

智能电网大数据技术发展研究

张东霞¹, 苗新², 刘丽平¹, 张焰³, 刘科研¹

(1. 中国电力科学研究院, 北京市 海淀区 100192; 2. 国网北京经济技术研究院, 北京市 昌平区 102209;
3. 上海交通大学电子信息与电气工程学院, 上海市 闵行区 200240)

Research on Development Strategy for Smart Grid Big Data

ZHANG Dongxia¹, MIAO Xin¹, LIU Liping¹, ZHANG Yan², LIU Keyan¹

(1. China Electric Power Research Institute, Haidian District, Beijing 100192, China; 2. State Power Economic Research Institute, Changping District, Beijing 102209, China; 3. School of Electrical and Electronic Engineering, Shanghai Jiao Tong University, Minhang District, Shanghai 200240, China)

ABSTRACT: Smart grid is one of the most potential field for big data application. Some basic concepts of big data for smart grid were expounded in the paper, including data resources and characteristics, application value presentation of big data in smart grid, difference in research approach between big data and traditional research method. The research and engineering application of big data in smart grid were overviewed. The theoretical basis and technology system for big data were presented. The primary technical fields and potential application benefits of big data were discussed. Finally, the research projects and roadmap for smart grid big data were proposed.

KEY WORDS: big data; smart grid; theoretical basis; technological system; application fields; technique road map

摘要: 智能电网是大数据最重要的应用领域之一。该文阐述了智能电网大数据的基本概念, 包括: 数据源及数据特征, 大数据应用于智能电网的价值体现, 大数据与传统研究方法的不同之处; 综述了国内外智能电网大数据研究和工程应用现状; 总结了大数据的理论基础和技术体系; 论述了智能电网大数据重点应用领域及应用价值; 提出智能电网大数据研究框架和技术发展路线建议。

关键词: 大数据; 智能电网; 理论基础; 技术体系; 应用领域; 技术路线图

0 引言

智能电网是电力工业发展的方向和趋势。智能电网利用先进的信息通信技术、计算机技术、控制技术及其他先进技术, 实现对发电、电网运行、终端用电和电力市场中各利益方的需求和功能的协调, 在尽可能提高系统各部分的高效率运行、降低成本和环境影响的同时, 尽可能提高系统的可靠

性、自愈能力和稳定性。

智能电网的最终目标是建设成为覆盖电力系统整个生产过程, 包括发电、输电、变电、配电、用电及调度等多个环节的全景实时系统。而支撑智能电网安全、自愈、绿色、坚强及可靠运行的基础是电网全景实时数据采集、传输和存储, 以及累积的海量多源数据快速分析^[1]。

大数据是近年来受到广泛关注的新概念, 是指通过对大量的、种类和来源复杂的数据进行高速地捕捉、发现和分析, 用经济的方法提取其价值的技术体系或技术架构^[2]。所以, 广义上讲, 大数据不仅是指大数据所涉及的数据, 还包含了对这些数据进行处理和分析的理论、方法和技术。

大数据早期主要应用于商业、金融等领域, 后逐渐扩展到交通、医疗、能源等领域, 智能电网被看作是大数据应用的重要技术领域之一。一方面, 随着智能电网的快速发展, 智能电表的大量部署和传感技术的广泛应用, 电力工业产生了大量结构多样、来源复杂的数据, 如何存储和应用这些数据, 是电力公司面临的难题; 另一方面, 这些数据的利用价值巨大, 不仅可将电网自身的管理、运行水平提升到新的高度, 甚至产生根本性的变革, 而且可为政府部门、工业界和广大用户提供更多更好的服务, 为电力公司拓展很多增值业务提供条件。

2012年以来, 国内外在智能电网大数据技术研究和工程应用方面做了一些有益的尝试, 奠定了一定的基础, 但总的来看, 这些工作尚处于探索起步阶段。智能电网大数据的研究和应用是一个长期而复杂的工作。客观上, 大数据的理论尚未形成, 大

数据的相关技术仍在快速发展中,还没有进入稳定时期;同时,智能电网通信信息系统的互操作问题仍然存在,数据模型尚未统一,给数据的获取和应用带来实际困难;主观上,电力公司在大数据的基本概念、研究方法、应用价值方面认识不足,没有达成共识,在思想认识和技术准备上存在不足,也给大数据在智能电网中的应用造成一定障碍。即使有些电力公司和电力研究者对大数据表示出很大兴趣,但由于缺少战略性研究和顶层设计的指导,影响了智能电网大数据研究和应用的有序推进^[3-4]。

文章阐述了智能电网大数据的数据特征、研究方法和价值体现;在综述智能电网大数据国内外研究和应用现状基础上,分析了智能电网大数据的发展驱动力和面临的挑战;系统总结了智能大数据的理论基础和技术体系,提出智能电网大数据的重点领域,分析了其应用价值;并在上述研究基础上提出智能电网大数据的研究框架和技术路线建议。

1 智能电网大数据概念

1.1 数据特征

随着电力信息化的推进和智能变电站、智能电表、实时监测系统、现场移动检修系统、测控一体化系统以及一大批服务于各个专业的信息管理系统建设和应用,数据的规模和种类快速增长,这些数据共同构成了智能电网大数据。

根据数据来源的不同,可以将智能电网大数据分为两大类:一类是电网内部数据;另一类是外部数据。内部数据来自用电信息采集系统(collection system information, CIS)、营销系统、广域监测系统(wide area measurement system, WAMS)、配电管理系统、生产管理系统(production management system, PMS)、能量管理系统(energy management system, EMS)、设备检测和监测系统、客户服务系统、财务管理系统等的数据库。外部数据来自电动汽车充换电管理系统、气象信息系统、地理信息系统(geographic information system, GIS)、公共服务部门、互联网等。这些数据分散放置在不同地方,由不同单位/部门管理,具有分散放置、分布管理的特性。

这些数据之间并不完全独立,其相互关联、相互影响,存在着比较复杂的关系。如气象条件和社会经济形势会影响用户的用电情况、用户用电数据影响电力市场交易情况,电力市场数据可以为相关公共服务部门决策提供依据,而电力企业的GIS数据必须以市政规划数据作为参考。此外,这些数据

