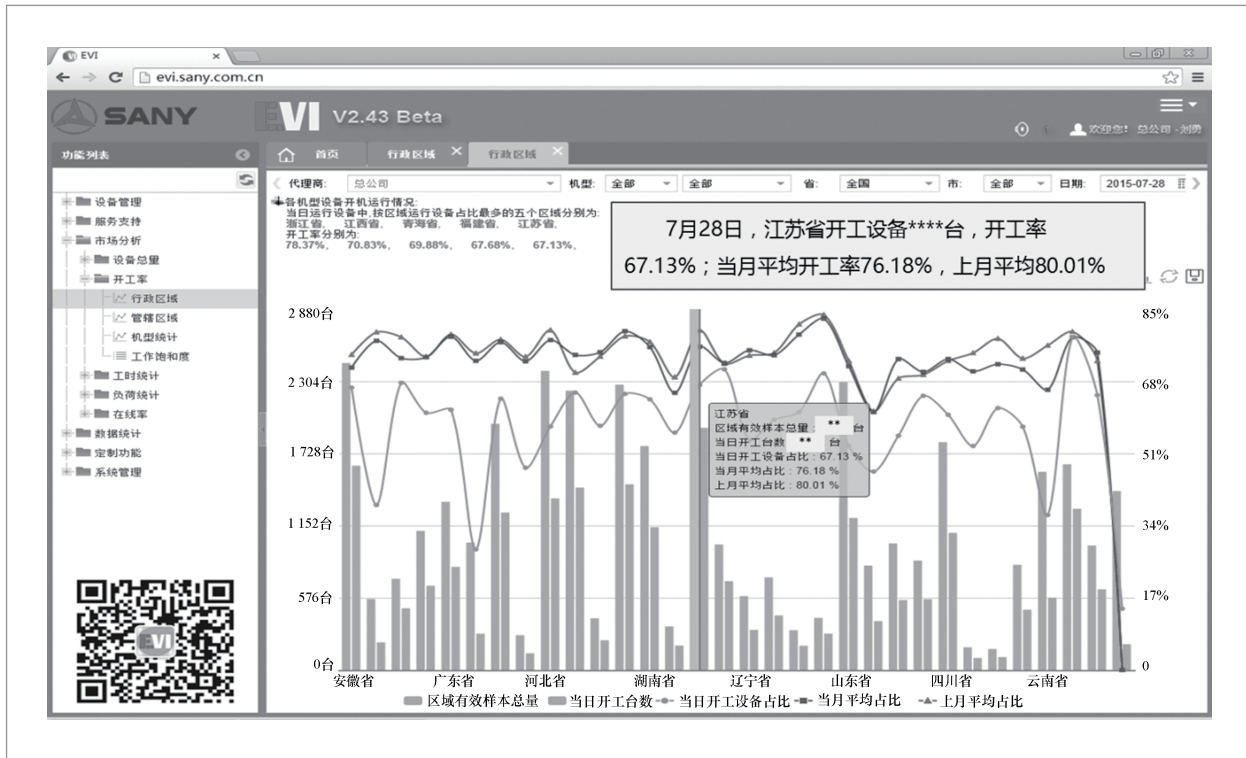


图9 市场分析——行政区域开工率统计



图10 设备管理——实时工况查看

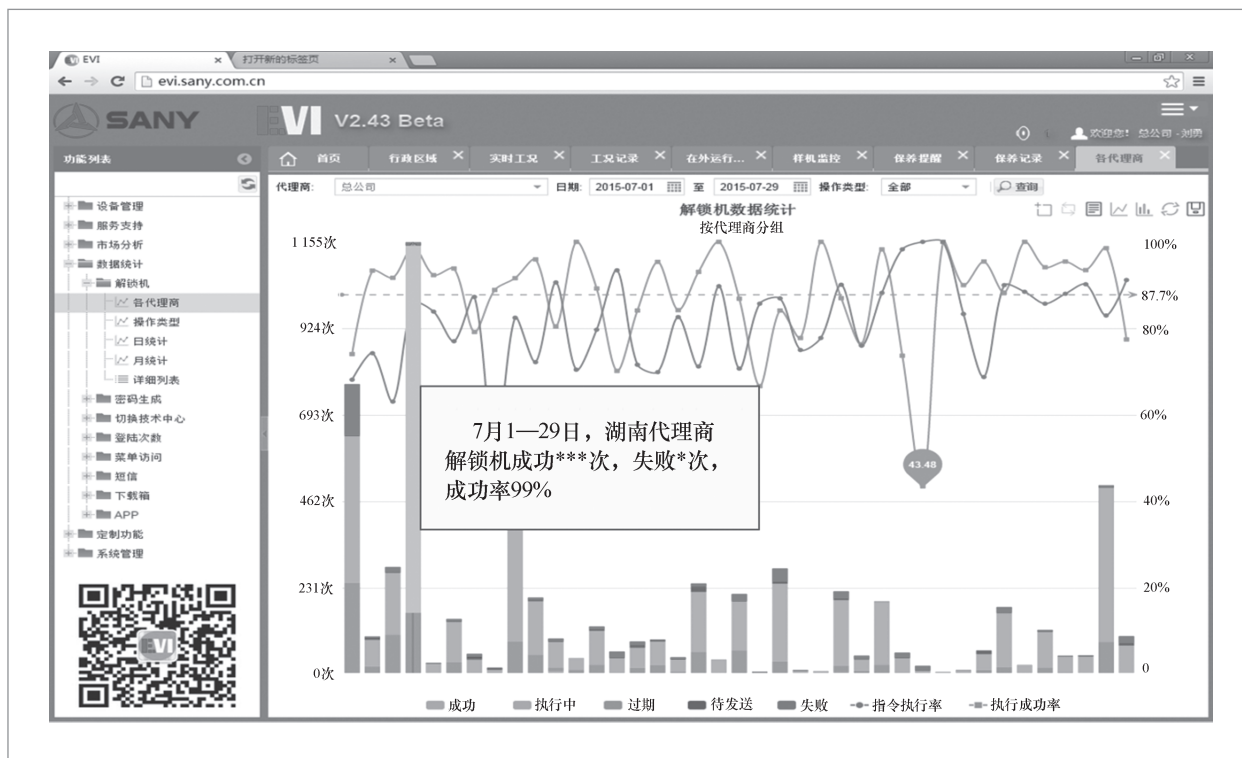


图 11 设备解锁机管理

山区工作的机器进行远程调试,以最快的速度解决客户的问题。

## 4.2 统计分析

### (1) “三一”指数

“三一”指数是以三一集团工程机械设备的海量工况数据为基础,选择一系列反映三一集团工程机械设备作业状况的指标,进行综合处理,用以反映工程机械经济景气活跃程度的综合指数,也是借助工业大数据平台打造出来的第一个面向国家,反映宏观经济和基础设施建设的统计值。大数据和云计算技术使一台台三一集团装备的运行数据实时反馈在易维讯平台上,每挥动一铲、行动一步,都形成数据痕迹。海量机械的应用场景和开工率等来自一线的真实情况,挖掘机、旋挖钻机、泵车、搅拌车等全国范围内的分布、开工率、

工时的热力图几乎就是一张基础设施建设行业的活力图。从2014年开始,每月定期向政府部门上报“三一”指数数据。

### (2) 辅助挖掘机升级

2015年西南地区某三一机型用户集体抱怨挖掘无力,而其他区域用户对挖掘力的评价很好。传统的办法是派遣多名服务工程师到各个使用者处调研,了解情况。现在采用大数据技术方法可以提高效率。由于用户很多,逐一调查成本太大,因此通过海量数据逼近问题的原因,事先找到解决问题的办法。

