

智能电网 大数据

王继业 主编



中国电力出版社
CHINA ELECTRIC POWER PRESS

智能电网 大数据

王继业 主编



中国电力出版社
CHINA ELECTRIC POWER PRESS

内容提要

智能电网大数据是大数据在电力系统内的应用实践，是大数据的理论、技术和方法与传统电力行业的融合，中国电力科学研究院在该领域开展了相关研究和应用工作，并取得了一定的成果。本书是中国电力科学研究院大数据研究团队的研究成果之一，介绍了大数据的基本概念、理论基础和研究方法，阐述了智能电网大数据的内涵、重点应用领域、技术体系及研究方法和应用实践，并通过七个不同领域的应用案例分享研究过程中的经验与收获，最后提出了智能电网大数据的实现路径与推进建议。

本书可供电力行业从事数据管理、分析和应用的相关工作人员、研究人员学习借鉴，也可为其他行业或高校大数据研究相关领域的读者提供指导和帮助。

图书在版编目(CIP)数据

智能电网大数据 / 王继业主编. —北京: 中国电力出版社, 2017.2

ISBN 978-7-5198-0143-4

I. ①智… II. ①王… III. ①智能控制—电网—数据处理—研究 IV. ①TM76

中国版本图书馆CIP数据核字(2016)第308064号

中国电力出版社出版、发行

(北京市东城区北京站西街19号 100005 <http://www.cepp.sgcc.com.cn>)

北京盛通印刷股份有限公司印刷

各地新华书店经售

*

2017年2月第一版 2017年2月北京第一次印刷

710毫米×1000毫米 16开本 9.125印张 118千字

定价: 65.00元

敬告读者

本书封底贴有防伪标签, 刮开涂层可查询真伪

本书如有印装质量问题, 我社发行部负责退换

版权专有 翻印必究

编写组

主 编: 王继业

副 主 编: 张东霞 朱朝阳 杨 锐

编 委: 邓春宇 王晓蓉 韩 笑

王新迎 张国宾 季知祥

刘 鹏 郑亚芹 史梦洁

杨国生 黄彦浩 陈振宇

狄方春 高云峰 胡航海

杨 硕 张玉天 肖 凯

刘凤魁 丁玉成 吴 茜

刘 威 刘沅昆 李 灿

序

面对化石能源日渐枯竭和全球气候、环境变化的现实压力，一场新能源革命在全球范围内正悄然兴起。随着风能、光能等可再生能源的发展及新能源汽车技术、能效技术等低碳技术的广泛应用，传统的电力系统正在向智能电网演变。与传统电力系统这一封闭的物理系统相比，智能电网更开放、更复杂，是具有可再生能源和分布式能源高渗透、用户广泛参与、高度智能化等特征的信息物理系统。一方面，智能电网每时每刻都在产生着体量大、结构复杂、彼此间存在着复杂关联性的数据，如何储存、利用这些数据，挖掘其价值，是电力界目前及未来所面临的重大挑战；另一方面，面对这一开放、复杂的信息物理系统，传统的“以物理原理为依据—建立数学模型—选定参数—计算—分析因果”的“物理模型驱动”方法，已难以完全满足要求，数据驱动的大数据方法作为一种辅助性手段，将在智能电网的发展中发挥越来越重要的支撑作用。

智能电网大数据是电力系统及相关领域多数据源、多领域、跨时空数据的有机融合。智能电网大数据将突破传统信息技术针对特定应用的局限性，贯通智能电网各环节间的数据，推进多业务间的高效协同；打通电力系统与一次能源、用电用户、自然环境、经济社会、管理政策之间的信息壁垒，构建起开放、融合、可扩展的分布式一体化数据服务体系，实现紊乱的数据资源向有效的数据资产的转化。借助大数据理论和技

术，基于智能电网大数据，可快速提炼出深层知识，挖掘出传统物理建模方法无法应对的诸多时空关联关系，支持用户参与，建立更加完善的电网监测、预测、风险管控系统，可支持更全面的分析、更准确的预测及更具价值的辅助决策。

智能电网大数据的研究和应用尚处于初级阶段，业界对智能电网大数据概念的认识还未形成共识，尚未形成一批具有说服力的研究成果来充分显现智能电网大数据的研究和应用价值。未来的路还很长，需要研究者完成从传统思维向大数据思维的转变，在深入开展理论研究的同时，从数据融合和应用入手，形成切实有效的技术解决方案，将智能电网大数据应用推向深入。

由中国电力科学研究院副院长王继业主编的这本《智能电网大数据》，就大数据的基本概念、理论基础和研究方法，智能电网大数据的内涵、重点应用领域、技术体系和研究方法与应用实践进行了分析和总结，分享了不同领域的应用案例，并提出了智能电网大数据的工作展望。这本书对于电力界同行及对大数据感兴趣的读者，均具有重要价值。



中国科学院院士

2016年12月

编者序

在众多的大数据研究观点和方法中，我们选择了以应用为核心和面向价值实现的实践路径。智能电网大数据应用研究目前主要分为两个方向：一是应用数据统计和数据挖掘方法，去发现数据所表现的电力系统物理本质和运行规律。与传统的统计分析和数据挖掘不同的是，在大数据时代我们注重多源数据融合基础上的高完整性有效信息的发现，数据说话代替经验决策，为驾驭和认识智能电网提供全景式、全过程的研究视角。二是应用新兴的数据驱动方法，如机器学习、深度学习、随机矩阵等，形成更智能化的解决方案。我们发现，传统的物理建模方法在面对日益开放、不确定性加剧的复杂大电网时，因不合理的假设和不得已的过度简化造成分析结果与实际情况不相符，而数据驱动方法无须假设和简化，依据的是反映系统真实情况的数据。因此，大数据方法作为传统方法的有效补充，在智能电网领域将发挥越来越重要的作用。

大数据的基本机理是通过数据透视事物的本质和关系。受国家电网公司委托，在融合300多个城市的营配调数据分析中，对城市配电设备的运行效率、供电能力以及分布情况进行了呈现，我们清晰地看到了针对性的方式优化、检修安排和投资规划的决策依据。时空多维因素的关联分析是大数据的特点。2015年，我们与国网山东省电力公司合作开展了配电变压器重过载预测研究，数据涉及历史数据、网络拓扑、配电变压器改

造、负荷特性、气象、节假日、用户数据等多维信息，集成了PMS、GIS、用电采集、配电自动化等多个系统，采用了3亿多条数据，分析了影响因素集合并建立重过载预测模型，经过历史数据验证，准确率达到80%，在国网山东省电力公司2016年春节保供电工作中提供了全新的业务支撑。在大数据分析中，数据融合是一个必经的环节。大数据研究过程中一个重要的贡献是智能电网的统一数据模型，没有统一的数据模型和交互机制无法进行多维数据的统一计算和分析。在此基础上，我们与国网福建省电力有限公司合作完成了D5000和OMS数据融合，在进一步完成PMS、WAMS、保护信息、气象、负荷等数据融合与分析后，将实现电网状态感知和风险预判，在基于大数据的智能搜索系统的配合下实现需求信息的全景呈现，比如当搜索“莫兰蒂”台风信息就可获得路径分析、气象、异常数据、失压数据、设备状态、日志、报表、决策提示等。这是全新的调控系统，将随着大数据应用的开发而建立。

大数据在继电保护领域中得到应用，通过对全网的继电保护设备数据统计分析给出了精准的设备分析评价与状态评价，设备的薄弱点和风险点、家族问题分布的充分呈现。这一点在电网资产大数据中有同样的研究方法。在智能电网中，电网运行、电网资产、电网用户是大数据主要应用的三大领域。用户信息有着丰富的挖掘价值，在对几个省的用电数据分析中，我