结构复杂、种类繁多,除传统的结构化数据外,还包含大量的半结构化、非结构化数据,如服务系统的语音数据,检测数据中的波形数据、直升机巡检中拍摄的图像数据等。而且这些数据的采样频率与生命周期也各不相同,从微秒级到分钟级,甚至到年度级^[5]。

综合各种对大数据的数据特征描述,考虑到智能电网数据的特点,智能电网大数据的数据特征可归结为如下几点:1)数据来自分散放置分布管理的数据源;2)数据量大、维度多、数据种类多;3)对公司、用户和社会经济均有巨大的价值;4)数据之间存在着复杂关系需要挖掘,且大多数情况下有实时性要求。

1.2 研究方法

大数据的研究方法是传统数据挖掘技术的提升、扩展甚至革命性改变,大数据为数据的处理和分析提供了新的思路和方法。且随着大数据理论和技术的发展,新的方法和技术也会随之产生。

1) 大数据方法适用情况。

大数据方法主要适用于如下情况:①由于数据规模大,数据处理时效性高,传统的数据处理技术无法满足技术要求或经济要求;②因数据类型多样,包含半结构化、非结构化数据以及空间矢量数据等,传统的处理技术不能满足要求。

2) 大数据研究思路。

智能电网大数据研究方法与电力系统传统的基于数据计算分析的方法相比,在解决问题的方法和研究过程方面都有很大不同,具体表现为:传统方法通常基于抽样数据,而大数据方法则采用尽可能多的数据;传统的电力数据分析通常基于某个部门或某个专业的数据,智能电网大数据分析则是在实现跨专业、跨部门数据融合基础上进行多维度数据分析。

3) 大数据研究过程。

传统的科学研究过程通常为:①科学假设,②科学实验,③实验结果分析,④科学假设得到证实或证伪。

而智能电网大数据的研究过程是:①科学假设,②数据获取与整合形成数据资源,③数据挖掘和分析,④数据结果分析,⑤科学实验,⑥实验结果分析,⑦科学假设的证实、证伪或新知识、新规律的发现。该过程可进一步细化为以下步骤。

1) 科学假设。

领域专家根据先验知识,同时考虑到数据分析

存在着超越先验知识的可能性,提出科学假设,设计应用场景,提出需要的数据源和数据量。

2) 数据获取与整合形成数据资源。

根据需要的数据源特点,提出数据获取与整合方案,形成统一的数据资源,为实现数据的融合,需要建立统一的数据模型。

3) 数据挖掘与分析。

领域专家、数据分析专家和计算机专家一起,根据需要解决的问题,研究需要的数学挖掘与分析方法,包括常规的数学挖掘与分析方法,以及领域专用的数学计算分析方法。

4) 数据结果分析。

领域专家对数据挖掘和分析结果进行专业分析,解释分析结果的内在原因。

5) 科学实验:

若有必要,应进行科学实验,验证数据挖掘和分析结果的合理性。

6) 科学结果分析。

对科学实验结果进行分析论证。

7) 科学假设得到证实、证伪或新知识和规律的发现。

对照分析结果和科学假设,得到如下3种可能结果:一是数据分析结果证明科学假设不成立;二是数据分析结果证明科学假设成立,某种事物存在着假设的关联性,且对假设的关联性给予了定量的描述;三是通过数据分析获得新的发现,并对新发现的知识或规律进行科学的解释、分析和验证。

2 国内外研究现状

2012年以来,国内外大学和研究机构、IT企业、电力公司均开展了智能电网大数据研究和工程应用。在国外,一些IT企业如IBM^[6]、Oracle^[7]等,陆续发布大数据白皮书,IBM和C3-Energy开发了针对智能电网的大数据分析系统,Oracle提出了智能电网大数据公共数据模型;美国电科院等研究机构启动了智能电网大数据研究项目;美国的太平洋燃气电力公司、加拿大的BC Hydro等电力公司基于用户用电数据开展了大数据技术应用研究。在国内,中国电机工程学会发布了电力大数据白皮书;国家科技部2014年下达了3项863项目^[7],支持智能电网大数据研究;自2012年以来,国家电网公司启动了多项智能电网大数据研究项目,江苏省电力公司于2013年初率先开始建设营销大数据智能分析系统,开展了基于大数据的客户服务新模式应

用开发研究;北京等电力公司也正在积极推进营配数据一体化基础上的智能电网大数据应用研究。

智能电网大数据研究和应用已取得了一些成效,但总的来看,研究成果仍比较粗糙,不成体系,研究和应用尚处于起步和探索阶段。在此起步阶段,充分借鉴国内外研究成果,分享知识经验,对于促进开展智能电网大数据的研究和应用,具有重要的意义。

智能电网产生了大量的数据,应用大数据技术,可有效提高电网的运行管理水平和为社会及用户的服务水平,与此同时也带来了很大的挑战。需要各方在充分认识大数据技术可带来的显著效益基础上,达成共识,产学研相结合,共同推动智能电网大数据的研究和技术发展。

3 发展驱动力和存在的障碍

3.1 驱动力

1) 电力公司部署了大量的智能电表及用电信息采集系统,其中包含着巨大的价值需要挖掘。例如,根据用户用电数据,可分析出用户的用电行为,为形成合适的激励机制、实施有效的需求侧管理(需求响应)提供依据。

2) 电力公司资产巨大,资产的监测和运维涉及到大量复杂的数据,通过数据分析,可提高网资产利用率和设备管理水平,存在着巨大经济效益。

3) 在实现营配数据一体化基础上,通过数据分析,电网公司可进行有效的停电管理,提高供电可靠性;也可进一步提高电能质量,减少线损;可防止用户窃电,以及避免造成其他非技术性损耗;经济效益显著。