首先,通过数据统计该型号挖掘机用户使用的档位,如图12所示。发现华东、中南地区7档使用频率最高,11档频率使用最少,载荷最轻。华北、西北、东北地区8档使用频率高,载荷较轻。西南地区9、10档使用频率高,11档使用频率也远高于其他地区,判断载荷最重。

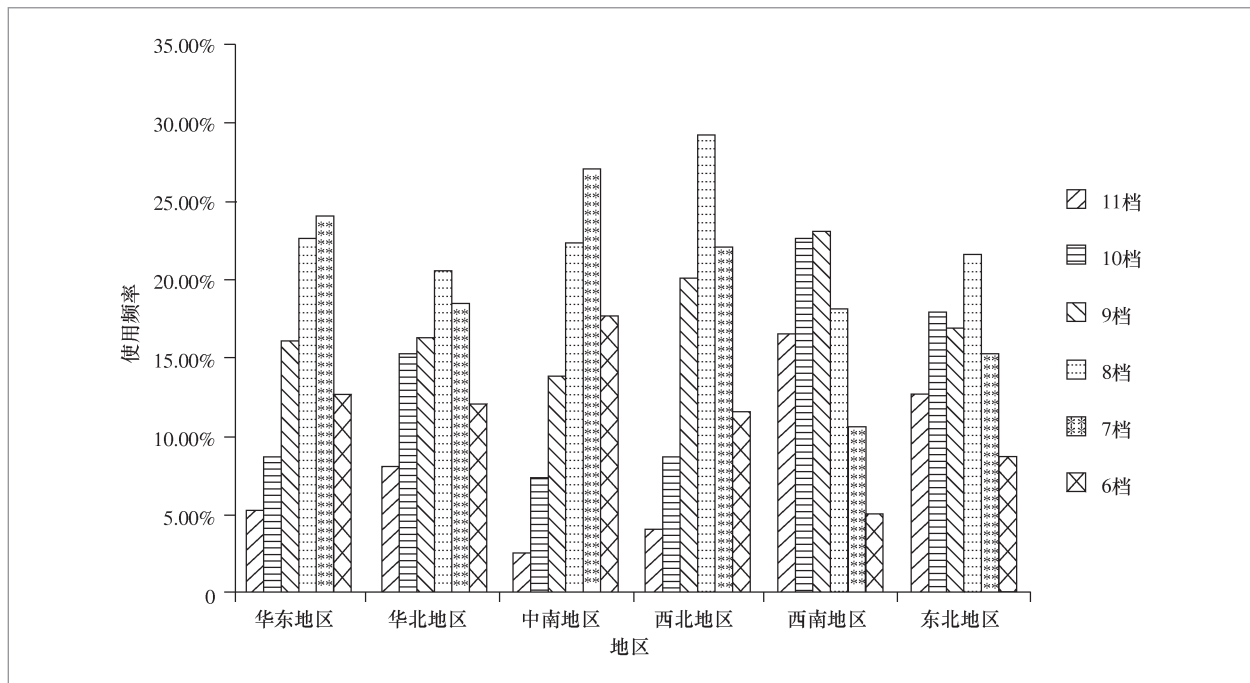


图12 档位分布

从台数上看, 该种机型主要卖给西南和西北地区, 占比最高; 从主压上看, 西南地区的主压也是最高的; 从输出功率上看, 西南地区功率消耗最大, 如图13所示。因此判断该地区用户注重效率而非油耗, 所以使用11档频率高。需要针对该机型的配置进行精准优化, 以适应西南地区的载荷。

华兴公司根据西南地区某三一机型的载荷特点, 确定“经济模式(S模式)降油耗, 重载模式(H模式)提效率”的原则, 采用精准控制算法, 提升H模式的匹配使其更贴近发动机的最佳经济区, 同时增大泵的流量与恒功率曲线, 使其更充分吸收发动机的功率, 最终达到节能增效的目的。

