

DOI : 10.3901/JME.2019.20.003

大数据分析技术在新能源汽车行业的应用综述 ——基于新能源汽车运行大数据*

余承其 张照生 刘 鹏 孙逢春

(北京理工大学电动车辆国家工程实验室 北京 100081)

摘要：为应对化石能源危机和环境污染问题，新能源汽车技术的发展与应用引起广泛重视。新能源汽车具有高信息化的特点和智能网联化的发展趋势，在日常运行中会产生大量行驶数据信息。利用海量多源异构数据进行安全预警与监管、车辆技术分析，是推动我国新能源汽车行业发展的关键。综述了大数据分析技术在新能源汽车行业的应用情况，概述了大数据分析技术的基础理论、发展历程，介绍了新能源汽车国家监测与管理平台的架构和功能，并着重阐述了新能源汽车大数据分析过程。分别从动力电池数据、汽车运行数据和充电数据的角度出发，分析了现有的研究方向和研究方法，列举了部分研究成果和应用情况。最后，对当前新能源汽车大数据分析领域存在的问题和发展应用前景进行了总结和展望。

关键词：新能源汽车；大数据平台；大数据分析技术；动力电池

中图分类号：U469

Overview of the Application of Big Data Analysis Technology in New Energy Vehicle Industry : Based on Operating Big Data of New Energy Vehicle

SHE Chengqi ZHANG Zhaosheng LIU Peng SUN Fengchun

(National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing 100081)

Abstract : New energy vehicle (NEV) has been widely used around the world in response to the fossil energy crisis and environmental pollution problems. NEV will generate massive real-world data during its daily operating which is contributed by high electrification and intelligent networking. Applying these multi-source heterogeneous data for a security warning and technical analysis will play a key role in promoting the development of NEV industry in China. The current situation of data-driven analysis technology in the NEV field is reviewed. Firstly, the basic theory of big data analysis techniques are introduced and the development of big data technology is depicted. The structure and function of the National Monitoring and Management Platform for New Energy Vehicles are introduced, and the particular process of big data analysis on NEV is emphasized. The previous data-driven research and methods in power battery, NEV daily operation and charging behavior are proposed for discussion respectively. Some representative research results and applications are displayed at the same time. Finally, the issues and prospects of the data-driven method on NEV application field are summarized and forecasted.

Key words : new energy vehicle ; big data platform ; data-driven methods ; power battery

0 前言

随着化石能源枯竭威胁的日益迫近以及大气污染等环境问题逐渐恶化，发展以电动汽车为代表的新能源汽车成为世界各国的普遍共识^[1-2]。我国政府在 2010 年发布的《国务院关于加快培育和发展战略性新兴产业的决定》，将新能源汽车列为七大战略性

新兴产业之一。受国家政策的支持和引导，2018 年我国新能源汽车产销量分别为 127 万辆和 125.6 万辆，连续第四年成为世界新能源汽车产销量第一大国。新能源汽车数量增长的同时，安全问题也日渐突出^[3]。仅 2018 年内，我国已发生新能源汽车自燃事件 40 余起，其中绝大部分事故由动力电池故障引发。同时，2018 年我国新能源汽车召回数量已超过 13 万辆，安全问题正阻碍新能源汽车产业的健康发展^[4]。

伴随新能源汽车车载信息系统的逐渐完善，车

* 科技部资助项目(2017YFC0840205)。20190607 收到初稿，20190906 收到修改稿

辆信息化程度日益提高,在车辆日常行驶过程中将会积累大量结构化数据、非结构化数据、历史准实时数据等^[5]。为保障居民日常出行安全和推动新能源汽车产业发展,新能源汽车国家监测与管理平台于 2016 年在北京建立。国家平台接收海量新能源汽车实时运行数据,配合车辆档案数据、气象数据、道路建设数据等形成新能源汽车多源异构数据库,并结合数据资源和大数据分析技术,挖掘数据背后蕴藏的丰富价值,以改善当前新能源汽车安全问题和促进新能源汽车产业发展。

大数据是当前最热门的研究领域之一,得到了全世界的广泛关注。在学术界,《Nature》和《Science》先后推出了数据处理专刊,讨论数据洪流为科研领域带来的全新机遇与挑战^[6]。在工业界,麦肯锡研究院指出大数据将会是下一个创新、竞争、生产力提高的风向标^[7],奥迪为代表的传统车企正加快大数据技术的应用,以求在产品更新、商业模式探索和数字化服务等方面取得提升^[8]。在政府层面,美国政府开放了超过 40 万条政府数据,同时颁布《大数据研究和发展计划》等政策来促进数据开放共享^[9]。欧盟也宣布将采取一系列具体措施发展大数据技术^[10]。我国政府也高度重视大数据技术的应用与发展。2015 年国务院印发了《促进大数据发展行动纲要》,指出数据已经成为国家重要的战略资源^[11]。2017 年,习近平主席就实施国家大数据战略发表重要讲话,强调推动实施国家大数据战略,更好服务我国经济社会发展和人民生活改善。

大数据技术在新能源汽车领域也展现出良好的应用前景,新能源汽车和大数据的融合是未来大趋势,也是我们国家汽车行业转型升级的战略重点方向。当前,新能源汽车运行大数据分析技术主要应用于动力蓄电池领域,主要包括动力电池故障诊断与热失控预警。此外,运行大数据在车辆运行状况监控、关键零部件检测、新能源汽车用户研究、基础设施建设布局分析等方面也有一定的应用基础。例如,对新能源汽车整车可靠性及故障规律的统计分析研究^[12],对电动出租车充电站合理选址方法的研究^[13],以及对新能源汽车充电需求与电网负荷预测的研究^[14]等。然而,早期新能源汽车领域的相关研究多基于严苛控制的实验室环境,缺乏对车辆实际运行环境的考量,同时对大数据处理的应对经验不足,在实车运行数据应用、真实环境下整车及动力电池性能研究等方面还亟待提高^[15]。

本文综述基于实车运行数据的新能源汽车大数

据分析技术的应用与发展情况,结合新能源汽车国家监测与管理平台的研究现状与工作基础,详细介绍当前新能源汽车大数据分析技术的成果和未来方向。本文简介了新能源汽车大数据基础,描述了大数据平台数据结构、数据采集与处理过程,全面剖析了动力电池大数据、行驶大数据、充电大数据的研究应用现状。最后,总结了当前研究工作中存在的不足,讨论了新能源汽车大数据分析技术的发展方向和应用前景。

1 新能源汽车大数据基础

传统数据分析领域主要基于表格数据等结构化数据进行研究,方式较为固化。而伴随着计算机、物联网等技术的兴起,大量非结构化数据,如图像、声音、视频等开始涌现,数据规模也呈现爆炸式增长趋势。充分了解大数据基础,准确把握大数据特征,将有助于挖掘海量数据的内在价值。

1.1 大数据基本特点

1890 年,美国统计学家 HOLLERITH 发明的一台电动机器来统计美国人口普查数据,并在一年内完成了预期八年的工作,这被认为是最早的大数据方法应用实例^[16]。20 世纪末,人类社会步入计算机与互联网时代,数据也进入爆炸增长时期,推特上一天产生的信息量相当于《纽约时报》100 多年产生的数据总量^[17]。2008 年全球数据量仅为 0.49 ZB,而 2017 年全球数据量为 21.6 ZB,已是 2008 年的 44 倍,研究人员预估,到 2020 年,全球数据量将达到 35 ZB^[18]。以当今著名互联网企业为例,Google 公司每月需要处理近 100 PB 的数据,淘宝网上每天的在线交易数据就达到了 10 TB^[19]。

然而,大数据不仅是规模庞大的数据集合,简单从数量上定义大数据有失偏颇。2001 年,META Group 的分析员 LANEY 提出,未来大数据管理主要有 3 个方面的挑战,数据规模(Volume)、数据处理时效性(Velocity)与数据多样性(Variety)^[20]。在此基础上,部分研究者补充了结果准确性(Veracity)和价值深度(Value)两个概念,形成了大数据的“5V”特点^[21]。大数据的“5V”特点主要体现在其处理、计算和储存过程中。然而,传统技术无法胜任大数据分析处理,也无法实现大数据实时在线计算,同时传统数据处理技术多基于结构化数据,无法应对文本、图片和媒体等非结构化数据。大数据处理需求和计算特征如表 1 所示,发展大数据处理技术是解决当前数据处理需求的有效途径。