4) 大数据将促进地球空间技术、天气预报数据在智能电网中的应用,提高负荷和新能源发电预测精确度,提高电网接纳可再生能源的能力。

5) 通过大数据分析,可探索新的商业模式,为电网公司带来效益。

3.2 需克服的障碍

1) 缺乏认同。

电力系统是一个传统的行业,电力专业人员对大数据的基本概念,大数据在智能电网中应用价值的认识不足,甚至存在着一定程度的怀疑,需要有令人信服的结果才能得到普遍认同。

2) 技术准备不足。

国内外虽然已在大数据方面开展了一些分散的研究和应用,但电力公司尚未制定出整个企业的

大数据战略，仍停留在初期探索阶段。

3) 资金投入。

大数据研发需要大量的投入，大数据应用获得的效益尚不明确，需要建立大数据应用的资金回收模型和评价方法，才能刺激资金的持续投入。

4) 互操作性。

数据交换中的互操作问题普遍存在，软硬件系统之间的技术差异和标准差异，也同样带来了互操作问题。

5) 通信信息架构的一致性。

当电力公司意识到利用大数据技术可以获得效益时，才发现现有的信息通信架构优势不适应大数据的发展需求，而通信信息系统的改造则需要投入新的资金。例如，虽然智能电表功能设计时包含了“超越计量”功能，但某些地区由于受到通信信息系统的限制，并不能真正实现这些数据的上传。

6) 数据获取。

获取数据的困难不仅来自硬件方面的限制，也来自现有的管理机制。目前，电力公司的数据多是在不同的数据库中采用不同的数据格式存储，形成竖井式的数据库，数据查询面临困难，并不具备用户友好性；运行管理方面也形成了竖井模式，不同部门、不同团队之间习惯了孤立工作，给交流和数据共享造成了困难。

7) 技术复杂性。

智能电网大数据是电网分析运行技术、数学挖掘和分析技术和 IT 技术的结合，综合性和先进性共同构成了技术的复杂性。

8) 数据的隐私和安全问题。

有些数据电力公司不能公开使用，有些数据涉及到安全问题，需在有限范围内使用，隐私保护和安全是大数据应用首先需要解决的问题。

4 大数据的理论基础

4.1 理论基础概述

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，大数据也被认为是一种思维方式，是一门科学。虽然大数据的理论体系尚未形成，但大数据与哲学、数学、物理学等学科的内在联系已有所显现，哲学、数学、物理学也将为大数据提供数理和哲学基础。此处仅简述大数据的哲学基础和数学基础。

4.2 大数据的哲学基础

哲学是关于世界观的学说，大数据存在着哲学

基础，因为大数据首先被认为是一种认识论，关于这一点，一些具有代表性的观点为：

1) 大数据代表了一种新的思维方式，也是一种新的方法论。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，而大数据作为一种新的认识论，通过物质世界的表现——数据的呈现，使物质和精神两个世界达到了统一。认识者获得的不是对象本身的绝对映像，而是从对象中抽象出来的、关于对象运动轨迹的数据。从这一角度看，同构关系是大数据的数理哲学基础^[9]；

2) 世界的规律表现为恒常规律、混沌规律和浮动规律。恒常规律的认识基于少量的数据即可实现，但对于混沌规律和浮动规律的认识，则需要以足够多的数据也即大数据为依据。所以，大数据更适合研究混沌规律和浮动规律。理论上，混沌规律是没有规律可寻的，应基于全部的数据，而浮动数据是有迹可寻的，只要有足够的数据即可。例如，当数据表现出一定的统计特性时，足够的数据可以通过统计特性展现其规律^[10-11]；

3) 浮动规律和混沌规律往往难以找到精确的因果关系，所以，大数据也更适合于从数据中挖掘关联关系。一种情况是，大数据挖掘出的关联关系，并没有超越人类的先验知识，而是对先验知识给予了定量的表征。另一种情况是，大数据挖掘出的知识，超越了人类的先验知识，为人类揭示了至今没有认识到的规律。

4.3 大数据的数学基础

数学是科学的基础，不能建立数学模型并用数学方法分析，不能称之为科学，但大数据统一的数学基础和系统的建模方法目前还没有形成。

“从数据中学习”是统计学的重要目标，正因如此，统计学也成为大数据最基本的数学基础。进行预测或分类时，在线性回归、支持向量机、人工神经网络的有指导或无指导学习过程中都体现了统计学方法。

在基于统计学的机器学习之外，大数据也存在许多基于全数据分析的数据分析方法。这些方法包括决策树方法、Apriori 算法、划分方法以及层次方法等。

同样基于大数据具有统计特征的假设，有学者尝试用随机矩阵作为某类大数据的统一数据模型。假设构成大数据的系统来自呈空间分布的 N 个数据源，每一个数据源可以表示为时变向量，由此可将整个系统的数据整合为一个单独的大型矩阵，也

即随机矩阵^[12]。

“熵”也被用来表征大数据所表现出来的特征。当大数据满足一定的统计特性时，大数据所表征的事物的特性可以用“熵”来表征^[13]。

针对大数据的数据类型复杂性，学术界目前鼓励发展一种一般性的结构化数据和半结构化、非结构化数据之间的转化原则，如，在对网络上多种来源的数据进行性质分析和规律探索中，有学者尝试运用图论和统计分析等方法对数据进行定量分析，并已有学者发现了复杂的网络大数据之中存在一些统计规律性。

综上所述，就大数据的数学基础而言，可总结如下：

1) 形成大数据的系统建模方法和统一数学结构，是形成大数据统一完备理论的前提和关键，也是大数据研究者的努力目标。由于大数据数据类型的复杂性，这一目标的实现看起来相当不易，即使是可行的，也必定要经历很长的路程；

2) 目前提出的一些数学方法主要是在统计学基础上发展的，“熵”的计算，随机矩阵理论等均是基于大数据具有统计特性假设基础上建立的；还有一些数学理论，是针对半结构化和非结构化数据的建模和分析而建立的。

5 智能电网大数据关键技术

5.1 分析架构

被广泛接受的大数据3层分析架构如图1所示^[14]，其中包含了数据访问和计算，数据隐私和领域知识，以及大数据挖掘算法。

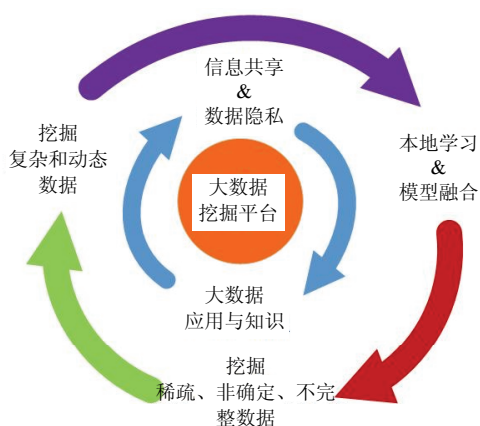


图1 大数据3层分析架构

Fig. 1 3 tier analytics architecture

对于内层架构，即大数据挖掘平台，其核心主要集中于数据访问和计算过程，随着智能电网中数据量持续增长，数据的分布存储将成为必然，而一

个高效的计算平台在计算时必须将分布式的大规模数据存储纳入考虑，将数据分析及处理任务分割成很多的子任务，并通过并行的程序在大量的计算节点上执行。在架构的外层，首先要对异构、不确定、不完备，以及多源的智能电网大数据通过数据融合技术进行预处理；其次，复杂和动态的数据在预处理之后被挖掘；之后，具有普适性的智能电网全局知识可以通过局部学习和模型融合获得；最终，模型及其参数需要根据反馈进行调整。分析架构的中间层对于内外2层起到重要的联系作用，智能电网大数据挖掘平台应该实现信息的共享与隐私的保护，而领域及应用知识的获取可以为数据挖掘工作提供参考。在整个过程中，信息共享不仅仅是每个阶段顺利进行的保证，同样是智能电网大数据处理和分析的目的所在。

5.2 关键技术

结合现有数据分析处理技术与智能电网的特点，本文将主要从大数据存储处理和数据解析2个方面总结大数据在智能电网中的应用技术。

1) 大数据存储及处理平台。

从大数据存储与处理之间相互关系的角度出发，主要的存储及处理模式可以分为流处理和批处理2种。流处理是直接处理，它将数据视为流，数据流本身具有大量、持续到达且速度快等特点，当新的数据到来时就立刻被处理并返回所需的结果，这种模式适用于电网中对实时性要求比较高的业务，如电源与负荷的联合调度以及设备的在线监测等。批处理是先存储后处理，其核心思想在于将问题分而治之，这种处理模式适合电网规划等对于实时性要求不高，但是数据量非常庞大繁杂的业务。由于智能电网是一个不断发展的系统工程，将来自方方面面的数据在逻辑上集中起来进行管控，无法保证其可行性、可靠性与可扩展性。而融合了分布式文件系统、分布式数据处理系统、分布式数据库等的云计算技术，可以作为大数据存储和处理的基础平台与技术支持，为大数据在智能电网中的应用服务。

2) 大数据的数据解析。

大数据解析，包含数据分析与解读2个方面。大数据分析是研究巨量的多种类型的数据，以发现其中隐藏的模式，未知的相互关系及其他有用信息的过程。为使分析结果被理解和应用，有必要进行大数据解读。大数据解读是对大数据本身及其分析过程进行深层次剖析以及多维度展示，并将大数据

分析结果还原为具体行业问题的过程。由于在解读的过程中伴随着对于数据本身的分析，因此大数据解读也可以看作是一种特殊的大数据分析方法。接下来，本文将介绍4种大数据解析技术^[14-18]。

①数据挖掘与融合。

数据挖掘是从大型数据集或数据流中提取有用的信息的过程，数据融合是将描述相同现实世界物体的多种数据或知识整合成为一致、准确、可解释的表达的过程。通过对不同来源的数据开展挖掘与融合，可以发掘出数据间潜在的关系，获得对事物更为准确的认知，如果坚持从片面、孤立的角度观察问题，只能如“盲人摸象”一般，由于局限的视角导致结论的偏颇。

通过对智能电网全景数据进行挖掘与融合，利用其在时间与空间上的冗余性与互补性，可得到对电网整体态势的实时、连续、宏观精准把握以及对电网局部状态的多角度、多层次、多模式精确感知。

电网在运行过程中会产生并积累大量的数据，如各种状态量数据、设备检修数据，用户用电数据等，大部分的数据在其利用周期结束之后就无法再继续创造价值，只能保存在各种存储介质当中无人问津。事实上，一些看似过期、陈旧且无用的数据仍有潜在价值。例如：电动汽车充电信息不仅是对一次次电能交互行为的记录，长期且大量的电动汽车充电数据对于充电站的布点极具参考价值；用户用电量长期以来主要作为负荷统计以及用户计费的依据而存在，而通过扩展分析就会发现它与住房空置率之间存在一定的联系；将部分电力电量数据免费或以估值出售的方式向部分用户开放并进行跟踪服务，有助于帮助这些用户实现节能降耗。

②领域普适知识挖掘。

传统的数据挖掘在面对大数据时已面临如下问题：a)知识数量庞大但可理解性不足；b)知识之间存在不一致性；c)对知识的普适性检验上存在欠缺，导致其可靠性受到质疑；d)与领域知识结合不足。针对以上问题，有学者提出领域普适知识挖掘概念：领域普适知识挖掘参考数据挖掘和复杂性系统科学研究的范式，旨在从大数据中发现统计意义上具有一定普遍性和不变性的知识与规律。

领域普适知识挖掘通过“大数据驱动的分析与挖掘→大数据呈现普适现象→预测、验证和评估”以及“系统抽象建模→理论分析和假设提出→仿真模拟和受控实验→涌现或发现普适现象”这两种范式的相互匹配与启发，螺旋式渐进地探索大数据层

面的普适规律，以及系统内部的工作机理。不同于传统数据挖掘方法得到知识的临时性和易变性，其应用优势在于：富有洞察性、可重复性和可预测性。

在智能电网的运行与规划过程中，存在一些普适性较强的经验或规律，但是由于某些原因，它们尚未或无法上升到理论的层面并以直观的形式展示在人们面前。而利用普适领域知识挖掘技术对智能电网大数据进行挖掘，获取隐藏于习惯或经验中的普适知识，可以帮助运行及规划人员提高对整个智能电网的认知。

