

面向智能电网应用的电力大数据关键技术

彭小圣¹, 邓迪元¹, 程时杰¹, 文劲宇¹, 李朝晖¹, 牛林²

(1. 华中科技大学电气与电子工程学院, 湖北省 武汉市 430074; 2. 国网技术学院, 山东省 济南市 250002)

Key Technologies of Electric Power Big Data and Its Application Prospects in Smart Grid

PENG Xiaosheng¹, DENG Diyu¹, CHENG Shijie¹, WEN Jinyu¹, LI Zhaohui¹, NIU Lin²

(1. School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, Hubei Province, China; 2. State Grid of China Technology College, Jinan 250002, Shandong Province, China)

ABSTRACT: Application of big data techniques in power system will contribute to the sustainable development of power industry companies and the establishment of strong smart grid. This paper introduced a universal framework of electric power big data platform, based on the analysis of the relationships among the big data, cloud computing and smart grid. Then key techniques of electric power big data were discussed in four aspects, including big data management techniques, big data analyzing techniques, big data processing techniques and big data visualization techniques. Finally, the paper presented three typical application examples of electric power big data techniques which were new and renewable energy integration, wind turbine condition monitoring and assessment, power system catastrophic failure prediction. The application of big data technologies in smart grid will bring deep change and brilliant future to power system, and will promote the development of power industry to a new generation.

KEY WORDS: big data; cloud computing; smart grid; data integration; data analysis; data processing; data visualization

摘要: 大数据为智能电网的发展注入新的活力, 掌握电力大数据的关键技术对电力行业的可持续发展和坚强智能电网的建立具有重要意义。在分析大数据、云计算、智能电网三者关系的基础上, 给出具有通用性的电力大数据平台总体架构, 并从电力大数据的集成管理技术、数据分析技术、数据处理技术、数据展现技术 4 个方面深入探讨符合电力企业发展需求的大数据关键技术的选择。最后通过 3 个典型案例, 分析了电力大数据关键技术在新能源并网、风电机组安全评估、电网灾难预警上的应用。大数据关键技术在电力行业的广泛应用必将带来行业的变革, 将智能电网的发展推向新的阶段。

关键词: 大数据; 云计算; 智能电网; 数据集成; 数据分析; 数据处理; 数据展现

0 引言

大数据这个术语最早期的引用可追溯到 Apache 软件基金会的开源项目 Nutch。当时, 大数据用来描述为更新网络搜索索引需要同时进行批量处理或分析的大量数据集^[1]。早在 2008 年 Nature 就出版了专刊 “Big Data”, 从网络经济学、超级计算、互联网技术、生物医药、环境科学等多个方面介绍了海量数据带来的挑战; 2011 年 Science 推出数据处理的专刊 “Dealing With Data”, 深入讨论了数据洪流(data deluge, DD)所带来的挑战, 并指出如果能够更有效地组织和利用这些海量数据, 人们将得到更多的机会发挥科学技术对推动社会发展的巨大作用; 2012 年奥巴马宣布美国政府投资 2 亿美元启动 “大数据研究和发展计划”, 将大数据比喻为 “未来的新石油”, 将对大数据的研究上升为国家意志, 掀起了世界各国大数据的研究热潮^[2-7]。

大数据的定义, 业界虽然有一些共识, 但是并未有统一的定义。麦肯锡认为 “大数据是指其大小超出典型数据软件抓取、储存、管理和分析范围的数据集合”; Gartner 认为 “大数据是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产”^[1]。在对大数据的定义中, 比较有代表性的定义是 3V 定义, 即规模性(Volume), 多样性(Variety)和高速性(Velocity), 规模性是指数据量庞大, 数据洪流已经从 GB、TB 级上升到 PB、EB、ZB 级; 多样性是指数据类型繁多, 并且包含结构化、半结构化和非结构化的数据; 高速性则是指数据以数据流的形态

基金项目: 国家自然科学基金项目(51228701)。

Project Supported by National Natural Science Foundation of China (51228701).

快速、动态的产生,数据处理的速度也必须达到高速实时处理^[7-12]。另外大数据第4V的讨论并没有取得一致的结论,国际数据公司(international data corporation, IDC)认为大数据应该具有价值性(Value),且价值密度稀疏;IBM则认为大数据的第4V特性是真实性(Veracity)^[8]。大数据的这些特点决定了在大数据时代,传统的数据处理技术必需有革命性的提升。

电力系统作为经济发展和人类生活依赖的能量供给系统,也具有大数据的典型特征。电力系统是最复杂的人造系统之一,其具有地理位置分布广泛、发电用电实时平衡、传输能量数量庞大、电能传输光速可达、通讯调度高度可靠、实时运行从不停止、重大故障瞬间扩大等特点,这些特点决定了电力系统运行时产生的数据数量庞大、增长快速、类型丰富,完全符合大数据的所有特征,是典型的大数据。在智能电网深入推进的形势下,电力系统的数字化、信息化、智能化不断发展,带来了更多的数据源,例如智能电表从数以亿计的家庭和企业终端带来的数据,电力设备状态监测系统从数以万计的发电机、变压器、开关设备、架空线路、高压电缆等设备中获取的高速增长的监测数据,光伏和风电功率预测所需的大量的历史运行数据、气象观测数据等。因此在电力系统数据爆炸式增长的新形势下,传统的数据处理技术遇到瓶颈,不能满足电力行业从海量数据中快速获取知识与信息的分析需求,电力大数据技术的应用是电力行业信息化、智能化发展的必然要求。