升级后的程序通过易维讯平台远程下发给各主机, 用户使用后, 均表示满意。

### 4.3 智能服务

#### (1) 挖掘机工况识别

挖掘机的工作原理是将发动机的机

械能转化为液压能, 再通过复杂的液压回路将液压能转化为机械能, 从而实现整机作业的过程<sup>[8]</sup>。挖掘机工作的过程中环境恶劣、负载随机、工况多变, 不同工况下的匹配模式不同, 设计要求不同, 元件的故障概率也不同。因此精确掌握在外挖掘机的工况对研发支持(如故障诊断和产品升级)、售后服务(如二手机评估和寿命预测)都有重要的作用。传统对工况的区分都是采用服务工程师现场确认的方式, 只能针对个别机器进行少量的采样。大数据技术的发展, 使全面了解各种机器的工况变成了可能。

由于在外运行的挖掘机没有工况标签, 无法进行训练, 因此在三一集团试验场内对大、中、小挖掘机各2款机型进行8种常见工况的测试。为了排除操作手的影响, 每个挖掘机都由多名操作手进行测试, 每种工况连续测试1 h, 同时为8种工况分别打上标签, 并保存测试数据, 供训练使用。由于训练好的模型需要应用在全

部在外运行的挖掘机上,因此数据采集时不能单独加装传感器,全部使用量产挖掘机自带的传感器。记录下来的数据按照均值类特征、波动类特征和占比类特征进行分类,归一化后进入训练模型。使用多种训练模型,如感知机、逻辑分类、决策树、神经网络、支持向量机、随机森林等,根据初步测试的结果选取模型进行细节调优,再用实际工况数据进行测试,最终模型工况识别的准确率在85%以上。

### (2) 风机健康评估

华兴工业大数据平台不仅接入了挖掘机等工程机械数据,还接入了三一集团风机的全部数据。以风机整体运行状态为例,采用健康值对其进行评估。健康值是0~1的连续值,其原理如图14所示,“1”表示当前运行数据分布完全与健康数据分布重合,随着系统之间衰退,测试数据分布逐渐偏移,两个分布的重叠部分逐渐减小,健康数值也随之减小。根据工程经验设定的阈值用于判断系统是否衰退到了需要进一步进行故障诊断的级别。

利用风机从切入风速到切出风速的功率曲线,使用多工况建模的手段建立风机在健康状态下的功率随风速变化的分布模型。在健康状态下,风机的风功率散点图应分布在设计的功率曲线周围,而当风机出现性能衰退或故障时,风机的功率分布将偏离设计的功率曲线,如图15(a)所示。另外,通过建立功率与风速的关系模型,对比当前风速下预期发电功率和实际发电功率之间的差异,就可以评估由于风机衰退造成的发电量损失,如图15(b)所示。通过对比维修或维护前后功率曲线的变化以及发电能力的恢复情况,就可以评估该维护或维修活动的效果。