③过程挖掘。

过程挖掘旨在建立处理模型和事件数据之间的联系，并将事件数据转化成有价值的见解。一方面，传统的业务流程管理(business process management, BPM)和工作流管理(workflow management, WfM)主要是模型驱动的，很少考虑到事件数据；另一方面，面向数据的分析技术(如数据挖掘、机器学习与商务智能)通常聚焦于简单的分类、聚类、回归，或规则学习问题，而没有考虑到端对端的业务模型。通过过程挖掘技术可以直接将事件数据和端对端的业务处理联系到一起。

智能电网中，过程挖掘可能会在引导用户错峰用电，为用户制订更为合理的节能方案，以及提升电网企业管理水平等方面发挥重要作用。

④数据可视化。

数据可视化技术是指运用计算机图形学和图像处理技术，将数据转换为图形或图像在屏幕上显示出来，并进行交互处理的理论、方法和技术，其基本思想是将数据库中每一个数据项作为单个图元素表示，大量的数据集构成数据图像，同时将数据的各个属性值以多维数据的形式表示，可以从不同的维度观察数据，从而对数据进行更深入地观察和分析。

通过将数据可视化技术与其他数据解析技术相配合，可能会为智能电网提供如下功能：a)概观与相关，对完整的电网数据集给出一个全貌，展示动态高维数据的发展趋势并对数据资产进行价值评估，对数据进行降维并基于各数据项彼此之间的相关性有选择地向业务部门或用户提供有价值的信息；b)放大、过滤与需求驱动，把用户的相关兴趣点放大并过滤掉不必要的信息，选择用户更为关注的浮动电价、不同居民用户电能消费特征、用户能耗等级，以及楼栋能源效率等内容进行细节化的展示；c)态势预估与展现，对电网中不断发展的具

有不确定性的变化点进行态势预估与宏观展现,如空间负荷增长趋势预测、网架扩展态势展现,以及极端天气的可视化应急响应等。可以说,数据可视化既是一种数据分析工具,又是一种结果展示方法,能够最为直观地体现智能电网大数据的应用方式与应用价值。

6 智能电网大数据重要领域

6.1 重要领域概述

参考近年来各国的研究,结合中国电科院各研究所开展的大数据应用需求分析,智能电网大数据目前重点在3个方面开展:一是为社会、政府部门和相关行业服务;二是为电力用户服务;三是支持电网自身的发展和运营,每个方向包含了若干技术领域,概括如表1所示。

表1 智能电网大数据重点方向和领域

Tab. 1 Key branches and fields

方向	重点领域
服务社会、 政府部门和相关行业	社会经济状况分析和预测
	相关政策制定依据和效果分析
	风电、光伏、储能设备技术性能分析
面向电力用户服务	需求侧管理/需求响应
	用户能效分析
	客户服务质量分析与优化
	业扩报装等营销业务辅助分析
	供电服务舆情监测预警分析
	电动汽车充电设施建设部署
支持公司运营和发展	电力系统暂态稳定性分析和控制
	基于电网设备在线监测数据的故障诊断与
	状态检修
	短期/超短期负荷预测
	配电网故障定位
	防窃电管理
	电网设备资产管理
	储能技术应用
	风电功率预测
	城市电网规划

限于篇幅,此处仅针对部分领域,介绍其应用需求和应用价值。

6.2 服务社会与政府部门类应用领域

1) 社会经济状况分析和预测。

电力关系经济发展、社会稳定和群众生活,电力需求变化是经济运行的“晴雨表”和“风向标”,能够真实、客观地反映国民经济的发展状况与态势。智能电网中部署的智能电表和用电信息采集系统,可获取详细的用户用电信息。用电信息采集系

统与营销系统所累积的电量数据属于海量数据,需要采用大数据技术,实现多维度统计分析、历史电量数据比对分析、经济数据综合分析等大数据量分析工作。对用户电量数据从分行业、分区域、分电价类别多维度开展用电情况统计分析,提取全社会用电量及相应社会经济指标,分析用电增长与相应社会经济指标关联关系,归纳总结各指标增长率与全社会用电情况的一般规律。通过对用户用电数据的分析,可为政府了解和预测全社会各行业发展状况和用能状况提供基础,为政府就产业调整、经济调控等做出合理决策提供依据。

2) 相关政策制定依据和效果分析。

通过分析行业的典型负荷曲线、用户的典型曲线及行业的参考单位 GDP 能耗,可为政府制定新能源补贴、电动汽车补贴、电价激励机制(如分时电价、阶梯电价)、能效补贴等国家和地方政策提供依据,也可可为政府优化城市规划、发展智慧城市、合理部署电动汽车充电设施提供重要参考,还可以评估不同地区、不同类型用户的实施效果,分析其合理性,提出改进建议。

6.3 面向电力用户服务类应用领域

1) 需求侧管理/需求响应。

根据不同的气候条件(如潮湿、干燥地带,气温高、低地区)、不同的社会阶层将用户进行分类;对于每一类用户又可绘制不同用电设备的日负荷曲线,分析其主要用电设备的用电特性,包括用电量出现的时间区间,用电量影响因素,以及是否可转移、是否可削减等,对于受天气影响的用电设备,如热水器、空调等,需分析其对天气的敏感性,当然,不同的季节、以及每日中的不同时间,用户用电对天气的敏感性都是不同的。分析不同用户对电价的敏感性,包括在不同季节、不同时间对电价的敏感性。在分类分析的基础上,通过聚合,可得到某一片区域或某一类用户可提供的需求响应总量,再分析哪一部分容量、多少时间段的需求响应量是可靠的,分析结果可为制定需求管理/响应激励机制提供依据。

2) 用户能效分析和管理。

对用户进行用电效率分析,首先需要采集到用户用电器分类用电数据。在智能电表部署之前,多采用侵入式方法,例如,在不同的用电设备接线处加装传感器。由传感器获取不同用电器的数据后,可以通过与典型数据、平均数据进行比对给出能效分析结论。在智能电表大量部署的情况下,由于智

