ISSN: 2454 - 4248
87 – 91

Role of Big Data Analytics in Smart Grid

Prof. C. E. Morkhade
Department of Electrical Engineering
MGI-COET Shegaon, India
morkhadechetan7@gmail.com

Prof. S. R. Sapkal
Department of Electrical Engineering
MGI-COET Shegaon, India
srsapkal925@gmail.com

Prof. V.A. Ghodeswar
Department of Electrical Engineering
MGI-COET Shegaon, India
ghodeswarvaibhav@gmail.com

Lect. C. P. Bhise
Department of Electrical Engineering,
STC-SPRT Khamgaon, India,
chaitaleebhise@gmail.com

***Abstract*—**Data analytics are now playing a important role in the current industrial systems. Driven by the progress of information and communication technology, an information cover is now added to the conventional electricity distribution and transmission network for data pool, storage and analysis with the help of extensive installation of smart meters and sensors. This paper reviews the big data analytics and corresponding applications in smartgrids. The features of big data, smart grids as well as huge amount of data collection components are also debated which help to get motivation and potential advantages of using unconventional data analytics in smart grids.

Keywords*-Smart grid; Big data; Computation platforms;*

*****

I. INTRODUCTION

With the fast development of digital technology and cloud computing, more and more data are generated through digital devices and sensors, such as computers, smart phones, advanced measuring devices as well as through human activities and communications. For occasion, the size of data on the internet is now measured in Exabyte's (10^{18}) and zetta bytes (10^{21}). Real and efficient analysis of these data carries huge value and benefit to our daily life activities. However, the composed data are mounting at an exponential growth, and the data structure becoming much more complicated. The dispensation and analysis method of these large volume data is a new challenge but opportunity at the beginning of this century.

Although big data is a novel term, the concept of inventing valuable information from huge collected data in commercial operation as aiding knowledge for business decision has already been proposed in 1989 by Howard Dresner as "business intelligence" (BI).

In power grid, the fossil fuels are facing the problem of depletion and the decarburization demands the power system to decrease the carbon emission. Smart grid and super grid are actual solutions for electrification of human society with high penetration of renewable energy sources. Although the rising consciousness of sustainable development have become the inspiration to the utilization of renewable energy sources, the erratic characteristics of wind and photovoltaic energies bring huge challenges to the safe and stable operation in power system. Outmoded electricity meters in distribution systems only produce a small amount of data which can be manually collected and analyzed for billing purpose. While the huge volume of data collected from two-way communication smart grids at different timeresolutionsinnowadaysneedprogressivedataanalyticsto extractvaluedinformation not only for billing information but also for monitoring status of the electricity network. For example, the high-resolution electricity consumption data can also be used for customer behavior analysis, demand predicting and energy generation optimization. Predictive maintenance and fault detection based on the data analytics with advanced metering structure are more important to the security of power system.

Thus, the great progress of information and communication technology (ICT) provides a new idea for engineers to control the traditional electrical system and makes it smart.

Big data analytics might help to expand new technologies or solutions in different components of power system. For example, one of the data linked component of smart grid is smart meter. This component measures the consumption sum of electricity of consumer. As mentioned in, smart meter measures electricity consumption data every 15 minutes. It is highlighted that the number of smart meter reading will rise from 24 million a year to 220 million per day for a large utility company which implements smart metering structure properly. In Table 1, the amount of data was collected by one million metering devices in a year is presented. It can be understood from the Table I, for

ISSN: 2454 - 4248
87 – 91

example, when the data is collected from 1 million devices every 15 minutes, 35.04 billion records are obtained with the volume of 2920 Tbdata [1].

TABLE I. THE CONNECTION OF COLLECTION FREQUENCY AGAINIST RECORD NUMBER AND DATA VOLUME FOR 1 MILLION METERING DEVICES IN A YEAR

Collection Frequency	1/day	1/hour	1/30 min	1/15 min
Records (billion)	0.37	8.75	17.52	35.04
Volume of Data (Tb)	1.82	730	1460	2920

This paper is to discuss the concepts of big data analysis and their applications in smart grids. The intent of this paper is to identify the sources of majority data, characterization of data, methods of analysis of data and benefits in smart grid.

II. Concept Of SG & Big Data

Smart grid is defined as electricity systems that can logically integrate the actions of all users connected to it in order to efficiently deliver sustainable, economic and secure electricity supplies. The U.S. defined the smart grid of future in a similar way that includes the digital technology to improve reliability, security and efficiency of the electric system through information exchange, distributed generation and storing resources for a fully automated power delivery network. Compared with outdated power systems, the extensive application of distributed generators under the call of green energy resources is shaking the position of large-scale centralized power plants, which makes the conservative centralized control strategy less effective due to the unidirectional power flow. Connection of limited power generations (typically in the range of 3 kW to 10 kW) to the public distribution grid requires two-directional operation and control of distribution grids. Faced with the trials of more complicated control and protection strategies, the conventional electro-mechanical electric grid is supposed to be enhanced with the help of development in the digital information and communications network to overcome the cost from power outages and power quality turbulences as billions of dollars annually.

Usually, the smart grid can be assessed with a Smart Grid Architecture Model (SGAM), which is a 3-dimensional outline that merges domains, zones and layers together. The conventional structure of power scheme can be found in the domains as generation, trans- mission, distribution, DER (Distributed Energy Resources) and customer premises. The zones which present the layout of power system management are composed of market, operation, station, field and process. On top of the first two extents, the layout of interoperability layers includes the component, communication, information, function and business layers. SGAM as an architectural overview can be used to find the lacunas and difficulties of existing smart grid morals.

The definition of big data is not very clear and unbrokentill date. But there is a harmony among different explanations- this is an emerging technical problem brought by a dataset of large volume, various categories and complicated structures which needs novel framework and methods to extract useful information effectively. Therefore, the definition of big data depends on the ability of data mining processes and the corresponding hardware equipment to deal with large volume datasets. It is a comparative concept instead of an absolute definition. The big data can be understood as amount of data beyond technology's capability to store, manage and process capably as the data size increasing along with the development of ICT technologies.

III. Big Data Sources & