

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, 包括发电、输电、变电、配电、用电及调度等环节的实时系统, 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能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御能力,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]。但若要有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 万个智能电表上收集超过 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] 张东鑫,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, ZHANG 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

大能源思维与大数据思维的融合

(二)应用及探索

薛禹胜^{1,2}, 赖业宁¹

(1. 南瑞集团公司(国网电力科学研究院), 江苏省南京市 211106;

2. 智能电网保护与运行控制国家重点实验室, 江苏省南京市 211106)

摘要: 前一篇论文诠释了大数据及电力大数据思维。一方面将能源系统作为信息物理系统(CPS)概念中的物理系统,以打通电力系统与一次能源系统及终端能源系统之间的物理藩篱;另一方面,将专用网及互联网共同组成通信环节,打通调度业务与经营业务之间的信息壁垒。通过因果关系型、统计关系型及博弈等行为关系型的数据及相应的大数据技术,物理系统与信息系统融合为信息物理能源系统(CPES)。作为两篇论文的续篇按此观点探索电力(能源)大数据的应用,并通过若干课题的研究,归纳大数据技术对提高能源流在不同时间尺度及空间中的经济性与可靠性的作用与途径。包括:①在原本完全依靠统计分析的过程中加入因果分析手段,以提高前者的适用性与精度;②在原本完全依靠数学模型分析的过程中加入统计分析手段,以提高前者的效率;③综合应用模型分析、相关分析、行为分析与实验经济学仿真分析等手段,开创那些无法单独采用一种分析手段有效探索的研究领域,例如大能源系统与自然环境、社会环境、市场经济、政策等非能源环节的交互影响。

关键词: 一次能源;终端能源;大数据技术;智能电网;信息物理系统(CPS);决策支持;应用研究

0 引言

电力系统是人类创造的最复杂的系统之一,也是社会发展的重要基础设施。为保障电力系统的绿色、高效、安全与充裕,各国都在大力建设覆盖发电、输电、变电、配电、用电及调度等环节的智能电网,即信息物理电力系统(cyber-physical power system, CPES),它是针对电力系统的信息物理系统(cyber-physical system, CPS)。其基础是电力系统广域全景实时数据的采集、传输、存储、分析及决策支持。

2011年日本核泄漏灾难以惨痛的代价昭示了电力(能源)系统与广义的环境系统之间的紧密关系^[1]。9.0级地震造成福岛核电站停堆,并失去外部电源,继发的海啸淹没了备用柴油发电机;由于移动电源没能正常供电,直升飞机又无处可降,故蓄电池耗尽后,冷却系统长期停止工作,造成核燃料棒熔

融,氢气爆炸,安全壳破损,导致核泄漏不断升级。这一连串小概率事件最终酿成了人类历史上第3次重大核事故。

2012年印度大停电再一次表明:一次能源、能源体制、环保约束、监管规则、规划设计、电力基础设施、调度运行、信息技术、市场运营,以及各参与方的博弈等因素都可能给电力可靠性埋下隐患并在停电灾难中推波助澜^[2]。

2015年,乌克兰电力监控系统遭到恶意代码攻击,导致23个35 kV变电站、8万用户停电长达数小时。黑客以邮件发送恶意代码控制数据采集与监控(SCADA)节点,下达断电指令并阻止恢复,攻击成本低,但后果严重,凸显了网络安全对能源、社会安全的重要性^[3]。

大能源思维就是为了确保电力、能源及环境的安全,而在协调的框架中研究整个能源流、信息流及相关资金流的可靠性,以及排放流与市场博弈的可控性。它将物理领域的范畴从电力系统扩展到能源系统及相关的非能源系统,推动一次能源的清洁替代与终端能源的电能替代,以确保可持续发展道路上的能源安全、市场安全与环境安全^[4],从而将

收稿日期:2016-03-11。

上网日期:2016-03-16。

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。

CPPS的概念拓展到信息物理能源系统(cyber-physical energy system, CPES)。

这就必然对数据采集的范畴与数据分析的技术提出新的挑战。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,从大量多源、多态、异构数据中快速提炼深层知识并发挥其实用价值^[5],成为复杂大系统(包括大能源系统)经济性与可靠性的重要保障技术。

大能源与大数据,这两种思维的融合提升了信息系统对能源系统的清洁智能及互联互通的支撑力度,包括对跨行业的广域多时间尺度数据的采集、传输、存储、交互、分析、展示和应用。

但是,当前大数据的成功案例基本上局限于由数据驱动的统计分析领域,例如网络商业、网络金融与面向消费的个性化服务,而在如何进入通常由因果关系数据一统天下的物理系统分析领域,进入传统上基于模型与常规数据的控制领域方面,则乏善可陈。大数据技术的另一使命,即在通常限于统计数据及经验决策的社会及企业管理问题中,引入数学模型及数值仿真来支持沙盘推演优化决策,则基本未能顾及。

为了从大数据中快速获得对具体物理系统有价值的知识,一方面要从数学与计算机技术的角度,研究大量群体的多维度数据及其融合;另一方面,必须紧密结合具体的物理系统及过程,甚至引入人的行为特征因素进行综合分析,才能有效挖掘知识。几乎所有关于大数据的论文都会强调不同类型数据之间的融合,但却鲜有认真讨论如何融合者。

在文献^[6]的基础上,本文探讨在原本基于统计分析的课题中融入数学模型及因果分析的可能性,以及在基于数学模型的分析与控制领域中融入统计型数据及博弈型数据的有效性。前者的例子为:大扰动受扰轨迹的解读;对风电等随机时间序列的超短期预测;充裕性从经验调度到基于数学模型的优化调度。后者的例子包括:在可信度控制下的相继开断潮流快速计算;机理判据与实证阈值相结合来加快受扰轨迹稳定裕度的求取;振荡模式易变性概念的提出;停电防御系统对灾害天气的自适应能力。此外,还介绍了博弈型数据的多代理建模与仿真,包括电动汽车(EV)的购买、出行、参与电网备用服务等行为实证的多代理建模及混合仿真;政策对需求响应及碳排放的影响。

本文分析的案例均为笔者亲历,虽有利于归纳,但也必然存在较大的局限性。还要说明的是,文中从大数据理念出发的解读,并非都是当时就有的认识。只是那时对知识挖掘的理解及对算法改进的努

力,在一定程度上符合了大数据思维。希望本文的思考与归纳有助于今后对先进方法论的自觉应用。

1 大能源系统研究中的大数据范式

实验曾是唯一的科学研究范式,然后出现基于定律和定理的理论研究,再后发展的计算机技术催生了计算科学。这3种研究范式针对的是因果关系型的数据。被认为是第4种范式的数据驱动型范式,或称大数据范式,以统计关系型数据为主要研究对象,以协同化、网络化与数据驱动为主要特征,包括知识的发现、标识与评估。

大数据是指具有复杂结构,且其数据量对于待求解的问题来说足够大到具备统计意义的数据集,包括那些不具有或尚未掌握其因果关系的大量无模型数据。但是,大数据绝不应该排斥因果关系型数据及博弈关系型数据。

大数据研究范式强调从大数据中发现关联关系。若仅仅凭借统计关系型数据,就只能回答事物间的相关关系“是什么”,但无法回答“为什么”的问题。此时不可能采用严格推导的分析思路与精确算法,而适用粗略的统计算法。但是,大数据研究范式也绝不应该排斥因果分析技术。

笔者认为不应该孤立看待上述不同的研究范式。大数据技术不能局限于数据驱动,而应该覆盖所有类型的数据,涉及异类数据间的协同,精确算法与统计算法的融合,以及不同研究模式的协调。事实上,复杂现象除了反映在统计关系上,也会反映在因果关系上,3维的洛伦兹系统就是生动的例子^[7]。由于基于因果关系的分析兼蓄了数学推导的严格性和物理概念的简明性,因此在揭示问题的内在关系和演化规律方面起着不可替代的作用。一方面,局部的因果关系往往有助于研究全局的统计关系;另一方面,在统计关系型数据中尽量挖掘出其中可能存在的(准)因果关系的数学模型,本身就是大数据技术的任务之一。

