ELSEVIER

BIG DATA APPLICATION IN POWER SYSTEMS

EDITED BY
REZA ARGHANDEH
YUXUN ZHOU

Big Data Application in Power Systems

Big Data Application in Power Systems

Edited by

Reza Arghandeh

Assistant Prof. in Electrical Engineering,
Department of Electrical & Computer Engineering,
Florida State University

Yuxun Zhou

PhD candidate, Department of Electrical Engineering
and Computer Sciences, UC Berkeley

ELSEVIER

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-811968-6

Publisher: Joe Hayton
Acquisition Editor: Lisa Reading
Editorial Project Manager: Ana Claudia A. Garcia
Production Project Manager: Vijayaraj Purushothaman
Cover Designer: Mark Rogers

Typeset by SPi Global, India

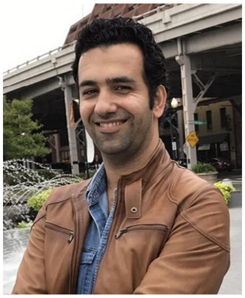
Contents

Contributors

Reza Arghandeh UC Berkeley and Florida State University, Tallahassee, FL, United States
Mohammad Babakmehr Colorado School of Mines, Golden, CO, United States
Ricardo J. Bessa INESC Technology and Science—INESC TEC, Porto, Portugal
Saverio Bolognani Automatic Control Laboratory ETH Zürich, Zürich, Switzerland
Angelo Cenedese University of Padova, Padova, Italy
Michael Chertkov Los Alamos National Laboratory, Los Alamos, NM, United States
Deepjyoti Deka Los Alamos National Laboratory, Los Alamos, NM, United States
Roy Dong University of California, Berkeley, Berkeley, CA, United States
Feng Gao Tsinghua University Energy Internet Research Institute, Beijing, China
Madeleine Gibescu Eindhoven University of Technology, Eindhoven, The Netherlands
Bri-Mathias Hodge National Renewable Energy Laboratory, Golden, CO, United States
Gabriela Hug ETH Zurich, Power Systems Laboratory, Zurich, Switzerland
Jeffrey S. Katz IBM, Hartford, CT, United States
Stephan Koch ETH Zurich; Adaptricity AG, c/o ETH Zurich, Power Systems Laboratory, Zurich, Switzerland
Hanif Livani University of Nevada Reno, Reno, NV, United States
Mehrdad Majidi University of Nevada, Reno, NV, United States
John D. McDonald GE Energy Connections-Grid Solutions, Atlanta, GA, United States
Sadaf Moaveninejad Polytechnic University of Milan, Milan, Italy
Elena Mocanu Eindhoven University of Technology, Eindhoven, The Netherlands
Ingo Nader Unbelievable Machine, Vienna, Austria
Behzad Najafi Polytechnic University of Milan, Milan, Italy
Phuong H. Nguyen Eindhoven University of Technology, Eindhoven, The Netherlands
Lillian J. Ratliff University of Washington, Seattle, WA, United States
Fabio Rinaldi Polytechnic University of Milan, Milan, Italy
Marcelo G. Simoes Colorado School of Mines, Golden, CO, United States
Matthias Stifter AIT Austrian Institute of Technology, Center of Energy, Vienna, Austria
Carol L. Stimmel Manifest Mind, LLC, Canaan, NY, United States
Gian Antonio. Susto University of Padova, Padova, Italy
Akin Tascikaraoglu Mugla Sitki Kocman University, Mugla, Turkey
Matteo Terzi University of Padova, Padova, Italy
Andreas Ulbig ETH Zurich; Adaptricity AG, c/o ETH Zurich, Power Systems Laboratory, Zurich, Switzerland

Yang Weng Arizona State University, Tempe, AZ, United States
Rui Yang National Renewable Energy Laboratory, Golden, CO, United States
Yingchen Zhang National Renewable Energy Laboratory, Golden, CO, United States
Jie Zhang University of Texas at Dallas, Richardson, TX, United States
Yuxun Zhou UC Berkeley and Florida State University, Tallahassee, FL, United States
Thierry Zufferey ETH Zurich, Power Systems Laboratory, Zurich, Switzerland

About the Editors



Reza Arghandeh is an Assistant Professor in the ECE Department in Florida State University. He is director of the Collaborative Intelligent Infrastructure Lab. He has been a postdoctoral scholar at the University of California, Berkeley's California Institute for Energy and Environment 2013–15. He has 5 years industrial experience in power and energy systems. He completed his PhD in Electrical Engineering with a specialization in power systems at Virginia Tech. He holds Master's degrees in Industrial and System Engineering from Virginia Tech 2013 and in Energy Systems from the University of Manchester 2008. From 2011 to 2013, he was a power system software designer at Electrical Distribution Design Inc. in Virginia. Dr. Arghandeh's research interests include, but are not limited to, data analysis and decision support for smart grids and smart cities using statistical inference, machine learning, information theory, and operations research. He is a recipient of the Association of Energy Engineers (AEE) Scholarship 2012, the UC Davis Green Tech Fellowship 2011, and the best paper award from the ASME 2012 Power Conference and IEEE PESGM 2015. He is the chair of the IEEE Task Force on Big Data Application for Power Distribution Network.