表 1 大数据处理需求和计算特征分类

分类角度	分类	
数据存储管理	半结构化/非结构化	结构化
数据分析类型	传统查询分析计算	复杂数据分析挖掘处理
数据获取处理方式	批处理	流式计算
数据处理响应性能	非实时/线下计算	实时/联机计算
数据关联性	简单关联数据	复杂关联数据
体系结构特征	基于集群的大规模分布式存储，并行化计算	

1.2 大数据处理技术

在现代化大数据处理架构出现前，技术人员借助 MPI(Message passing interface)编程模型和方法处理大规模数据^[22]。MPI 是一种高性能并行消息传递接口，是当时主要的数据编程与计算载体，它能够充分利用硬件资源进行并行计算，被广泛应用于物理、气象等领域。

由于 MPI 方法缺少良好架构支撑，且有自动化程度低、程序设计复杂、程序员任务重等缺点，研究人员开发了 Hadoop MapReduce 处理系统。MapReduce 主要面向大规模数据的并行处理计算，最先由 Google 的研究团队开发，以供内部员工进行数据处理^[23]。之后，Apache Nutch 的技术团队将 MapReduce 拓展为基于 Java 语言的开源并行计算框架系统 Hadoop MapReduce，降低该技术的门槛以便大众使用^[24]。Hadoop MapReduce 凭借其出色的任务调度、数据恢复和系统优化功能，成为了目前主流的大数据处理系统，普遍应用于学术和工业界。

MapReduce 是为数据线下批处理而设计的，当需要在线快速处理数据时，MapReduce 效率较低。2013 年开发的 Spark 大数据处理系统吸收了 Hadoop MapReduce 的优点，并大幅提高了并行计算性能，弥补了后者在数据实时计算方面的不足，使得现代大数据分析技术更趋完整^[25]。Spark 诞生之初，采用较专业的函数式编程语言 Scala 作为开发语言，给 Spark 的使用和推广带来了限制。随着多种常用编程语言(例如 Python 和 R)支持功能的加入，以及数据结构 Dataset 的更新，Spark 逐渐为广大数据研究人员所接受^[26]。

除 Spark 之外，源自欧洲的 Flink 也是常用的并行化大数据处理系统。Flink 同时支持流式计算和批处理计算，有丰富的数据转换接口^[27]。与 Spark 不同，Flink 有着独特的存储管理机制，可大量节约运算空间，同时它能自动对程序进行优化，避免冗余结果缓存。它提供 Java、Scala、Python 等多种编程语言接口，为进一步方便用户使用，Flink 还提供表格计算、复杂事件处理等大数据计算库，可与其他主流处理系统集成使用，用户能根据自身实际需求

灵活、便捷地选择相应处理系统。

以上大数据处理系统按照处理对象和处理形式可分为四类。批处理系统、流式实时处理系统、实时交互式查询系统、图数据处理系统^[28]。具体分类情况如表 2 所示。

表 2 主流大数据处理系统分类

分类	代表系统
批处理系统	MapReduce, Hadoop
流式实时处理系统	Flink, Storm
实时交互式查询系统	Spark, Dremel
图数据处理系统	GraphLab

尽管当前大数据处理技术发展迅速，但仍存在以下难点。

(1) 大数据集成技术需解决数据的广泛异构性和数据质量两个关键问题。首先是数据的广泛异构性，传统结构化数据转化为结构化、半结构化、非结构化数据的融合；数据源扩展为手机、笔记本电脑等便携式移动设备，赋予了数据时空特性^[29]；多源异构数据的数据转化过程复杂，数据储存与管理难度增加。其次是数据的质与量之间的平衡，数据量无序扩张可能引起垃圾信息的泛滥，增加数据处理成本，干扰数据分析过程^[30]；数据清洗颗粒度过粗将不能有效去除噪声数据，降低数据清洗效果，颗粒度过细则导致清洗过度，易丢失有价值的信息。

(2) 当前大数据分析技术仍存在三方面不足，即数据库架构单一、数据计算能耗大和先验知识缺乏。由 MapReduce 和 Hadoop 为代表的数据处理系统多用于离线数据线下批处理，Spark 和 Flink 为代表的数据处理系统推广程度不够，需研发更高效便捷的数据架构以应对海量数据在线处理的需求；全世界以数据处理业务为核心的企业产生的碳排放量以每年 6% 的速率增加，到 2020 年将占到全球碳排放总量的 12%^[31]，由海量数据存储、计算及通信产生的能耗和碳排放问题亟待重视；针对非结构化数据的实时处理分析仍缺乏相关研究经验。例如，传统公交车等候时间研究以结构化等待时间数据为基础，当结合视频信息分析后，研究人员发现实际等待时间是调查结果的 1.21 倍，与单纯的结构化数据研究结果并不相符^[32]。

(3) 大数据时代，数据资源利用与隐私保护间的矛盾必将长期存在。虽然在 21 世纪初，已有研究者通过决策树分类方法掩盖数据中的敏感信息，平衡个人隐私保护和数据挖掘研究^[33]。但计算机的普及和互联网的发展扩大了数据传播途径和速度，隐私数据暴露的问题逐渐增多，社交网络上的“人肉

搜索”事件即是这种问题的体现^[34]；数据公开与隐私保护的矛盾也在激化，企业需要从相关数据中了解客户以实现个性化服务，科研人员需要利用数据进行研究，政府机构也需要统计数据来掌握国民经济运行状态。虽然 SMITH 和 XU^[35-36]等提出了一些解决方案，在满足数据分析需求的同时兼顾隐私保护，但由于大数据的维度和内容时刻发生变化，动态数据脱敏和隐私保护技术仍然是未来大数据研究难点之一。

1.3 新能源汽车国家监测与管理平台

大数据技术与新能源汽车结合的前提是建立大数据平台，以高效收集海量数据资源。为解决我国新能源汽车安全问题，完善新能源汽车行业监管，促进新能源汽车产业发展，工业和信息化部于 2016 年在北京建立新能源汽车国家监测与管理平台(以下简称国家平台)。截至 2019 年 5 月，国家平台接入车辆突破 220 万辆，预计 2020 年接入 700 万辆，2025 年接入 8000 万辆。国家平台的建立为政府部门加强新能源企业及车辆的安全监管起到重要支撑作用。

国家平台架构主要基于 Linux 系统和 Java 编程语言，采用 Hadoop 体系构建。Hadoop 是当前主流的大数据处理架构，国内外有众多采取 Hadoop 架构构建大数据平台的先例，涵盖医疗、银行、轨道交通、电力系统等各个领域，其模式已经非常成熟^[37-41]。

国家平台现有数据类型主要分为静态数据和动态数据。静态数据又被称为档案数据，由车辆基本信息组成，如车牌号、车辆 VIN 码、车辆厂商、车辆类型、销售区域等。动态数据的数据类型分为在线实时运行数据和离线存储历史数据，该两种数据的区别在于数据所处阶段与存储位置不同。实时运行数据即当前传递数据，该数据不断被更新替换，存储在实时缓存中，便于工作人员监控车辆运行安全。被替换的数据将转化为历史数据，存储于专用服务器中，以备研究人员调用和查验。

实时运行数据的数据帧间隔有 1 s、10 s 和 30 s 三种，数据项按照 GB/T 32960 的要求，主要采集自以下几大系统，动力电池系统、电机驱动系统、车辆控制系统以及其他部分。动力电池系统数据主要包括电池系统总电压和总电流、SOC、电池单体电压、电池系统特征点温度等。电机驱动系统数据主要有电机电压和电流、转速、扭矩以及温度等。车辆控制系统数据主要包含车辆速度、挡位信息、加速踏板行程、GPS 位置等。此外还有空调信息和轮胎胎压状态等其他信息数据^[42]。传递至国家平台的

部分实时数据项如图 1 所示。



图 1 部分实时数据示意图

基于隐私保护的原则，私人领域的新能源汽车只在故障预警时传递完整监控数据。而在公共交通领域，新能源公交车、出租车及物流车均全天候传递完整数据，以保障公共交通安全。国家平台主要履行行业监管职责，同时科研人员利用运行数据，对电池系统、驾驶行为、车辆能耗、充电行为等进行分析研究，以推动新能源汽车行业整体发展。

2 新能源汽车运行大数据分析技术

新能源汽车运行大数据分析流程主要分为大数据采集、大数据预处理与存储、大数据统计分析、大数据挖掘等步骤^[43]。现针对各步骤，以国家平台相关工作为背景进行介绍。

2.1 运行大数据采集

传统的大数据主要采集自互联网，采集方式包括日志文件采集和网页爬虫采集，在工程应用领域也广泛利用传感器采集数据^[44-45]。新能源汽车信息化程度高，车载传感器数量众多，在车辆运行过程中，车载终端将收集各传感器数据和电池管理系统数据，按照国标进行数据标准化及数据加密，再通过 GPRS 无线网络将加密信息传递至大数据平台，完成运行大数据采集工作。