们对售电量的预测、中短期负荷预测精准度均达到了90%。通过已经完成的用电数据标签化，对即将开展的用户类别、征信、行为分析做好了技术准备。

智能化建立在信息感知的响应之上，而智能电网的状态感知，须以大数据技术为支撑和基础，未来电网以数据为核心将成为必然，两年多的研究和实践表明了智能电网大数据极具应用价值和广阔的发展前景。

大数据是什么？从其哲学基础上讲，大数据是一种新的认识论。数据是事物内部机理特性或关联特性的一种外在表现，当数据超出表现理论的临界值时，就可依据数据认识事物。大数据适合认识统计规律和混沌规律，依据足够的样本到全量数据，适合于挖掘关联关系。从数学的角度看，从数据预处理需要的运筹学到数据分析使用的统计学，以及深度分析挖掘或机器学习的算法和模型中使用的计算数学和微分方程理论等，数学知识和思想贯穿于大数据的整个过程。IT领域有观点认为大量的数据集合就是大数据，还有观点认为是处理大量数据的一系列技术统称。而我们除了这两方面，还注重数据广泛关联后隐藏因果关系的发现和探索，这种数据处理的思想，收集和整合相关联的数据，试图发现常识以外的逻辑关系。我认为，大数据是具有一定体量的数据集合，是一系列对数据处理和利用的理论、方法与技术的集合，同时也是对数据分析挖掘和价值萃取利用思想的集合。这三个方面，缺一不可。

研究中发现，人们对大数据认识上存在一些误区。误区一，技术的创新能解决业务问题。有些IT厂商以技术的角度，说服企业通过技术创新解决生产和管理中的问题，而事实是，技术的创新只能解决效率问题，对具体的业务内容无法干预。误区二，用了Hadoop就叫大数据。大数据是业务主导的多层体

系架构，Hadoop仅是底层技术部分，代表处理能力，在数据利用思想方面没有质的变化。误区三，希望找到一键式数据分析工具来解决分析问题。数据分析最根本的是要解决业务问题，需要带着问题去寻找、观察、分析数据，分析工具只能起到辅助作用，对分析的贡献仅限于流程规范和效率提升，无法代替人解决分析问题。误区四，平台只针对某一项业务而搭建。研究发现，智能电网大数据全景性关联，必须打造全业务、全流程的数据集成与储备，通过数据融合技术而形成分析基础，否则将难以形成精准的结论。大数据的平台应该是一个数据利用的综合框架，包含着众多的产品和组件，可以支撑数据处理和利用各个层面的工作。组件间可以灵活地组织搭配使用，以应对纷繁复杂的业务需求，而平台需不断地迭代创新。

大数据能干什么？大数据的能力体现在五个阶段。第一阶段是数据的抽取与整合。将来自不同数据源和时间片段的数据进行关联和处理，放在统一的数据模型和同一个时间维度上进行分析。第二阶段是对数据基本特征的观察和提炼。以统计分析为手段，从不同视角、维度、时间颗粒度对数据进行全面观测。这一阶段能发现很多的数据规律，再加上业务解读，可形成业务分析报告，但这只是数据利用的初级阶段。第三阶段是非常规手段的深度分析。利用关联分析、分类聚类、回归预测等手段对数据的潜在因果关系、群体特征、趋势走向进行深度挖掘，找寻统计分析无法发觉的业务规律。在这一阶段大数据的处理计算能力得以充分展现，是解决具体业务问题必要的探索。第四阶段是为解决业务问题进行业务建模。由业务模型转化为数据模型，最终转变为数学模型。这一过程从多重的业务场景中提炼出核心问题和规律，将其转变为业务数据的处理流程，最终落到了一个个可用于计算的数据模型算法上。第

五阶段是模型的固化实现。将分析的思路和方法与业务系统对接，伴随有直观的可视化展现，至此完成了一个大数据的完整应用。期间伴随业务人员、计算机技术人员、数据分析师、决策者等不同角色的参与和互动。而能否将分析流程进行程序固化并与业务环境融合，决定了一次数据分析的生命周期和最终成果的大小。因此，大数据能干什么不仅取决于从事大数据分析的技术人员，更重要的是想让大数据干什么，需要业务需求策划与大数据能力的融合。

