

大能源思维与大数据思维的融合

(一)大数据与电力大数据

薛禹胜, 赖业宁

(南瑞集团公司(国网电力科学研究院), 江苏省南京市 211106)

摘要:大能源思维将电力视为能源生产与消费全流程中的枢纽环节,藉此推动上游一次能源的清洁替代与下游终端能源的电能替代,支撑能源的可持续发展。大数据思维将各种数据资源从简单的处理对象转变为生产的基础要素。这两种思维的融合,使电力大数据成为大能源系统广泛互联、开放互动及高度智能的支撑,包括:广域多时间尺度的能源数据及相关领域数据的采集、传输和存储,以及从这些大量多源异构数据中快速提炼出深层知识并发挥其应用价值。作为两篇论文中的开篇,在演绎大数据基本概念、结构类型及本质特征的基础上,归纳电力大数据的特点。针对综合能源,通过基于数学模型的因果型数据、无因果关系的统计型数据以及参与者博弈型数据的融合,构建信息能源系统的知识挖掘平台。其续篇将讨论信息能源系统,并通过若干案例,反映大数据思维对提高大能源经济性与可靠性的贡献。

关键词:信息能源系统; 能源互联; 因果型数据; 统计型数据; 博弈型数据; 非结构型数据; 知识提取

0 引言

现代产业与信息技术的发展使数据资源成为越来越重要的生产要素。爆炸式增长的数据量对多源、异构、高维、分布、非确定性的数据及流数据的采集、存储、处理及知识提取提出了挑战。大数据思维就是此环境中的产物,它并不局限于传统的基于因果关系的逻辑推理研究,甚至更多地通过统计型数据的搜索、分类、比较、聚类、关联等的分析和归纳,关注数据集内隐藏的相关性(支持度、可信度、关注度)。图灵奖得主吉姆·格雷将这种数据密集型的研究范式从理论科学、计算科学、实验科学中分离出来,成为“第四范式”^[1]。

2008年《Nature》推出Big Data专刊^[2]后不久,文献[3]将大数据计算称为商务、科学和社会领域的革命性突破。2011年,《Science》的Dealing with Data专刊阐述了大数据对科学研究的重要性^[4]。同年,麦肯锡公司发布关于大数据的竞争力、创新力和生产力的调研报告^[5],分析了大数据的影响、关键

技术及应用领域,进一步燃起学术界和产业界对讨论大数据的热情。美国政府于2013年3月宣布投资2亿美元启动“大数据研究和发展计划”^[6],将其上升为国家战略。大数据被认为是继物联网、云计算之后,IT行业又一次颠覆性的技术变革。

云计算是利用互联网实现随时、随地、按需、便捷地访问共享资源池(如计算设施、应用程序、存储设备等)的计算模式^[7],关注计算能力,并与关注知识提取的大数据技术相得益彰。

电力是社会发展的基础。随着全球能源形势的日益严峻,各国大力开展了智能电网的研究和实践^[8-9]。其目标是建设一个覆盖整个电力生产过程的实时系统,以支撑电网的安全、自愈、绿色、坚强及可靠运行^[10]。而其基础则是电网广域全景实时数据的采集、传输、存储、分析及决策支持。

而愿景中的电网则是网架坚强、广泛互联、高度智能、开放互动的智能电网。文献[11]分析了智能电网大数据的产生与特点,指出已有的数据处理技术在智能电网的数据存储、处理和展示等方面所面临的挑战,已成为智能化发展的制约因素。文献[12]探索了基于云计算的智能电网信息平台的可靠存储和高效管理。文献[13]研究了用户侧电力消

收稿日期: 2015-12-08。

国家自然科学基金重点资助项目(61533010);中英合作研究项目(513111025-2013);中泰合作研究项目(5151101161);国家电网公司科技项目。

费的大数据,从中挖掘其用电行为,以改进需求响应系统。

以电为核心的大能源体系正在成为全球能源的发展战略^[14]:一次能源的清洁替代与终端能源的电替代都必须依靠电网的输送才能大规模地实现。文献[15]指出,电力可靠性问题的顶层设计应该以大能源观为指导,不能局限于电力系统自身,还要分析其与一次能源、外界环境、管理政策、用户需求与方式变化间的交互,研究广义阻塞对电力系统安全性与充裕度的影响。

电力、能源及广义环境的多源、多态及异构数据的数量呈指数级增长,需要有相应的广域采集、高效存储和快速处理技术予以支撑。而从这些数据中挖掘知识及价值应用则使电力大数据的话题融入大数据的研究热潮。本文及其续文是笔者关于如何将电力大数据的思维应用到电力的广义可靠性、大能源安全及环境安全方面的思考。