Yuxun Zhou is currently a PhD candidate at Department of EECS, UC Berkeley. Prior to that, he obtained the Diplome d'Ingenieur in applied mathematics from Ecole Centrale Paris and a BS degree from Xi'an Jiaotong University. Yuxun has published more than 30 refereed articles, and has received several student awards. His research interest is on machine learning theories and algorithms for modern sensor rich, ubiquitously connected cyber-physical systems, including smart grid, power distribution networks, smart buildings, etc.

Preface: Objective and Overview of the Book

The term "big data" is fairly new in power systems. Yet, its application and methodologies applied to massive data sets were developed a long time ago for electricity load consumption forecasting. The recent developments in monitoring, sensor networks, and advanced metering infrastructure (AMI) dramatically increase the variety, volume, and velocity of measurement data in electricity transmission and distribution networks. Moreover, the progress in advanced statistics, machine learning (ML), database structure, and data mining methodologies marked by increasing the availability of open source platforms for data analytics is transforming the power system area and turning utilities into data-driven enterprises.

In order to discuss the big data analytics applications for power systems, this book brings together experts from all organizations and institutions impacted including academia and industry. We focus on rapidly modernizing monitoring and analytical approaches to process the high dimensional, heterogeneous, and spatiotemporal data. This book discusses challenges, opportunities, success stories, and pathways for utilizing big data value in smart grids. The dramatic change in the field of scientific computing, microprocessors, and data communications is a burden for electric utilities to understand, follow, and adopt the advanced statistics, computer science, and mathematics concepts. Today's utility engineers need to be more informed of the basic concepts and applications for massive field data analysis. This book's goal is to facilitate the transition to data-driven utilities by providing a comprehensive view on big data issues, methodologies, and their various applications in the power systems area.

Much like the authorship of the chapters in this volume, the intended audience for this book extends from researchers, graduate students, and faculty working in electricity networks and smart grid area to industrial scientists, engineers, data analysis experts, and software developers who are working on electricity networks and advanced technologies for smart grids. This book is also useful for people with less technical expertise in scientific computing. We expect that the reader will have some proficiency in power systems fundamentals and that he/she has had at least one elementary course in statistics. This book can also be useful for senior undergraduate students who have passed courses on power systems.

This book has three sections as follows: I. *Harness the Big Data From Power Systems*, II. *Harness the Power of Big Data*, and III. *Put the Power of Big Data Into Power Systems*. The opening section is an overview of the opportunities and challenges for data-driven utilities in the era of distributed technologies and resources such as Internet of Things (IoT), flexible demand, distributed generation, and energy storage. The second section reviews research trends on ML and artificial intelligence for the power system industry. The final section provides examples of the advanced data analytic applications for the grid operation. Taken together, these three book sections provide an overview of the entire cycle of data analysis in power systems. The book begins with the utility enterprise structure, business model, and privacy issues, then delves into research trends in advanced data analysis, and ends full circle with real-world examples of actual applications of data analytics used daily by utilities.

SECTION ONE: HARNESS THE BIG DATA FROM POWER SYSTEMS

To provide a big picture for electric utilities, this section describes the current and future trends for data mining and data processing in electric utilities. The move toward data-driven utility is possible by a fundamental shift in organizational culture and business processes, as well as data-related technology and practices. Moreover, enriching electric utilities with data requires interoperability across all operational and enterprise units and recognition that maintaining the data privacy, security, and the seamless data flow is highly challenging. The interoperability in holistic data-driven utilities expands to customers through their engagement and continues demand-side management for higher reliability, service quality, and efficiency. Aligning customers' needs and expectations with utilities' business drivers will shape the roadmap to generate, process, and access the data in utilities.

The information and communication technology (ICT) platform is at the heart of the roadmap to data-driven utilities which supports the data flow from customers all the way to the transmission and generation operators. Utilities made aggressive steps toward smartness by adopting the distribution automation (DA) solutions followed by AMI platforms. DA and AMI made a revolution in grid operation. However, the data flood from DA and AMI has created a nightmare for the utilities' ICT infrastructure. A holistic approach for data-driven utilities is needed to openly discuss and clarify the foundational ICT requirements to serve all functions of the electric grid.

Becoming a data-driven utility is inevitable in the age of internet, cloud computing, smart phones, and distributed resources. The advanced data analytics make continuous innovation possible by unlocking insights never seen before. The ML, deep learning, and statistical inference are tools that help utilities to keep up with the torrent of data from different resources. Advanced big data analytics provide estimation, predication, diagnostics, and prognostics conclusions from historical and real-time data flows. As more data becomes available to utilities over time, the ML algorithms provide more refined insights on grid operation planning. However, the synergies between ICT networks, grid components, operators, and customers run the power system into a complex giant for ad hoc data-driven approaches and policies. This section of the book endeavors to deliver the message that a holistic approach based on a foundation of open architecture and standards will ensure the open flow of data and interoperability between devices, systems, databases, and people in order to make data-driven utilities. The all-inclusive approaches to generate, transfer, and handle data also bring tremendous opportunities to break traditional barriers in utility organizations for delivering safe, reliable, and affordable power to their customers.

Chapter 1 by John McDonald introduces these concepts through three case studies. He explores the value of a data-driven utility in terms of asset management and safety, the fundamentals of standards and interoperability, and the enterprises of increased visibility into the transmission and distribution network. The chapter illustrates the holistic data-driven utility and its fundamental business drivers to establish information and communications technology foundation, human resource, customer relation, and data-oriented organizational cultural data in the grid operation and planning.

The data-driven utilities now face greater and more frequent risk of intrusion and/or interruption due to the fact that these networks are merging with cyber networks, resulting in sociotechnical and cyber-physical systems that are creating an infrastructural IoT where all grid components can interact and collaborate. Integrating cyber components into the electric grid also means an incredible increase in security vulnerability and interdependencies among infrastructure components that create the risk of cascading effects after attacks. Moreover, the enhanced observability of the grid thanks to the smart meters' high granular data is making more customers concerned and uncomfortable about data privacy. Carol L. Stimmel discusses state-of-the-art data privacy and security in Chapter 2. She lists a number of actual cases for cyber security attacks on the grid and explains the impact of data-driven approaches in enhancing the data security and privacy. Data-driven utilities function as much more than the operators of the physical grid; utilities are also responsible for massive enterprise systems with financial information, customer data, and a growing network of digital operations under human control. Thus, security strategies must become more nuanced and complex, and should include privacy and other internal information technology controls.

The big data era is changing the utility workforce paradigm. Several major utilities are adding more software developers and data scientists to their R&D and operating groups, as well as power system experts. The power of data in innovation is seen in more smart grid projects and AMI implementations. Some of these projects applied Big Data and Analytics even without adding any new sensors, demonstrating the power of knowing more about what information was already available to the utility through the SCADA systems. The utility innovation movement came from the foresight that discarded data may prove useful. This includes data discarded during the process of developing an analytics strategy, including predictive maintenance programs, thought to be valuable as the design phase began, even though there was no known need for all of the data at the time. Analytics has moved from replicating alarm limits already available, to deep learning for customer behavioral studies and cognitive computing for renewable adoption optimization, as well as numerical methodologies for dynamic electricity market forecasting. Jeffrey Katz from IBM contributes Chapter 3, "The Rule of Big Data and Analytics in Utilities Innovation" that explains how data analytics pave the ground for innovation in utilities by pointing to a number of successful projects in different utilities.

To harvest the advantages of big data, utilities need to employ platforms that can handle high volume, velocity, and volatility of the data. There are commercial and ready-to-use platforms that serve the big data community. It is time for utilities to take the lead in shaping power systems-specific data platforms. The in-memory calculation engine and parallel computing framework, Hadoop/MapReduce and Spark, are ready for handling an extremely large scale of dataset; on the other hand, the stream processing engine, Storm, Streams, and Spark Streaming are built to analyze data in motion and act on information as it is happening. The architecture of big data platforms includes data integration, warehousing, analytics, and combining the demand of smart grids to put forward a set of frameworks such as the Apache Hadoop ecosystem which has excellent computing ability and can adapt to various business requirements. Chapter 4 "Frameworks for Big Data Integration, Warehousing, and Analytics" by Feng Gao discusses different tools and techniques to support the growth of smart grid and big data with high performance computing, with a focus on the platform, data integration, warehousing, and analytics that are particularly adaptive to handle a variety of characteristics of energy industry data within the data lifetime cycle.

SECTION TWO: HARNESS THE POWER OF BIG DATA

This theory-oriented section focuses on big data analytics. In particular, it discusses ML and data mining algorithms, methods, and implementation that are adaptable for data visualization, representation, exploratory analysis, regression, and pattern recognition in power systems. The objectives of this section are twofold. On one hand, both classical and status quo ML paradigms are reviewed and discussed, motivating the proper usage of traditional supervised/unsupervised learning tools and the recent developments of semi-supervised learning, multitask, multiview learning, sparse representation, deep learning, etc., for various tasks in power systems. The hope is that the dramatic progress in ML can be fully harnessed to reform the solution of power system state estimation, load forecasting, event detection, and structure identification. On the other hand, the reversed direction, i.e., the challenges and new problems brought by power system data to ML, is discussed. Similarly to the impact of computer vision, natural language processing, speech recognition, or robot control on the advancement of ML, it is expected that the complexity of the interconnected system, the behavior-related data generating process, as well as the unique sensing and measurement techniques in power systems, would inspire novel theoretical and methodological results for ML.

It is worth pointing out that in this section, the term ML is used in a broader sense, generally referring to a task to improve some performance metric, by executing a series of computation (algorithm) with some training experience (in the form of collected sensor measurement, expert knowledge, survey entries, etc.). Lying at the crossroads of statistics, computer sciences, artificial intelligence, and applied mathematics, the ML methods discussed in this section deserve a comprehensive description from diverse perspectives, including, but not limited to, their underlying probabilistic assumption, theoretical/empirical generalization performance, model selection (hyper-parameter selection), computational complexity, numerical implementation, etc. Although a mathematically rigorous treatment of the above topics is not the focus of this book, useful references are provided to interested readers. More often than not, the proper usage of the state-of-the-art ML algorithm, or a desire to advance ML driven by power system applications, would surprisingly progress both research fields.

More specifically, Chapter 5 starts with a brief discussion of classical supervised and unsupervised learning paradigms. The focus is not to give an extensive review of the field, which is impossible due to its many ramifications, but rather to equip the readers with popular approaches for regression, classification, dimension reduction, among other fundamentals. The chapter then focuses on two important issues, feature engineering and model selection, in some depth to demonstrate the proper usage and systematic tuning of those off-the-shelf ML tools. The rest of this chapter is devoted to the introduction of some recent schemes of ML that seem promising for power system data analysis applications. The topics discussed include semi-supervised learning, multitask learning, transfer learning, multiview learning, information representation, etc.

Following the discussion, Chapter 6 provides a case study on the use of the clustering algorithms for enhanced visibility of the electrical distribution system. Based on smart meter data of more than 30,000 loads in the city of Basel, Switzerland, the authors demonstrate the power of exploratory data analysis using unsupervised learning methods, which successfully reveals hidden structure, property, and geographical consistency from the measurement data. The rich information mined from this analysis can be leveraged by DSOs to support the grid operation.

The rest of the chapters in this section discuss in detail several advanced ML methods for power system applications. Motivated by the unprecedented high volumes of data made available by the growth of home energy management systems and AMI, Dr. Mocanu et al. in Chapter 7 present the deep learning framework to automatically extract knowledge and use it to improve grid operation. The chapter starts with a moderate introduction to the most well-known deep learning concepts, such as deep belief networks and high-order restricted Boltzmann machine, followed by a discussion on their theoretical advantages and limitations, such as computational requirements, convergence, and stability. As a concrete application, two case studies involving building energy prediction using supervised and unsupervised deep learning methods are presented. The chapter concludes with a glimpse into future trends highlighting some open questions as well as new possible applications.

Chapter 8 “Compressive Sensing for Power System Data Analysis,” focuses on the applications of another state-of-the-art ML framework, namely compressive sensing-sparse recovery (CS-SR), which has enjoyed great success in other fields like bio-engineering, signal processing, and computer vision, among others. The adaptation of CS-SR in smart power networks monitoring, data analysis, security, and reliability should expect similar successes. The sparse nature of the electrical power grids, as well as electrical signals, can be exploited to introduce alternative mathematical formulations to address some of the most challenging system modeling, that of sparse identification problems in power engineering. The chapter begins with a concise presentation on the theoretical and technical background of CS-SR. Next, the discussion moves to innovative CS-SR applications in smart grid technology. Finally, the CS-SR techniques are explored in depth to propose novel methods for distribution system state estimation (DSSE), single and simultaneous fault location in smart distribution, and transmission networks, and partial discharge (PD) pattern recognition.

The rapid advancement of sensing and measurement technology in power systems has given researchers access to real-time records of system dynamic states. In particular, development of phasor measurement unit (PMU) technology has allowed the continuous monitoring of the transmission line and the connected power systems, and can be complemented with utility monitoring devices, smart meters, and insulation monitoring units to build a thorough picture of the whole grid structure, health, and dynamic behavior. The data collected from these real-time measuring procedures is usually in the form of time series (TS). Hence, in Chapter 9 of this section, Dr. Gian Antonio Susto et al. present an overview about the most recent ML techniques used for TS pattern recognition. The chapter first summarizes existing methods of TS classification and highlights the issue of computational complexity, and then provides discourse on the various dimension reduction and numerosity reduction techniques for a more parsimonious and informative representation of TS data. The chapter concludes with a comprehensive comparison of diverse classification methods in terms of their underlying assumption, performance, computational complexity, flexibility for decentralized execution, and other categories.

SECTION THREE: PUT THE POWER OF BIG DATA INTO POWER SYSTEMS

This final section of the book presents the data-driven approaches unique to the design, operation, and planning of utilities. Moreover, data-driven utilities need new business models for knowledge extraction from data. Some examples are analysis of the demand response (DR) potential of grid users, big data preprocessing from grid sensors, large-scale simulation of electricity markets, and predictive maintenance of electrical equipment. Forecasting of real-time and day ahead market price, load, and renewable generation TS present huge business value for utilities' stakeholders and customers. The big data applications in the distribution and transmission networks are mainly driven by two objectives: firstly, to increase the monitoring and situational awareness capability and develop fast decision-making methods for operators, and secondly, to implement predictive active management strategies that take advantage of flexibility from various technologies in the electricity supply and demand such distributed energy resources, energy storage, and DR.

However, exploiting the full potential of big data in utilities is challenged by lack of statistics and data analytics knowledge in utilities workforce. Moreover, the "ready-to-use" and industry-level ML tools and solutions are not wildly available to utilities which may increase the learning curve and utilities' modernization time. This section provides a collection of modern data-driven solutions such as distributed learning and optimization, spatial-temporal modeling of TS, data reduction, assimilation, and visualization methods for classic power system problems including state estimation, topology detection, fault detection, and load disaggregation. The author hopes this book brings more interests in ML and deep learning applications in power system operation and planning.

Chapter 10, "An Overview of Big Data Application in Power Transmission and Distribution Networks" provides a comprehensive overview of data-driven trends such as feature extraction/reduction and distributed learning to extract knowledge from the power system and market data. Furthermore, it describes the data-driven techniques for dynamic and steady-state analysis and control of distribution and transmission systems.

In Chapter 11, "On Data-Driven Approaches for Demand Response," Akin Tascikaraoglu presents a detailed investigation of the applications and benefits of big data analytics in demand-side management or DR and their roles in providing higher saving potential for both system operators and end users. He also shows some examples of real-world implementations of DR.

Chapters 12 and 13 are devoted to topology detection. Knowledge of the exact topology, the open or closed status of switches and circuit breakers throughout the network, is essential for all aspects of the power system operation. Chapter 12, "Topology Learning in Radial Distribution Grids" presents an acquisitive algorithm to learn the grid topology using voltage measurements collected at a subset of the buses in power distribution networks. Chapter 13, "Grid Topology Identification via Distributed Statistical Hypothesis Testing," proposes an algorithm based on the identification of Markov random fields (graphical models) and conditional correlation properties that characterize voltage measurements in power distribution networks. It shows the correlation of voltage magnitude measurements in a radial distribution feeder with the topology of the grid.

In Chapter 14 entitled "Supervised Learning-Based Fault Location in Power Grid," Dr. Livani, Hanif suggests an SVM network for the classification, identification, and localization of faults in a complex power transmission grid. Based on the high-resolution/high-volume data made available by the proliferation of intelligent electronic devices (IEDs) in smart grids, this method is able to achieve efficient and accurate fault diagnosis for system operators. The lesson learned from this chapter, in particular, is to combine the effort to modify existing ML algorithms with signal processing, and to increase our knowledge about the system itself for handling new problems arising from the complex power system and grid.

To introduce cutting edge tools, packages, and information technology for readers who are interested in developing real-world power system data analysis platforms, the authors of Chapter 15 investigate the usage of recent big data tools and methods in the context of power distribution networks. This chapter illustrates the use of MapReduce functions within R or Java, which is combined with commercial distributed analytics database, the application of affinity graphs for representing collaborative filters, a performance comparison to conventional database concepts, and many other features.

Being able to forecast energy resources, load patterns, and system state are key features of next-generation smart grid technology. An accurate predictive platform would greatly benefit the planning, scheduling, and unit commitment in terms of both efficiency and security. Chapter 16 entitled "Predictive Analytics for Comprehensive Energy System State Estimation" provides an overview and a thorough discussion on predictive ML methods for wind, solar energy forecasting, load prediction, power system state estimation, etc. The ML tools included in the chapter range from classical regression, TS analysis, to kernel method such as support vector regression and Gaussian process.

Finally, Chapters 17 and 18 are devoted to a particular yet important application of big data analytics method to smart grid, namely energy disaggregation or nonintrusive load monitoring (NILM). In essence, the goal is to estimate the power usage of individual appliances from an aggregate electricity consumption measurement. Provided with more precise information including itemized energy consumption profiles, both end users and grid managers can improve their utility in terms of energy consumption prediction, demand side management, and user segmentation. Chapter 17 surveys the existing literature for background, ML methods, and possible applications of energy disaggregation, while Chapter 18 discussed the issue of privacy in the energy disaggregation framework. Both chapters are witness to the combination of cutting edge ML methods and a deep understanding of the system characteristics for the advancement of smart grid technologies.

Acknowledgments

The idea for this book goes back to a few years ago when we were analyzing smart meters and SCADA data from some Californian electric utilities using different machine learning and statistical inferences. Later on, we started to work on phasor measurement units (PMU) and micro-PMU data streams which have much more resolution than the smart meters. The PMU and power quality recording data (120 Hz to 30 kHz and beyond) plus highly spatial distributed data from smart meters marked the advent of big data in power systems. Utilities are already dealing with big data challenges considering the lack of knowledge in workforce and the lack of suitable infrastructure to handle and process the massive data. We are sure that some of our readers have a similar experience. On top of that, in the near future every house may have rooftop solar panels, controllable loads, smart appliances, electric vehicles, and various software-enabled hardware that will be more connected in the era of Internet of Things.

This book is a step toward data-driven utilities by presenting a combination of the high-level view on utility enterprise architecture, data analysis methodology, and various applications of data analytics in power transmission and distribution networks.

We have been lucky enough to have great maestros in our lives. Our parents Ali & Soodabeh Arghandeh and Yanping & Suxue Zhou, our advisers Prof. Robert Broadwater and Prof. Saifur Rahman at Virginia Tech and Prof. Costas Spanos and Prof. Alexandra von Meier at UC Berkeley.

In this book, we have a collection of highly recognized experts in academia and industry in the field of power systems and data analysis from all around the world. We would like to thank them all for their outstanding contributions. We would like to thank Dr. Heather Paudler for her valuable input on the book. We extend special thanks to Renata R. Rodrigues and Ana C. A. Garcia from the Elsevier editorial team for their countless help and advice during the different stages of preparation for this book. We also appreciate Honoka

Hamano's efforts in designing the book cover, icons for each section, and various other creative graphics inside the book.

Finally, we would like to thank several reviewers for valuable comments on preliminary drafts of this book: Jeffrey S. Katz, Ricardo Bessa, John D. McDonald, Carol L. Stimmel, Mohammad Babakmehr, Elena Mocanu, Madeleine Gibescu, Mehrdad Majidi, Gian Antonio Susto, Deepjyoti Deka, Fabio Rinaldi, Feng Gao, Han Zou, Ming Jin, Ruoxi Jia, Yingchen Zhang, Behzad Najafi, Amin Hassanzadeh, Mihye Ahn, Hanif Livani, Matthias Stifter, Saverio Bolognani, Michael Chertkov, Amirhessam Tahmassebi, Madhavi Konila Sriram, Roy Dong, and Jose Cordova.

We look forward to hearing from our readership; please contact us with any comments, suggestions, and questions.

Reza Arghandeh
Florida State University, Tallahassee, FL, United States

Yuxun Zhou
University of California, Berkeley, CA, United States

CHAPTER 1

A Holistic Approach to Becoming a Data-Driven Utility

John D. McDonald

GE Energy Connections-Grid Solutions, Atlanta, GA, United States

CHAPTER OVERVIEW

The ultimate goal of harnessing big data is to improve customer service and achieve enterprise business goals while increasing the reliability, resiliency, and efficiency of operations. Thus, business drivers should dictate data needs and the technology roadmap to achieve ongoing improvements in these areas. A data-driven utility should first identify its fundamental business drivers to understand precisely what intelligence is needed for operations and the enterprise and what specific technology supports the creation of intelligence and value, both for current business challenges and for future business needs and technology functionalities. Intelligence, and automation, relies on a two-way, integrated communication system based on standards; thus a utility must first develop a "strong" grid by establishing an information and communications technology foundation based on an open architecture and standards. This first step requires that information technology and communications groups work together to understand and support the functional requirements such as network response requirements, bandwidth, and latency, of each disparate data path—from sensor to end user—for current and future systems and applications. Then a data-driven utility should develop a "smart" grid, which requires the convergence of information technology and operations technology and their respective staffs—the beginning of an operations- and enterprise-wide cultural shift to holistic utility management that focuses on value creation and eliminates organizational silos. On the technology side, integration of data-producing devices and systems precedes automation. Determining substation automation applications relies on observing the behavior of data over time (daily, seasonally) and diverse conditions (weather patterns). On the organization side, all operations and enterprise groups should cooperate to identify their data needs to create a data requirements matrix. Information and operations technology personnel can then determine the least number of platforms and the most efficient paths to route data from device to end user, taking security into account. Access and authentication rules ensure that only the right person gets the right data at the right time. A key concept in a data-driven utility is that every internal stakeholder who can create value from data should have secure access to that data. Operational data is routed to the control center in real time, while nonoperational data is extracted from intelligent electronic devices, concentrated and sent across the operations firewall to be stored and processed in a data mart for on-demand access by enterprise groups and their applications. Three case studies illustrate the value of a data-driven utility in terms of asset management and safety, the fundamentals of standards and interoperability, and the enterprise value, in dollars, of increased visibility into the transmission and distribution network.

Big Data Application in Power Systems. https://doi.org/10.1016/B978-0-12-811968-6.00001-2

1 INTRODUCTION

In this digital age, power utilities *must* harness data to achieve the operational and enterprise efficiencies, insights, and flexibility to thrive amid emerging technologies and disruptive market forces. The question is not whether to become a data-driven utility, but how to do so. The opportunities and challenges are many. In the simplest terms, harnessing data in a comprehensive manner will require a transformational journey that will remake every power utility that undertakes the challenge. The process of becoming a data-driven utility requires a fundamental shift in organizational culture and business processes as well as data-related technology and practices. The desired result is not limited to the creation of a more reliable, resilient, and efficient grid. This transformation should also enable enterprise flexibility that supports new utility business models. Becoming a data-driven utility is an endeavor in which philosophy and technology go hand in hand.