因此,广义的大数据范式应该包含实验、理论研究、计算科学、数据驱动等范式,而由数学模型、多代理、实验经济学仿真参与者动态交互驱动的混合仿真,就是数据驱动与实验、理论、计算等范式的结合。为此,需要探讨如何利用建模技术、仿真技术、算法及指标,来推动对统计关系型数据的研究。例如:因果关系中的不确定因素需要用统计关系来表达;统计关系中挖掘出来的机理则也可能表达为传递函数之类的因果关系^[8]。

对于电力系统,其复杂性不但源自其高维时变的非线性,还反映在多领域(物理、经济、资源、环境、

信息)、不同稳定性(静态、动态、暂态、周期、结构)、多时间尺度(电磁暂态、机电暂态、中期、长期)、多物理量(电流、电压、频率、功角)及多空间尺度(局部模式、全局模式)等方面^[9-10]。

美国能源部对北美“8·14”大停电的调查报告指出:由于电力公司和电网监测中心未及时做出准确的分析和判断,没有在相继事故的间隔时段内及时采取有效措施,因此系统在多重相继开断下变得脆弱,最后发展为大停电^[11],这涉及信息监测、在线分析和控制决策等3个任务。不论是预防控制、继电保护、紧急控制、校正控制,还是恢复控制,都需要自适应能力,即不断跟踪电网拓扑和工况的变化,及时更新最优决策。为此,需要将目前相互孤立的各种广域数据采集监控系统集成为统一、开放的广域信息平台。实时采集的静态/动态信息可以及时反映已发生的现象,但若要在推演场景下优化控制决策,就必须依靠仿真分析。

物理框架一旦从电力系统拓展到大能源系统,问题的复杂性必将成倍增加。为了支持更全面的分析、更准确的预测及更具价值的辅助决策,需要构筑多领域跨时空的数据价值链,建立完善的市场监测、预测和预警系统,管控能源金融风险及信息安全风险。

2 大能源系统对信息系统的要求

大能源系统的大数据分析对数据的采集、存储、集成、挖掘、管理、展现及安全的要求包括:①突破时空限制,快速处理跨物理空间和网络空间的各类数据资源;②打通各领域系统的业务流、资金流和信息流之间的藩篱,从价值的传递提升为价值的创造;③构建大能源数据中心及其运行管理平台,形成多业务多维度协同分析的数据价值链,支撑大能源的安全及高效利用;④确保网络安全和信息可靠。

在数据架构方面,需要突破传统信息技术针对特定应用的局限性。一方面要贯通各种能源系统内部各环节间的数据,推进多业务间的高效协同;另一方面要打通电力系统与一次能源、用户需求、自然环境、经济社会、管理政策之间的信息壁垒,为能源安全及环境安全提供信息支持。实现开放、融合、可扩展的分布式一体化数据服务体系。

在数据获取与集成方面,需要融合多源数据,大幅提高数据采集的准确性、时效性和可靠性。构建横向协同、纵向贯通的数据平台,支撑地理信息系统(GIS)、电力系统分析、运维检修、电力市场(包括发电侧、需求侧、输电权、辅助服务、碳资产)、电力营销、移动应用、云环境下的业务应用及集成服务模式,

提升数据的完整性、及时性、有效性和一致性^[12]。

在数据存储方面,需要有效、快速、可靠地存取甚至达到海量的数据,包括大容量、分布式集群存储设备,及数据归仓、云存储、虚拟存储等硬软件的研究,协同提高计算能力。

在数据管理方面,在对各类数据,特别对非结构化数据的组织、清洗、融合、集成、抽取、转换、装载、检索、维护时,减少数据冗余,提高数据共享、服务及管理效率。既要充分体现其内在联系,便于数据修改与扩充,又能保证数据的独立性、可靠性、安全性与完整性。

在数据分析与决策支持方面,需要深入探索大数据的挖掘技术,重视信息的深度加工和有效表达。文献^[13]列举的大数据分析技术有:遗传算法、神经网络、数据挖掘、回归分析、分类分析、聚类分析、关联规则学习、数据融合与集成、机器学习、自然语言处理、情感分析、空间分析、时间序列分析等。

在信息安全方面,需要将信息安全从被动防御向多层次、主动防御转变。包括应用可信计算、并行加密、标识认证、智能分析等新技术,在支撑多类型终端开放互动能力的同时,保障信息的机密性、完整性、可用性和可控性,特别是云环境及移动互联环境下的安全性。要可靠防御黑客攻击及各种灾害对信息系统的破坏,正确处理大数据的开放与安全的矛盾。

在数据展现方面,需要把握数据的内在特征和规律,综合运用计算机图形学、图像处理、计算机辅助设计、历史流、空间信息流、计算机视觉及人机交互等领域技术,灵活地为不同的物理系统提供直观互动的可视化展现与虚拟现实,把复杂的数据转化为可以交互的图形,提升电力数据的可利用价值。

3 大能源系统中的大数据应用前景

大数据的研究从数据的挖掘、分析、处理到应用,其中应用是整个大数据的出发点和归宿,应该遵循需求牵引、问题导向、案例推理的原则。只有把大数据转化成思想及知识,才能带动大数据技术的产业。

3.1 智能电网

能源革命推进了智能电网及各种新能源及用电技术的关联;大规模风光电的接入大大增加了电力生产的不确定性 & 电网运行的困难;大规模 EV 的充放电又增加了电力消费的随机性;发、输、配、用、储,各环节内的不确定因素及其交互影响越来越复杂。大数据技术成为提高电力流效率及防御大停电

灾难的基础。

2013年美国电力科学研究院(EPRI)启动了输电网及配电网的现代化示范项目;国际电工委员会(IEC)定义的智慧能源系统是以智能电网为核心,以先进的信息技术为支撑,实现电网与其他能源系统交互,推动大规模新能源接入^[14]。

中国电机工程学会2013年发布了中国电力大数据发展白皮书^[15]。国家电网公司也启动了设备状态预警、营销、用电、客户服务等方面的大数据示范应用研究,融合传感量测、信息通信、计算机、数据分析与领域技术。旨在:①通过各种渠道获得覆盖电力生产、输送、消费、交易、社会服务的电力大数据;②融合因果分析、统计分析、行为分析、智能分析等手段来挖掘知识,实现智能预测、精准调度、辅助服务、需求侧管理、个性化服务、状态检修、灾害预警、故障诊断、突发事件应对、行为分析、能效管理、投资管理、电力金融平台、市场交易;③提升对基础设施规划、设计、建设、运行、维护、营销、交易、监控、管理、资本运作的科学决策水平,持续推进能源体制创新。

3.2 大能源系统

以电为中心、以清洁化与智能化为特征的能源革命正在推动整个能源的生产、消费和管理模式的重大变化。促进各类数据资源整合,覆盖电力系统、供热(冷)系统、燃气系统、燃油系统,以及气象、经济、交通等非能源系统,而在各能源子系统内则覆盖了能源的生产、转换、传输、交易、存储、消费各环节。建设基于互联网的智慧用能的量测与交易平台,基于智能楼宇与智能工厂的能源综合服务中心,实现多种能源的智能定制;鼓励个人、家庭用户与分布式发电、储电、储热、储冷、储氢等多类型的分布式储能资源之间通过微平衡市场进行局部自主交易,推动紧急备用、调峰调频等增值服务。

3.3 环境安全

人类活动产生的排放若超过生态系统的自修复能力,就会破坏其平衡。反过来,后者将以频发的极端气候来增加停电风险等方式以惩罚人类,甚至引起气候系统崩溃而毁灭性地打击人类。因此,必须重视生态工程,并通过排放权市场及其监管来管理风险。为此需要获取环境信息,包括污染程度、类型、交叉程度、分布区域、危害、波及人数等数据,并在大数据支持下综合分析传统风险与新型风险。