目前，我国新能源汽车大数据监管体系形成了企业监测平台、地方监管平台和国家监管平台的三级架构。车辆运行数据实时传递至企业监测平台，企业主要负责本企业新能源汽车安全监管，同时将公共领域车辆运行数据、统计信息以及故障处理信息实时上报给地方和国家平台。其具体架构如图 2 所示。

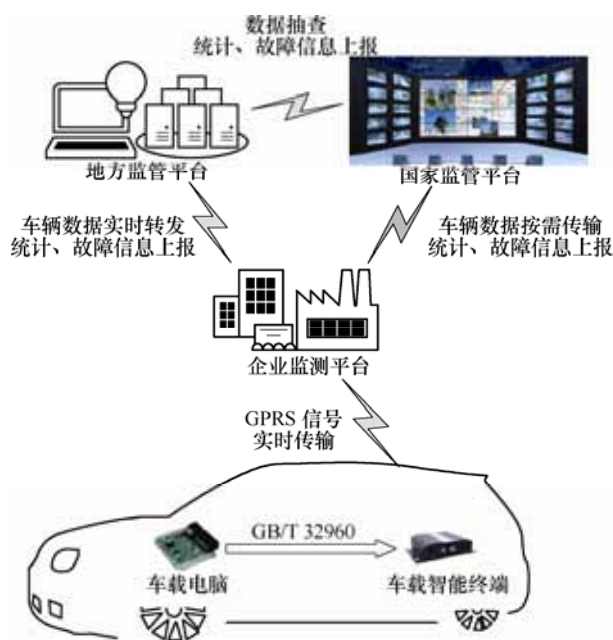


图2 新能源汽车大数据平台三级架构

2.2 运行大数据预处理与存储

新能源汽车日常运行环境复杂，数据在采集、传递、解码等步骤均有可能产生误差，除提升软硬件水平、加强数据质量监管之外，合理地运用大数据预处理技术也可改善数据质量，有助于提升数据挖掘过程的准确率和效率。大数据预处理技术主要包括数据集成、数据清洗、数据规约、数据降维、数据变化与离散化^[46]。新能源汽车运行数据预处理分为三个阶段，在原始数据采集时，根据车载终端程序设计，初步处理数值正负、数据精度、异常数据过滤等方面的问题；国家平台接收数据之前，将通过数据实时计算模块，对数据进行合法性检查和故障预判分析，保障数据的有效性和准确性；接收的数据仍可能存在掉帧、错误帧等问题，研究人员在进行数据分析前，将根据实际计算需求，选择缺失值填充或数据清洗的方式进一步处理数据。

在数据存储方面，当前国家平台数据存储空间按照200万辆新能源汽车3年数据存储量进行设计，测算带宽资源500 Mbps、服务器154台。在存储方式上，利用分布式HDFS文件数据存储服务，采用parquet列式格式进行存储，满足灵活化数据扩展及挖掘需求，未来可横向扩展至满足百万PB级别的数据存储规模。

2.3 运行大数据统计分析和可视化

大数据统计分析是指用统计学方法初步分析数据资源，萃取和提炼数据中的内涵信息，找出所研究对象的内在规律，指导数据挖掘工作^[47]。传统的统计分析方法有聚类分析^[48]，相关性分析^[49]，回归

预测分析^[50-51]，分类分析^[52]等。此外，可视化分析也是重要的数据分析方法，其旨在利用计算机自动化分析能力，同时充分挖掘人的认知优势，将人、机长处有机结合起来，辅助研究人员直观高效地洞悉数据背后的信息和规律^[53]。

国家平台当前已有成熟的数据统计分析模式，并建立了可视化大屏展示界面。自2016年3月开始，国家平台统计和分析首都电动汽车的运行状况，形成“首都公共领域新能源汽车统计与分析月报”；自2018年4月开始，进行了全国车辆整体运行情况的统计与分析，形成了“全国新能源汽车统计分析月报”，图3展示了北京市月报中关于电动出租车日均行驶里程的统计分析情况。

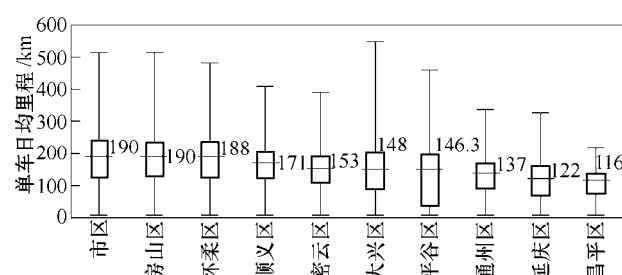


图3 北京市不同区域电动出租车日均行驶里程

同时，国家平台建立了大数据可视化系统和大数据实时展示大屏。大屏主要展示内容有国家平台宏观展示界面，包括入网车辆统计总数、累积运行里程、节能减排统计信息等；企业入网信息界面，展示注册企业数、过检企业数、过检车型数等，并针对全国企业的分布情况进行汇总及分类；全国地图展示界面，提供全国录入车辆数、车辆在线数、累计节油量、累计减少碳排放量等信息，并可按照省和直辖市为单位进行查看；区域平台展示界面，包含充电热力图、节能减排分布图、车辆里程核查情况、行驶轨迹图等信息；此外，还有单车全方位信息展示界面和车辆全生命周期展示界面。

2.4 运行大数据信息挖掘

简单数据统计不能满足当前新能源汽车行业的数据分析需求，数据背后的信息挖掘和规律探索更值得关注，例如新能源汽车安全预警、车辆技术分析、充电基础设施建设与管理等。为解决上述需求痛点，需紧密结合新能源汽车实际运行情况，利用大数据挖掘方法，针对性建立相关模型，以分析新能源汽车运行情况和性能变化。

计算机运算能力和存储能力的提升为大数据分析提供了基础，使得研究人员可以开展整体数据研究而非局限于少量样本数据研究，研究方向也从追求数据和算法的精度，转向融合多源异构数据，探

索事物的发展和变化规律^[54]。后续章节将对新能源汽车大数据分析方法和应用情况进行详细介绍。

3 新能源汽车大数据分析应用

3.1 动力电池大数据分析应用

动力电池是新能源汽车的核心零部件,动力电池性能直接影响整车性能。动力电池大数据分析主要利用 BMS 采集到的电流、电压、温度等电池单体及系统工作参数。随着 BMS 智能化的提高,其采集数据的频率和精度可以受远程监控系统的控制,方便研究人员选取颗粒度适合的数据进行动力电池大数据分析研究。以下将从当前主要研究方向对动力电池大数据分析应用进行介绍。

3.1.1 电池不一致性分析及故障诊断

电池的不一致性是指同一规格、同一型号的电池,其电压、内阻、容量等方面参数的差别^[55]。电池不一致性是影响电池系统性能的主要因素之一,也是诱发电池故障的重要原因。许多关于电池不一致性的早期研究,表明了不一致性与动力电池组使用寿命的关系^[56-57]。随着相关研究的进行,电压不一致性开始作为电池故障诊断的重要指标之一。LIU 等和 SIDHU 等^[58-59]都提出了通过扩展卡尔曼滤波的方法来获得电池系统故障表征参数,结合统计分析方法,判断电池故障发生率。YAO 等^[60]通过试验获得电压数据,利用集成香农熵预测了由电池单体连接错误造成的电池故障。YAN 等^[61]详细介绍了基于勒贝格取样的锂离子电池组故障诊断和预测模型,以及一种“按需进行”的创新诊断方法,可大大减少模型计算压力。在利用实车大数据进行电池故障诊断研究方面,WANG 等^[5]阐述了一种基于模糊改进香农熵的电压故障诊断方法,并在 Z-分数法的基础上提出了一种安全管理策略。该文献根据车辆运行数据,成功地找出了电动出租车上的电压异常单体。ZHAO 等^[42]利用机器学习算法和 3σ 多重筛选策略,基于新能源汽车运行大数据,检测出电池包中电压变化异常的单体,并成功在实车上进行了验证(图 4)。