1 大数据概念的演绎

1.1 定义

对“大数据”还没有统一的定义,通常指量大、多源、异构、复杂、增长迅速,无法用常规的方法处理的数据集合^[16]。许多数据往往只在统计学观点上具有某种相关性,而不一定像传统应用的数据那样具有严谨的因果关系。对这样的统计关系型数据,只有当反映一个真实问题的数据量达到能在一定程度的统计意义上描述其真实面貌时,才能有效地提取知识,支持决策。而对于常规的因果关系型数据来说,数量的大小往往仅影响到计算资源,而与提取知识的方法关系不大。

因此,大数据与小数据之间并无绝对的界限,而是相对于目标问题而言的。大规模的数据量只是大数据概念的特征之一,也不应该用海量的规模作为大数据的必要条件。

大数据技术涵盖了从信息产生、采集、存储、转换、集成、挖掘、分析、计算、展示、应用及维护等数据全生命周期管理过程^[17],需要具备从不同类型的多源异构数据中,快速提炼出有价值信息的能力。

1.2 基本特征

数据是以编码形式对自然现象、社会现象、试验仿真结果及经验的记录,包括数字、文字、图像、声音等形式。与传统数据相比,大数据具有四大显著特征4V,即量(Volume)、类(Variety)、速(Velocity)、值(Value)^[18]。“量”是指数据容量足够大;“类”是

指数据种类呈现多源多态特性;“速”是指实时性要求高;“值”是指数据价值密度相对较低。在数月的监控视频中,可能仅1~2 s的画面有用,但通过关联数据的挖掘、分析和提取,却可能获得很高的信息价值。

一般认为,数据规模或复杂程度超出了常用技术按照合理的成本和时限来处理的能力,就可称为大数据。笔者认为如果统计型数据量对于具体目标问题来说,已经具有统计意义,就可以用大数据思维来处理。若为大数据设立数百TB的阈值,必将大数据思维束之高阁,扼杀了其广泛应用的生命力。

除了从因果、统计及博弈等数据关系的视角来对数据类型分类外,还可以按数据结构的视角将其分为三种类型:①结构化数据,即可以用二维表结构来逻辑表达的行数据,关系型数据库是其有效处理工具;②非结构化数据,是一类不能用有限规则完全表征与刻画的异构性数据,如图像、音频和视频等信息,它们之间不存在直接的因果关系,存在不相容性及认知的不一致性;③半结构化数据,介于上述两种结构之间的数据,可以用树、图等模型表达,如Web、办公文档及电子邮件等信息。据统计,随着社交网络、传感物联及移动计算等新技术、新渠道的不断涌现,企业中的半结构化或非结构化数据占比越来越大,有的已达80%。

1.3 思维方式

大数据技术的思维方式是:将采集到的经验与现象实现数据化与规律化,在继承传统的统计学、计算数学、人工智能、数据挖掘等方法的基础上,从单一维度转向多维度统筹融合,开发知识处理的新方法,从更深刻的视角,以更高的时效发掘多源异构数据,从而发现新知识和新规律,并实际应用的方法学。

若可以用简单的表达式直接描述某问题的结果与变量之间的关系,那么即使其数据量很大,也不一定属于大数据思维。当数据之间没有或还不能严格描述其因果关系,而数据集已相对地具备统计意义,就需要并可以采用大数据思维来处理。例如:对于一个市场调查的数据集,如果仅仅求取人群对某种商品的需求分布,那么思维方法并不会因为数据集是否海量,而改变常规的数据处理思维。但如果要从人群对该商品众多技术、经济、全程服务特性等各方面要求中提炼出有助于提高企业竞争能力的决策支持,那就需要有新的思维方法,包括信息挖掘与利用的思维、模式及方法,以帮助人类获得更深刻的洞

察力。

关注统计关系的思维方法,同样可以应用于有限数据集上^[19],只是其结论的可信度可能受到影响。

1.4 大数据技术的挑战

1.4.1 大数据的采集与集成

利用各种传感器及终端,采集反映物理世界、市场经济与人类行为等现象的静态/动态的异构数据,成为人、机、物三元世界的信息纽带^[20]。其挑战主要体现在:①从结构化数据为主,向结构化、半结构化与非结构化的三者融合的变革;②数据来源的多样化,特别是移动终端的广泛应用,使大量数据需要同时带有时间与空间的标志;③有用信息被淹没在大量无关或弱相关的数据中,或需要处理低质量及局部缺损的数据。

1.4.2 大数据的存储

大量多态异构数据的高效、可靠、低成本存储模式是大数据的关键技术之一,对多源多态数据流之间的交互索引与转换效率影响很大。

数据压缩技术可以减少数据传输量及提高存储效率,但也增加了数据处理环节及计算负担^[21]。分布式存储要权衡对存储空间及实时性的影响;对实时数据采用实时数据库或内存数据库;对核心业务数据采用传统的并行数据仓库系统;对大量的历史和非结构化数据采用分布式文件系统。

