EXAMPLE CHAPTER – Contributed chapter

Chapter 1: Introduction

Author(s):
A. F. Zobaa (Brunel University London, United Kingdom, azobaa@ieee.org)
*T. J. Bihl (Wright State University, United States, Trevor.Bihl@wright.edu)

1.1 Introduction

As a concept, big data and power systems might appear unrelated; however, the Smart Grid and advances in general computing power have made power systems a data driven industry. The result of the ability to collect endless data is the emergence of big data. However, power systems are connected to physical devices and critical infrastructure (CI) and thus additional research problems and concerns exist in power system big data.

1.2 Big Data

Big data involves more than the size of the data itself and extends to the complexity and speed at which it is collected. Frequently, big data is defined with vague and self-referencing definitions and naturally big data logically extends from data (Bihl, Young II, & Weckman, 2016). While data is generally any sensed output, big data involves data that is too big, complex or overwhelming to be analyzed by traditional methods (Bihl, Young II, & Weckman, 2016).

The primary attributes of big data are the 3 "V's" of *volume*, *variety*, and *velocity* (Bihl, Young II, & Weckman, 2016). While 42+ attributes have been defined by some researchers in describing big data, see (Shafer, 2017), the 3 V's capture of the gist of big data problem. As attributes, *volume* relates to the overall size of the data, *variety* indicates that big data can contain various types of data (text, strings, numbers, and etc.) all within one dataset, and *velocity* indicates that big data is collected in real time (Bihl, Young II, & Weckman, 2016).

Critically, *velocity* is an attribute frequently associated with big data. Given enough time, any given large volume and highly various data set could eventually be analyzed using traditional methods. However, when this data is continuously being collected, a *velocity* problem exists whereby the growing size and complexity preclude traditional methods. Thus, advanced analytics and data management methods are both necessary, c.f. (Gutierrez, Boehmke, Bauer, Saie, & Bihl, 2018) (Najafabadi, et al., 2015).

1.3 Future Power Systems

Future power systems implies power systems that differ from today's due to increased decentralization, expanded communication and monitoring abilities, and wider variety of sources (Hebner, 2017). Multiple thrusts exist in power system research to accommodate this future; these include expanding the smart grid, increasing penetration of the IoT, expanding renewable sources, and microgrid considerations.

Expanding penetration of the Smart Grid is not only expected but already underway (Amin & Wollenberg, 2005). Along with the Smart Grid comes a multitude of logged customer and power grid data which can be analzed to find power theft (Jiang, et al., 2014) and improve operating conditions of the grid at large (Fan, et al., 2013). The IoT further expands upon the Smart Grid by enabling communication with any and all devices (Gubbi, Buyya, Marusic, & Palaniswami, 2013). An IoT enabled power grid thus enables the monitoring of the critical infrastructure, while posing both big data and security problems (Sajid, Abbas, & Saleem, 2016).

Increasing decentralization through more microgrids, and nanogrids, can be also expected in the future power grid. While these have the ability of providing local resiliency (Hebner, 2017), they introduce uncertainty in larger grid planning (Khodaei, Bahramirad, & Shahidehpour, 2015). Added to this, is the expected increase in the use of renewables, which also increase power system planning problems due to their general availability uncertainty (Polatidis, Haralambopoulos, Munda, & Vreeker, 2006) (Atwa, El-Saadany, Salama, & Seethapathy, 2010).

1.4 Book Organization

To examine these problems, this book examines various intersections of big data and future power systems. For this goal, this book provides 9 chapters, including the introduction, which focus on the primary themes of big data in future power systems. Overall, this book discusses big data analysis methods, big data problems in future power systems, IoT concerns, security concerns related to big data, and various associated complexities.