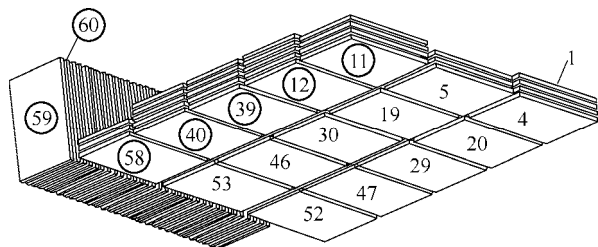


图4 运用大数据方法检测出异常单体

3.1.2 电池热失控预警

车用锂电池具有高比能量的特点,车辆碰撞或电池过充后,容易引发电池热失控、起火爆炸等严重事故,因此,电池热失控研究是当前电池领域的重点^[62]。在试验数据应用方面,PANCHAL 等^[63]进行了方形和袋装锂离子电池的放电试验,利用热传感器大量采集试验数据,研究不同试验条件下的电池产热速率,以周围温度、放电电流和电池容量作为模型输入,电池产热速率作为模型输出训练神经网络模型,该模型预测结果良好,成功反映了输入因素与电池产热速率的关系。LIU 等^[64]结合神经网络模型和卡尔曼滤波算法,提出了一种可估计电池内部温度的数据驱动方法,研究者采用电压、电流和电池表面温度数据作为输入,测得的电池内部温度数据作为输出训练模型,测试结果显示模型精度高,且具有可推广性。在实车数据应用方面,HONG 等^[65]利用新能源汽车运行大数据,结合熵判断理论,提出了一种电池热失控诊断方法。该方法实现了热失控的预测和诊断,能精确预报电池包热失控发生的时间和热失控单体位置,HONG 等还基于 Z 分数法提出了一种电池热安全管理策略。

3.1.3 电池健康状态估计

作为新能源汽车关键零部件,电池占据了整车成本的三分之一^[66]。准确估计动力电池健康状态,有利于评估退役动力电池残值,为二手新能源汽车和废旧动力电池估值提供依据。在电池健康状态(State of health, SOH)估计方面,已有大量的学者从理论角度出发,在实验室条件下从事相关研究。FRISK 等^[67]基于车载铅酸电池更换记录,利用随机生存森林的方法对电池的剩余使用寿命进行预估。文章分析数据来源于欧洲 5 个国家的 33 603 辆车,包含 291 个变量,数据类型包括温度、压力等结构化数据,以及电流、电压概率分布直方图等非结构化数据,在大数据的应用上极具创新性。RICHARDSON 等^[68]利用基于高斯过程回归的贝叶斯无参数估计方法对电池容量衰退进行估计。除传统的电流、电压和温度变化数据之外,研究人员提炼测量数据,得到新的数据标签如温度和电流分布情况等,使得预测模型更为高效准确。数据源自与现实类似的随机充电过程也是该文章的一个亮点。WENG 等^[69]和 LI 等^[70]采用容量增量分析法(Incremental capacity analysis, ICA)对电池 SOH 衰退进行分析和预测。根据充电时电压和容量的变化关系,绘制 IC 曲线。比较不同健康状态下 IC 曲线的峰值、谷值移动情况,建立 IC 曲线峰谷值与 SOH 间的关系方程,实现对电池 SOH 的预估。该方法基

于数据驱动,无需建立复杂的电池模型,具有较大应用前景。WENG 等成功将该方法应用于电池包上,证明其建立的模型的广泛适用性,LI 等的研究中发现 SOH 与容量增量值变化成线性关系。WANG 等^[71]在容量增量分析法的基础上进行拓展,提出了基于多岛遗传算法和高斯过程回归的锂离子电池 SOH 估计方法。

3.2 车辆运行大数据分析应用

新能源汽车上具备多种传统传感器,用于测量车辆运行数据,包括电机信息、车辆速度信息、GPS 信息等,这些数据同样具有丰富的科研价值,可开展驾驶行为分析、能耗分析、用户特征分析等研究。

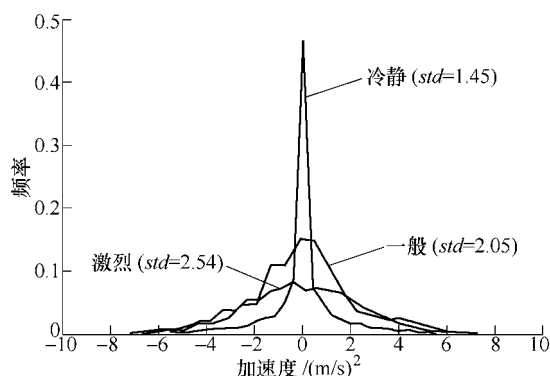


图5 三种驾驶模式的加速度分布曲线及标准差

3.2.1 驾驶行为分析

新能源汽车续航里程除受电池性能水平的影响之外,与用户的驾驶行为也密切相关,良好的驾驶习惯能够增加新能源汽车的实际续航里程、降低能量消耗。MCGORDON 等^[72]研究发现,即使平均车速一样,司机操作差异导致踏板变化的频率不同,也会影响最终能耗。踏板变化轻柔的测试组节能效果较优,且电池寿命较长,同时,该作者指出驾驶行为分析是能耗分析的前提和基础。PIAO 等^[73]在混合动力汽车模型的基础上,开发了驾驶行为识别模型。研究者基于车辆速度、加速度和踏板行程之间的关系,利用回归分析方法建立了驾驶员驾驶行为跟踪模型,并对模型进行了合理性验证。VATANPAVAR 等^[74]提出了一种基于情景感知的驾驶行为估计方法,该方法可写入电动汽车 BMS 中,并辅助优化 BMS 功能。研究者利用人工神经网络和非线性自回归算法(Nonlinear autoregressive, NARX),建立了驾驶行为模型,根据真实驾驶员的历史驾驶数据、当前驾驶动作反应以及从谷歌地图中获得的该路线平均速度来训练 NARX 模型,以实现特殊驾驶行为的建模,进行驾驶习惯的长期评估。部分研究者利用新能源汽车实车监控数据进行驾驶行为分析,毛正涛等^[75]通过新能源汽车运行监控数

据,定义和识别了紧急制动、加速启动等几种不良驾驶行为。同时,基于不良驾驶行为发生频率,应用模糊综合评价法建立了评价模型,利用非参数相关分析中的 spearman 方法对影响能耗的驾驶行为进行了挖掘分析。

3.2.2 实时能耗分析

目前,国内外电动汽车能耗模型研究主要有基于动力学方法和基于数据驱动方法。基于车辆动力学建模是进行车辆能耗研究的经典方法,但伴随电动汽车的行驶数据逐渐丰富,以及对实时能耗预测的需求日益增强,电动汽车能耗研究方法逐渐以数据驱动方法为主。LEE 等^[76]利用一种基于大数据的方法预测纯电动汽车续航里程。研究者首先基于试验数据建立了电池组 1 600 次循环寿命模型,该模型使用了电池内阻和 SOH 等电池寿命表征参数,利用该模型可判断电池剩余寿命。同时,作者利用自组织映射机器学习方法,对行驶数据进行聚类,结合道路情况,分析驾驶行为变化及其对电动汽车能耗的影响。最后,综合以上步骤,考虑电池寿命和驾驶行为的影响权重,建立可实时估计车辆续航里程的模型。FIORI 等^[77]利用车辆速度、加速度和道路等级数据作为输入变量,建立了电动汽车瞬时能耗的计算模型。该研究针对不同电动汽车进行了比较,并通过建立的能耗模型,对包括空调和供暖系统在内的辅助系统能耗影响进行了量化。结果表明,使用供暖和空调系统显著降低了电动汽车的效率和续航里程。陈燎等^[78]为提高续航里程估算精度,在行驶工况识别基础上,提出一种基于电池剩余电量和当前车辆能耗的续航里程估算模型,该模型能有效地消除里程估计误差,并具有较好的收敛性及鲁棒性。该文章利用 Matlab / Simulink 建立电池模型及整车能耗模型,基于以上模型建立特征参数与能耗之间的模糊规则库,再单独计算空调系统对续航里程的影响,最后利用卡尔曼滤波法对输出结果进行优化。

3.2.3 用户特征分析

互联网公司多依赖用户特征分析进行用户喜好推送、定制化服务和用户关系网络管理等,尤其在新用户首次登录时,快速、准确的用户定位十分重要^[79]。与传统汽车行业不同,新能源汽车行业与互联网结合紧密,同时,新能源汽车也是未来的人-车-路智能网络的重要组成部分^[80]。在这一网络中,结合车辆行驶数据,进行用户行为和需求分析,一方面可为用户提供个性化、定制化的便捷服务,另一方面,通过用户画像还可为保障社会公共安全提供支持。HIGGINS 等^[81]建立了适用于不同类型消费

者的电动汽车多准则购买决策模型, 该研究表明, 出行距离、身份特性、家庭收入是影响消费者选择电动汽车的主要因素。KIECKHÄFER 等^[82]提出一种估计电动汽车市场份额的混合模型。采用系统关联性模型从宏观角度研究消费者选择行为、消费者意愿、电动汽车性能、服务质量等因素间的相关性, 并利用离散选择模型优化以上结果。研究者基于真实市场数据, 对德国电动汽车市场进行分析, 结论指出, 消费者的特性因素在电动汽车市场份额估计中不可忽视。CUI 等^[83]利用出租车 GPS 大数据, 基于历史出行信息对居民出行模式进行识别, 并研究交通网络空间维度的相关特征, 如出行需求、出行速度、出行频率、交通网络流量等, 为进一步分析城市交通网络运营效率以及供需匹配问题提供决策支持。LI 等^[84]结合实证数据, 归纳总结了消费者的过往体验、感知回报、感知风险等因素, 同时分析了这些因素与消费者购买意愿之间的关系, 并进一步研究了消费者个性特征对上述关系的影响。LI 等^[85]通过详细调研 40 篇文献, 从消费者自身因素、环境客观因素、心理主观因素三个方面, 系统性地阐述了影响消费者对新能源汽车接受程度的原因。

3.3 充电大数据分析应用

充电桩和充电站等充电基础设施作为电动汽车的能量提供者, 是电动汽车发展与推广的根本, 快捷、便利和经济的充电设施可以增强消费者购买电动汽车的意愿。

3.3.1 充电行为分析及充电基础设施选址指导

选址是充电设施建设的第一步, 也是最重要的环节, 可直接影响到充电站的服务质量和运行效率, 分析城市中用户充电行为及其分布情况, 有利于充电站的合理选址。DARABI 等^[86]基于美国国家公路交通安全管理局统计数据, 对插电式混动汽车的充电行为进行了分析。研究人员根据车辆到站时间、车辆日均行驶里程、剩余能量、充电电流大小及站中充电车辆总数等参数, 对充电站带来的电网负荷进行了预测。SOARES 等^[87]提出了一种电动汽车充电需求估计方法, 利用交通信息系统中地理位置、电网接入位置、历史数据以及实时需求数据, 采用蒙特卡洛模拟方法和模糊 Logistic 模型, 根据充电站位置信息和电池剩余电量信息, 预测电动汽车的充电需求。在新能源汽车充电基础设施选址方面, CAI 等^[88]基于 10 000 辆出租车的行驶大数据, 结合车辆停车点坐标和车辆停车时长进行分析, 结果表明车辆停车热点可以作为充电站选址的参照, 良好的充电站选址可有效减少资源浪费和二氧化碳排放。GUO 等^[89]从可持续性发展的角度, 构建了新能

源汽车充电站选址的评价指标体系, 综合考虑了环境、经济和社会三方面影响因素。研究者采用多准则决策方法和模糊 TOPSIS 法从多个候选点中选择最优充电站位置。ZHANG 等^[90]根据当前充电站容纳水平和充电需求增长速度, 建立了未来充电站选址以及站内充电桩数目的优化模型。作者还利用华盛顿特区、纽约和波士顿的路网数据, 研究不同情景下的充电需求增长情况, 并指出当前充电设施建设计划中的潜在问题, 为政策制定者提供建议。

3.3.2 电网负载分析及储能电站建设

新能源汽车在公共领域的充电行为具有较大的随机性, 且通常以高功率的快充充电为主, 随着新能源汽车保有量的增长, 这种随机性大功率放电现象对电网的影响不可忽视。CLEMENT-NYNS 等^[91]研究了插电式混合动力汽车充电时的电网功率损失和电压不稳定现象, 通过一种均衡充电策略, 提升了电网强度和车辆的充电效率, 实现电网能量的高效和安全利用。HAIDAR 等^[92]利用纯电动汽车实车运行数据, 基于 ZIP 模型和参数建立可反映实车充电行为的精确模型。该模型可分析纯电动汽车充电行为, 并进一步探究充电行为对电网的影响。结果表明, 在研究过程中, 应采用贴近真实情况的负载变化模型, 而非采用恒定功率负载模型代替纯电动汽车充电行为。还有部分学者研究了充电经济性的问题, ALHAZMI 等^[93]首先利用 OPF 方法分析现有充电基础设施的服务能力, 综合考虑了私人充电和公用充电等多种方式的占比。利用模型预测未来充电站建设情况, 使得新能源汽车与快速充电站的数量处在动态平衡之中。同时, 研究者分析了充电等待时间和充电服务计费等因素, 提出了充电经济性和服务质量的评估方法, 实现充电服务与电网建设的综合考量。

4 当前主要问题及展望

新能源汽车和大数据的融合已经成为一种发展大趋势, 基于大数据的智能网联汽车, 将是我们国家汽车行业转型升级的战略重点方向。在当前新能源汽车大数据应用方面, 还存在着许多待解决的关键问题。

(1) 新能源汽车与大数据深度结合。当前新能源汽车相关领域研究多基于精确控制的实验室环境进行, 与新能源汽车现实运行环境有较大的差别, 现实工况非常复杂, 影响因素众多且存在耦合关系, 当前研究中却缺少该方面的考量。研究成果最终将被实际应用, 因此, 合理利用实车运行数据, 剖析

现实环境中的影响因素,构建适合于实车条件下的计算模型,是当前研究成果落地实施的关键。例如在驾驶员特性及能耗预测的相关研究中,需要使用实际行驶数据来反映驾驶员特性,若采用模拟或平均数据进行模型训练,将难以得到准确的结果^[94]。

此外,现有研究多基于单一、具体的研究对象展开。比如,动力电池方面多以电池单体为主,驾驶行为分析多基于车辆速度数据。然而电池系统是由数量众多的电池单体串并联构成,驾驶员的行为也体现在转向、路线选择、目的地选择等方面。新能源汽车是多个零部件和多种因素的复杂集合体,将实车运行大数据与新能源汽车紧密结合,才能准确反映车辆的真实情况,这是推动新能源汽车发展的必要手段。

(2) 大数据集成及大数据分析研究方法研究。传统的数据分析只运用少量、低维的数据,例如在传统汽车定价中只考虑需求与价格的关系,而大数据方法则需考虑高维观测变量与需求之间波动联系,比如天气、服务、舆情等因素,建模和计算过程更为复杂。在另一方面,大数据的数据类型已拓展到数字、声音、文本、图像等非结构化数据,多源异构数据的整合、存储与处理的相关研究现还处于初级阶段。此外,大数据中包含大量静态数据和动态流数据,需要开发出合适的分布式计算模型,兼容多类型数据的处理架构,同时兼顾运算速度以及交互作用。

大数据中95%的数据都是非结构化数据,传统数据统计分析技术已不能适应当前分析需求,利用与非结构化数据相对应的大数据分析技术,如文本分析、音频分析、视频分析、社交媒体分析以及预测分析等,才能有效挖掘出大数据中的潜在价值^[95]。在对多源异构数据进行针对性处理之后,应运用数据驱动算法,如自适应模糊推理、自回归滑动平均、人工神经网络、遗传编程、支持向量机等,来进行各具体研究方向的模型建立、结果预测、故障诊断等工作^[96]。该类算法可以充分利用海量数据优势,发挥大数据的作用。

(3) 以人为本的新型服务模式探索。随着经济水平提高,人们也开始注重自我价值的表达和实现。传统制造业也借机开始向服务型制造业转型,以人为本、以客户为本的理念逐渐深入人心。研究表明,商家在出售商品时提供更多个性化选择,能大大增强消费者占有欲,有利于商品销售^[97]。因此,许多汽车生产企业开始根据消费者的个人喜好,为消费者定制个性化的汽车产品,以提升销量和口碑。新能源汽车是新兴事物,其信息化程度高的特点也非

常适合开展面向用户的定制服务。

在新能源汽车相关的以往研究中,例如出行和充电规律方面,并未充分考虑车辆行驶与充电相结合的特性,也未能将具有时空性质的交通、充电大数据与多维用户身份特征数据进行联合挖掘。同时,在分时租赁、用户充电指导方面,缺乏满足用户个性需求的相关解决方案。结合大数据分析,给用户及时、专业的定制化服务,将是未来新能源汽车商业领域的发展趋势。