虽然难以严格确定自然界的可恢复域,但人们按大数据及经验给出了温室气体排放的安全阈值。各国正在通过协商各自的排放额度,以保证全球气温上升幅度不超过 2°C ,并在实施气候工程以提高

自然界修复能力的同时,监控各行各业,特别是能源流对应的排放。但要使更多民众积极参与到能源的清洁替代与电能替代措施中,则除了完善监控体制外,还必须制定适度的补贴及惩罚政策来引导人们参与能源流的优化行为。政策不到位就难以实现减排目的,但过激的政策又会使减排成本过高。两者都可能颠覆气候工程。

碳排放的动态过程受到众多参与者,包括各种化石类及可再生能源的生产商、发电商、终端能源商、监管者、政策制定者博弈行为的影响。他们的行为可能为有限理性,也可能为非理性。分析涉及各种一次能源的采集、输送和储藏,物理系统,经济系统,环境系统,及监管政策机制等领域的统计关系数据、因果关系数据及包括博弈数据在内的各种社会行为数据,使决策从当前基于主观经验的模式,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。

3.4 市场经济

电力生产、电网运行与其营销、管理、资产全寿命管理等任务不但影响到能源经济、政策制定和经济决策,也受到经济社会状况及政策等外部环节的影响。大能源概念下出现的虚拟电厂、负荷集成商、第三方增值服务供应商等新兴市场主体,大大增加了灵活性资源供应。相应地,需要完善市场机制,兼容用户以直接、间接等多种方式自主参与灵活性资源市场交易的渠道;建立合理的灵活性资源补偿定价机制,保障灵活性资源投资拥有合理的收益回报。但是,当前的信息系统基本上局限于电力生产与调度的业务层,故难以有效地促进EV与智能电网间能量和信息的双向互动,推动智能充放电业务,鼓励用户参与电力需求响应,自主提供能量响应、调频、调峰等灵活的能源服务,实现EV与新能源的协同优化运行,以互联网平台为依托实现实时交易。

需要学习沃尔玛利用大数据思维从“啤酒与尿布”商业案例中受益的经验,挖掘市场参与者的行为特征,预测市场风险,支撑更灵活的交易,也促使各管理者和参与者转向采用量化的推演范式及决策机制。大数据思维将经典的市场理论与创新的数据驱动方法结合,在分析与决策中引入信息论和行为学,使宏观决策与微观决策之间的互动更加迅速高效。

基于客户服务系统、用户信息采集系统、营销业务系统等营销数据,评估客户信用,分析电能替代潜力、用电行为,预测售电量,评估市场风险,制定竞争策略,挖掘在电子商务、节能服务等领域的商业价值。

3.5 信息系统

在信息系统本身的建设与完善中,也应该充分发挥大数据技术的作用。包括加强信息安全体系及安全技术,大数据共享与隐私的协调,开放与防御的平衡,数据被窃或篡改的防御,个人信息开放和保护平衡,高级可持续攻击的应对。

在分析与预警方面,由于发、输、变、配、用电需要保持瞬时平衡,在稳定性分析、可再生能源的预测及备用容量的调度、大停电风险分析与综合防御等方面均需突破。大数据算法的思路强调全部数据而不是个别子集,更加关注效率与相关性,关注混沌规律(无规律可寻)和浮动规律(有迹可寻,但要求数据足够多),而不是恒常规律(通常基于少量数据即可发现),从价值密度低、异源异构异质的大数据中有效提取价值并增值。此外,还需要建立参与者博弈行为的多代理模型;建立可同时支撑数学模型、多代理及真实参与者的混合仿真平台;克服知识挖掘中的瓶颈。利用统计分析技术提高模型仿真的速度,或利用(准)模型分析技术提高统计分析的质量^[16-20]。

在决策与控制方面,需要提高监测预警、状态评估、故障分析、检修决策、风险管控、资源利用、信息安全等水平。例如,在非自治和非线性因素下,如何摆脱数学模型而仅根据实测轨迹来量化系统动态。需要在算法与受扰轨迹的知识提取上取得重大突破,才能真正发挥相量测量单元(PMU)在广域保护中的作用。

3.6 管理

重要的管理决策应该从当前基于主观意愿的经验模式,转为基于数学模型、实证多代理模型及真实参与者的混合仿真,通过沙盘推演来优化决策的科学模式。

4 利用因果分析的手段改进统计分析的适用性

对传统上依靠统计分析的经验判据来说,并非都是因为客观上不存在因果关系。至少下述案例说明:随着人们探索自然规律的不断深入,这些客观关系有可能用数学模型来表达。大数据技术不仅是实现该目标的重要手段,也将是受益的技术领域之一。

4.1 将同步稳定性的分析从统计关系的定性判断提升为因果关系的量化指标

电力系统状态变量与时变参数的维数、多时间尺度及实时性要求等使其稳定分析和控制极为困难。当前,求取超大电网的受扰轨迹并无本质困难,但在得到由仿真给出的关于受扰轨迹的大量数据

后,却只能凭经验来定性地判断系统是否稳定。为了避免将稳定轨迹误判为失稳轨迹,必须将该限值取得很大;而为了不漏判失稳轨迹,又必须将积分区间取得很长,从而进一步增加了计算量。有关的阈值是根据统计数据与经验而设定的,既无法估计稳定的充要条件,又无法比较参数对稳定性的影响程度;既难以探索稳定机理,又难以优化决策。

大扰动稳定性方面唯一成功的定量理论是为模型最简单的定常单机系统建立的等面积解析准则。它虽然不能计入耗散项和控制器等复杂模型,也不能分析多摆稳定性或多机系统稳定性,但该准则基于解析表达式来处理因果关系,具有清晰的物理概念、简明的解析性和直观的量化能力。

扩展等面积准则(EEAC)理论^[21],先用数值积分法求取基于全模型的受扰全过程轨迹,再将观察空间从积分空间中分离出来,通过因果关系来识别主导模式,并定量分析受扰轨迹的稳定裕度和参数稳定裕度。为此,将高维轨迹严格地通过线性保稳降维映射,得到多个正交的单机轨迹,并将只适用于定常哈密顿单机系统的等面积解析准则拓广到复杂模型的单机系统。

EEAC证明了:多机系统分岔的充要条件是至少有一对互补群的群间相对动能超过了相应的势垒;多机系统的临界模式和稳定极限则取决于最临界的映象。由此,在原本完全依靠经验与统计分析来解读受扰轨迹稳定性的过程中,加入了受扰轨迹集(大数据)与系统稳定裕度(单个标量数据)之间的因果关系分析,从而将经验的解读方式提升为严格的机理分析。不妨认为,EEAC是在处理统计数据中发现了客观上存在,但一直却鲜为人知的因果关系。笔者坚信,随着大数据研究的深入开展,类似的成果将大量涌现。

4.2 在风电的超短期预测中加入因果关系以获取自适应能力

风电预测技术并不会减少风电随机性的影响,但可降低其模糊性对电力系统的扰动。传统的时序预测方法按唯一的统计公式来外推时间序列,而未考虑其他众多相关因素对外推公式适应性的影响。其预测模型是以历史数据拟合的整体误差最小为准则而建立,既然是对所有样本集的折中结果,就不可能顾及实际预报时段内的具体特征。

文献^[22]提出了“离线分类优化,在线选配模型”的自适应预测方法,按照相关因素及风电动态特征,建立有限的因果关系,并据此选用不同的样本集,分别优化预测模型及参数。这使风电等随机时间序列的超短期预测从历史数据的笼统统计改进为