能电表可以获得较短时间间隔的用电数据，无需再加装传感器，可以通过电表数据，识别用户端的不同类型负荷比例，并与典型数据的比对得出能效分析结果。

从海量用户的负荷曲线，采用数据挖掘技术，按照特定的函数算法，按行业、季度聚合成行业的典型负荷曲线模型，然后将所有的用户的负荷曲线与行业的典型负荷曲线进行对比，分析出与典型负荷曲线变化趋势不一致的用户，由此对用户的能效给出评价，并提出改进建议。

3) 业扩报装等营销业务辅助分析。

业扩报装辅助分析以营配集成为纽带，将用电信息采集系统、营销系统和 PMS 及 SCADA 系统的数据相融合，实现对变电站、线路及下挂用户和台区的负荷、电量监测分析，为加快业扩报装的速度和提高供电服务水平提供技术支撑。同时极大地提高电网设备运行的可靠性，为优化配电网结构，降低电网生产故障，提高公司用电营销管理精益化水平提供了手段。

4) 供电服务舆情监测预警分析。

通过与微博、微信等互联网新媒体的服务对接机制收集海量用电信息、用户信息以及互联网舆论信息，建设大数据舆情监测分析体系，利用大数据采集、存储、分析、挖掘技术，从互联网海量数据中挖掘、提炼关键信息，建立负面信息关联分析监测模型，及时洞察和响应客户行为，拓展互联网营销服务渠道，提升企业精益营销管理和优质服务水平。

5) 电动汽车充电设施建设部署。

融合电动汽车用户信息、居民信息、配电网数据、用电信息数据、地理信息系统数据、社会经济数据等，可利用大数据技术预测电动汽车的短中长期保有量、发展规模和趋势、电量需求和最大负荷等情况。参照交通密度、用户出行方式、充电方式偏好等因素，依据城市与交通规划以及输电网规划，建立电动汽车充电设施规划模型和后评估模型，对电动汽车充电设施的部署方案制定和建设后期的效能评估提供依据。

6.4 支持公司运营和发展类应用领域

1) 电力系统暂态稳定性分析和控制。

在线暂态稳定分析与控制一直是电力运行人员追求的目标，随着互联电网规模越来越大，“离线决策，在线匹配”和“在线决策，实时匹配”的暂态稳定分析与控制模式已不能满足大电网安全

稳定运行要求，因而逐渐向“实时决策，实时控制”的方向发展^[19-20]。

基于 WAMS 数据的电力系统暂态稳定判据和控制策略决策已有很多研究成果，但目前主要停留在理论研究阶段，并没有付诸实施^[13-14]。在大数据理论和技术指导下，需要将现有的分析方法与数据的处理技术相结合，不仅需要计算速度能否满足需求，还需要考虑数据的缺失和错误对分析结果的影响等问题。此外，如何将分析结果用直观的方法展示出来，有效指导运行人员做出科学的决策，也是需要解决的问题。

2) 基于电网设备在线监测数据的故障诊断与状态检修。

在实现 GIS、PMS、在线监测系统等各类历史数据和实时数据融合的基础上，应用大数据技术进行故障诊断，并为状态检修提供决策，可实现对电网设备关键性能的动态评估与基于复杂相关关系识别的故障诊断，为解决现有状态维修问题提供技术支撑。

3) 短期/超短期负荷预测。

分布式能源和微网的并网增加了负荷预测和发电预测的复杂程度。负荷预测也必须考虑到天气的影响以及能源交易状况，包括市场引导下的需求响应等。传统的预测方法无法体现某些因素对负荷的影响，从根本上限制了其应用范围和预测精度。应用大数据技术，建立各类影响因素与负荷预测之间的量化关联关系，有针对性地构建负荷预测模型，可更加精确地预测短期/超短期负荷。

4) 配电网故障定位。

利用大数据技术，配合故障投诉系统，融合 SCADA、EMS、DMS、D-SCADA 等系统中的数据作出最优判断，建立新型配电网故障管理系统，可以快速定位故障，应对故障停电问题，提高供电可靠性。此外，随着分布式电源在系统中比重的逐渐增加，其接入会影响到系统保护的定值及定位判据。对于带分布式电源的配电网故障定位也要根据不同的并网要求选择合适的定位策略。

5) 防窃电管理。

电力公司通过电量差动越限、断相、线损率超标、异常告警信息、电表开盖事件等数据的综合分析，建立窃电行为分析模型，对用户窃电行为进行预警；通过营配系统数据融合，可比较用户负荷曲线、电表电流、电压和功率因数数据和变压器负载，结合电网运行数据，实现具体线路的线损日结算，

通过线损管理功能不仅可以知道实施窃电用户所在的具体线路,并且可以定位至某一具体用户,克服目前检查范围广,查处难度大的问题。

6) 电网设备资产管理。

基于电网设备信息、运行信息、环境信息(气象、气候等)以及历史故障和缺陷信息,从设备或项目的长期利益出发,全面考虑不同种类、不同运行年限设备的规划、设计、制造、购置、安装、调试、运行、维护、改造、更新直至报废的全过程,寻求寿命周期成本最小的一种管理理念和方法。依据交通、路政、市政等可能具备的外部信息,如工程施工、季节特点、树木生长、工程车 GPS 等外部信息,关联电网设备及线路 GPS 坐标,对电网外力破坏故障进行预警分析。

7) 储能技术应用。

由于储能系统大多是由数量庞大的电池单体组成(动辄以万计),每个电池单体又包含单体电压、电流、功率、电池荷电状态、平均温度、故障状态等相关信息,汇总起来整个电站监测信息可能达到数十万个点,储能相关数据量十分庞大。利用大数据分析技术,可对储能监控系统相关数据进行有效采集、处理与分析,为储能应用提供依据。