计算过程中,首先进行数据的过滤、分割和归一化,然后基于专家知识从数据采集与监视控制(supervisory control and

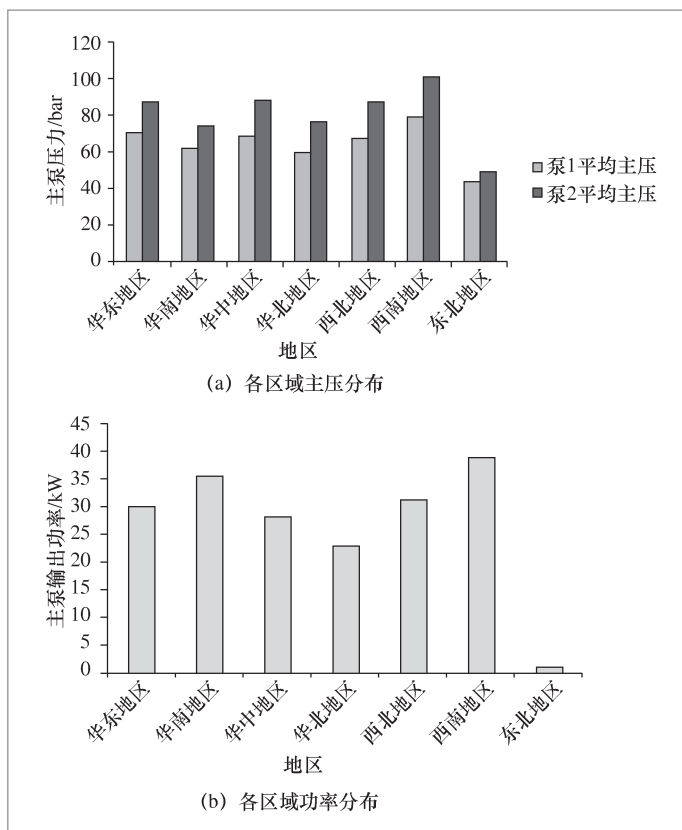


图 13 负载分布

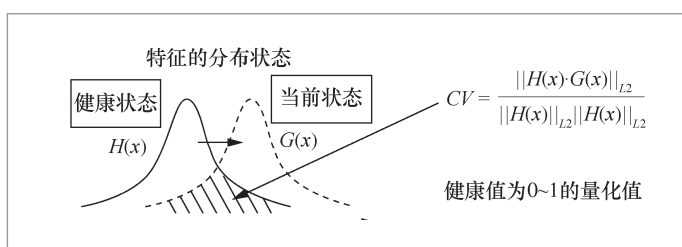


图 14 健康值计算方法

data acquisition, SCADA) 系统的多个变量中选取与功率曲线分析相关的参数,输入多工况建模算法,如高斯混合模型(Gaussian mixture model, GMM)、自组织图(self-organizing map, SOM)、神经元网络(network of neuron, NN)等,再使用相对应的距离算法计算出评估性能差异的模型距离值,如L2、MQE和残差分析等。

部件级别的健康管理采用的方法与上面类似,只是针对不同部件选择的参数和

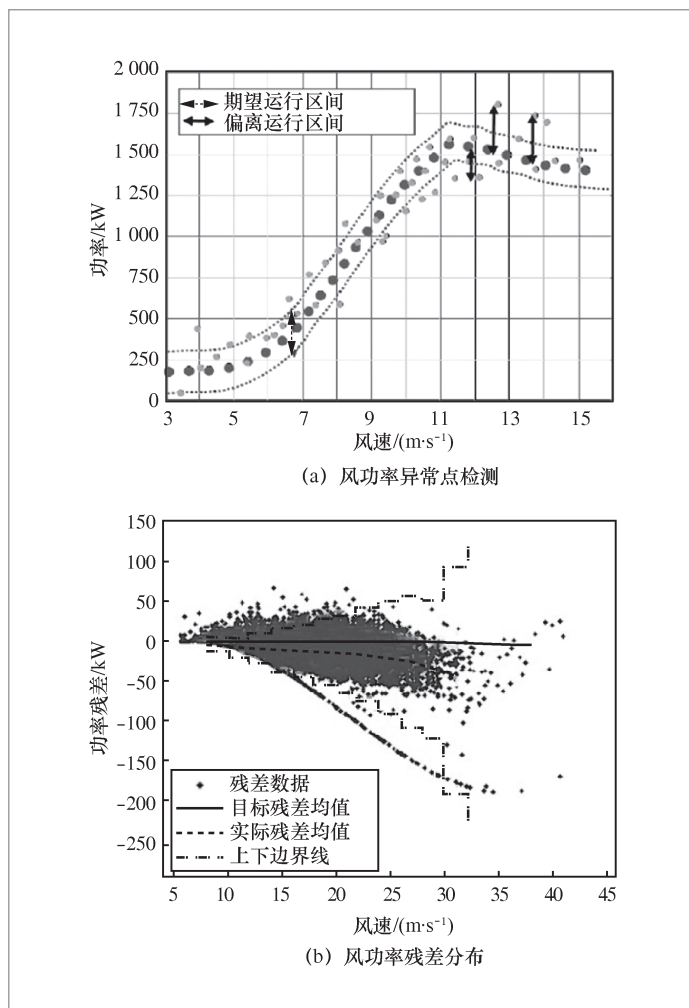


图 15 风功率健康值计算方法

健康分析建模方法有所不同。如传动链的健康值计算需要振动状态监测系统中的振动数据以及SCADA中的转速、温度等信息，变桨系统则需要SCADA数据中的转速、风速、桨距角等信息。