1.4.1 Overview

This book is organized as follows:

- Chapter 2 discusses analytics and machine learning methods in general and those applicable to big data in power systems
- Chapter 3 discusses additional big data analytics relative to smart grid components
- Chapter 4 discusses optimization methods which are suitable for big data models in power systems
- Chapter 5 extends the discussion of Chapter 4 by considering various cyber security issues that exist in IoT-enabled future power systems
- Chapter 6 discusses electricity theft detection and mitigation which is enabled by big data collection from the Smart Grid
- Chapter 7 discusses renewable energy planning concerns which are associated with planned future power systems that have high renewable penetration
- Chapter 8 discusses transformer protection methods which are enabled by big data collection on transformers.

1.4.2 Big Data Application and Analytics in a Large-Scale Power System

To analyze big data, a variety of machine learning methods are generally employed. Machine learning is broadly synonymous with pattern recognition, statistics, and data mining (Mannila, 1996) (Hand, 1998). However, due to emergence of big data, a variety of new methods have recently emerged, e.g. large scale neural network known as “deep learning,” which are capable to analyze and exploit the bigness of big data. While these methods have achieved significant advancements in image recognition, see (LeCun, Bengio, & Hinton, 2015), they have begun to see use in power system big data analysis.

1.4.3 The Role of Big Data Analytics in Smart Grid Communications

Because a smart grid can be described as a huge sensor network, with a lot of intelligent devices, the growth in the number of devices will produce a considerable amount of measured data. How to quantify and to analyze these data to enhance grid operation arises as one big concern. Advances of the Smart Grid promise to give operators and utilities a better understanding of customer behavior, demand consumption, weather forecast, power outages, and failures. However, it is vital to quantify the volume of sampled data to take advantage of them. Therefore, this chapter aims to characterize and to evaluate the emerging growth of data in communications network applied to smart grid scenario. A future active distribution system will serve as an example to demonstrate the data requirements for monitoring and controlling the grid.

1.4.4 Big data optimization in electric power systems: A review

Traditional data-processing applications have difficulties operating effectively due to the complexity, velocity and voluminously of big data. This chapter presents a review of big data optimization problems in electric power systems. The chapter starts with scientometric mapping methods that show variety and diversity of large-scale optimization problems in today’s power system networks. An electrical grid power system could be categorized into generators which provide the required electric power, transmission systems that carry the electricity from the generating units, and distribution systems that feed the power to nearby industries and homes. The optimization issues such as logistics optimization in power system, as well as some optimization techniques including non-smooth, nonconvex, and unconstrained large-scale optimization, are presented. Additionally, some metaheuristic methods in large-scale power system optimization problems have been reviewed.

1.4.5 Security Methods for Critical Infrastructure Communications

The proliferation of communication devices in Critical Infrastructure (CI) applications presents security challenges. A variety of security approaches are used to prevent unauthorized access to CI networks. This chapter will review 1) the communication devices used in critical infrastructure, especially power systems, 2) security methods available to vet the identity of devices, and 3) general security threats in CI networks. Device identity verification methods will be discussed and range from bit-level, e.g. encryption keys, to physical layer, e.g. radio frequency fingerprinting, methods.

1.4.6 Data Mining Methods for Electricity Theft Detection

Electricity theft is a major concern for utilities in both the developed and developing world. Although the United States has a low electricity theft rate, an estimated $4 billion of revenue is lost per year in the United States alone; the developing world generally sees much higher losses. Detecting potential electricity thieves is thus of interest to mitigate losses. Check meters and usage analysis have been used primarily to identify possible electricity thieves. However, advances in computing, the smart grid, smart meters and in data mining have enabled more analysis to be conducted in this area. This chapter will review the wide variety of techniques and applications developed for electricity theft detection.

1.4.7 Unit Commitment Control of Smart Grids

Future power grids are planned to have significant renewable energy penetration. However, these sources of energy are unpredictable in nature. The unit commitment (UC) problem is the problem of producing power by collaboration of sources in order to achieve demand. This chapter discusses and presents a centralized approach to solve the UC problem for energy systems that contain a variety of generating components (traditional to renewable).