大数据应该怎么干？这是本书编写之初进行调研时目标读者提出最多的问题。开展大数据工作需要具备四个方面的基础条件：一是软硬件基础，即视业务需求和数据体量而定的服务器和数据处理软件；二是技术基础，运用基本数据提取、处理和解读能力的关系型数据库、分析挖掘工具及数据展示工具，建立一个分析流程，开展分析基础工作；三是数学基础，对数据的基础处理包括数据质量提升的处理、数据统计方式转变、统计指标和维度设计等，对数据的分析挖掘包括数据分类、预测算法的选择与实现、深度挖掘及机器学习的算法应用；四是业务基础，当前大数据分析的两个流派分别是业务驱动和数据驱动，最终还是要汇集到解决业务问题的角度，对业务的理解越深，对业务数据掌握越熟，就越具备成功开展大数据研究的基础。

智能电网大数据将向什么方向发展？随着物联感知、云计算、大数据、深度学习等创新技术的发展，未来智能电网的主要特征是信息物理融合，而人工智能在智能电网中将得到应用。智能电网大数据发展的最终目标是利用机器处理复杂的全景状态感知、预测、诊断与决策，从而支撑电网的自我调节及其与外界的互动。从架构上首先是全面的数据产生与共享机制，需要电网的IT化和泛在感知，这也是电网信息通信及其产业发展的方向，同时需要创新数据

共享的管理与思维模式；其次是不断创新和面向业务发展的平台处理技术，今天的处理技术要向更大的承载与处理能力和建模智能化发展；最后是业务分析经典化和建模固化，并由深度学习来形成超预置的智能与认知。

我们正在经历着电网发展模式的变革，以特高压为主干的大电网广域互联形成电网一体化；以新能源消纳为主导的源荷储等多元素广泛接入形成电网平台化；以市场机制改革引导的电力消费与交易形成电网互动化，电网的发展模式将从以技术和安全为核心，转向以技术和安全支撑下的功能需求为核心。智能电网也将从发电、输电、变电、配电、用电、调度环节和设备的智能化走向系统的智能化。未来的电网智能化过程，就是应用先进的信息通信技术，以数据为媒介，以数字化为手段对电网进行重塑的过程。数据是电网创新变革中的重要驱动力。

希望本书对电力行业从事数据管理、分析和应用的相关工作人员、研究人员和广大读者在大数据理解认识和分析工作等方面有所借鉴和助益。在编写中得到周孝信院士和郭剑波院士的指导和帮助，同时也得到了来自系统内外、高校和企业多位专家的支持与关注，在此表示由衷的感谢！

A handwritten signature in black ink, consisting of stylized Chinese characters that appear to read '王康' (Wang Kang).

2016年12月于中国电力科学研究院

目 录

序 编者序

1	认识大数据	1
	1.1 大数据的基本概念	2
	1.2 大数据的理论基础	3
	1.3 大数据研究方法	4
2	智能电网大数据	9
	2.1 智能电网发展趋势	10
	2.2 智能电网大数据定义及其应用前景	11
	2.3 智能电网大数据研究和应用现状	15
	2.4 面临挑战	19
3	智能电网大数据重点应用领域	21
	3.1 电网运营和发展	22
	3.2 电力用户	26
	3.3 社会与政府	30
4	智能电网大数据技术体系	35
	4.1 智能电网大数据技术架构	36
	4.2 智能电网大数据关键技术	37
	4.3 智能电网大数据关键技术对比	52