8) 城市电网规划。

通过实现用户用电数据、用户停电数据、城市电力服务数据(95598 客服电话)、基于 GIS 的城市配电网拓扑结构和设备运行数据、城市供电可靠性数据、气候数据和天气预报数据、电动汽车充电设施建设及利用数据、人口数据、城市社会经济数据、城市节能和新能源政策及实施效果数据、分布式能源建设和运行数据等、社交网站数据的整合,识别城市电网薄弱环节,辅助城市电网规划。在上述数据融合的基础之上,利用人口调查信息、用户实时用电信息和地理、气象等信息绘制“电力地图”,可以街区为单位,反映不同时刻的用电量,并将用电量与人的平均收入、建筑类型等信息进行比照。通过“电力地图”,能以更优的可视化效果反映区域经济状况及各群体的行为习惯,为电网规划决策提供直观依据支撑。

7 研究框架和技术发展路线图建议

7.1 研究框架

智能电网大数据研究体系可分为 3 个层次:第 1 层是大数据相关理论,理论是认知的必经途径,也是大数据技术发展和相关新技术衍生的基础;第

2 层是技术层面,包含了现有的和即将发展和产生的数据存储、处理和分析技术,技术是大数据价值体现的手段和前进的基石;第 3 层是大数据的应用层面,应用层面是大数据的最终价值体现。

智能电网大数据研究和技术应围绕着大数据研究体系的 3 个层次展开:1)基础理论研究;2)智能电网中的大数据关键技术研究;3)技术领域大数据应用研究。

大数据的基础理论正在形成中,新的理论的形成又将带动新技术的发展。为支持智能电网大数据技术的发展,需跟踪大数据基础理论研究,逐步形成智能电网大数据的理论体系、知识体系和技术体系。大数据是典型的“大科学”,不是一个专业领域的研究人员和工程技术人员可以独立研究的单一学科,是需要多个专业人员协同合作的综合性学科。为有效开展智能电网大数据应用研究,融合不同专业人员的知识技能,需要研究形成智能电网大数据的系统方法论。

智能电网中的大数据关键技术包括数据的获取、集成和融合、数据质量控制、存储、处理和分析等多个方面。数据的获取和融合以实现互操作为基础,互操作有着不同的层次,包括管理机制、通信系统、数据模型等几个层次。为了获取需要的数据以及保证数据的质量和数量,需要研究新的数据管理机制,若有必要,还需要建设和改造信息通信系统。以用户用电信息采集系统为例,由于局域网的通讯并没有采用通用的协议,各厂商在物理层、数据链路层等并未遵循统一的规范和标准,导致了不同厂商生产的通讯模块之间无法实现互联互通,设备的兼容性和互换性无法得到解决,也由于最后 1 公里通讯的可靠性难于保障,用户用电信息采集数据的获得仍得不到保证,同时还存在着营配数据不统一问题,影响了数据价值的挖掘;在通信系统完善、营配数据一体化基础上,可实现很多高级应用,如网络重构、窃电分析、故障预测、稳定运行和资产管理。此外,为实现数据的融合,还需要建立公共数据模型,这是因为属于同一个实体或概念的数据往往在多个数据源中以不同的形式表示,数据集成和融合技术被用于将这些不同形式的数据进行统一和集成。传统的数据融合和质量控制技术主要针对规模较小且语义清楚的结构化数据,而面向大数据的数据融合和质量控制仍然是一个需要深入研究的问题。具体而言,需要在现有的数据集成为融合技术的基础上,结合大数据的异构性、冗

余性和相关性等特性,研究大数据的数据融合和集成方法,以有效地解决大数据获取的全面性和一致性问题。智能电网大数据分散存储不同的地方、不同的系统中,大量的、实时性的数据集中放置在一个平台上,常常是技术上不可行或不满足经济性,需要研究如何实现分散存储的数据之间的融合及同步处理分析技术。

在数据获取和融合基础上,需要结合不同的应用进行数据的管理、处理,并选择适宜的数据分析方法。在这个层次,既包含通用技术,也包含智能电网专用技术。构成智能电网大数据的不同数据源产生的数据具有不同的采样周期和生命周期,需要研究智能电网中不同数据源之间的数据接入和融合、处理技术。

在应用层面,需要针对智能电网相关技术领域,深入开展大数据技术应用需求分析、场景设计、分析模型、专业方法的研究。应用领域包含但不限于第6节的技术领域。

7.2 技术发展路线建议

基于大数据技术发展和应用现状,结合我国的实际情况,对我国智能电网大数据发展路线建议如下。

2014年:

1) 确定智能电网大数据的知识体系、技术体系和研究框架;

2) 着力培养一支从事智能电网大数据研究的专业队伍;

3) 继续深入开展智能电网技术领域大数据技术需求分析、场景设计、应用价值分析;

4) 完成智能电网大数据平台研究和开发。

2015年:

1) 形成智能电网大数据研究方法论。

2) 重点开展智能电网大数据的数据接入和融合相关关键技术研究,包括:

①完成智能电网大数据标准体系研究,重点完成智能电网大数据统一数据模型研究,参与国际智能电网大数据标准化研究。

②研究可实现智能电网数据融合的技术改造措施和相关关键技术,包括通信信息系统建设和完善、数据安全性和隐私保护相关技术研究,提高智能电网大数据质量。

③重点在营配数据一体化基础上开展以用户用电信息采集数据为核心的大数据应用研究,为更为广泛的技术研究奠定基础,积累经验。

2016年:

1) 完善智能电网大数据标准体系;

2) 完成营配数据一体化大数据应用研究;

3) 研究如何进一步提高智能电网大数据的数据质量和数据融合程度,扩大数据融合范围。

2017—2020年:

在智能电网其他技术领域全面开展大数据应用研究。

8 结论

1) 智能电网是大数据的重要应用领域。一方面,随着智能电网的建设,产生了大量的量测、监测数据,如何处理这些数据,挖掘其价值,是电力公司面临的问题;另一方面,利用大数据技术,不仅可充分利用电网自身的数据,还可以充分利用外部数据,大力提升电网的发展和运营水平,提高电网公司服务社会、服务用户的水平,扩展增值业务。