The philosophy piece is simple and three-fold. First, data should drive improvements in a power utility's raison d'être. The ultimate, traditional goal of a power utility is to serve customers by delivering power safely, efficiently, and affordably. We are likely to see this fundamental mandate broaden to include customer service options, enabled by data. Harnessing data can support improvements in customer service, enhance customer and stakeholder value and increase the reliability, resiliency, and efficiency of operations. This is true whether a utility is cooperatively owned, municipally owned, or investor owned. Second, the organizational and technological transformations required to become a data-driven utility are so far-reaching that only a holistic approach will serve. Third, and most broadly, current and near-term societal and market trends pose a challenge to utilities' historic, regulated monopoly business and regulatory model. If a utility wants to determine its own fate, it must be proactive. Data is the new enabler of value and its opportunities and challenges must be actively embraced with a sense of urgency.

2 ALIGNING INTERNAL AND EXTERNAL STAKEHOLDERS

One fundamental concept in becoming a data-driven utility is that every internal stakeholder who can create value from data should have secure and timely access to that data. The very process of identifying useful data, collecting, processing, and presenting it or making it accessible on-demand will drive cultural and business process change throughout a utility. Creating a data-driven utility requires cooperation and coordination across all operational and enterprise units and the recognition that silos are obsolete legacies of past practices.

One should not underestimate the fundamental transformation unleashed by pursuing the goal of becoming a data-driven utility.

This observation holds true for external stakeholders as well. On the customer side, data has also become a valuable commodity. Customers are no longer passive ratepayers. Their energy use data belongs to them and, increasingly, they expect value for it. Public utility commissions recognize that customers own their energy use data, that utilities must secure it, and that the individual customer has the prerogative to say how that customer-specific data is used or shared. Whether utilities use data to create service options with value to both utility and customer may well determine their future success as an enterprise. Today, emerging technologies, third parties, and disruptive market forces abound, seeking to provide utility customers with value and service options based on their energy use data. For utilities, data has become not only the means to thrive but also the means to survive.

3 TAKING A HOLISTIC APPROACH

A holistic, methodical approach to becoming a data-driven utility has several common, recognizable steps, though the outcome for any individual utility will likely be unique, due to its existing customer base, business model, and legacy infrastructure. In this introductory chapter and overview of the topic, we will examine the implications of a holistic approach, the technology-related phases it requires, and connect the dots between data-producing sensor and data-reliant end user. A brief synopsis of three case studies will illustrate many of these points.

A holistic approach to becoming a data-driven utility literally takes everything into account. It views transmission and distribution as a single integrated entity. It encompasses the operations and business of delivering power to customers in a manner that achieves customer engagement and satisfaction based on increased system reliability, resiliency, and efficiency. Built on a foundation of open architecture and standards, a holistic approach ensures interoperability between devices, systems, and databases. It enables value creation at operational and enterprise levels. It enables forward and backward compatibility to derive full value from current and future investments in technology while maintaining the value of legacy equipment. In terms of an end-to-end system, a holistic approach provides a means by which all data-producing devices—increasingly, nearly every device in a T&D system—can be mapped to communication channels and networks with the appropriate response requirements, routed to both operations and enterprise sides of the organization, and presented and/or made accessible on-demand to the right people in the right time and place for value creation.

A holistic approach aligns customer needs and expectations with utility business drivers and depends on a technology roadmap for grid modernization that supports this alignment. In terms of utility culture and organization, a holistic approach eliminates silos and demands utility-wide cooperation and coordination to avoid redundant systems and costs. Thus it provides the basis for prudent, well-vetted investments that will create customer and stakeholder value and benefits that increase over time, meet future needs, and are likely to win regulatory approval.

In an era in which the utility business model requires review and transformation and digital technology produces an increasing granularity, quality, and quantity of data, a holistic approach to becoming a data-driven utility offers the richest opportunity for success.

4 "STRONG" FIRST, THEN "SMART"

Aligning customer needs and expectations with utility operational and business drivers should dictate how data is generated, collected, stored, processed, presented, or accessed, and how actionable intelligence is applied. A data-driven utility should review its current and mid-term operational and business models and identify its customer needs and fundamental business drivers. This will help in understanding precisely what actionable intelligence—and, thus, data—is needed for both operations and the enterprise to meet its self-determined goals of improving customer service and pursuing value creation.

To optimize current practices and enable future flexibility in reaching operations and enterprise goals, a utility must first develop a "strong" grid before pursuing a "smart" grid. This can only be achieved by establishing an information and communications technology (ICT) foundation based on open architecture and industry standards. The development of operational intelligence (and automation) and enterprise value relies on a two-way, standards-based, integrated communication system [1].

This first step requires that information technology (IT) and communications groups work together to understand and support the functional requirements (response requirements, bandwidth, latency) of each disparate data path—from sensor to end user—for current and future systems and applications. This approach requires organization-wide cooperation, which is no small feat. Enabling this fundamental cultural shift requires executive leadership, potentially third-party facilitation, and incentives that reward personnel for organization-wide and customer value creation rather than for individual staff and bailiwick-level achievements.