中国电机工程学会信息化专委会在2013年3月发布了《中国电力大数据发展白皮书》,将2013年定为“中国大数据元年”,掀起了电力大数据的研究热潮^[10]。根据白皮书描述,电力大数据的特征可概括为3V和3E。3V为体量大(Volume)、速度快(Velocity)和类型多(Variety);3E为数据即能量(Energy)、数据即交互(Exchange)和数据即共情(Empathy)^[10]。其3V的描述和其他行业的描述比较接近,3E的描述具有典型的电力行业特征,体现了大数据在电力系统应用中的巨大价值。数据即能量简而言之,就是指通过大数据分析达到节能的目的,电力大数据应用的过程,就是电力数据能量释放的过程;数据即交互是指电力大数据与国民经济其他领域数据进行交互融合,才能发挥其更大价值;数据即共情是指电力大数据紧密联系千家万户、厂矿企业,只有情系用电客户,满足客户需求,电力企

业方能以数据取胜^[5]。

电力大数据贯穿发、输、变、配、用等电力生产及管理的各个环节,是能源变革中电力工业技术革新的必然过程,不仅是技术上的进步,更是涉及电力系统管理体制、发展理念和技术路线等方面的重大变革,是下一代电力系统在大数据时代下价值形态的跃升^[10]。对建设坚强智能电网而言,亟需开展大数据相关技术的研究,为电力大数据时代的到来奠定理论基础和技术积累。

1 智能电网、云计算和大数据的关系

1.1 智能电网与大数据的关系

智能电网就是将信息技术、计算机技术、通信技术和原有输、配电基础设施高度集成而形成的新型电网,具有提高能源效率、提高供电安全性、减少环境影响、提高供电可靠性、减少输电网电能损耗等优点^[11]。智能电网的理念是通过获取更多的用户如何用电、怎样用电的信息,来优化电的生产、分配及消耗,利用现代网络、通信和信息技术进行信息海量交互,来实现电网设备间信息交换,并自动完成信息采集、测量、控制、保护、计量和监测等基本功能,可根据需要支持电网实时自动化控制、智能调节、在线分析决策和协同互动等高级功能,因此相关研究者指出:可以抽象的认为,智能电网就是大数据这个概念在电力行业中的应用^[12]。

1.2 大数据与云计算的关系

根据美国国家标准与技术研究院的定义,云计算是一种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^[13]。

从技术上看,大数据根植于云计算,云计算的数据存储、管理与分析方面的技术是大数据技术的基础。利用云计算强大的计算能力,可以更加迅速地处理大数据,并更方便地提供服务;通过大数据的业务需求,可以为云计算的发展找到更多更好的实际应用^[14]。云计算使大数据应用成为可能,但是没有大数据的信息沉淀,云计算的功能将得不到完全发挥,所以从整体上看,大数据与云计算是相辅相成的^[15]。

云计算和大数据的侧重点不同,因此也有较大的差别。大数据关注重心在于数据背后的信息沉淀与业务分析,因此其推动力量来源于拥有大数据的企业和软件厂商;云计算关注重心在于计算能力,偏重于技术解决方案,因此其推动力量来自于计算

资源和存储资源的生产厂商^[13-15]。

云计算技术的发展早于大数据技术的发展，但是大数据的业务需求又为云计算技术的发展带来新的机遇，一方面促进了云计算技术向更加贴近用户需求的方向发展，另一方面带来了更高处理速度、更大存储容量的要求。

1.3 智能电网、云计算、大数据的相互关系

图1简要描述了智能电网、云计算、大数据三者之间的相互关系。

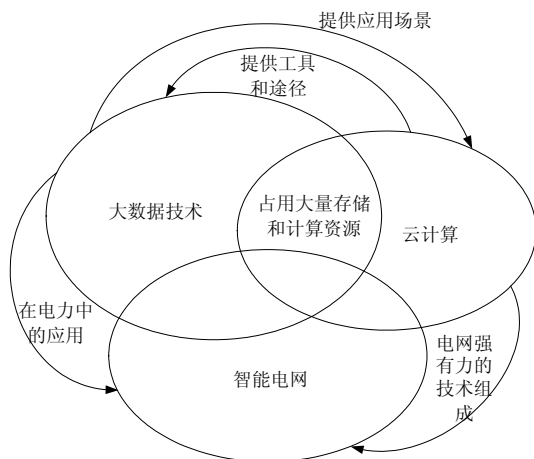


图1 大数据技术，云计算，智能电网三者的相互关系

Fig. 1 Relationships among smart grid, big data and cloud computing technologies

云计算能够整合智能电网系统内部计算处理和存储资源，提高电网处理和交互能力，成为电网强有力的技术组成^[16]；大数据技术立足于业务服务需求，根植于云计算，以云计算技术为基础；智能电网可以抽象的认为是大数据这个概念在电力中的应用^[12]，所以三者是彼此交互的关系。

智能电网、云计算、大数据三者之间的关系，从更加深层次来讲，是电力系统发展到不同阶段的产物，具有代际传承的特点。图2从代际传承的角度描述了三者之间的相互关系。

智能电网是信息技术、计算机技术、人工智能技术等在传统电网上应用沉淀的结果，满足电网信息化、智能化、清洁化等高层次的运营和管理需求，既是对传统电网的继承，也是对传统电网的发扬，所以其发展必然与新技术同步。来自于计算机和信息技术领域最前沿的云计算技术和大数据技术，正是其发展阶段技术层面和应用层面两个具有划时代意义的新技术。云计算技术中的分布式存储技术和并行计算技术，满足了电网海量数据的存储和计算需求，因此云计算技术推出不久，电力云的概念就提出来，云计算技术在电力系统中的应用也逐渐

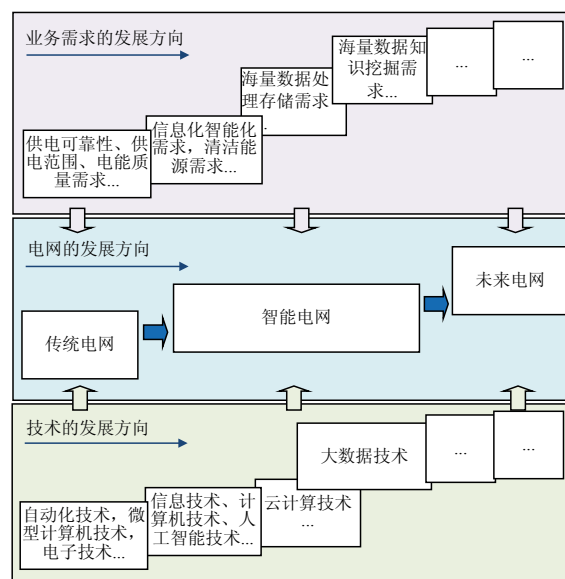


图2 大数据技术，云计算，智能电网三者的代际传承关系

Fig. 2 Evolution relationships among big data, cloud computing and smart grid

呈现出百花齐放的态势，推动了智能电网的发展。大数据技术既是传统数据分析与挖掘技术的延续，也是数据量级增长到一定阶段时知识挖掘与业务应用需求的必然产物，因此大数据技术的大部分应用都以云计算的关键技术或者与云计算类似的分布式存储和处理技术为基础。电力大数据技术的发展从某种意义上讲，可以看成是云计算技术在智能电网中，高级业务需求的实现过程。

2 电力大数据平台的总体架构

Apache 基金会开源技术通用的大数据平台整体架构如图3所示^[17]。该架构具有较好的通用性，适用于电力企业大数据的规划，其主要思想是利用基于 Hadoop 文件系统(Hadoop distributed file system, HDFS)的分布式文件处理系统作为大数据

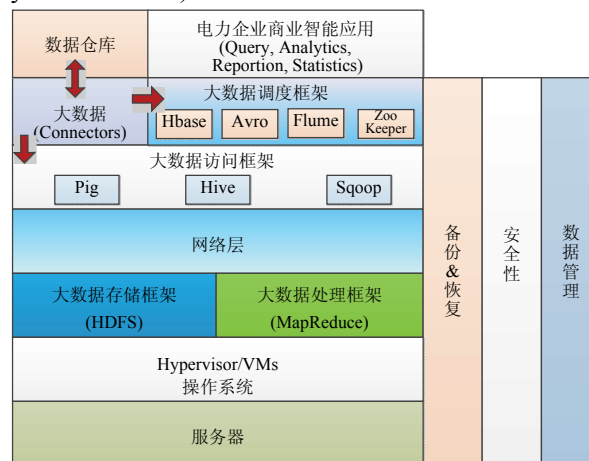


图3 电力大数据平台的总体架构

Fig. 3 Universal framework of big data platform

的存储框架,利用基于 MapReduce 的分布式计算技术作为大数据的处理框架。以分布式文件处理技术为基础,使 PB、ZB 级的数据存储成为可能;以分布式计算技术为基础,使得 PB、ZB 级数据的查询分析成为可能。另外该框架中还包含商业智能应用、传统的数据仓库、大数据访问框架、大数据调度框架、网络层、操作系统、服务器、备份和恢复、数据管理等模块。

大数据存储框架和大数据处理框架通常构建在通用的服务器、操作系统或者虚拟机上,使得该架构所需的硬件具有低成本和高扩展性的特点,标准的普通服务器或者 PC 机即可成为基于该架构的终端构成单元。

大数据存储框架和大数据处理框架之上是通过网络层连接的大数据访问框架,该访问框架包含并行计算机编程语言 Pig、数据仓库工具 Hive、开源数据传递工具 Sqoop 等子模块。

大数据调度框架包含基于列存储的开源非关系型数据库 Hbase、数据序列化格式与传输工具 Avro、日志收集系统 Flume、分布式锁设施 ZooKeeper 等模块。大数据调度框架实现了对大数据的组织与调度,为数据分析提供了必要条件。在大数据调度框架之上是企业级商业智能应用系统,可以开展查询、分析、统计、报表等高级应用^[17]。大数据的管理、安全和备份恢复框架帮助进行大数据的治理和保护^[17]。

该框架几乎涵盖了大数据技术的所有环节,值得指出的是,通过该访问框架不仅可以实现对分布式文件存储系统的访问,而且通过大数据连接器和开源数据传递工具 Sqoop 可以实现对传统数据仓库的访问。

大数据的处理流程可以定义为在合适工具的辅助下,对广泛异构的数据源进行抽取和集成,按照统一的标准对结果进行存储,利用恰当的数据分析技术对存储的数据进行分析,达到从中提取出有价值的知识的目的,并用合适的方式将结果展现给终端用户^[8]。对电力企业来讲,电力大数据的基本处理流程与传统数据处理流程并无太大差异,主要区别在于:电力大数据需要处理大量、非结构的数据,所以在各个环节都可以采用 MapReduce 等方式进行并行处理^[14]。

图3所述的电力大数据平台总体架构,应与电力系统中来源丰富的数据流结合,催生具体的产业应用。电力系统是一种高维非线性的复杂系统,其

内部的数据流包含电力流、信息流、业务流、故障流、气象流等不同的数据流向。图4描述了在电力系统总体平台架构之上的电力企业商业应用与电力系统内部数据流的可能结合点,包括发、输、变、配、用、调等环节的负荷控制系统、管理信息系统(management information system, MIS)、监测控制和数据采集(supervisory control and data acquisition, SCADA)系统、电能计量系统、风电光伏功率预测系统、电力设备在线监测系统等。