## 5 结束语

工业大数据价值创造的序幕刚刚开启，不仅在于对现有业务的优化，更在于支撑企业、行业乃至全社会的创新、转型和发展。虽然华兴公司在工业大数据的

应用方面已有了一些基础，但目前整体尚处于起步阶段。一方面数据的采集还需要根据应用目标提升数量和质量，另一方面更多零部件故障预测模型还有待建立，以形成可以实用的大数据模型。工业大数据的特点是数据分析要与工业知识体系和逻辑相结合，未来的知识体系不是依靠专业人士的经验，而是以数字的方式保存下来，需要在持续的运营中将知识积累、分享并传承下去。同时不管三一集团的规模有多大，华兴公司记录的数据有多少，一个企业积累的数据都很难定义为“大数据”，需要有更多外部数据和应用模型的加入，在行业内形成共享，从而基于数据的应用才会更好。

## 参考文献:

- [1] 王建民. 工业大数据实践与思考[N]. 中国信息化周报—大数据, 2016-05-30.  
WANG J M. Practice and thinking of industrial big data[N]. China Information Weekly—Big Data, 2016-05-30.
- [2] 张曙. 工业4.0和智能制造[J]. 机械设计与制造工程, 2014, 43(8): 2-4.  
ZHANG S. Industrial 4.0 and intelligence manufacturing[J]. Machine Design and Manufacturing Engineering, 2014, 43(8): 2-4.
- [3] 王喜文. 工业大数据的四种用途和两大价值[J]. 物联网技术, 2016(4): 7-8.  
WANG X W. Four uses and two major values of industrial big data[J]. Internet of Things Technologies, 2016(4): 7-8.
- [4] 李杰. 工业大数据—工业4.0时代的工业转型与价值创造[M]. 北京: 机械工业出版社, 2015.  
LI J. Industrial big data—the revolutionary transformation and value creation in industry 4.0 Era[M]. Beijing: China Machine Press, 2015.
- [5] 陈尚义. 百度大数据应用与实践[J]. 大数据, 2015(1): 2015009.

- CHEN S Y. Big data applications and practices of Baidu[J]. Big Data Research, 2015(1): 2015009.
- [6] 郑树泉, 覃海焕, 王倩. 工业大数据技术与架构[J]. 大数据, 2017(4): 67-80.  
ZHENG S Q, QIN H H, WANG Q. Industrial big data technologies and architecture[J]. Big Data Research, 2017(4): 67-80.
- [7] 李杰. 从大数据到智能制造[M]. 上海: 上海交通大学出版社, 2016.
- LI J. From big data to intelligent manufacturing[M]. Shanghai: Shanghai Jiao Tong University Press, 2016.
- [8] 岳滨楠. 液压挖掘机整机性能的数字化匹配研究[D]. 上海: 同济大学, 2012.  
YUE B N. Research on digital match of hydraulic excavator overall performance[D]. Shanghai: Tongji University, 2012.

## 作者简介



**袁爱进** (1959-), 男, 三一集团有限公司教授、副总裁, 主要研究方向为智能装备、智能控制、工业大数据。



**岳滨楠** (1984-), 男, 博士, 上海华兴数字科技有限公司算法工程师, 主要研究方向为数据挖掘算法。



**闫鑫** (1981-), 男, 上海华兴数字科技有限公司副总经理, 主要研究方向为智能控制算法、工业大数据实现。



**黄健** (1972-), 男, 上海华兴数字科技有限公司工业大数据部门经理, 主要研究方向为工业大数据平台。

收稿日期: 2017-09-04