(4) 大数据安全及数据储存。汽车行业将进一步向智能网联化发展,以求建立更先进的智慧交通网络。智能网联化是以新能源汽车为基础,同时新能源汽车也是最好的智能网联发展平台,因此,数据通讯安全技术尤为重要。已有大量电影、电视作品描绘了黑客利用数据技术漏洞,远程操控智能车辆进行恐怖袭击,扰乱人们日常生活秩序的镜头。当前新能源汽车运行大数据初具规模,虽然数据价值开发尚未成熟,但已发现少量有目的性的盗取行为。可以预见在不久的将来,随着新能源汽车大数据的使用价值被进一步挖掘,以及人们隐私保护意识的增强,大数据安全将成为新能源汽车产业发展的关键性问题。

此外,大数据压缩存储和检索技术也亟需提高。新能源汽车国家监管平台接入车辆数已超过200万辆,每天产生T级的行驶数据,数据存储的压力随时间逐步积累增加。数据的压缩、存储、通讯以及快速检索将会成为制约大数据技术发展的瓶颈,是必须攻克的难关。海量数据带来了机遇,也带来了压力和责任,在大数据存储方式、存储形式、存储内容方面,有很多重要工作亟待开展。

5 结论

本文综述了大数据分析技术的发展情况,及车辆运行大数据在新能源汽车行业的应用现状,重点介绍了大数据分析技术在新能源汽车方面的最新研究成果。详细描述了大数据的特点及发展历程,以及新能源汽车运行大数据的相关信息;简要介绍了新能源汽车国家监测与管理平台的架构和功能,着重阐述了当前国家平台的数据情况及数据项组成;针对大数据分析技术在新能源汽车中的应用,依照动力电池大数据、车辆运行大数据、充电大数据的顺序,全面展示了当前各个领域的研究现状以及与大数据分析方法的结合情况;最后,讨论了大数据分析技术在新能源汽车行业的发展方向和应用前景,从五个方面总结当前研究中存在的主要问题,

并提出对未来研究方向的展望。

参 考 文 献

- [1] LI L, YOU S, YANG C, et al. Driving-behavior-aware stochastic model predictive control for plug-in hybrid electric buses[J]. *Applied Energy*, 2016, 162(1): 868-879.
- [2] 张雷, 胡晓松, 王震坡. 超级电容管理技术及在电动汽车中的应用综述[J]. *机械工程学报*, 2017, 53(16): 32-43.
ZHANG Lei, HU Xiaosong, WANG Zhenpo. Overview of supercapacitor management techniques in electrified vehicle applications[J]. *Journal of Mechanical Engineering*, 2017, 53(16): 32-43.
- [3] CHEN Z, XIONG R, TIAN J, et al. Model-based fault diagnosis approach on external short circuit of lithium-ion battery used in electric vehicles[J]. *Applied Energy*, 2016, 184: 365-374.
- [4] KIM G H, SMITH K, IRELAND J, et al. Fail-safe design for large capacity lithium-ion battery systems[J]. *Journal of Power Sources*, 2012, 210(4): 243-253.
- [5] WANG Z, HONG J, LIU P, et al. Voltage fault diagnosis and prognosis of battery systems based on entropy and Z-score for electric vehicles[J]. *Applied Energy*, 2017, 196: 289-302.
- [6] 王华伟. 铁路运输设备技术状态大数据平台研究[D]. 北京: 中国铁道科学研究院, 2017.
WANG Huawei. Research on big data platform for railway transportation equipment technical condition[D]. Beijing: China Academy of Railway Sciences, 2017.
- [7] MANYIKA J, CHUI M, BROWN B, et al. Big data: The next frontier for innovation, competition and productivity [EB/OL]. [2012-10-02]. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.
- [8] DREMEL C, WULF J, HERTERICH M M, et al. How AUDI AG established big data analytics in its digital transformation[J]. *MIS Quarterly Executive*, 2017, 16(2): 81-100.
- [9] KIM G H, TRIMI S, CHUNG J H. Big-data applications in the government sector[J]. *Communications of the ACM*, 2014, 57(3): 78-85.
- [10] FILIP F G, HERRERA-VIDEIRA E. Big data in the European Union[J]. *The Bridge*, 2014, 44(4): 33-37.
- [11] 张月茹. 我国政府数据开放与安全政策协同研究[D]. 哈尔滨: 黑龙江大学, 2018.
ZHANG Yueru. Research on the data opening and security policy synergy of government data in China[D]. Harbin: Heilongjiang University, 2018.
- [12] 杨小娟, 阳冬波, 贾红, 等. 新能源汽车可靠性大数据分析技术研究[J]. *装备维修技术*, 2018, 165(1): 25-31.
- [13] TU W, LI Q, FANG Z, et al. Optimizing the locations of electric taxi charging stations: A spatial-temporal demand coverage approach[J]. *Transportation Research Part C: Emerging Technologies*, 2016, 65: 172-189.
- [14] ARIAS M B, BAE S. Electric vehicle charging demand forecasting model based on big data technologies[J]. *Applied Energy*, 2016, 183: 327-339.
- [15] 张文, 王东, 郑静楠, 等. 电动汽车领域的大数据研究与应用[J]. *大众用电*, 2016(S2): 64-68.
ZHANG Wen, WANG Dong, ZHENG Jingnan, et al. Research and application of big data in electric vehicle field[J]. *Popular Utilization of Electricity*, 2016(S2): 64-68.
- [16] 张兰廷. 大数据的社会价值与战略选择[D]. 北京: 中共中央党校, 2014.
ZHANG Lanting. Research on social value and strategic choice of China based on big data[D]. Beijing: Central Party School of the Communist Party of China, 2014.
- [17] LEETARU K, WANG S, CAO G, et al. Mapping the global Twitter heartbeat: The geography of Twitter[J]. *First Monday*, 2013, 18(5): 18-20.
- [18] ZHANG Q, YANG L T, CHEN Z, et al. A survey on deep learning for big data[J]. *Information Fusion*, 2018, 42: 146-157.
- [19] MANOGARAN G, THOTA C, LOPEZ D. Human-computer interaction with big data analytics[M]// HCI challenges and privacy preservation in big data security. Hershey: IGI Global, 2018: 1-22.
- [20] LANEY D. 3D data management: Controlling data volume, velocity and variety[J]. *META Group Research Note*, 2001, 6(70): 1.
- [21] 顾荣. 大数据处理技术与系统研究[D]. 南京: 南京大学, 2016.
GU Rong. Research on techniques and systems for big data processing[D]. Nanjing: Nanjing University, 2016.
- [22] GROPP W D, GROPP W, LUSK E, et al. Using MPI: portable parallel programming with the message-passing interface[M]. Boston: MIT Press, 1999.
- [23] DEAN J, GHEMAWAT S. MapReduce: Simplified data processing on large clusters[J]. *Communications of the ACM*, 2008, 51(1): 107-113.
- [24] BHANDARKAR M. MapReduce programming with apache Hadoop[C]//2010 IEEE International Symposium on Parallel & Distributed Processing (IPDPS), April