算法在处理复杂结构数据时相对低效,故希望事先为复杂结构的数据建立索引结构来帮助搜索,并合理地将非结构型数据结构化。

1.4.3 大数据的分析

需要关注大数据的形态描述、基本运行规律及其可控性。其中,最复杂的是人类社会行为信息,其决策行为必须与物理系统及信息系统相结合。

目标领域的信息若能与关联领域的信息相结合,则不但知识量得以增加,而且更有可能揭示综合性、交叉学科的未知知识,甚至从统计型数据中发现其(准)因果关系。知识的涌现性反映在模式、行为和智慧上。例如:将提高风机效率的研究与气象信息、电网信息联系,其效果将大大提升。大数据与云计算为之提供了数据资产的保管、搜索的技术,但也不是数据越多越好,而信息的提炼与知识的发现一般很难通过直观方式的分析和解读来获取。

数据驱动的分析方法包括:聚类、判别、回归、识别、隐变量、主因分析、时间序列。需要处理大规模的不定解问题,及必要时信息的补充问题,利用特征

的相关性来发展统计学,有效地表达高维随机变量函数的联合概率分布。

用以发现知识的技术有:遗传算法、神经网络、数据挖掘、专家识别、回归分析、聚类分析、关联分析、数据融合、机器学习、自然语言处理、情感分析、空间分析、时间序列分析及其他计量学方法。一个例子是 Google 采用机器学习和神经网络来分析服务器群的数据,掌握大量变量之间的交互。神经网络在无显式编程下自适应学习,大大提高服务器群的效率,发现普通人注意不到的复杂互动关系。

随着大数据的应用从离线走向在线,甚至实时,所面临的系统复杂性、数据复杂性和计算复杂性挑战更为严峻。目前虽然出现了一些较为有效的方法,如流处理模式、批处理模式,及两者的融合^[22],但仍未有一个相对普适的(准)实时的分析处理框架,在合理精确性的前提下实现快速的随机优化。

1.4.4 大数据的易用性

易用性应该贯穿在大数据的集成、存储、计算、分析,乃至展现等整个业务流程^[23]。从技术层面看,可视化、人机交互及数据起源技术都可有效提升易用性,但仍存在元数据的高效管理的难点。元数据是关于数据的组织、数据域及其关系的信息,是描述信息资源的重要工具。

1.4.5 大数据的安全性

数据可靠性和通信网络安全性至关重要。必须研究各种广域量测数据和仿真数据的完备性、适时性和价值的评估技术,研究在广域信息不完全条件下的分析、控制技术。

“8·14”大停电前的几个月,北美电力可靠性委员会(NERC)警告说,随着电力业务的复杂化,越来越多的电力公司不遵循 2002 年发布的非强制性的安全导则,致使一些数据采集与监控(SCADA)网络被蠕虫破坏。大数据安全是一项包括技术层面、管理层面、法律层面的社会系统工程,其保障体系的框架由组织管理、技术保障、基础设施、产业支撑、人才培养、环境建设组成。应该研究数据源和传输的可靠性,研究信息系统故障或受到攻击时的行为,以及信息的阻塞、淹没、丢失和出错对大能源可靠性的影响。

移动互联时代中,人们在不同场合产生各种数据足迹;大数据技术将大量行为信息聚集在一起,就很可能暴露其隐私。由于尽可能地获取信息本来就是应用大数据技术的目的,故与隐私权的保护存在着矛盾。如何协调共享与隐私,需要制度与监管的

保证。

现有的数据安全保护技术主要针对静态数据集^[24],包括开放与防御的平衡,防止数据被窃取或篡改。但仍需要解决动态信息的安全性问题。

1.4.6 大数据的应用

大数据学科的兴起与人类需求密切相关,其中的信息挖掘及知识提炼环节必须与待解决问题紧密结合,而应用环节则是大数据技术发展的动力与归宿。因此,大数据研究应该遵循问题导向、需求牵引及数据共享的原则。必须结合具体的目标问题,将采集到的低价值的大数据加工成高价值的思想或知识,大数据技术才有生命力。若没有应用企业的深度参与,若不能按照商品的规律运行,大数据技术就难以真正取得收效。此外,若没有各管理部门对数据共享的落实,大数据技术也只能是纸上谈兵。