与因果关系结合的分类统计,不但提高了预测精度,而且提供了对预测值的可信度评估信息。

“离线优化”是从历史风电功率时间序列中提取特征量并选择相关阈值,据之将风电功率时间序列划分为不同形态的子集,外加一个囊括所有不具有排他性分类特征的“非形态子集”。分别按对应的训练样本优化各子集专用的预测模型及参数。“在线匹配”是根据前一短时段内的形态识别结果,调用相应的模型预测。

时间序列预测法根据目标变量本身随时间变化的趋势外推,但输入数据单一、强壮性差。空间相关性是时序预测的有效补充,特别有利于应对风速骤变的不确定性。但有关文献仅考虑临近时间窗口的空间相关性,未将数据样本与相关特征勾连。上述时间序列自适应预测的思路也可用来处理空间相关性,以计入日历时辰、地貌、气象和距离等因素的影响。离线分类建模以参考风电场和目标风电场的全部历史数据的时间序列作为输入,抽取空间相关特征,分类并离线建立对应的预测模型;在线模型选取环节根据参考风电场的预测数据和气象等特征信息,选取合适的模型和参数预测,再结合本地的时序预测,与空间预测相互校验和修正。若能同时考虑本地风速的时序相关性,多地风速的空间相关性,以及数值天气预报的大尺度结果,则可实现更有效的风电时空预测。

4.3 充裕性调度从经验方式提升为协调优化方式

充裕性是系统在不同的时间尺度及潜在事件下,满足用户对功率、电量及电能质量需求的能力。传统上将充裕性视为静态问题,忽略冷备用升为热备用所需的时间延迟,不考虑博弈与环境等非电因素,从而粗放地按经验取负荷总量的某个固定比例来配置备用容量。但是,大规模可再生能源与充放电负荷的入网,需求驱动与博弈行为的大量增加,使充裕性问题甚至比稳定性问题更加复杂。备用容量及备用电量的调度方式必须加入更多的因果关系型数据及数学规划方法。

充裕性控制的内容包括合约决策及执行决策两个步骤,前者通过市场手段获取备用资源的控制权,后者在系统运行需要时根据各种备用资源的技术特性优化实施控制权。合约决策是运营部门与发电商或用户在控制权交易上的博弈;执行决策是调度部门按风险动态配置已获控制权的备用资源。决策过程还包括待调控出力的预报、充裕性事件的设定、系统功率平衡与备用功率的动态行为、充裕度量化指标的在线评估、备用资源的调度与控制^[23]。

为此,需要采集涉及物理、经济、环境、行为方式

等领域的大数据,充裕性决策框架基于数学模型、少量实验参与者与大量计算机代理的动态交互仿真。将获取的动态响应曲线作为合约市场的输出,提供给决策支持环节在线优化调度,包括冷备用与热备用等不同等级备用之间的动态升降,同时还要兼顾系统对安全性和电能质量的要求。为此,需要将充裕控制问题按事件序列、相继事件、非预防性控制决策、预防性调度决策逐步解耦;基于防控代价及剩余风险之和最小的原则,从下到上逐层扩大优化范围。

5 利用统计分析的手段提高因果分析的效率

即使在传统上基于数学模型,且已相当成熟的因果算法中,也不乏利用中间结果的统计分析来提升算法效率的空间。

5.1 可信度控制下的相继开断潮流快速计算

若要精确而强壮地反映电网普适的物理规律及对象电网的行为特性,就必须基于数学模型的数值仿真。但相继开断的分析面临着维数灾和交流潮流计算量大的瓶颈。分布因子法以潮流方程在运行点的局部线性化为基础,大大减少了计算量。但由于其忽略了非线性因素,故其计算误差不但难以控制,而且将随着多重开断而迅速增大,严重阻碍其实际应用。显然,此处中间结论的可信度比精准度更重要。

通过大数据的相关性分析,可以确认分布因子法的大误差算例与交流潮流方程的海森阵的二范数之间呈现强相关性。为了克服直接计算该二范数的困难,提出了在解析求取一阶灵敏度阵的基础上,通过数值摄动方式快速求取近似二阶灵敏度阵,并取其二范数来评估分布因子法的可信度。据此,文献^[24]提出在相继开断仿真中优先采用分布因子法,并估计其结果的可信度,当后者超过阈值时,自动运行交流潮流来消除累积误差。然后,以此为新起点,继续评估后续的开断。其中的阈值需要通过大数据的统计处理得到。由于新方法融合了分布因子法的快速性和交流潮流法的精确性,大大减少了因果分析的计算量。

5.2 大扰动受扰轨迹稳定裕度的智能快速求取

运动稳定性量化分析算法 EEAC 由完全忽略时变因素的静态 EEAC(SEEAC)、部分计及时变因素的动态 EEAC(DEEAC)、严格保留全部时变因素的集成 EEAC(IEEAC)这 3 种算法组成。三者的精度及计算量均依次增加。EEAC 算法顺序调用三者,以协调精度和计算量。

长期的困惑是:为什么相当多的算例可以用解析型的 SEEAC 得到精确的结果,但另外一些却不行。显然,若能可靠识别前一类算例,就可节省大量计算量。对大量算例的中间结果的详尽分析与验证,归纳出 SEEAC 及 DEEAC 产生大误差的原因是映射步长太大而不能很好地反映映象系统的时变性。

这样,在用 EEAC 分析一个算例的过程中,尽可能地利用中间结果,自动判断是否可以提前终止从 SEEAC,经 DEEAC,到 IEEAC 的计算流程。例如,当 DEEAC 给出的结果与 SEEAC 充分接近时,就不必再调用 IEEAC,从而显著减少了计算量^[25]。

5.3 振荡模式的易变性概念及其应用

大扰动稳定性的传统研究方法缺乏动态模式对参数敏感性的概念,故难以研究参数空间中的混沌机理。小扰动稳定性的传统研究方法基于线性化的系统模型,故长期来并不了解动态模式随受扰过程而变的可能性。

利用 EEAC 的模式解耦理论及对动态模式的识别能力,在对大量实际电力系统的穷尽式仿真结果的统计分析中发现:在一些情况下,多机系统的动态模式对于参数极其敏感,甚至在失稳区域中出现稳定的孤岛^[26]。而上述模式易变性与互补群各群内部的非同调性之间存在强相关。模式易变性概念的提出及其机理的发现,对于非线性运动系统稳定性的基础研究及工程应用意义重大。

通过参数摄动,分别考察各受扰轨迹的动态模式,可以判断系统的动态模式是否易变。但更希望可以不必摄动参数,而只根据目标参数点所对应轨迹的动态特征来识别系统的模式易变性。研究发现通过挖掘较大故障清除时间下的各机转子角间隙信息,就可完美地实现之,有关研究成果将另文发表。

不论是后验型的易变性定性分析,还是先验型的易变度定量分析,都是在因果型分析的 EEAC 中引入了统计关系型数据及对应的统计分析技术后得以成功。

5.4 停电防御系统对自然等外部环境的自适应能力

2008 年的冰灾深刻揭示了极端外部灾害的及时预警和群发性相继故障的量化分析的重要性;大灾难往往由偶然故障引发,并经过一系列相继故障和人为错误的推波助澜演化而成。要避免由小概率灾害带来的高风险,必须将自然环境和社会环境信息纳入风险分析的信息采集范围,从技术、规划、措施、管理、公共危机处理等多维度实现时空协调的综合防御。

当前停电防御系统仅采集电力信息,并按固定的预想故障集分析。为了有效防御自然灾害引发大停电,需要利用卫星遥感数据,跟踪气象数据,掌握输电走廊的地形、地质及植被、电气设备的位置等非电信息。据此预测山火、台风、暴雨、山洪、滑坡和泥石流时空演变,在线掌握断线、倒塔、闪络等故障率的变化,计算不同支路在自然灾害下的故障率及风险,在线动态修改安全分析的预想故障集。这样,跟随外部环境的变化来调整需要重点关注的故障,从而赋予停电防御系统一定的自适应能力^[27]。