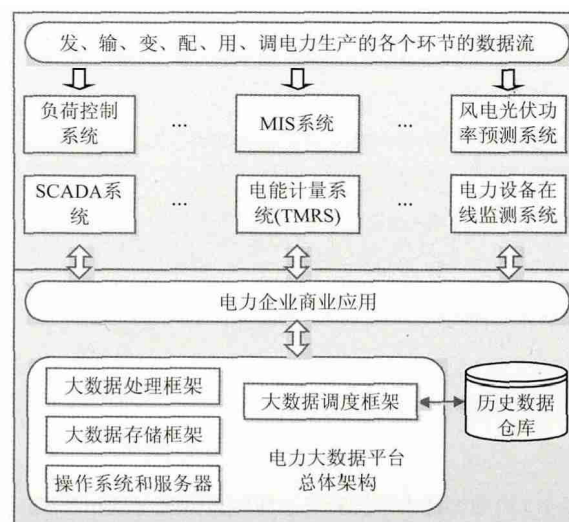


图4 电力大数据平台架构与电力系统数据流的结合点

Fig. 4 Link between universal framework of big data platform and data flow of power system

对这些电力系统的子信息源而言,一方面可以单独应用大数据技术,提升其产业价值,例如电力设备在线监测系统本身就是一个大数据系统,另一方面可以融合不同的子信息源,在更高的层面上构件大数据平台,例如融合电能计量系统、SCADA系统、MIS系统、负荷控制系统,可以构建基于大数据平台的网损分析系统,实现网损的自动统计与分析。基于该网损分析大数据平台,还可开展基于自动网损统计的用户窃电行为挖掘,实现更深层次的应用。

3 电力大数据关键技术

3.1 电力大数据的集成管理技术

电力企业数据集成管理技术是合并来自2个或者多个应用系统的数据,创建一个具有更多功能的企业应用的过程。从集成的角度来说,就是把不同来源、格式、特点、性质的数据在逻辑上或者存储介质上有机地集中,为系统存储一系列面向主题的、集成的、相对稳定的、反映历史变化的数据集,从而为系统提供全面的数据共享。电力企业集

成管理技术就是为解决电力企业内部各系统间的数据冗余和信息孤岛而产生的^[18]。

电力大数据的数据集成管理技术，包含关系型和非关系型数据库技术、数据融合和集成技术、数据抽取技术、过滤技术和数据清洗等^[10]。大数据的一个重要特点就是多样性，这就意味着数据来源极其广泛，数据类型极为繁杂，这种复杂的数据环境给大数据的处理带来极大的挑战，要想处理大数据，首先必须对数据源的数据进行抽取和集成，从中提取出实体和关系，经过关联和聚合之后采用统一的结构来存储这些数据，在数据集成和提取时需要数据清洗，保证数据质量及可靠性^[8]。

大数据存储管理中一个重要的技术是 NoSQL 数据库技术，它采用分布式数据存储方式，去掉了关系型数据库的关系型特性，数据存储被简化且更加灵活，具有良好的可扩展性，解决了海量数据的存储难题。有代表性的 NoSQL 数据库技术有 Google 的 BigTable 和 Amazon 的 Dynamo 等。

3.2 电力大数据的数据分析技术

大数据技术的根本驱动力是将信号转化为数据，将数据分析为信息，将信息提炼为知识，以知识促成决策和行动^[12]。借助电力大数据的分析技术可以从电力系统的海量数据中找出潜在的模态与规律，为决策人员提供决策支持。麦肯锡认为可用于大数据分析的关键技术源于统计学和计算机科学等学科，包含关联分析、机器学习、数据挖掘、模式识别、神经网络、时间序列预测模型、遗传算法等多种不同的方法^[19]。

大数据研究不同于传统的逻辑推理研究，是对巨大数量的数据做统计性的搜索、分类、比较、聚类等的分析和归纳，因此继承了统计科学的一些特点，如统计学关注的变量相关性或称关联性，所谓“相关性”是指 2 个或 2 个以上变量的取值之间存在某种规律性^[7]。“相关分析”的目的是找出数据集里隐藏的关系网，一般用支持度、可信度、兴趣度等参数反映相关性^[7]。

牛津大学网络学院教授维克多在他的著作《大数据时代》中指出：在大数据时代，相关关系分析法将大放异彩，通过找到一个良好现象的关联物，相关关系分析可以捕获现在和预测未来。大数据相关关系分析法，建立在海量样本的基础上，不采用随机分析法这样的捷径，而采用分析所有数据的方法；大数据的简单算法比小数据的复杂算法更有效，其结果更快、更准确而且不易受到干扰，因此

他指出建立在相关关系分析法基础上的预测是大数据的核心。

大数据这种不注重因果关系侧重于相关关系的分析方法，带来了科学研究思维方式的重大转变，已故图灵奖得主吉姆·格雷提出的数据密集型科研“第四范式”，将大数据科研从前 3 种范式(理论科学、计算科学、实验科学)中分离出来，单独作为一种科研范式，正是因为其研究方式不同于基于数学模型的传统研究方式^[7]。

电力大数据分析技术，从根本上来讲，属于传统数据挖掘技术在海量数据挖掘下的新发展，但由于大数据海量、高速增长、多样性的特点，并且不仅包含结构化数据，还含半结构化和非机构化数据，因此传统的很多处理小数据的数据挖掘方法已经不再实用^[20]。大数据环境下的数据挖掘与机器学习算法，可以从 3 个方面着手：1) 从大数据的治理与抽样、特征选择的角度入手，将大数据小数据化^[21-22]；2) 开展大数据下的聚类、分类算法研究，例如基于共轭度的最小二乘支持向量机(least squares support vector machine, LS-SVM)^[23]，随机可扩展 Fuzzy C-Means (FCM)^[24]等；3) 开展大数据的并行算法，将传统的数据挖掘方法通过并行化，应用到大数据的知识挖掘中，例如基于 MapReduce 的机器学习与知识挖掘^[25-26]。

3.3 电力大数据的数据处理技术

电力大数据的数据处理技术包括分布式计算技术，内存计算技术，流处理技术等，这 3 种技术适用的对象和解决的主要问题如图 5 所示。分布式计算技术是为了解决大规模数据的分布式存储与处理。内存计算技术是为了解决数据的高效读取和处理解决在线实时计算。流处理技术则是为了处理实时到达的、速度和规模不受控制的数据。

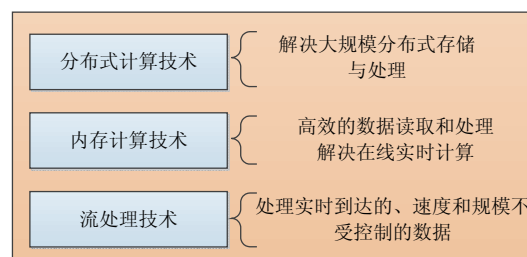


图 5 大数据处理技术适用的对象

Fig. 5 Data processing techniques of electric power big data

分布式计算是一种新的计算方式，研究如何将一个需要强大计算能力才能解决的问题分解为许多小的部分，然后再将这些部分分给多个计算机处

理,最后把结果综合起来得到最终结果。分布式计算的一个典型代表是 Google 公司提出的 MapReduce 编程模型,该模型先将待处理的数据进行分块,交给不同的 Map 任务区处理,并按键值存储到本地硬盘,再用 Reduce 任务按照键值将结果汇总并输出最终结果。分布式技术适用于电力系统信息采集领域的大规模分散数据源。

内存计算技术是将数据全部放在内存中进行操作的计算技术,该技术克服了对磁盘读写操作时的大量时间消耗,计算速度得到几个数量级的大幅提升。内存计算技术伴随着大数据浪潮的来临和内存价格的下降得到快速的发展和广泛的应用,EMC、甲骨文、SAT 都推出了内存计算的解决方案,将客户以前需要以天作为时间计算单位的业务降低为以秒作为时间计算单位,解决了大数据实时分析和知识挖掘的难题。

流处理的模型是将源源不断的数据组视为流,当新的数据到来时就立即处理并返回结果,其基本理念是数据的价值会随着时间的流逝而不断减少,因此尽可能快地对最新的数据做出分析并给出结果,其应用场景主要有网页点击的实时统计、传感器网络、金融中的高频交易等^[8]。随着电力事业的发展,电力系统数据量不断增长,对实时性的要求也越来越高,将数据流技术应用于电力系统可以为决策者提供即时依据,满足实时在线分析需求^[27]。

3.4 电力大数据的数据展现技术

电力大数据的数据展现技术包括可视化技术、空间信息流展示技术、历史流展示技术等^[10]。可视化技术、空间信息流展示技术、历史流展示技术从3个不同的方面诠释了电力大数据展现技术的丰富内涵。借助电力大数据的数据展现技术可帮助管理人员更直观、准确地理解电力系统数据表达的意义,了解电力系统的运行状态。

可视化技术广泛应用于电网状态的实时监控,显著提高了电力系统的自动化水平。未来电力系统可视化还可结合复杂网络中的相关理论在电网自动分层分区、自动布点等方面展开深入研究,发掘电网更深层次的规律和联系^[28]。

空间信息流展示技术主要体现在电网参数与已有地理信息系统的结合上,包含变电站三维展示、虚拟现实等技术^[28-29]。将电力配电设备管理与地理信息系统紧密结合起来,有利于电网管理人员直观地了解设备情况,从而为其决策提供最新的地

理信息^[30-31]。在变电站工程设计中用空间信息流展示技术可以节约时间、资源、成本,为电力企业带来巨大的效益。

历史流展示技术体现在对电网历史数据的管理与展示上^[32]。在电力系统中,深层次的应用分析往往以历史数据为基础。对生产现场的实时监测数据、电网的规划数据和负荷预测数据,通过历史流展示技术,可以绘制出数据的发展趋势并预测出未来的数据走势;通过历史流回放展示技术,可以模拟历史重大事件发生、演变,挖掘历史事件潜在的知识与规律。

4 大数据在智能电网中的应用案例

电力大数据的应用场合涵盖发、输、变、配、用、调等电力行业的各个环节,在风电场选址、降低网损、风电并网、电网安全监测、大灾难预警、电力企业精细化运营、电力设备状态监测等领域有非常强的可实现性。随着智能电网建设的进一步推进,大数据技术在智能电网中将发挥越来越大的作用。下面通过几个典型应用案例进一步阐述大数据在智能电网中的应用前景。

1) IBM 大数据技术在新能源接入中的应用。

在电力生产环节,随着新能源大量接入,打破了相对静态的传统电力生产,使得电力生产的管理和计量变得日趋复杂^[10]。大数据技术能为电力企业做出更好的预测。

丹麦的维斯塔斯风力技术集团,通过在世界上最大的超级计算机上部署 IBM 大数据解决方案,得以通过分析包括 PB 量级气象报告、潮汐相位、地理空间、卫星图像等结构化及非结构化的海量数据,从而优化了风力涡轮机布局,提高风电发电效率。

IBM 公司针对风电企业在风电场微观选址中面临的挑战,提出基于高精度数值天气预报的微观选址解决方案,来解决风资源捕捉利用的最大化问题和风机维护成本的最小化问题^[33]。IBM 微观选址方法如图6所示。该选址方案结合风资源精细化评估、气象灾害风险评估、风机经济效益评估、风场施工标准评估等因素,通过基于先进成熟的高精度数值天气模型,能够考虑更多的大气动力过程和物理过程,大幅提高模型的精确度和可靠性,从而把风电场备选区域和整个大气有机耦合在一起,做到更长期、全面地分析备选区域各个点的四维风资源分布情况,以及风资源的季节、年、年代变化情况,从而避免由单点推断整个风电场的资源分布,或由

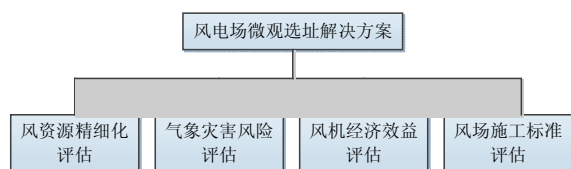


图6 风电场微观选址

Fig. 6 Micro-siting solution for location selection of wind farm

于使用时间过短的测风塔资料而无法刻画年和年代变化所带来的偏差问题^[33-34]。

2) 大数据技术在风电机组安全状态评估中的应用。

风电机组运行环境非常恶劣，受雷雨、盐雾、冰雪等因素的影响^[35]。采用基于大数据挖掘技术的风电机组安全经济运行状态综合评估系统，监测零部件磨损、疲劳等原因引起的状态变化信息，并由此识别和预测风电机组设备或者零部件的故障，提高机组的运行安全性，避免早发故障导致的更为严重的故障，并降低运维费用。基于大数据分析的风电机组潜在故障早期预警流程图如图7所示^[35]。该预测系统以设备历史数据和在线实时监测数据为基础，通过设备属性配置定义创建设备超球模型，并用该模型在在线超球模型中评估设备状况。在线评估模型的输出是相似度曲线，用于关联点排序、状态预警和测点预测值的预测。

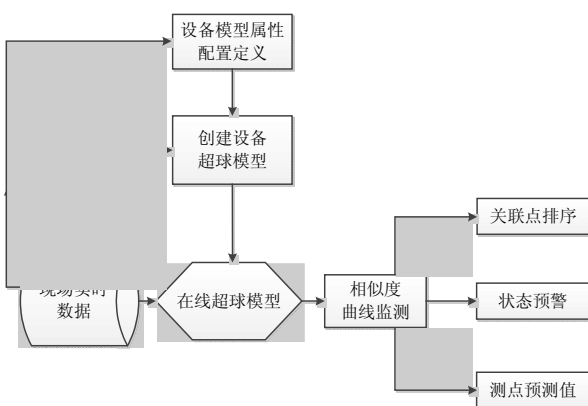


图7 风电机组潜在故障早期预警图

Fig. 7 Flowchart of the condition monitoring and assessment of wind turbine

3) 大数据技术在电网灾难预警中的应用。

随着电网日益增加的复杂性和不断变化的自然环境，电力系统中的灾难性连锁事故频繁发生，这些灾难性连锁事故大多数始于系统某个元件故障^[36]。大规模停电事故初期往往是少量元件相继故障，在事故扩大阶段则与电力系统中的脆弱环节有紧密的联系，因此从整体预防的角度出发，通过大数据技术辨识电力网络中的脆弱环节对提高电力

系统的可靠性，降低大规模停电事故的发生概率有重要意义^[37-38]。一个典型的基于大数据的电网灾难预警系统框架如图8所示^[38]。该框架以电网拓扑数据、地理信息数据、电网运行数据为数据基础，在算法模型层通过判据指标计算获得判据指标库，通过原因寻找引擎获得原因分析树，通过评价指标计算获得评价指标库。以此为基础，在业务层通过薄弱区域识别、薄弱区域原因分析，并联合判据层的评价指标库，给出综合预警结果^[38]。

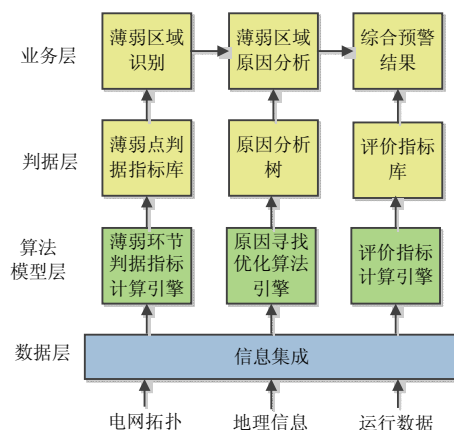


图8 基于大数据的电网灾难预警系统框架

Fig. 8 Disaster warning system framework of power system based on big data

5 结论与展望

本文探讨了大数据、云计算、智能电网三者之间的相互关系，并给出电力大数据的总体可执行框架，对电力企业大数据平台的搭建具有参考价值。本文阐述了电力大数据平台中4个核心的关键技术，即集成管理技术、分析技术、处理技术、展现技术。大数据的集成管理技术应该考虑关系型和非关系型数据库技术的综合应用，并重点考虑 NoSQL 数据库技术；大数据的分析技术应从大数据的治理与抽样、大数据特征选择、大数据小数据化、大数据的分类算法、大数据挖掘并行化等方面开展研究；大数据的处理技术应根据具体的应用需求考虑分布式计算、内存计算、流处理等技术；电力大数据的数据展现技术可考虑可视化技术、空间信息流展示技术、历史流展示技术等。文章最后的典型应用案例，对电力企业开展相关研究具有参考价值。