2) 大数据不仅是一项综合性技术,也是一门科学。大数据有广泛、深厚的哲学和数学理论基础,但其理论体系尚未形成,随着新的理论和方法的形成,也会催生新的技术。电力工作者应紧密跟踪大数据理论和技术的发展,为有效利用大数据技术,做好思想和技术上的储备。

3) 国内外智能电网大数据研究和工程应用刚刚起步,尚处于探索阶段,为推动智能电网大数据发展,需多方努力,达成共识,共同消除存在的障碍,制定各个层面的发展战略,才能使智能电网大数据得到健康有序发展。

4) 智能电网大数据的研究和工程应用,是一个复杂的工程,需要经过一个漫长的发展过程,文中提出的研究框架和技术发展路线,可供领导者和研究者参考。

参考文献

- [1] 刘振亚. 智能电网技术[M]. 北京: 中国电力出版社, 2010.
Liu Zhenya. Smart grid technology[M]. Beijing: Chinese Electric Power Press, 2010(in Chinese).
- [2] Gantz J, Reinsel D. Extracting value from chaos [J]. Proceedings of IDC iView, 2011: 1-12.
- [3] GTM Research. The soft grid 2013-2020: Big data & utility analytics for smart grid-research excerpt [R/OL]. GTM, 2013. <http://www.giiresearch.com/report/gm257044-soft-grid-big-data-utility-analytics-smart-grid.html>.

- [4] EPRI. Big data challenges for the grid: EPRI survey results and analysis Data analytics and applications demonstration newsletter[R/OL]. 2013. http://www.smartgridnews.com/artman/publish/Business_Analytics/Big-Data-challenges-for-the-grid-EPRI-survey-results-and-analysis.
- [5] EPRI. The whys, whats, and hows of managing data as an asset[R]. USA: EPRI, 2014.
- [6] IBM. Managing big data for smart grids and smart meters[R/OL]. IBM Software White Paper. http://www.smartgridnews.com/artman/publish/Business_Strategy/Managing-big-data-for-smart-grids-and-smart-meters-5248.html.
- [7] Oracle Utilities. Utilities and big data: A seismic shift is beginning[R]. An Oracle Utilities White Paper.
- [8] 科技部. 国家高技术研究发展计划(863 计划)2015 年度项目申报指南[EB/OL]. <http://program.most.gov.cn/htmledit/639A0448-5482-4F63-42C4-95368125A2F8.html>. MOST. National High Technology Research and Development Program of China (863 Program) 2015 application guide[EB/OL]. <http://program.most.gov.cn/htmledit/639A0448-5482-4F63-42C4-95368125A2F8.html>(in Chinese).
- [9] 李德伟. 同构关系: 大数据的数理哲学基础[N]. 光明日报, 2012-25-12.
Li Dewei. Isomorphous relationship: the mathematic and philosophical foundations of big data[N]. Guangmin Daily, 2012-25-12(in Chinese).
- [10] 陈俊欢. 论大数据理论的基本原理及其创造价值[EB/OL]. <http://www.docin.com/p-714878917.html>. Chen Junhuan. View on the basic principle and create value of big data theory[EB/OL]. <http://www.docin.com/p-714878917.html>(in Chinese).
- [11] 李国杰. 大数据研究的科学价值[J]. 中国计算机学会通讯, 2012, 8(9): 8-15.
Li Guojie. Scientific value of big data research[J]. Chinese Computer Academy Newsletter, 2012, 8(9): 8-15(in Chinese).
- [12] Qiu R, Antonik P. Big data and smart grid: A Random Matrix[M]. John Wiley and Sons, 2014: 33-60.
- [13] Wu Xindong, Zhu Xingquan, Wu Gongqing, et al. Data mining with big data[J]. IEEE Transactions on knowledge and data engineering, 2014, 26(1): 97-107.
- [14] Hu Han, Wen Yonggang, Tat Seng Chua, et al. Toward scalable systems for big data analytics: a technological tutorial[J/OL]. Digital Object Identifier 10.1109, 2014: 657-668.
- [15] 中国电机工程学会. 中国电力大数据发展白皮书[R]. 北京: 中国电机工程学会, 2013.
Chinese Society of Electrical Engineering. China utility big data development whitepaper[R] Beijing: Chinese Society of Electrical Engineering, 2013(in Chinese).
- [16] 陈俊欢. 论大数据理论的基本原理及其创造价值[EB/OL]. <http://www.docin.com/p-714878917.html>.
Chen Junhuan. Fundamental principle and created value of big data theory[EB/OL]. <http://www.docin.com/p-714878917.html>(in Chinese).
- [17] 赵国栋, 易欢欢, 糜万军, 等. 大数据时代的历史机遇[M]. 北京: 清华大学出版社, 2013.
Zhao Guodong, Yi Huanhuan, Mi Wanjuan, et al. Historic opportunity in the era of big data[M]. Beijing: Tsinghua University Press, 2013(in Chinese).
- [18] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6): 647-657.
Li Guojie, Cheng Xueqi. Big data research: Important strategy area of science & technology and economy & society development in the future: status quo and scientific thinking of big data[J]. Bull Chin Acad Sci, 2012, 27(6): 647-657(in Chinese).
- [19] 吴为. 基于响应的电力系统暂态稳定性实时判别与控制技术的研究[D]. 北京: 中国电力科学研究院, 2014.
Wu Wei. Studies on real time transient stability determination of power system and control technology based on response data[D]. Beijing: China Electric Power Research Institute, 2014(in Chinese).
- [20] 顾卓远. 基于响应的电力系统暂态稳定控制技术研究[D]. 北京: 中国电力科学研究院, 2014.
Gu Zhuoyuan. Research on response-based power system transient stability control technique[D]. Beijing: China Electric Power Research Institute, 2014(in Chinese).



张东霞

收稿日期: 2014-10-08。

作者简介:

张东霞(1964), 女, 博士, 教授级高级工程师, 主要研究方向为电力系统分析与规划、智能电网技术及其标准化, zhangdx@epri.sgcc.com.cn。

(编辑 李泽荣)