6 融合因果分析、统计分析与行为分析,开拓全新的研究领域

与前两节不同,在本节的案例中,两种分析方法融合的作用不再是改进原有算法,而是开辟了两种分析方法都无法单独完成的任务领域。

6.1 研究范式的深度融合

6.1.1 根据实证数据建立参与者博弈行为的多代理模型

将电力系统放入包括一次能源、环境、经济的大能源框架中,综合分析能源流、物质流、信息流、资金流的广义阻塞因素。针对节能减排及碳交易、用电侧 EV、一次能源等具体问题建立动态仿真模型,研究决策支持,就必然遇到大量的博弈行为。

国际能源署(IEA)开发的 Market Allocation Model(MARKAL)模型将参与者决策因素作为给定的边界条件,难以反映决策行为与其他模型之间的交互。基于有限理性经济人的假设研究决策行为,认为决策者的偏好和一系列心理规则决定了决策空间的边界。实验经济学通过真实人参与的实验,可探究博弈行为的影响,但需要长期占用大量合格的实验者,且难以分析决策对参数的敏感性。

笔者提出:通过对问卷调查,实验经济学仿真,或者参与者实际博弈数据的挖掘分析,提取描述其博弈行为中的多维随机变量的联合概率分布函数,就可以建立不同参与者行为的多代理模型,代替对应的参与者加入到反复的实验经济学仿真,实现对包括真实决策者在内的大规模复杂系统的仿真模拟,建立具有工程实用价值的决策大数据沙盘推演新方法^[28]。

需要进一步研究自动学习及智能抽取特征的方法,对数据进行更加准确、有效的表达,并将其融入建模过程,减少人为设计中的不完备。

6.1.2 基于数学模型、实证多代理模型、真实参与者的混合仿真

实验经济学通过真实人参与的仿真,探究在一

定的社会环境中支配经济行为的规律,以及有限理性或非理性博弈行为对系统行为的影响。但是,当通过摄动方式分析系统行为对某指定参数的敏感程度时,希望每个参与者仅根据该参数的变化来改变其决策,而对其他参数的态度则在多次仿真中保持一致,但真实人却难以符合此要求。此外也难以长期邀请到合格的专家参与仿真,故希望能尽量用多代理来代替参与仿真的真实人。其前提是该代理决策的概率分布在任一场景下都符合相关真实人(群)决策的概率分布。为了能尽量用多代理模型代替大部分(有限)理性的真实参与者的行为,根据上一节介绍的从大数据中提取多层次相关的统计信息,建立能真实反映特定样本集多维概率分布特性的多代理模型。

为此需要建立支持数学模型、多代理模型,以及反映特殊人群或非理性行为的少量真实人之间的动态交互混合仿真的平台,通过仿真可以计及各种不确定性因素对决策后效与系统态势的影响。

大数据思维将经典的市场理论与创新的数据驱动方法结合,在分析与决策中引入信息论和行为学,挖掘参与者的博弈行为,使宏观决策与微观决策之间的互动更加真实。包括预测市场风险,支撑更灵活的交易,完善市场监测、预测和预警系统,管控能源金融风险以及信息安全风险,同时也促使各参与者转向量化的推演范式及决策机制,支撑对人、机、物交互中引起的涌现性(emergence)的研究。

由于关注全部数据,关注效率及精度,关注统计关系及因果关系,关注混沌规律(无规律可寻)及浮动规律(有迹可寻,但要求数据足够多),故特别适合于(部分)缺乏严格因果关系的场合。

6.2 通过混合仿真,实现沙盘推演

6.2.1 EV购车愿望、出行及参与电网辅助服务行为的沙盘推演

EV是在终端能源侧实现电能替代的重要手段。文献[29]等根据特定人群给出的一组关于购买偏好(或出行意愿)的问卷答案,构建对应的多代理的概率模型。对于给定的具体场景,该多代理多次决策的概率分布可以非常接近于问卷答案的概率分布。

为确保多代理模型的输出与某个指定人群的行为在统计学上的一致性,需要从学习样本集挖掘有关的多特征量心理阈值的联合概率分布。为此,利用关注度排序方法来反映学习样本人群对各因素的重视程度,利用树型频次矩阵反映相邻重要层决策之间的关联性,同时通过引入规则,以减少对问卷数

的依赖。通过蒙特卡洛方法按上述数据结构反映的权重完成意愿的随机抽取,从而构建重现人群意愿分布的多代理模型。学习样本集可来自仔细设计的调查问卷,或实验经济学仿真结果或实际行为的记录。

这样,就可用得到的多代理模拟该人群的决策行为或意愿,用不同的多代理代替行为特征不同的人群参与大量仿真。同时,由少量的真实人参与实验经济学仿真,以反映特殊人群的主观意愿。两者结合可以穷尽式地研究涉及大量参与者的社会问题,提供科学决策支持。例如出行距离、续驶里程焦虑、行程充电次数等因素对EV出行影响的贡献度,以及影响参与电网辅助服务意愿的主要因素。

6.2.2 碳排放市场的沙盘推演

为保障能源可持续发展与环境安全的分段目标合理化,并优化对应的动态路径,希望依靠混合仿真平台,计及广义环境及博弈行为的不确定性,推演整个博弈过程的轨迹。人类的社会行为信息非常复杂,而分析其与物理系统的交互影响则更加困难。文献[30]根据参与者实际博弈数据建立了对应的多代理模型,进而从混合仿真得到的轨迹中提取各种量化指标,有效评估不同方案的减排效果及风险防控能力,支撑碳市场的构建与运行决策。

6.3 跨领域CPS的协同分析

要实现物理系统的高效、高可靠和自愈,需要实现信息在物理系统内的广泛共享和有效利用,提高对系统的感知和控制能力。两者深度融合成为CPS,是物理系统智能化的关键。

普适的大数据技术要与具体业务配合。在数据的获取、存储、处理、安全、隐私保护的技术层面和管理层面都存在着新的挑战。在技术层面上包括:对有效数据的可靠采集;数据种类交叉、冗余、不一致、不完整、异步采集等壁垒;数据完整性、质量、一致性与多源的矛盾;开放性与隐私保护的矛盾;广度与深度的矛盾;大数据与深层知识的矛盾;统计分析与个性化服务的矛盾;数据的大量与知识的精准的矛盾。在管理层面上包括:消除多领域跨部门之间的数据孤岛与藩篱;共享交互与知识产权的矛盾;数据价值的评估体系。

6.3.1 CPPS领域的研究

电力系统内部数据源包括广域测量系统(WAMS)、SCADA、用电信息采集系统(AMI)、生产管理系统(PMIS)、能量管理系统(EMS)、配电管理系统、电力设备在线监测系统、客户服务系统、财务管理系统;电网外部来源包括气象信息系统、

GIS、互联网、公共服务部门、经济运行发展及用电需求等。

现代电力系统要求统筹考虑经济、社会、环境等方面对电力的要求与影响,解决广义阻塞^[31],综合信息、物理以及人的行为因素,优化能源、经济及环境的整体解决方案。为此,电力大数据除了对电力系统内部的发电、输电、变电、配电、用电及调度等不同环节之间数据的互联互通提出更高要求外,电力流、信息流^[32]、资金流、排放流、综合能源流、非能源的业务流之间的协调优化也需要在更广泛的多领域之间共享数据,提升数据价值利用水平^[33]。这反映在电力流的以下不同环节中。①发电领域:提升能源转换效率,提高风光电预测精度,协调消纳大规模新能源及分布式能源,深度调节火电。②输电领域:提升能源输送能力及效率,提升在线监测、抵御自然灾害风险及自愈能力。③变电领域:提升变电侧状态感知、广域全景实时信息共享和网络优化、实时控制和智能化运行及资产全寿命管理。④配电领域:分布式电源及储能的友好接入,提升供电可靠率及配电网控制水平,推进配电网能量流、信息流和业务流的融合及双向互动。⑤用电领域:实时电价及能耗,节能等用电服务模式,优化用电管理、用户能效诊断等增值服务。⑥调度领域:建设以数据驱动的智能调度体系,拓展对电力领域可靠性问题影响的研究至非电力领域,推进自动化、协调互动、科学组织生产,确保电网安全、经济、优质运行。