- 19-23, 2010, the Downtown Sheraton Atlanta, Atlanta, Georgia. New York : IEEE , 2010 : 1-1.
- [25] ZAHARIA M , XIN R S , WENDELL P , et al. Apache spark : A unified engine for big data processing[J]. Communications of the ACM , 2016 , 59(11) : 56-65.
- [26] MENG X , BRADLEY J , YAVUZ B , et al. Mllib : Machine learning in apache spark[J]. The Journal of Machine Learning Research , 2016 , 17(1) : 1235-1241.
- [27] CARBONE P , KATSIFODIMOS A , EWEN S , et al. Apache flink : Stream and batch processing in a single engine[J]. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering , 2015 , 36(4) : 28-38.
- [28] 徐华. 基于云的大数据处理系统性能优化问题研究[D]. 合肥 : 中国科学技术大学 , 2018.
- XU Hua. Performance optimization for big data progressing systems in the cloud[D]. Hefei : University of Science and Technology of China , 2018.
- [29] XU Y , SHAW S L , ZHAO Z , et al. Understanding aggregate human mobility patterns using passive mobile phone location data : A home-based approach[J]. Transportation , 2015 , 42(4) : 625-646.
- [30] CHU X , ILYAS I F , KRISHNAN S , et al. Data cleaning : Overview and emerging challenges[C]//Proceedings of the 2016 International Conference on Management of Data , June 26-July 01 , 2016 , San Francisco , California. New York : ACM , 2016 : 2201-2206.
- [31] RONG H , ZHANG H , XIAO S , et al. Optimizing energy consumption for data centers[J]. Renewable and Sustainable Energy Reviews , 2016 , 58 : 674-691.
- [32] FAN Y , GUTHRIE A , LEVINSON D. Waiting time perceptions at transit stops and stations : Effects of basic amenities , gender , and security[J]. Transportation Research Part A : Policy and Practice , 2016 , 88 : 251-264.
- [33] AGRAWAL R , SRIKANT R. Privacy-preserving data mining[C]//ACM Sigmod Record , May 16-18 , 2000 , Dallas , Texas. New York : ACM , 2000 , 29(2) : 439-450.
- [34] KHANDPUR R P , JI T , JAN S , et al. Crowdsourcing cybersecurity : Cyber attack detection using social media[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management , November 06-10 , 2017 , Singapore. New York : ACM , 2017 : 1049-1057.
- [35] SMITH M , SZONGOTT C , HENNE B , et al. Big data privacy issues in public social media[C]//2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST) , June 18-20 , 2012 , Campione d'Italia , Italy. Piscataway : IEEE , 2012 : 1-6.
- [36] XU L , JIANG C , WANG J , et al. Information security in big data : privacy and data mining[J]. Ieee Access , 2014 , 2 : 1149-1176.
- [37] LAM C. Hadoop in action[M]. Greenwich : Manning Publications , 2011.
- [38] MEHRAGHDAM S , KELLER M , KARL H. Specifying and placing chains of virtual network functions[C]// 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet) , October 8-10 , 2014 , Luxembourg. Piscataway : IEEE , 2014 : 7-13.
- [39] MURDOCH T B , DETSKY A S . The inevitable application of big data to health care[J]. JAMA , 2013 , 309(13) : 1351.
- [40] 王丽, 王苹, 沈俊辉. 基于 Hadoop 的中医药大数据平台基础架构的设计与研究[J]. 中国医药导报 , 2018 , 15(6) : 158-162.
- WANG Li , WANG Ping , SHEN Junhui. Design and research on the infrastructure of Chinese medicine big data platform based on Hadoop[J]. China Medical Herald , 2018 , 15(6) : 158-162.
- [41] 王磊, 陈青, 高洪雨, 等. 基于大数据挖掘技术的智能变电站故障追踪架构[J]. 电力系统自动化 , 2018 , 42(3) : 84-91.
- WANG Lei , CHEN Qing , GAO Hongyu , et al. Intelligent substation fault tracking architecture based on big data mining technology[J]. Automation of Electric Power Systems , 2018 , 42(3) : 84-91.
- [42] ZHAO Y , LIU P , WANG Z , et al. Fault and defect diagnosis of battery for electric vehicles based on big data analysis methods[J]. Applied Energy , 2017 , 207 : 354-362.
- [43] 曹军威, 袁仲达, 明阳阳, 等. 能源互联网大数据分析技术综述[J]. 南方电网技术 , 2015 , 9(11) : 1-12.
- CAO Junwei , YUAN Zhongda , MING Yangyang , et al. Review of energy net based on big data analysis[J]. Southern Power System Technology , 2015 , 9(11) : 1-12.
- [44] 李学龙, 龚海刚. 大数据系统综述[J]. 中国科学 : 信息科学 , 2015 , 45(1) : 1-44.
- LI Xuelong , GONG Haigang. Review of big data systems[J]. Scientia Sinica Informationis , 2015 , 45(1) : 1-44.
- [45] 雷亚国, 贾峰, 周昕, 等. 基于深度学习理论的机械装备大数据健康监测方法[J]. 机械工程学报 , 2015 , 51(21) : 49-56.
- LEI Yaguo , JIA Feng , ZHOU Xin , et al. A deep learning-based method for machinery health monitoring with big data[J]. Journal of Mechanical Engineering , 2015 , 51(21) : 49-56.
- [46] GARCÍA S , LUENGO J , HERRERA F. Data preprocessing in data mining[M]. New York : Springer , 2015.

- [47] 张引,陈敏,廖小飞. 大数据应用的现状与展望[J]. 计算机研究与发展, 2013, 50(S2): 216-233.
ZHANG Yin, CHEN Min, LIAO Xiaofei. The status and prospects of big data application[J]. Journal of Computer Research and Development, 2013, 50(S2): 216-233.
- [48] AHLQVIST E, STORM P, KÄRÄJÄMÄKI A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables[J]. The Lancet Diabetes & Endocrinology, 2018, 6(5): 361-369.
- [49] LEE D, QUADRIFOGLIO L, TEULADA E S D, et al. Discovering relationships between factors of round-trip car sharing by using association rules approach[J]. Procedia Engineering, 2016, 161: 1282-1288.
- [50] 朱雪初,乔非. 基于工业大数据的晶圆制造系统加工周期预测方法[J]. 计算机集成制造系统, 2017(10): 95-102.
ZHU Xuechu, QIAO Fei. Machining cycle prediction method of wafer manufacturing system based on industrial big data[J]. Computer Integrated Manufacturing Systems, 2017(10): 95-102.
- [51] WAMBA S F, GUNASEKARAN A, AKTER S, et al. Big data analytics and firm performance: Effects of dynamic capabilities[J]. Journal of Business Research, 2017, 70: 356-365.
- [52] HAFEZ M M, SHEHAB M E, EL FAKHARANY E, et al. Effective selection of machine learning algorithms for big data analytics using apache spark[C]//International Conference on Advanced Intelligent Systems and Informatics, October 24-26, 2016, Cairo, Egypt. Cham, Switzerland: Springer, 2016: 692-704.
- [53] 任磊,杜一,马帅,等. 大数据可视分析综述[J]. 软件学报, 2014(9): 1909-1936.
REN Lei, DU Yi, MA Shuai, et al. Review of visual analysis of big data[J]. Journal of Software, 2014(9): 1909-1936.
- [54] 邵景峰,贺兴时,王进富,等. 大数据环境下的纺织制造执行系统设计[J]. 机械工程学报, 2015, 51(5): 160-170.
SHAO Jingfeng, HE Xingshi, WANG Jinfu, et al. Design of textile manufacturing execution system based on big data[J]. Journal of Mechanical Engineering, 2015, 51(5): 160-170.
- [55] 王震坡,孙逢春,张承宁. 电动汽车动力蓄电池组不一致性统计分析[J]. 电源技术, 2003, 27(5): 438-441.
WANG Zhenpo, SUN Fengchun, ZHANG Chengning. Study on inconsistency of electric vehicle battery pack[J]. Chinese Journal of Power Sources, 2003, 27(5): 438-441.
- [56] 王震坡,孙逢春. 电动汽车电池组连接可靠性及不一致性研究[J]. 车辆与动力技术, 2002(4): 11-15.
WANG Zhenpo, SUN Fengchun. Research on connection reliability and inconsistency of battery pack on electric vehicle[J]. Vehicle & Power Technology, 2002(4): 11-15.
- [57] 王震坡,孙逢春,林程. 不一致性对动力电池组使用寿命影响的分析[J]. 北京理工大学学报, 2006, 26(7): 577-580.
WANG Zhenpo, SUN Fengchun, LIN Cheng. An analysis on the influence of inconsistencies upon the service life of power battery packs[J]. Transactions of Beijing Institute of Technology, 2006, 26(7): 577-580.
- [58] LIU Z, HE H. Model-based sensor fault diagnosis of a lithium-ion battery in electric vehicles[J]. Energies, 2015, 8(7): 6509-6527.
- [59] SIDHU A, IZADIAN A, ANWAR S. Adaptive nonlinear model-based fault diagnosis of li-ion batteries[J]. Industrial Electronics IEEE Transactions on, 2015, 62(2): 1002-1011.
- [60] YAO L, WANG Z, MA J. Fault detection of the connection of lithium-ion power batteries based on entropy for electric vehicles[J]. Journal of Power Sources, 2015, 293: 548-561.
- [61] YAN W, ZHANG B, WANG X, et al. Lebesgue sampling-based diagnosis and prognosis for lithium-ion batteries[J]. IEEE Transactions on Industrial Electronics, 2016, 63(3): 1804-1812.
- [62] ZHU X, WANG Z, WANG C, et al. Overcharge investigation of large format lithium-ion pouch cells with Li (Ni_{0.6}Co_{0.2}Mn_{0.2}) O₂ cathode for electric vehicles: degradation and failure mechanisms[J]. Journal of The Electrochemical Society, 2018, 165(16): A3613-A3629.
- [63] PANCHAL S, DINCER I, AGELIN-CHAAB M, et al. Experimental and theoretical investigations of heat generation rates for a water cooled LiFePO₄ battery[J]. International Journal of Heat and Mass Transfer, 2016, 101: 1093-1102.
- [64] LIU K, LI K, DENG J. A novel hybrid data-driven method for Li-ion battery internal temperature estimation[C]// 2016 UKACC 11th International Conference on Control (CONTROL), August 31-September 2, Belfast, Northern Ireland, UK. Piscataway: IEEE, 2016: 1-6.
- [65] HONG J, WANG Z, LIU P. Big-data-based thermal runaway prognosis of battery systems for electric vehicles[J]. Energies, 2017, 10(7): 919.
- [66] 袁博. 新能源汽车技术发展与趋势综述[J]. 现代商贸工业, 2018, 39(35): 16-20.
YUAN Bo. Review of development and trend of new energy vehicle technology[J]. Modern Business Trade Industry, 2018, 39(35): 16-20.