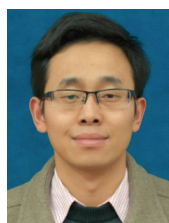
大数据技术可以在5个方面创造价值，分别是：创建透明度，让利益相关方更加容易的获取信息；通过试验来发现需求，暴露可变因素并提高业绩；根据客户需求细分人群；通过自动化算法替换或者支持人为决策；创新商业模式、产品和服务。大数

据技术在商业领域已经获得较为广泛的应用并创造出巨大的商业价值,但是在电力系统中的应用才刚刚起步,因此结合大数据的技术优势和电力系统的应用需求,发挥电力大数据的价值,将为智能电网的建设带来新的发展契机。电力企业应该牢牢抓住这个契机,从数据政策、人才培养、关键技术研发等层面,全面促成电力大数据技术的发展。

参考文献

- [1] Gartner. Top ten strategic technology trend for 2012 [EB/OL]. (2011-11-05)[2014-08-17]. <http://www.gartner.com>.
- [2] Birney E. The making of ENCODE: lessons for big-data projects[J]. Nature, 2012(489): 49-51.
- [3] 张东霞, 苗新, 刘丽萍, 等. 智能电网大数据技术发展研究[J]. 中国电机工程学报, 2015, 35(1): 2-12.
Zhang Dongxia, Miao Xin, Liu Liping, et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE, 2015, 35(1): 2-12(in Chinese).
- [4] 张素香, 赵丙镇, 王风雨, 等. 海量数据下的电力负荷短期预测[J]. 中国电机工程学报, 2015, 35(1): 37-42.
Zhang Suxiang, Zhao Binzhen, Wang Fengyu, et al. Short-term power load forecasting based on big data[J]. Proceedings of the CSEE, 2015, 35(1): 37-42(in Chinese).
- [5] UN Global Pulse. Big data for development: challenges & opportunities[EB/OL]. (2012-11-05)[2014-08-10]. <http://www.unglobalpulse.org/projects/Bigdata Development>.
- [6] Agrawal D, Bernstein P, Bertino E, et al. Challenges and opportunities with big data[EB/OL]. (2012-02-01)[2014-10-15]. <http://www.cra.org/ccc/resources/ccc-led-white-papers>.
- [7] 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(6): 647-657.
Li Guojie, Chen Xueqi. Research status and scientific thinking of big data[J]. Bulletin of the Chinese Academy of Sciences, 2012(6): 647-657(in Chinese).
- [8] 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(in Chinese).
- [9] 李贵兵, 罗洪. 大数据下的智能数据分析技术研究[J]. 科技资讯, 2013(30): 11-12.
Li Guibing, Luo Hong. Intelligent data analysis under the background of big data[J]. Science & Technology Information, 2013(30): 11-12(in Chinese).
- [10] 中国电机工程学会信息化专业委员会. 中国电力大数据发展白皮书[M]. 北京: 中国电力出版社, 2013: 10-15.
- [11] 张文亮, 刘壮志, 王明俊, 等. 智能电网的研究进展及发展趋势[J]. 电网技术, 2009, 33(13): 1-11.
Zhang Wenliang, Liu Zhuangzhi, Wang Mingjun, et al. Research status and development trend of smart grid [J]. Power System Technology, 2009, 33(13): 1-11(in Chinese).
- [12] 孙柏林. “大数据”技术及其在电力行业中的应用[J]. 电气时代, 2013(8): 18-23.
Sun Bolin. Big data technology and its application in power industry[J]. Electric Age, 2013(8): 18-23(in Chinese).
- [13] 罗军舟, 金嘉晖, 宋爱波, 等. 云计算: 体系架构与关键技术[J]. 通信学报, 2011, 32(7): 3-21.
Luo Junzhou, Jin Jiahui, Song aibo, et al. Cloud computing: architecture and key technologies[J]. Journal on Communications, 2011, 32(7): 3-21(in Chinese).
- [14] 李志刚, 朱志军. 大数据: 大价值、大机遇、大变革[M]. 北京: 电子工业出版社, 2012: 182.
Li Zhigang, Zhu Zhijun. Big data: big value, big chance, big reform[M]. Beijing: Publishing House of Electronics Industry, 2012: 182(in Chinese).
- [15] 姚宏宇. 大数据与云计算[J]. 信息技术与标准化, 2013(5): 21-22.
Yao Hongyu. Big data and cloud computing [J]. Information Technology & Standardization, 2013(5): 21-22(in Chinese).
- [16] 饶威, 丁坚勇, 路庆凯. 智能电网云计算平台构建[J]. 华东电力, 2011, 39(9): 1493-1496.
Rao Wei, Ding Jianyong, Lu Qingkai. Cloud computing platform for smart grid[J]. East China Electric Power, 2011, 39(9): 1493-1496.
- [17] 赵刚. 大数据技术与应用实践指南[M]. 北京: 电子工业出版社, 2013: 56-58.
Zhao Gang. Big data technology and application practice [M]. Beijing: Publishing House of Electronics Industry, 2013: 56-58(in Chinese).
- [18] 苏文博, 林亮成, 杨景龙. 电网数据集成与管理系统研究[J]. 中国电力教育, 2008(S3): 369-370.
Su Wenbo, Lin Liangcheng, Yang Jinglong. Research on data integration and management system for power grid[J]. China Electric Power Education, 2008(S3): 369-370(in Chinese).
- [19] McKinsey & Company. Big data: the next frontier for innovation, competition, and productivity[M]. New York: McKinsey Global Institute, 2011: 1-28.
- [20] 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.
- [21] Li Y H, Maguire L. Selecting critical pattern based on local geometrical and statistical information[J]. IEEE Transactions on Pattern Analysis and Machine

