Guest Editorial Big Data Analytics for Grid Modernization

ADVANCED analytics plays a vital role in the era of big data, such as managing smart cities, predicting crime activities, optimizing medicine formula based on genetic defects, detecting financial frauds, and personalizing marketing campaigns. Information extracted from the big data benefits many industries in their day-to-day operations. The deployment of phasor measurement units (PMUs), smart meters and other smart devices has offered engineers the access to a large variety of data at an unprecedented granularity and volume. However, the old data management systems and applications are not designed to handle the big data. Therefore, how to extract actionable information and values out of the big data and how to integrate the information into grid operations and planning to ensure the secure, reliable and economical supply of electricity are becoming increasingly critical.

This special section aims to publish original research papers and visionary reviews on the technologies, algorithms and case studies that are associated with big data analytics for modernizing the electric power grid. The key word "big data" for this special section can be interpreted from the following four aspects: data, problem, methodology, and technology. Firstly, the **data** involved in the analysis is big in at least one of its three defining dimensions: volume, variety or velocity. The big data used in the utility industry includes but is not limited to smart meter data, phasor measurement unit data, weather data, and social media data. Secondly, the **problem** under investigation is to prepare for analyzing the big data, such as data compression and data security issues. Thirdly, the **methodology** requires customized modeling of individual components of a system, or leads to in-depth understanding of the individual components. For instance, estimating the invisible solar generation belongs to this category. Last but not least, the **technology** can be used to help reach the answer faster, or answer the questions otherwise difficult to answer. For example, a distributed platform can be used to speed up the analytic tasks.

After a careful peer review process, 17 papers are collected in this special section. Most of them cover multiple aspects from the aforementioned list. Based on the smart grid applications, they are categorized into five groups.

Preparing for Big Data Analytics

Zhou *et al.* propose a distributed data analytics platform that is suitable for large volume of streaming data such as synchrophasor.

Gadde *et al.* propose a generalized compression technique that utilizes both spatial and temporal redundancies exhibited in the synchrophasor data.

Tong *et al.* propose the general extreme value distribution characteristic for household load data and then utilize it to identify load features including load states and load events. In this work, a highly efficient data compression format is designed to store key information of load features.

Vasilakos and Hu present a literature survey on big energy data analytics and security. The authors propose several taxonomies to express the intriguing relationships of various variables in the field.

Meter-Level Demand Clustering and Forecasting

Y. Wang *et al.* propose an approach to clustering electricity consumption behavior dynamics, where "dynamics" refer to transitions and relations between consumption behaviors, or rather consumption levels, in adjacent periods.

Ben Taieb *et al.* address the need within the energy industry for probabilistic forecasts of household electricity consumption. Those forecasts are necessary to quantify the uncertainty of future electricity demand in order to undertake appropriate planning of generation and distribution.

Renewable Integration

Shaker *et al.* contribute two papers to this special section. The first paper establishes the problem of estimating the aggregated power generation of behind-the-meter solar panels. Those panels are not monitored and thus are "invisible" to power system operators. The authors discuss several fundamental issues arising from inviable solar systems and establish a base line model for estimating the aggregated output energy of a large number of invisible solar systems. In the second paper, the authors propose models and methods to practically estimate the aggregated solar power generation of a large number of invisible, small solar panels that are distributed in a large geographical area.

Zhang and Grijalva propose a data-driven approach for the detection, verification, and estimation of residential PV system installations to allow utilities to detect and monitor PV installations in their network for maintain grid reliability and safety.

Wang and Huang propose a theoretical framework for the cooperative planning of renewable generations in a system of interconnected microgrids. Simulations based on realistic data demonstrate that all microgrids benefit through the cooperation in the proposed framework, and the overall system cost

Date of current version August 19, 2016.
Digital Object Identifier 10.1109/TSG.2016.2593358

is reduced by 35.9% compared to the non-cooperative planning benchmark.

Power System Modeling

Peppanen *et al.* present a computationally efficient distribution system parameter estimation algorithm to leverage the data collected by smart meters and photovoltaic (PV) micro inverters to calibrate the model parameters of distribution feeders.

Chen *et al.* present a measurement-based method to use high-speed synchronized voltage and current phasor data collected from PMUs for computing Jacobian matrix. This method can be used for near real-time system topology estimation, because it readily adapts to changes in system operating points and topology.

Event Detection and Analysis

Sun *et al.* propose a novel method to detect and locate power outages based on the information collected from the social media. In particular, the authors design a probabilistic framework to integrate the textual, temporal, and spatial information to identify the events.

H. Jiang *et al.* investigate big data characterization for smart grids and its applications in fault detection, identification and causal impact analysis. The spatial-domain and temporal-domain characterizations are combined to reduce the volume of the big data substantially while preserving the comprehensive information in the data.

Rafferty *et al.* propose a real-time detection and classification algorithm that applies the moving window principal component analysis (PCA) to frequency measurements.

T. Jiang *et al.* propose a new method to estimate the dominant modes by monitoring the inter-area oscillations in the China Southern power grid using PMU data under both the ringdown and ambient conditions. The state space model is identified by the data driven stochastic subspace identification algorithm. To accurately identify the dominant modes, repetitive results are calculated with model order variation. Clustering analysis and stepwise refinement are then applied to discriminate the dominant modes from trivial ones for improving the estimation accuracy.

B. Wang *et al.* present a PMU data application. By mapping an online power system Transient Stability Assessment (TSA) problem as a two-class classification problem, the authors propose to solve the problem using a novel data mining algorithm, the core vector machine. As long as online PMU big data are received, TSA can be done simultaneously.

The guest editorial board solicited a long list of research topics, of which many have appeared in this special section. We hope that these exemplary papers contributed from industry and academia can offer the readers some state-of-the-art case studies. One the other hand, we hope that the readers can also recognize the research opportunities from those important topics not addressed in this special section, such as data visualization, data-driven dynamic pricing, predictive asset management, and customer analytics.

Acknowledgment

The Guest Editorial Board would like to thank the authors for their contributions to this special section, and the reviewers for their timely and valuable comments during the peer review process. Special thanks to Dr. Mohammad Shahidehpour, the past Editor-in-Chief, Dr. Jianhui Wang, the Editor-in-Chief of the IEEE Transactions on Smart Grid, and the staff of the journal editorial office for their timely advice and support.

Tao Hong, *Guest Editor-in-Chief*
University of North Carolina at Charlotte
Charlotte, NC, USA

Chen Chen
Argonne National Laboratory
Lemont, IL, USA

Jianwei Huang
The Chinese University of Hong Kong
Hong Kong, China

Ning Lu
North Carolina State University
Raleigh, NC, USA

Le Xie
Texas A&M University
College Station, TX, USA

Hamidreza Zareipour
University of Calgary
Calgary, AB, Canada