1.4.8 Data-based Transformer Differential Protection

This chapter uses pattern recognition and dimensionality reduction methods for differential protection of power transformers. Both the linear principal component analysis (PCA) and the nonlinear, and neural network based, curvilinear component analysis (CCA) are considered. Both PCA and CCA use the differential current from current transformers at transformer terminals. By using two techniques, this chapter illustrates how pattern recognition methods can be used to preprocess differential current to discernment internal faults currents (transformer differential protection zone) from inrush and over-excitation currents. Both PCA and CCA are employed with the PSCAD electromagnetic simulation software in a three-phase power system, for distinct scenarios. The results show the feasibility to develop a differential protection to power transformers using data pattern recognition algorithms.

1.5 Conclusions

Overall, a wide variety of challenges exists in the future power grid, ranging from cyber security to data handling to planning. This book aims to discuss and present a variety of approaches to handling each of these challenges, in addition to discussions and reviews of the various topics and domains. To this aim, each chapter focuses on one specific topic and minimal overlap exists between chapters. However, the underlying theme in all chapters is big data made by possibly by future power system infrastructure.

1.6 References

Amin, S., & Wollenberg, B. (2005). Toward a smart grid: power delivery for the 21st century. *IEEE power and energy magazine*, *3*(5), 34-41.

Atwa, Y., El-Saadany, E., Salama, M., & Seethapathy, R. (2010). Optimal renewable resources mix for distribution system energy loss minimization. *IEEE Transactions on Power Systems, 25*(1), 360-370.

Bihl, T., Young II, W., & Weckman, G. (2016). Defining, Understanding, and Addressing Big Data. *International Journal of Business Analytics (IJBAN), 3*(2), 1-32.

Fan, Z., Kulkarni, P., Gormus, S., Efthymiou, C., Kalogridis, G., Sooriyabandara, M., . . . Chin, W. (2013). Smart grid communications: Overview of research challenges, solutions, and standardization activities. *IEEE Communications Surveys & Tutorials, 15*(1), 21-38.

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems, 29*(7), 1645-1660.

Gutierrez, R., Boehmke, B., Bauer, K., Saie, C., & Bihl, T. J. (2018). Cyber Anomaly Detection: Using Tabulated Vectors and Embedded Analytics for Efficient Data Mining. *Journal of Algorithms and Computational Technology*.

Hand, D. (1998). Data mining: Statistics and more? *The American Statistician, 52*(2), 112-118.

Hebner, R. (2017). The power grid in 2030. *IEEE Spectrum*, *54*(4), 50-55.

Jiang, R., Lu, R., Wang, Y., Luo, J., Shen, C., & Shen, X. (2014). Energy-theft detection issues for advanced metering infrastructure in smart grid. *Tsinghua Science and Technology, 19*(2), 105-120.

Khodaei, A., Bahramirad, S., & Shahidehpour, M. (2015). Microgrid planning under uncertainty. *IEEE Transactions on Power Systems, 30*(5), 2417-2425.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436.

Mannila, H. (1996). Data mining: machine learning, statistics, and databases. *Eighth International Conference on Scientific and Statistical Database Systems*, 2-9.

Najafabadi, M., Villanustre, F., Khoshgoftaar, T., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data, 2*(1).

Polatidis, H., Haralambopoulos, D., Munda, G., & Vreeker, R. (2006). Selecting an appropriate multi-criteria decision analysis technique for renewable energy planning. *Energy Sources, Part B, 1*(2), 181-193.

Sajid, A., Abbas, H., & Saleem, K. (2016). Cloud-assisted iot-based scada systems security: A review of the state of the art and future challenges. *IEEE Access*, *4*, 1375-1384.

Shafer, T. (2017, Apr). *The 42 V's of Big Data and Data Science*. Retrieved from KD Nuggets: https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html