借助于数据统计、机器学习、传统数据挖掘、领域知识挖掘,以及数据可视化等技术,洞察并展现一些隐藏于大数据中的分布、关系、趋势、模式、规律及性质,从中获悉用户个性化用能需求与特性、能耗的时空分布等。

数据驱动无模型分析适用于相对简单、直观、数据性明显的业务。分析智能表计数据,可获得用户的分类、行为特性及价格敏感度信息,有助于差异化服务需求。对于因果关系强的业务则必须与领域知识相结合,而涉及博弈行为时则需要依靠数学模型、多代理与实验经济学的混合仿真^[34]。由计算设备、通信网络、传感设备与各种固定或移动的物理设备协同,在计算机虚拟环境中同步构造了物理系统的镜像,从物理系统和信息系统的整体视角预警风险,推演决策。许多物理部件都将植入嵌入式控制设备并通过通信网络与控制中心互联,控制中心可以根据传感器网络收集到的系统内外信息,在线自适应地修改分散控制器的参数或流程,甚至信息系统的结构。在反复的沙盘推演基础上,协调众多个

体和系统之间的优化,提高整个系统的功效。

6.3.2 CPES 领域的研究

全球能源安全和气候变化问题迫切需要科学合理的能源整体解决方案。大能源观以电能为核心,将一次能源、二次能源、终端能源等环节视为整体,并计及与经济发展、社会文明、民生福利、气候环境、政策法规等的交互影响,来研究能源安全与能源经济。它以系统论的方法,可持续发展的理念,从全局、整体、历史、开放、普遍联系的视角统筹考虑能源发展与经济、社会、环境、外交等各方面的关系,推动能源发展方式转变。为此,需要构建一个融合开放的大能源数据平台,支撑经济环境、一次能源、管理政策、能源需求及用电方式变化等因素的统筹协调,协同解决广义阻塞问题^[35]。

在一次能源侧构建多元互补的综合能源系统,优化制氢等多类型、大容量、低成本、高效率、长寿命的综合储能系统以及就地可调负荷,以应对可再生能源的随机性及模糊性,减少弃风弃光。鼓励化石类能源加工的优化,分质分级利用,促进可再生能源与化石能源的协同生产。在能源输送环节协调建设及运行电网、热网、石油管网、天然气网、交通网。在终端能源侧构建与互联网深度融合、与分布式能源友好的冷热电三联供微网。通过能源流监测、管理和调度的网络化,推动分布式可再生能源的发展。

单领域信息所包含的知识量有限,而相关的多领域的信息结合,知识量将大大增加。但如何找到并揭示未知的知识,则是关键。CPES 是以整个能源系统为物理基础的 CPS。它通过传感测量、专用网、互联网、大数据、计算机和控制技术与各种一次能源、二次能源及终端能源深度融合而成。在先进的信息通信技术支撑下,以电能为核心枢纽,经济而有效地支撑一次能源的清洁替代,可靠而开放地推动终端能源的电能替代。多少有些令人遗憾的是,人们将其概念范畴的某个子集称为能源互联网(Energy Internet)^[36]。

一次能源、排放、自然灾害、电力市场、资金约束、通信等领域的事件都会通过影响电力系统内部环节而引发充裕性问题,故与广义阻塞概念的引入类似,有必要将充裕性的概念扩展到广义范畴。文献^[37]讨论了一次能源对电力系统充裕度的影响,并建议将停电防御框架拓展至一次能源领域。

CPES 将网络技术深度嵌入到能源系统中,感知和改变后者的状态,从多种形式的中提取知识及决策支持^[38]。提供更加细致的保护和防御以抵御对数据、硬件及功能的威胁。包括基于模型的

程序指引,从数据中学习并预测,及根据例子预测和判断。由于相关人员的参与面及程度的加大,行为数据及其影响大大增加。

需要将广域信息的采集范围从电力系统内部扩大到自然环境和社会环境,对冻雨、台风、雷击等自然灾害早期预警,将预警目标从单故障风险拓展到群发性相继故障的风险,跟踪一次能源和环境条件等的变化态势,动态评估信息系统的风险及其对停电防御的影响,留出更长的时间来准备应对预案。将自适应优化的决策范畴从减少停电规模(通过预防控制和紧急控制)扩展到同时考虑减少停电时间(通过恢复控制)。尽量避免突发的小概率事件演化为大范围的公共危机,并通过对灾难的善后处理尽快地回归正常秩序。

泛在信息网络包括因特网与各种专用网。一方面,综合能源系统的调度与控制需要依靠专用网来保证其可靠性及实时性。对各种能源系统的专用信息网络建立有序的共享,实现整个能源流的实时监测、动态控制,实现人与物理进程的交互控制。另一方面,转变能源生产和消费模式都需要发挥市场在资源配置中的决定性作用,提高用户参与能源交易的广度与深度,建立新型能源市场交易体系和商业运营平台,发展储能和EV应用、智慧用能和增值服务、绿色能源灵活交易,破除物理系统及信息系统各自内部及相互之间的藩篱,促进人才、资金、资源、技术等要素的流通。

建立多方参与、充分竞争,以能源、辅助服务、新能源配额、碳排放、虚拟能源货币等为标的物的多元能源市场交易体系,鼓励众多参与者灵活自主地参与能源市场,并通过沙盘推演,研究参与者的博弈行为对行动计划效果的影响。这些都需要在保障信息安全和市场参与者的合法权益的前提下,依靠互联网的开放性、商业性开展^[39]。

7 讨论

“CPES”的名称较贴切地反映了智能电网的发展方向,电力与能源及环境的关联,以及大数据等信息技术及大能源物理技术深度融合的需求^[40]。为此,不但需要将物理目标从电力流扩充到整个能源流,而且需要将相关的信息系统从企业专用网拓展到电子商务、社交网络、物联网、移动互联网、数据终端。所涉及的数据范畴则在物理系统内部数据上增添了覆盖环境、能源、市场、金融、社会等方面的数据。由于计及了各种能源系统(或相关的非能源系统)与电力系统之间的交互影响,可以更真实地描述

电力安全、能源安全及环境安全;按大数据思维预测非电的能源系统及非能源的外部环节的演变态势,显著增加了风险预警的提前时间。

大数据包括来自物理世界的传感(或实验)数据,以及来自人类社会的行为数据、媒体数据及日志数据。后者不但非结构化数据多,而且价值密度很低,故比科学实验的数据分析更困难。发展趋势是:不断将常规的小数据拓广为更广泛的非常规的大数据,同时,又不断地从大数据中抽取能用小数据表达的深层知识。如果未与信息提炼及知识发现过程紧密结合,大数据往往成为负资产。这是大数据与小数据之间的辩证关系。

严格来讲,统计学无法检验逻辑上的因果关系。文献[41]中的例子很生动:统计结果表明“吸烟人群的肺癌发病率会比不吸烟的人群高几倍”,但无法得出“吸烟致癌”的逻辑结论。因为至少有另一种假设也能解释该统计结果,即有一种基因能使人喜欢抽烟,同时又使其容易得肺癌。在这种假设下,该基因和肺癌就是因果关系,而吸烟和肺癌则仅有相关性。统计学的相关性还可能产生错觉:例如统计结果表明,下雨之前常见到燕子低飞。若从时序看,得到的结论可能是“燕子低飞是下雨的原因”,而客观上的因果关系却是“即将下雨是燕子低飞的原因”。