知识作为资源,需要像物质资源那样分配及流通。为此,需要对知识产品定价,并从数据使用的视角揭示信息流与科学活动的内在规律。

目前对大数据应用的关注,主要集中在商业与服务业这些以统计关系为主的领域^[25],较少涉及具有较强因果关系的领域,如电力系统及能源市场等领域。笔者认为,是否能成功地将统计关系、博弈关系与因果关系取长补短,是大数据技术能否扩大应用领域的一个关键。

2 电力大数据

2.1 问题的提出

从能源系统看,它包括电网和非电的能源网两部分;电力系统应该在能源的清洁替代与电能替代中有效地承上(一次能源)启下(终端能源),兼顾左右(与其他二次能源协调)。

从信息系统看,它包括专用通信网与互联网(Internet)两部分。专用网用于对信息安全性或实时性要求高处,但由于非开放式的接入,故不适合与广泛用户的互动;互联网用于开放场合,但信息安全及实时性差。显然,两者对于信息系统来说都不可或缺。

能源革命需要的是能源系统的全部组分与信息系统的全部组分的深度融合^[26],即大能源系统与大信息系统组成的信息能源系统,而不是局限于它们的某个子集之间。但是,智能电网的研究范畴基本上局限于能源系统中的电力系统与通信系统中的专用网^[27],因此其信息物理融合的概念也就局限于“电力生产信息+电力系统”。作为能源革命重要环

节的“一次能源系统及终端能源系统”,以及作为信息革命重要环节的“网络金融及需求侧参与信息”却并没有得到应有的重视。

显然,目前的智能电网框架中的专用通信网的功能需要从电力系统扩展,不但涉及各种非电的能源环节,而且涉及相关的非能源环节,以更好地支持对电力可靠性及经济性的研究,并支撑综合能源安全、能源经济安全及环境安全。另一方面,开放的网络经济及广泛的用户参与都需要互联网的支持,互联网数据的管理与挖掘成为非常重要的任务。

整个能源行业在转型发展中面临的机遇和挑战,来自一次能源的压力、环境安全的紧迫性、电力系统内外复杂性的增加、运行环境及扰动事件的不确定性与风险的增大、经济与技术的发展、市场改革的要求。为了应对上述挑战,必须提高数据的及时性、完整性、一致性及信息安全防护能力,提高对数据的管控能力,消除数据壁垒、存储无序且不一致的现象,完善对外部环境、基础设施及人才队伍的掌控。

从传统的电力数据演变为电力大数据,跨领域的时空扩展将电力系统的界面条件从确定性变为时变性,同时也增加了多时间尺度的动态复杂性^[28]。涉及各类数据的采集、集成、存储、管理、知识挖掘、决策支持、可视展现等技术,也反映了电力及综合能源数据的管理、知识的挖掘和应用等一系列推进能源生产、转换、输送、消费方式的创新思维。其中的互联网数据大多为传统数据库不支持的非结构化类型,包括图片、视频、音频、地理位置信息、网络日志、博弈行为、金融动态、政策法规。数据的在线或实时处理,往往呈现出突发及涌现等非线性状态演变现象,难以预测。

为此,信息创新必须与能源革命在更高层次上深度融合,特别是在一次能源中的清洁替代及终端能源中的电能替代上,不但将协调优化的概念提升到综合能源流的范畴,而且推动电能更主动发挥在一次能源与终端能源之间的核心纽带作用,实现能源生产模式与消费模式的革命。

但是,将大数据技术局限于互联网数据也是片面的。通过专用网或仿真网获取的数据,包括智能电网概念中已涉及者,以及有待外拓的非电能源领域与非能源领域中的数据问题,同样存在大量的挑战。既然讨论的是信息物理融合问题,那么其框架就更不应该在信息系统内部或能源系统的内部设立藩篱。

当前的研究都在一定程度上受限于物理系统中的藩篱或信息系统中的孤岛。例如:①稳定性与经济性的研究都针对给定的边界条件,不能真实反映上下游环节的变化对电力系统的影响;②充裕性问题被粗犷处理为固定比率的备用容量,无法适应大规模不确定性的可再生能源及充放电用户的入网;③决策过程基本无法考虑博弈行为的影响;④忽视了信息系统本身可靠性的影响;⑤缺乏对电力系统外部环节的自适应能力。

要突破上述藩篱与孤岛,就必须遵循以电力系统为核心环节的大能源系统在大数据时代下的发展理念,顺应管理体制及技术路线的变革。

2.2 电力大数据的特征

一方面,电力大数据具有大数据的共性,包括目标领域向其他相关领域的扩展,以及数据类型向非结构型及非因果型数据扩展,时间维度向多尺度的流数据扩展。由此形成大量的异构异质数据,包括数字、文本、图形、图像、音频、视频等静态和不同时间尺度的动态数据,以及大量统计关系与博弈关系的数据,都需要快速处理。