5	智能电网大数据研究方法与应用实践	57
5.1	平台建设	58
5.2	数据收集与管理	65
5.3	数据汇聚与融合	67
5.4	分析挖掘	73
6	智能电网大数据应用案例	81
6.1	基于关联因素分析的台区重过载预测	82
6.2	大用户日用电负荷预测	88
6.3	基于频繁项集的电网调控系统设备事故关联分析	91
6.4	基于失稳故障辨识的大电网快速判稳	96
6.5	基于大数据的继电保护设备家族可靠性分析	99
6.6	新能源大数据的快速提取与实时可视化	105
6.7	随机矩阵在智能电网大数据分析中的应用	109
7	智能电网大数据工作展望	123
7.1	发展趋势	124
7.2	实现路径	126
7.3	推进建议	127
7.4	展望	128
	参考文献	131

智能电网大数据 BIG DATA

认识大数据

大数据是当今最热门的话题之一，智慧城市、智慧交通、智慧医疗、智慧能源等与“智慧”相关的领域都离不开它的技术支撑。总统大选、奥运会、世界杯、春节等很多重大事件、重要活动、重大节日前后都有大数据的分析和预测，大数据与云计算、物联网技术的深度结合已极大地改变了人们的生产生活方式。大数据的基本概念、理论基础和研究方法成为人们认识和应用大数据需要了解和思考的首要问题。

1.1 大数据的基本概念

自2012年以来，大数据在世界范围内掀起持续的热潮，受到广泛的关注，应用范围越来越广。大家从不同的视角去理解和看待大数据，并没有形成统一的定义。总体来看，大数据的定义可分为三个层次。狭义来讲，大数据指量大、复杂、增长迅速，无法在可容忍的时间内、用常规的方法处理的数据集合。广义来看，大数据是以这个数据集为研究对象的一项综合性技术，是传感测量技术、信息通信技术、计算机技术、数据分析技术与专业领域技术的结合。从更为广阔的视角看，大数据是一种新的方法论，它认为世界上的一切事物，包括人的心理和行为都可由数据表征或体现，透过数据可以更好地认识世界和人类自身，指导人们的生产和生活；它直接从数据出发，采用数据驱动的方式去分析和解决问题，着眼点不仅限于事物本身或有限范围，而是从更宽的视野、更广的范围、更深的层次去寻找关联和规律，认识事物、预测未来。

1.2 大数据的理论基础

1.2.1 大数据与认识论

大数据不仅是一项由多学科、多领域结合而成的综合性技术，也是一种哲学观、一种方法论，有其哲学基础。

大数据是协调人类精神和外在物质世界的媒介和手段。数据和信息是连接物质与精神的媒介。如果人类用思想序列运动指挥物质序列运动，达到预期的目的，这就说明人类正确地认识了物质序列的运动。无论事物存在着怎样复杂或隐秘的内在规律，只要这一事物存在着或发生了，它就一定会有所表现，而数据就是事物内部机理特性或关联特性的外在表现。当数据量不足时，不足以准确完整地反映该事物，因此这样的数据不能满足表现理论。当数据量的规模足够大，超过了表现理论所需的临界值时，我们就能够依据数据描述和认知该事物了。大数据概念的提出是人类认识发展过程的又一次进步，大数据使我们认识到，人类认识和实践的历史就是一部数据搜索、处理、挖掘和创新的历史。人类对世界的认识，是精神世界和物质世界两个同构系统之间的相互作用，而大数据这一新的认识论，通过物质世界的的数据呈现，使物质世界和精神世界实现了统一。

大数据适合认识统计规律和混沌规律。世界上万事万物的运行总是遵循着一定的规律，即自然规律。人类认识的目的在于认识世界万物的规律性。规律性可以分为三种：第一种是恒常规律，如欧姆定律，当阻抗一定时，电压和电流呈线性关系，表述其规律的数学模型基于物理特性，而线性关系只需要两个参数便可确定，无须更多数据；第二种是统计规律，需要足够的随机样本数据；第三种是混沌规律，需要更多的数据甚至全量数据。统计规律和混沌规律往往难以找到精确的因果关系，更适合利用大数据进行研究并从中挖掘关联关系。