基于实证数据的统计分析方法既不依赖于数学模型及参数,也不苛求场景细节,但是却难以回答“为什么”的问题,也难以分析特征对于参数及行为的敏感度。基于模型仿真的因果分析则相反:有清晰的物理概念,有利于机理分析。但由于针对具体案例,每次仿真必须事前确定全部数学模型、参数及仿真场景,故计算量大,存在维数灾,且难以计及非常规数据的影响。

因此,大数据的统计分析与数学模型分析不可能相互取代。数据量再大,数据驱动技术再发展,也不能代替基础理论研究。人类探索未知世界不仅想知道“是什么”,也想知道“为什么”。事实上,大数据技术的目的之一就是为了揭示事实上存在,但目前并不掌握的(准)因果关系。

不论是认为当前电力领域没有研究大数据的空间,还是将所有的电力系统问题都冠以“大数据”,这两个极端都是对电力大数据的误解。其共同点是没有把数据转化成思想或知识。笔者认同文献[42]指出的:在通常具有较完备理论体系的电力系统中引入大数据技术,应针对常规方法无法解决的问题和潜在的需求,而不是希望以大数据技术完全取代现有技术。要充分发挥大数据技术的作用,就要结合

具体问题,协调好因果分析与统计分析各自的优点。

在大能源领域中应用大数据技术,应该将物理系统的范畴扩大到整个能源流。针对要解决的具体问题的特点,充分利用实时网、管理网、互联网、仿真网等传送的各类数据。同时,还应该拓广研究的范式,包括:在原本完全依靠经验与统计分析的过程中加入数学模型及因果分析的手段;在原本完全依靠数学模型的分析方法中加入统计分析的手段;将涉及博弈因素大数据,建立反映多维联合概率分布函数的多代理模型,通过实验经济学仿真实现沙盘推演。

文献[43]预测了大数据发展趋势。为了使知识资源能像物质资源一样有效流通与分配,并能从数据利用的视角揭示信息流与物理流之间的交互规律,需要定量地对知识及其提取方法做恰当的评价。

8 结语

大能源观覆盖了包括电力系统与各种一次能源与广义的终端能源在内的整个能源流,并反映了能源流与外界环境、管理政策、用户需求、信息及人的行为等系统之间的交互影响,是对孤立研究电力系统的思维模式的重大革新。

大数据思维是协同处理统计型、因果型及博弈型等不同类型数据的知识获取新范式。大数据技术并不是常规数据技术的简单演化,而是在数据管理、知识提取和价值应用各方面的思维革新。

大能源思维与大数据思维的碰撞,将信息演变成新型生产资料,并与能源深度融合,大大提升了电力与能源数据价值应用的层次和水平。这不仅有助于能源的高效利用,而且为能源资源评估、用能预测、能源规划、运行、协同调度、转换、管理的高效协调和可持续发展奠定了科学与技术基础。

参考文献

- [1] 薛禹胜,肖世杰.综合防御高风险的小概率事件——对日本相继天灾引发大停电及核泄漏事件的思考[J].电力系统自动化,2011,35(8):1-11.
XUE Yusheng, XIAO Shijie. Comprehensively defending high risk events with low probability[J]. Automation of Electric Power Systems, 2011, 35(8): 1-11.
- [2] 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.
- [3] 乌克兰电力系统遭受攻击事件综合分析报告[EB/OL].[2016-02-26]. http://mp.weixin.qq.com/s?__biz=MjM5MTA3Nzk4MQ==&mid=401994100&idx=1&sn=ebbe4e2dd089c3a12b914546367678b5&scene=2&srcid=0226JvA9wUCKeXGPiGPNsaiQ&from=timeline&isappinstalled=0#wechat_redirect.
- [4] 刘振亚.中国电力与能源[M].北京:中国电力出版社,2012.
- [5] 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.
- [6] 薛禹胜,赖业宁.大能源思维与大数据思维的融合:(一)大数据与电力大数据[J].电力系统自动化,2016,40(1):1-8. DOI: 10.7500/AEPS20151208005.
XUE Yusheng, LAI Yening. Integration of macro energy thinking and big data thinking: Part one big data and power big data[J]. Automation of Electric Power Systems, 2016, 40(1): 1-8. DOI: 10.7500/AEPS20151208005.
- [7] XUE Y. The stability-preserving trajectory-reduction methodology for analyzing nonlinear dynamics [C]// International Conferences on Info-tech & Info-net (ICH2001), October 29–November 1, 2001, Beijing, China; 7-19.
- [8] QIU R, ANTONIK P. Big data and smart grid: a random matrix[M]. Hoboken, NJ, USA: John Wiley and Sons Ltd, 2014: 33-60.
- [9] 宋亚奇,周国亮,朱永利.智能电网大数据处理技术现状与挑战[J].电网技术,2013,37(4):927-935.
SONG Yaqi, ZHOU Guoliang, ZHU Yongli. Present status and challenges of big data processing in smart grid[J]. Power System Technology, 2013, 37(4): 927-935.
- [10] 苗新,张东霞,孙德栋.在配电网中应用大数据的机遇与挑战[J].电网技术,2015,39(11):3122-3127.
MIAO Xin, ZHANG Dongxia, SUN Dedong. The opportunity and challenge of big data's application in power distribution networks[J]. Power System Technology, 2015, 39(11): 3122-3127.
- [11] US-Canada Power System Outage Task Force. Final report on the August 14th blackout in the United States and Canada: causes and recommendations[EB/OL]. [2016-03-01]. <http://www.nerc.com/>.
- [12] 曲朝阳,陈帅,杨帆,等.基于云计算技术的电力大数据预处理属性约简方法[J].电力系统自动化,2014,38(8):67-71. DOI: 10.7500/AEPS20130601001.
QU Zhaoyang, CHEN Shuai, YANG Fan, et al. An attribute reducing method for electric power big data preprocessing based on cloud computing technology[J]. Automation of Electric Power Systems, 2014, 38(8): 67-71. DOI: 10.7500/AEPS20130601001.
- [13] 严霄凤,张德馨.大数据研究[J].计算机技术与发展,2013,23(4):168-172.
YAN Xiaofeng, ZHANG Dexin. Big data research [J]. Computer Technology and Development, 2013, 23(4): 168-172.
- [14] DORR D, GRAY G. The whys, whats, and hows of managing data as an asset[R]. California, USA: Electric Power Research Institute, 2014.
- [15] 中国电机工程学会信息化专委会.中国电力大数据发展白皮书[R].北京:中国电机工程学会,2013.
- [16] CHEN Y, SHEN C. A Jacobian-free Newton-GMRES(m)