另一方面,电力大数据必然继承了能源行业数据的特征,包括大量的因果关系数据、高维的时空数据,广域的监测控制,快速的时间响应及实时控制数据。除了电力系统的状态外,还需要获取并分析相关领域的的数据,并处理部分数据缺失时的不确定性。

运行工况或故障场景都会影响系统的稳定性和控制策略。中国的电网现在已普遍实现了在线的量化分析功能,按实测工况和典型故障来指导预防控制,并向自适应的紧急控制与系统保护发展^[29]。但若有效地应对极端自然灾害环境,则还需要采集并处理大量非结构型的视频、语音、图片,并与电力系统分析功能有机结合。大规模间歇性可再生能源(RES)与电动汽车(EV)的入网对电力充裕性与备用调度提出挑战,要求大大提高对风电、太阳能发电及EV充放电的预测精度,这就需要处理大量非结构型的地理及交通动态数据。为了应对相继故障,基于常规因果关系统型数据的分析算法也应该解决多米诺效应的演化路径及实时评估的经典难题。

2.3 电力大数据的类型

除了按数据结构等大数据概念来划分电力(能源)系统大数据,还存在其他视角,列举如下。

1)按业务领域,电力大数据可分为:①规划运行类,包括电力规划、电能生产、运行监控、设备检修等数据;②企业运营类,包括企业发策、运营、电力市

场、用户信息等数据;③企业管理类,包括人财物资、资本运作、企业资源计划(ERP)管理、协同办公等数据;④非电的能源类,包括各种一次能源、非电的二次能源、终端能源使用模式等数据;⑤非能源类,包括气象、环境、碳资产、宏观经济政策等数据。

2)按时间维度可分为:①背景及法规数据;②历史数据;③调查及预测数据;④在线实测数据;⑤仿真推演数据。

3)按数据采集来源可分为:①书本及文档;②专用网数据;③互联网数据;④问卷及调查;⑤交互博弈及仿真结果。

2.4 电力系统的数据量

常规 SCADA 系统按采样间隔 3~4 s 计算,10 000 个遥测点每年将产生 1.03 TB 数据(12 B/帧×0.3 帧/s×86 400 s/d×365 d×10 000 遥测点);广域相量测量系统的采样率为 100 次/s,10 000 个遥测点每年将产生 495 TB 数据。美国 PGE 公司每个月从 900 万个智能电表收集超过 3 TB 的数据。国家电网公司的 2.4 亿块智能电表,年产生数据量约为 200 TB,而整个国家电网公司信息系统灾备中心的数据总量,接近 15 PB。以 2004 年山东系统(97 台机、462 条母线、702 条支路)为例,设在线每 5 min 对 220 kV 以上线路的三相永久故障分析一次暂态稳定性,仿真时长 10 s,仿真步长 0.01 s,则一年内将产生 1 092 TB 的数据量。当大规模的间歇性分布式可再生能源入网后,发电侧计及气象数据的实时监控的数据量比传统发电成倍增加。智能配电、智能电表、电动汽车和用电技术的发展也大大增加了需求侧的数据量。电力企业的精细化管理,与一次能源、环境、交通、市政等外部系统的联系日益深化,对数据量的依存度也越来越高。

2.5 电力大数据对电力可靠性的支撑

整合各种广域系统采集的静态和动态数据,包括雷电场、台风风力场、山火场等非结构型数据与常规的电力数据及仿真信息相结合,从所获的现场实测或仿真得到的时间响应曲线中提取深层知识。以支持稳定性与充裕性的量化分析及自适应控制,协调故障前的预防型博弈,故障后瞬时的故障隔离和预测型紧急控制,检测到违约症状后的校正控制,以及系统崩溃后的恢复控制。在此基础上,综合防御框架还应该实现多道防线的时空协调,通过协同各道防线和各种控制手段,最大限度地减少大停电的风险。

信息的可靠性是分析和决策的物质基础;机理

分析和量化分析则通过数据挖掘来深刻掌握电网的行为特性^[30];正确的决策是为了能以最小的风险代价来最大限度地满足电力需求。互为补充的相量测量单元(PMU)、远方终端设备(RTU)和仿真数据共同满足系统在空间中和时间上的可观性和可控性^[31]。需要研究在不完备数据下,如何将信息论和系统论相结合,从数据中挖掘深层信息。

在电力的长期可靠性分析与控制方面,需要考虑社会、经济、科技、能源与政策等发展的不确定性,以及各种博弈行为的影响,将目前依靠主观判断的方式逐步提升为基于混合仿真推演的方式。

3 信息能源系统的大数据平台

3.1 电力(能源)系统运行数据的采集

关系数据库强调完整性及一致性,云计算强调可扩展性,但难以保证实时性。Internet 的 TCP/IP 协议在流量控制和数据纠错时会产生 10 ms 量级的时延,难以满足实时控制,特别是紧急控制的要求。