- [67] FRISK E, KRYSSANDER M, LARSSON E. Data-driven lead-acid battery prognostics using random survival forests[C]//Proceedings of the Annual Conference of The Prognostics and Health Management Society, September 29-October 2, 2014, Fort Worth, Texas, USA. Scottsdale, Arizona: PMH Society, 2014: 92-101.
- [68] RICHARDSON R R, OSBORNE M A, HOWEY D A. Battery health prediction under generalized conditions using a Gaussian process transition model[J]. Journal of Energy Storage, 2019, 23: 320-328.
- [69] WENG C, FENG X, SUN J, et al. State-of-health monitoring of lithium-ion battery modules and packs via incremental capacity peak tracking[J]. Applied Energy, 2016, 180: 360-368.
- [70] LI Y, ABDEL-MONEM M, GOPALAKRISHNAN R, et al. A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter[J]. Journal of Power Sources, 2018, 373: 40-53.
- [71] WANG Z, MA J, ZHANG L. State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the gaussian process regression[J]. IEEE Access, 2017, 5: 21286-21295.
- [72] MCGORDON A, POXON J E W, CHENG C, et al. Development of a driver model to study the effects of real-world driver behaviour on the fuel consumption[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2011, 225(11): 1518-1530.
- [73] PIAO C H, DUAN C X, LI Y S, et al. Research on the driver's following behavior based on hybrid electric vehicle model[J]. Applied Mechanics and Materials, 2013, 281: 159-162.
- [74] VATANPARVAR K, FAEZI S, BURAGO I, et al. Extended range electric vehicle with driving behavior estimation in energy management[J]. IEEE Transactions on Smart Grid, 2018, 10(3): 2959-2968.
- [75] 毛正涛, 张金喜. 基于新能源汽车远程监控数据的驾驶行为识别建模与应用[J]. 汽车与配件, 2017(17): 80-83. MAO Zhengtao, ZHANG Jinxi. Driving behavior recognition modeling and application based on new energy vehicle remote monitoring data[J]. Automobile & Parts, 2017(17): 80-83.
- [76] LEE C H, WU C H. A novel big data modeling method for improving driving range estimation of EVs[J]. IEEE Access, 2015, 3: 1980-1993.
- [77] FIORI C, AHN K, RAKHA H A. Power-based electric vehicle energy consumption model: Model development and validation[J]. Applied Energy, 2016, 168: 257-268.
- [78] 陈燎, 杨帆, 盘朝奉. 基于电池能量状态和车辆能耗的电动汽车续驶里程估算[J]. 汽车工程学报, 2017(2): 113-122. CHEN Liao, YANG Fan, PAN Chaofeng. A driving range estimation model for electric vehicles based on battery energy state and vehicle energy consumption[J]. Chinese Journal of Automotive Engineering, 2017(2): 113-122.
- [79] ELKAHKY A M, SONG Y, HE X. A multi-view deep learning approach for cross domain user modeling in recommendation systems[C]//Proceedings of the 24th International Conference on World Wide Web, May 18-22, 2015, Florence, Italy. New York: ACM, 2015: 278-288.
- [80] MAPLE C. Security and privacy in the internet of things[J]. Journal of Cyber Policy, 2017, 2(2): 155-184.
- [81] HIGGINS A, PAEVERE P, GARDNER J, et al. Combining choice modelling and multi-criteria analysis for technology diffusion: An application to the uptake of electric vehicles[J]. Technological Forecasting and Social Change, 2012, 79(8): 1399-1412.
- [82] KIECKHÄFER K, VOLLING T, SPENGLER T S. A hybrid simulation approach for estimating the market share evolution of electric vehicles[J]. Transportation Science, 2014, 48(4): 651-670.
- [83] CUI J, LIU F, HU J, et al. Identifying mismatch between urban travel demand and transport network services using GPS data: A case study in the fast growing Chinese city of Harbin[J]. Neurocomputing, 2016, 181: 4-18.
- [84] LI W, LONG R, CHEN H, et al. Household factors and adopting intention of battery electric vehicles: a multi-group structural equation model analysis among consumers in Jiangsu Province, China[J]. Natural Hazards, 2017, 87(2): 945-960.
- [85] LI W, LONG R, CHEN H, et al. A review of factors influencing consumer intentions to adopt battery electric vehicles[J]. Renewable and Sustainable Energy Reviews, 2017, 78: 318-328.
- [86] DARABI Z, FERDOWSI M. Aggregated impact of plug-in hybrid electric vehicles on electricity demand profile[J]. IEEE Transactions on Sustainable Energy, 2011, 2(4): 501-508.
- [87] SOARES J, BORGES N, GHAZVINI M A F, et al. Scenario generation for electric vehicles' uncertain behavior in a smart city environment[J]. Energy, 2016, 111: 664-675.
- [88] CAI H, JIA X, CHIU A S F, et al. Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet[J]. Transportation Research Part D: Transport and Environment, 2014, 33: 39-46.

- [89] GUO S , ZHAO H. Optimal site selection of electric vehicle charging station by using fuzzy TOPSIS based on sustainability perspective[J]. *Applied Energy* ,2015 ,158 : 390-402.
- [90] ZHANG A ,KANG J E ,KWON C. Incorporating demand dynamics in multi-period capacitated fast-charging location planning for electric vehicles[J]. *Transportation Research Part B : Methodological* , 2017 , 103 : 5-29.
- [91] CLEMENT-NYNS K , HAESSEN E , DRIESEN J. The impact of charging plug-in hybrid electric vehicles on a residential distribution grid[J]. *IEEE Transactions on Power Systems* , 2010 , 25(1) : 371-380.
- [92] HAIDAR A M A , MUTTAQI K M. Behavioral characterization of electric vehicle charging loads in a distribution power grid through modeling of battery chargers[J]. *IEEE Transactions on Industry Applications* , 2016 , 52(1) : 483-492.
- [93] ALHAZMI Y A , SALAMA M M A. Economical staging plan for implementing electric vehicle charging stations[J]. *Sustainable Energy ,Grids and Networks* ,2017 ,10 :12-25.
- [94] PLÖTZ P , GNANN T , WIETSCHEL M. Modelling market diffusion of electric vehicles with real world driving data—Part I : Model structure and validation[J]. *Ecological Economics* , 2014 , 107 : 411-421.
- [95] GANDOMI A , HAIDER M. Beyond the hype : Big data concepts ,methods ,and analytics[J]. *International Journal of Information Management* , 2015 , 35(2) : 137-144.
- [96] WANG W C , CHAU K W , CHENG C T , et al. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series[J]. *Journal of Hydrology* , 2009 , 374(3-4) : 294-306.
- [97] DAWKINS S , TIAN A W , NEWMAN A , et al. Psychological ownership : A review and research agenda[J]. *Journal of Organizational Behavior* , 2017 , 38(2) : 163-183.
-
- 作者简介:余承其,男,1994 年出生,博士研究生。主要研究方向为电动汽车动力电池健康管理,新能源汽车大数据分析。
E-mail : shechengqi@bit.edu.cn
- 张照生(通信作者),男,1984 年出生,博士,讲师,硕士研究生导师。主要研究方向为新能源汽车大数据分析。
E-mail : zhangzhaosheng@bit.edu.cn
- 刘鹏,男,1983 年出生,博士,副教授,硕士研究生导师。主要研究方向为新能源汽车大数据分析。
E-mail : bitliupeng@bit.edu.cn
- 孙逢春,男,1958 年出生,博士,教授,博士研究生导师。主要研究方向为电动车辆系统技术、电驱动系统关键技术、充/换电站技术、清洁交通技术等。
E-mail : sunfch@bit.edu.cn