- method with adaptive preconditioner and its application for power flow calculations[J]. IEEE Trans on Power Systems, 2006, 21(3): 1096-1103.
- [17] 覃智君, 侯云鹤, 吴复立. 大规模交直流系统潮流计算的实用化模型[J]. 中国电机工程学报, 2011, 31(10): 95-101.
QIN Zhijun, HOU Yunhe, WU Fuli. Practical model for large-scale AC-DC system power flow calculation[J]. Proceedings of the CSEE, 2011, 31(10): 95-101.
- [18] 孙宏斌, 王康, 张伯明, 等. 采用线性决策树的暂态稳定规则提取[J]. 中国电机工程学报, 2011, 31(34): 61-67.
SUN Hongbin, WANG Kang, ZHANG Boming, et al. Rule extraction in transient stability study using linear decision trees [J]. Proceedings of the CSEE, 2011, 31(34): 61-67.
- [19] 于之虹, 郭志忠. 基于数据挖掘理论的电力系统暂态稳定评估[J]. 电力系统自动化, 2003, 27(8): 45-48.
YU Zhihong, GUO Zhizhong. A novel approach for transient stability assessment based on data mining theory [J]. Automation of Electric Power Systems, 2003, 27(8): 45-48.
- [20] 马世英, 仲悟之, 汤涌, 等. 基于多信息源的电压安全全过程预警及防控系统[J]. 电网技术, 2013, 37(9): 2459-2465.
MA Shiyong, ZHONG Wuzhi, TANG Yong, et al. A multiple information sources based whole process early warning and preventive control system for voltage security [J]. Power System Technology, 2013, 37(9): 2459-2465.
- [21] 薛禹胜. 运动稳定性量化理论——非自治非线性多刚体系统的稳定性分析[M]. 南京: 江苏科学技术出版社, 1999.
- [22] 郁琛, 薛禹胜, 文福拴, 等. 按时序特征优化模型后在线选配的超短期风电预测[J]. 电力系统自动化, 2015, 39(8): 5-11. DOI: 10.7500/AEPS20141230007.
YU Chen, XUE Yusheng, WEN Fushuan, et al. An ultra-short-term wind power prediction method using "offline classification and optimization online model matching" based on time series features [J]. Automation of Electric Power Systems, 2015, 39(8): 5-11. DOI: 10.7500/AEPS20141230007.
- [23] 薛禹胜, 谢东亮, 薛峰, 等. 智能电网运行充裕性的研究框架: (一)要素与模型[J]. 电力系统自动化, 2014, 38(10): 1-9. DOI: 10.7500/AEPS20140416001.
XUE Yusheng, XIE Dongliang, XUE Feng, et al. A research framework for operation adequacy in smart grid: Part one elements and models [J]. Automation of Electric Power Systems, 2014, 38(10): 1-9. DOI: 10.7500/AEPS20140416001.
- [24] 薛禹胜, 戴元煜, 于继来, 等. 可信度控制下的相继开断潮流快速计算[J]. 电力系统自动化, 2015, 39(22): 37-45. DOI: 10.7500/AEPS20150214002.
XUE Yusheng, DAI Yuanyu, YU Jilai, et al. A credibility index based algorithm for fast estimating power flow in cascading failures[J]. Automation of Electric Power Systems, 2015, 39(22): 37-45. DOI: 10.7500/AEPS20150214002.
- [25] XUE Y S, HUANG T G, LI K, et al. An efficient and robust case sorting algorithm for transient stability assessment[C]// IEEE Power and Energy Society General Meeting, July 26-30, 2015, Denver, USA: 5p.
- [26] XUE Y S. The mechanisms of unstable modes evolving with fault clearing time[C]// Proceedings of the 16th Symposium on Electrical Power Engineering, November, 1995, Taiwan, China.
- [27] 薛禹胜, 吴勇军, 谢云云, 等. 停电防御框架向自然灾害预警的拓展[J]. 电力系统自动化, 2013, 37(16): 18-26.
XUE Yusheng, WU Yongjun, XIE Yunyun, et al. Extension of blackout defense scheme to natural disasters early-warning [J]. Automation of Electric Power Systems, 2013, 37(16): 18-26.
- [28] XUE Y S, WU J A, XIE D L, et al. Experimental study on EV purchases assisted by multi-agents representing a set of questionnaires[C]// International Conference of Life System Modeling and Simulation (LSMS) and International Conference on Intelligent Computing for Sustainable Energy and Environment (ICSEE), September 20-23, 2014, Shanghai, China: 11p.
- [29] 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.
- [30] HUANG J, XUE Y S, JIANG C, et al. An experimental study on emission trading behaviors of generation companies[J]. IEEE Trans on Power Systems, 2015, 30(2): 1076-1083.
- [31] XUE Y S, LI T R, YIN X, et al. Managements of generalized congestions[J]. IEEE Trans on Smart Grid, 2013, 4(3): 1675-1683.
- [32] YU W J, XUE Y S, LUO J B, et al. A UHV grid security and stability defense system: considering the risk of power system communication[J]. IEEE Trans on Smart Grid, 2016, 7(1): 491-500.
- [33] 赵俊华, 文福拴, 薛禹胜, 等. 电力 CPS 的架构及其实现技术与挑战[J]. 电力系统自动化, 2010, 34(16): 1-7.
ZHAO Junhua, WEN Fushuan, XUE Yusheng, et al. Cyber physical power systems: architecture, implementation techniques and challenges[J]. Automation of Electric Power Systems, 2010, 34(16): 1-7.
- [34] 徐伟, 薛禹胜, 陈实, 等. 从实测轨迹提取知识时的困难及展望[J]. 电力系统自动化, 2009, 33(15): 1-7.
XU Wei, XUE Yusheng, CHEN Shi, et al. Difficulties and prospects of knowledge extracting from measured trajectories [J]. Automation of Electric Power Systems, 2009, 33(15): 1-7.
- [35] 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.
- [36] XUE Y S. Energy Internet or comprehensive energy network? [J]. Journal of Modern Power Systems and Clean Energy, 2015, 3(3): 297-301.
- [37] XUE Y S, CAI B, JAMES G, et al. Primary energy congestion

- of power systems[J]. Journal of Modern Power Systems and Clean Energy, 2014, 2(1): 39-49.
- [38] XUE Y S, NI M, YU J, et al. Study of the impact of communication failures on power system[C]// IEEE Power and Energy Society General Meeting, July 26-30, 2015, Denver, USA.
- [39] 陈启鑫,刘敦楠,林今,等.能源互联网的商业模式与市场机制(一)[J].电网技术,2015,39(11):3050-3056.
CHEN Qixin, LIU Dunnan, LIN Jin, et al. Business models and market mechanisms of Energy Internet (1)[J]. Power System Technology, 2015, 39(11): 3050-3056.
- [40] YU X, XUE Y. Smart grids: a cyber-physical systems perspective[J]. Proceedings of the IEEE, January 2016: 1-13. DOI: 10.1109/JPROC.2015.2503119.
- [41] 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.
- [42] 黄彦浩,于之虹,谢昶,等.电力大数据技术与电力系统仿真计算结合问题研究[J].中国电机工程学报,2015,35(1):13-22.
- HUANG Yanhao, YU Zhihong, XIE Chang, et al. Study on the application of electric power big data technology in power system simulation[J]. Proceedings of the CSEE, 2015, 35(1): 13-22.
- [43] 潘柱廷,程学旗,袁晓如,等.CCF 大专委 2016 年大数据发展趋势预测——解读和行动建议[J].大数据,2016(1):105-113.
PAN Zhuting, CHENG Xueqi, YUAN Xiaoru, et al. Developing trend forecasting of big data in 2016 from CCF TFB: interpretation and proposals[J]. Big Data, 2016(1): 105-113.

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 张焱)

Integration of Macro Energy Thinking and Big Data Thinking

Part Two Applications and Explorations

XUE Yusheng^{1,2}, LAI Yening¹

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

2. State Key Laboratory of Smart Grid Protection and Control, Nanjing 211106, China)

Abstract: In part one of this series paper, the big data thinking and the macro energy thinking have been explained. On one hand, the energy system is considered in the context of the cyber-physical system, to break down the physical barriers among power system, primary energy system and end-use energy system. On the other hand, the private network and the Internet are combined together to constitute the communication system, to break down the barriers between operation and business activities. Thus, the above-mentioned physical and information systems can be integrated into a cyber-physical system, through causal data, statistical data, gambling data and corresponding big data technology. Along this direction, this paper explores the applications of power (energy) big data. Based on several topics that have been studied, the significance and approaches of big data technology in enhancing the economy and reliability of energy flow in different spatial and temporal scales are induced, which include: ① introduce casual analysis measures to topics that usually relied entirely on statistical analysis, to improve the applicability and accuracy; ② introduce statistical analysis measures to topics that usually relied entirely on casual analysis, to enhance the efficiency; ③ combine casual analysis, statistical analysis and experimental economics simulation analysis, to create a new research area of what a single analysis paradigm cannot solved, such as the interactions between the macro energy system and the natural environment, the social environment, market economy, policy and other non-energy sectors.

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: primary energy; end-use energy; big data technology; smart grid; cyber-physical system (CPS); decision support; application research