RTU 和 SCADA 系统可以用低成本采集大量的广域数据,并满足静态可观性,但没有统一时标,采样周期长、时延大,只能抽取慢速动态特征。故障录波和保护信息系统就地记录大量暂态信息,可支持故障的事后分析,但上传的滞后时间长,难以实时应用。

PMU 数据不依赖于系统模型和参数,可在统一时标下快速采样,提供系统实际的时间响应曲线,用来校核仿真模型和参数,并提供仿真的初值。但受价格限制较难满足高可观性。

20 年来的研究热点之一是将 PMU 用于实时稳定性分析及开环控制,根据受扰后短期实测数据预估稳定性,并在失稳前选择并执行控制措施,但至今未能用于稳定分析和开环控制。文献[32]明确指出,在非自治因素下单纯依靠轨迹的外推来预测并不可靠,并且即使在没有时间约束的离线环境下,要对有数学模型的仿真曲线进行稳定性量化分析也非易事。若在计算已有轨迹的稳定裕度时没有模型参数可用,其量化分析就更难。更重要的是,只有通过仿真才能在措施实施前确认其效果。不与仿真工具结合,就难以按合理的保守性来确保控制的充分性。其根本原因是:在事前,PMU 无法观察到控制措施的效果,也无法在不依靠系统模型及参数的前提下求取实测轨迹(特别是稳定轨迹)的稳定裕度。PMU 在应用上的这个瓶颈是本质性的,不可能真正突破。

现场伪量测是对广域数据加工中挖掘出来的数据;仿真伪量测是在仿真中产生的数据。它们的正确性受到模型、参数、初值和算法等的限制。

3.2 电力(能源)系统的大数据平台

随着数据采集环节的质量提高、成本降低,以及各种广域数据采集系统的数据共享,可以通过更多数据来全面掌控电力与能源。为此,需要将各种广域监测系统和数值仿真系统置于统一的平台中,有助于数据挖掘,识别相继故障的先兆特征、临界条件和传播模式,从而提高预警水平,增强防御停电灾难和事后分析的能力。

由硬件资源、基础软件、网络通信、数据集成、计算支撑、应用支撑、安全管控等环节构成的大数据平台支持多源多渠道异构数据的即插即用、融合与管理,支撑各种开发、应用及信息展示功能的即插即用。通过基于数学模型的因果型数据、无因果关系的统计型数据以及参与者博弈行为的融合,管理各类能量的生产、转换、输送及使用,调度与仿真数据的集成与管理。

4 结语

虽然当前关于大数据的应用案例大多发生在互联网企业中,但传统的电力及能源企业也在思考如何从关于大数据的空泛介绍走向实际应用。特别是除了直接依赖互联网的电力金融业务及面向消费的个性化服务以外,在基于传统数据的系统分析与控制领域中,如何融合电力及能源的统计关系数据、因果关系数据及博弈行为数据,发挥大数据的价值。例如:间歇性能源及负荷预测,引导需求响应及节能减排,降低停电风险,反窃电,堵塞经营漏洞,优化资产全寿命周期管理等方面。特别是:如何使企业决策从当前基于常规数据及主观经验的模式,发展为基于数学模型、参与者及多代理模型的混合仿真的沙盘推演模式。其中的多代理模型就需要大数据技术的支撑。这关系到电力大数据技术是否能进入到通常由因果关系数据一统天下的物理系统分析领域。为此,思维方式需要重大变革。

参考文献

- [1] TOLLE K M, TANSLEY D S W, HEY A J G. The fourth paradigm: data-intensive scientific discovery[J]. Proceedings of the IEEE, 2011, 99(8): 1334-1337.
- [2] Nature. Big data[EB/OL]. [2008-09-03]. <http://www.nature.com/news/specials/bigdata/index.html>.
- [3] BRYANT R E, KATZ R H, LAZOWSKA E D. Big-data computing: creating revolutionary breakthroughs in commerce,