A foundational ICT platform that links all operational and enterprise aspects of a utility is a prerequisite for enterprise-wide data management. This ICT platform should support full information flow, data management and analytics, and grid monitoring and control. It also comprises the basis for future functionalities that potentially include new consumer services, the integration of distributed energy resources (DERs) and other, yet-to-be-determined needs. The efficacy of this phased approach—seeking a “strong” grid before a “smart” grid—has been affirmed by lessons learned from the stimulus-funded work accomplished under the American Recovery and Reinvestment Act (ARRA) between 2009 and the present. One simple example illustrates this point.

ARRA funding opportunities allowed many utilities to adopt advanced metering infrastructure (AMI). Some of these utilities took a traditional approach by assigning AMI implementation to their metering group alone. As these same utilities later contemplated the implementation of distribution automation (DA), they compounded their original mistake by assigning DA to a distribution engineering group in operations [2].

The direction is positive, but the execution is flawed. DA is the next logical step in grid modernization after AMI and it has the most attractive, stand-alone (i.e., nonsubsidized) business case. But these utilities are finding that their earlier decisions on data networks and IT infrastructure to support AMI do not support DA integration or that implementing DA requires a costly, disruptive workaround. In a holistic approach to data management, all operational and enterprise units would openly discuss their future direction and related projects and set foundational ICT requirements to serve them all. This fundamental step would eliminate redundant efforts and costs—and the creation of two separate data streams—because two or more systems in this example share a need for a service territory-wide communication network. Extrapolate this single example across a utility’s many networks, systems, and applications and extend it into the future along a well-plotted technology roadmap. Although it requires daunting cultural change and significant up-front time and effort, a holistic approach ultimately saves time, effort, and money and provides ever-increasing benefits to a future-facing, data-driven utility. In contrast, as this example illustrates, a fragmented, piecemeal approach is likely to result in stranded assets or, at best, time-consuming, costly workarounds at each step in a technology roadmap.

Once a strong ICT foundation has been established, a data-driven utility can proceed to develop a “smart” grid and to map data from sensor to end user. This next step requires the convergence of IT and operations technology (OT) and their respective staffs—the beginning of an operations- and enterprise-wide cultural shift to holistic utility management that focuses on customer-stakeholder-centric value creation and eliminates organizational silos and siloed thinking.

Guidelines for a holistic approach to becoming a data-driven utility:

- Align internal and external stakeholders.
- Think in terms of holistic solutions across the organization.
- Build a strong grid first, with robust ICT performance, then build a smart one.

5 INCREASING VISIBILITY WITH IEDs

As readers know, sensors, processing, and the visibility they produce have been applied to the transmission system for some time. The real growth in the need for visibility is downstream in the distribution system, where data-producing sensors and devices in the form of intelligent electronic devices (IEDs) are proliferating. The proliferation of IEDs in the distribution system is enabling utilities to treat T&D as a single entity and is a major enabler for the transformation to a data-driven utility. Yet a lack of visibility in the distribution system remains widespread; for example, only two-thirds of the distribution substations in the United States currently have automation.

IEDs can take the form of standalone sensors or they can be data-producing substation protection and control equipment such as protective relays, load tap changers, and voltage regulators. They produce two streams of data: operational and nonoperational. Operational data is routed in real time to operators in control centers for monitoring and control purposes (see Fig. 1, Types of data: "operational" data). Nonoperational data can provide significant insights

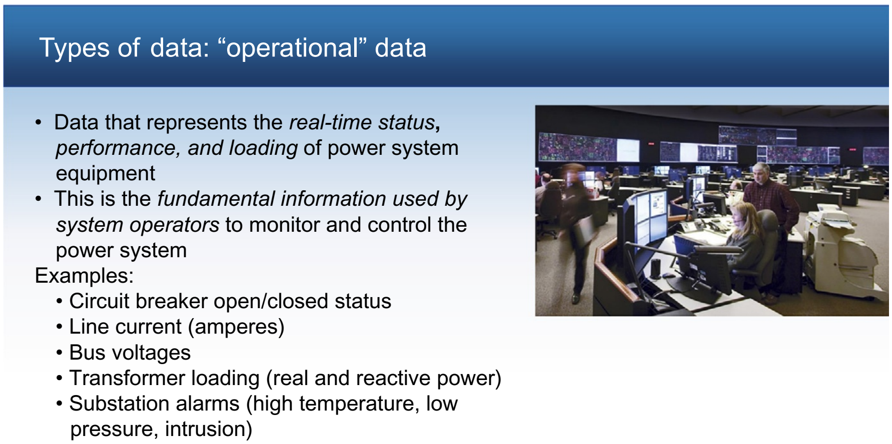


FIG. 1
Types of data: "operational" data. From J.D. McDonald, Powerpoint presentation, Enterprise Data Management, slide # 5.