- Intelligence, 2011, 33(6): 1189-1201.
- [22] Hoi C H, Wang J L, Zhao P L. Online feature selection for mining big data[C]// Proceedings of 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining. Beijing, China: ACM, 2012: 93-100.
- [23] Kim B J. A classifier for big data[C]//Proceedings of the 6th International Conference on Convergence and Hybrid Information Technology. Daejeon, Republic of Korea: ACM, 2012.
- [24] Havens T C. Fuzzy c-means algorithms for very large data[J]. IEEE Transactions on Fuzzy System, 2012, 20(6): 1130-1146.
- [25] Shim K. Algorithms for big data analysis[J]. Proceedings of the VLDB Endowment, 2012, 5(12): 2016-2017.
- [26] Zhang J B, Li T R, Pan Y. Parallel rough set based knowledge acquisition using Map Reduce from big data[C]//Proceedings of 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining. Beijing, China: ACM, 2012.
- [27] 孔英会. 数据流技术及其在电力信息处理中的应用研究[D]. 河北: 华北电力大学, 2009.
- Kong Yinghui. Study on data stream techniques and its application in electric power information processing [D]. Heibei: North China Electric Power University, 2009(in Chinese).
- [28] 郭崇军, 洪峰, 陈金富, 李勇, 徐友平, 奚江惠, 汪剑波. 可视化技术在电力系统中的应用探讨[J]. 水电能源科学, 2011, 29 (2): 146-149.
- Guo Chongjun, Hong Feng, Chen Jinfu. Application of Scientific visualization in smart grid[J]. Water Resources and Power, 2011, 29(2): 146-149(in Chinese).
- [29] 王先兵, 张学东, 何涛, 等. 三维虚拟变电站数字可视化管理与监控系统[J]. 武汉大学学报: 工学版, 2011, 44(6): 786-791.
- Wang Xianbing, Zhang Xuedong, He Tao, et al. Digital visualization management and monitoring system for 3D virtual transformer substations[J]. Engineering Journal of Wuhan University, 2011, 44(6): 786-791(in Chinese).
- [30] 白文光. 可视化供电能力计算软件在地调中的应用研究[D]. 北京: 华北电力大学, 2005.
- Bai Wenguang. Study on the application of visible power supply capability computation for district electric power dispatching center[D]. Beijing: North China Electric Power University, 2005(in Chinese).
- [31] 胡之武. 基于 GIS 电网运行数据可视化的方法[D]. 浙江: 浙江大学, 2005.
- Hu Zhiwu. Study on the visualization of power system running data based on GIS[D]. Zhejiang: Zhejiang University, 2005(in Chinese).
- [32] 陈启买, 刘海, 贺超波, 等. 实时监控数据图形展示与历史回放关键技术研究[J]. 电脑编程技巧与维护, 2009(22): 95-98.
- Chen Qimai, Liu Hai, He Chaobo, et al. Research on key technology of graph display and history playback for real time monitoring data[J]. Artificial Intelligence and Identification Techniques, 2009(22): 95-98(in Chinese).
- [33] IBM. 风电场微观选址. IBM[EB/OL]. (2014-08-17). http://www31.ibm.com/solutions/cn/industries/energy/thankyou/energy_wp.shtml.
- IBM. Power wind site down time IBM's solution in power industry[EB/OL]. (2014-08-17). http://www31.ibm.com/solutions/cn/industries/energy/thankyou/energy_wp.shtml(in Chinese).
- [34] 刘俊卿. 大数据驯服风能[J]. 中国经济和信息化, 2014(2): 99-100.
- Liu Junqing. Big data tame wind[J]. Informatization, 2014(2): 99-100(in Chinese).
- [35] 吕庭彦, 李亚冬, 蒋维, 等. 基于大数据挖掘技术的风电机组安全经济运行状态综合评估系统[C]//中国电力企业联合会议. 北京: 中国电力企业联合会, 2013: 10.
- Lu Tingyan, Li Yadong, Jiang Wei, et al. Comprehensive assessment system of safe and economic operation state for wind turbines[C]//Beijing: China Electricity Council. China Electricity Council, 2013: 10(in Chinese).
- [36] 曹一家, 陈晓刚, 孙可. 基于复杂网络理论的大型电力系统脆弱线路辨识[J]. 电力自动化设备, 2006, 26 (12): 1-5.
- Cao Yijia, Chen Xiaogang, Sun Ke. Identification of vulnerable lines in power grid based on complex network theory[J]. Electric Power Automation Equipment, 2006, 26 (12): 1-5(in Chinese).
- [37] 白加林, 刘天琪, 曹国云, 等. 电力系统脆弱性评估方法综述[J]. 电网技术, 2008(S2): 26-30.
- Bai Jialin, Liu Tianqi, Cao Guoyun, et al. A survey on vulnerability assessment method for power system [J]. Power System Technology, 2008(S2): 26-30(in Chinese).
- [38] IBM. IBM 智慧电力解决方案(电网部分)[EB/OL]. (2014-09-20). IBM. http://www.ibm.com/smarterplanet/cn/zh/smart_grid/ideas/index.html.
- IBM. IBM smart grid solution (power grid) [EB/OL]. (2014-09-20). http://www.ibm.com/smarterplanet/cn/zh/smart_grid/ideas/index.html (in Chinese).



彭小圣

收稿日期: 2014-09-30.

作者简介:

彭小圣(1983), 男, 博士, IEEE 会员, 主要研究方向为电力系统主设备智能监测与控制维护、局部放电在线监测与模式识别、粗糙集理论为基础的数据挖掘, XiaoshengPeng@hust.edu.cn;

邓迪元(1992), 男, 硕士研究生, 主要研究方向为电力系统主设备智能监测。

(编辑 李泽荣)