- science, and society[R]. 2008.
- [4] Science. Special online collection: dealing with data[EB/OL]. [2011-02-11]. <http://www.sciencemag.org/site/special/data>.
- [5] MANYIKA J, CHUI M, BROWN B, et al. Big data: the next frontier for innovation, competition, and productivity[R]. 2011.
- [6] United States Department of Energy Office of Electric Transmission and Distribution. The smart grid: an introduction [EB/OL]. [2015-11-03]. [http://www.oe.energy.gov/DocumentsandMedia/DOE_SG_Book_Single_Pages\(1\).pdf](http://www.oe.energy.gov/DocumentsandMedia/DOE_SG_Book_Single_Pages(1).pdf).
- [7] MELL P, GRANCE T. The NTST definition of cloud computing[R]. 2011.
- [8] 姚建国,赖业宁.智能电网的本质动因和技术需求[J].电力系统自动化,2010,34(2):1-5.
YAO Jianguo, LAI Yening. The essential cause and technology requirements of smart grid construction [J]. Automation of Electric Power Systems, 2010, 34(2): 1-5.
- [9] 张东霞,姚良忠,马文媛.中外智能电网发展战略[J].中国电机工程学报,2013,33(31):1-14.
ZHANG Dongxia, YAO Liangzhong, MA Wenyuan. Development strategies of smart grid in China and abroad [J]. Proceedings of the CSEE, 2013, 33(31): 1-14.
- [10] 余贻鑫,刘艳丽.智能电网的挑战性问题[J].电力系统自动化,2015,39(2):1-5.DOI:10.7500/AEPS20141204007.
YU Yixin, LIU Yanli. Challenging issues of smart grid [J]. Automation of Electric Power Systems, 2015, 39(2): 1-5. DOI: 10.7500/AEPS20141204007.
- [11] 宋亚奇,周国亮,朱永利.智能电网大数据技术现状与挑战[J].电网技术,2013,37(4):927-935.
SONG Yaqi, ZHOU Guoliang, ZHU Yongli. Present status challenges of big data processing in smart grid [J]. Power System Technology, 2013, 37(4): 927-935.
- [12] 王德文,宋亚奇,朱永利.基于云计算的智能电网信息平台[J].电力系统自动化,2010,34(22):7-12.
WANG Dewen, SONG Yaqi, ZHU Yongli. Information platform of smart grid based on cloud computing [J]. Automation of Electric Power Systems, 2010, 34(22): 7-12.
- [13] WANG P J. Dynamic data center operation with demand-responsive electricity price in smart grid [J]. IEEE Trans on Smart Grid, 2012, 3(4): 1743-1754.
- [14] 刘振亚.中国电力与能源[M].北京:中国电力出版社,2012.
- [15] XUE Y S, XIAO S J. Generalized congestion of power systems: insights from the massive blackouts in India [J]. Journal of Modern Power Systems and Clean Energy, 2013, 1(2): 91-100.
- [16] KATAL A, WAZID M, GOUDAR R H. Big data: issues, challenges, tools and good practices [C]//2013 Sixth International Conference on Contemporary Computing (IC3), August 8-10, 2013, Noida, India: 404-409.
- [17] 孟小峰,慈祥.大数据管理:概念、技术与挑战[J].计算机研究与发展,2013,50(1):146-169.
MENG Xiaofeng, CI Xiang. Big data management: concepts, techniques and challenges [J]. Journal of Computer Research and Development, 2013, 50(1): 146-169.
- [18] 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. The scientific value of big data [J]. Research Communications of the CCF, 2012, 8(9): 8-15.
- [19] XUE Y S, WU J A, XIE D L, et al. Multi-agents modelling of EV purchase willingness based on questionnaires [J]. Journal of Modern Power Systems and Clean Energy, 2015, 3(2): 149-159.
- [20] 王珊,王会举,覃雄派,等.架构大数据:挑战、现状与展望[J].计算机学报,2011,34(10):1741-1752.
WANG Shan, WANG Huiju, QIN Xiongpai, et al. Architecting big data: challenges, studies and forecasts [J]. Chinese Journal of Computers, 2011, 34(10): 1741-1752.
- [21] 金培权,郝行军,岳丽华.面向新型存储的大数据存储架构与核心算法综述[J].计算机工程与科学,2013,35(10):12-24.
JIN Peiquan, HAO Xingjun, YUE Lihua. A survey on storage architectures and core algorithms for big data management on new storage [J]. Computer Engineering & Science, 2013, 35(10): 12-24.
- [22] WU X D, ZHU X Q, WU G Q, et al. Data mining with big data [J]. IEEE Trans on Knowledge and Data Engineering, 2014, 26(1): 97-107.
- [23] 李建中,刘显敏.大数据的一个重要方面:数据可用性[J].计算机研究与发展,2015,50(6):1147-1162.
LI Jianzhong, LIU Xianmin. An important aspect of big data: data usability [J]. Journal of Computer Research and Development, 2015, 50(6): 1147-1162.
- [24] 冯登国,张敏,李昊.大数据安全与隐私保护[J].计算机学报,2014,37(1):246-258.
FENG Dengguo, ZAHNG Min, LI Hao. Big data security and privacy protection [J]. Chinese Journal of Computers, 2014, 37(1): 246-258.
- [25] CHEN M, MAO S, LIU Y. Big data: a survey [J]. Mobile Networks and Applications, 2014, 19(2): 171-209.
- [26] 董朝阳,赵俊华,文福拴,等.从智能电网到能源互联网:基本概念与研究框架[J].电力系统自动化,2014,38(15):1-11.DOI:10.7500/AEPS20140613007.
DONG Zhaoyang, ZHAO Junhua, WEN Fushuan, et al. From smart grid to energy internet: basic concept and research framework [J]. Automation of Electric Power Systems, 2014, 38(15): 1-11. DOI: 10.7500/AEPS20140613007.
- [27] Versant, NoSQL and the smart grid big data challenge [EB/OL]. [2012-08-13]. <http://www.greentech-media.com/articles/read/versant-nosql-and-the-smart-grid-big-data-challenge>.
- [28] 薛禹胜.时空协调的大停电防御框架:(二)广域信息、在线量化分析和自适应优化控制[J].电力系统自动化,2006,30(2):1-10.
XUE Yusheng. Space-time cooperative framework for defending blackouts: Part II reliable information, quantitative analyses and adaptive controls [J]. Automation of Electric Power Systems, 2006, 30(2): 1-10.
- [29] 薛禹胜,费圣英,卜凡强.极端外部灾害中的停电防御系统构

- 思:(二)任务与展望[J].电力系统自动化,2008,32(10):1-5.
- XUE Yusheng, FEI Shengying, BU Fanqiang. Upgrading the blackout defense scheme against extreme disasters: Part II tasks and prospects [J]. Automation of Electric Power Systems, 2008, 32(10): 1-5.
- [30] 赵俊华,文福拴,薛禹胜,等.电力信息物理融合系统的建模分析与控制研究框架[J].电力系统自动化,2011,35(16):1-8.
- ZHAO Junhua, WEN Fushuan, XUE Yusheng, et al. Modeling analysis and control research framework of cyber physical power systems[J]. Automation of Electric Power Systems, 2011, 35(6): 1-8.
- [31] 李强,周京阳,于尔铿,等.基于混合量测的电力系统状态估计混合算法[J].电力系统自动化,2005,29(19):31-35.
- LI Qiang, ZHOU Jingyang, YU Erkeng, et al. A hybrid algorithm of power system state estimation based on PMU measurement and SCADA measurement[J]. Automation of Electric Power Systems, 2005, 29(19): 31-35.
- [32] 薛禹胜,徐伟,DONG Zhaoyang,等.关于广域测量系统及广域控制保护系统的评述[J].电力系统自动化,2007,31(15):1-5.
- XUE Yusheng, XU Wei, DONG Zhaoyang, et al. A review of wide area measurement system and wide area control system [J]. Automation of Electric Power Systems, 2007, 31(15): 1-5.
- 薛禹胜(1941—),男,博士,中国工程院院士,国网电力科学研究院名誉院长,博士生导师,主要研究方向:电力系统自动化。E-mail: xueyusheng@sgepri.sgcc.com.cn
- 赖业宁(1975—),男,通信作者,博士,高级工程师,主要研究方向:智能电网、电网调度自动化及新能源控制。E-mail: laiyening@sgepri.sgcc.com.cn

(编辑 章黎)

Integration of Macro Energy Thinking and Big Data Thinking

Part One Big Data and Power Big Data

XUE Yusheng, LAI Yening

(NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing 211106, China)

Abstract: The macro energy thinking which regards electricity as a hub between energy production and consumption, can promote “clean energy substitution” of the upstream primary energy and “electricity substitution” of the downstream end-use energy, in order to support the sustainable development of energy. Meanwhile the big data thinking regards various data resources as fundamental elements of production rather than simple process objects. The integration of these thinking will make the big data on power become the foundation of an extensively interconnected, openly interactive and highly intelligent macro energy system. Key elements of this integration include the acquirement, transmission and storage of wide-area power data with different timescales, the data from related domains, as well as the fast and in-depth knowledge extraction from the multi-source heterogeneous data and its applications. As the first part of a series of paper, this paper summarizes unique features of big data based on the deduction of the basic concept, data structures and essential characteristics of big data. For the comprehensive energy network, a knowledge extraction platform is constructed by integrating the causal data (based on mathematical models), the statistic data (with non-causal relationship) and the gambling data (of human participants). More case studies will be proposed in the subsequent paper, which will show the contributions of big data thinking to enhance the economy and reliability of macro energy systems.

This work is supported by National Natural Science Foundation of China (No. 61533010), NSFC-EPSRC Collaborative Project (No. 513111025-2013), China-Thailand Cooperation Fund Project (No. 5151101161), and State Grid Corporation of China.

Key words: cyber energy system; energy interconnection; causal data; statistic data; gambling data; unstructured data; knowledge extraction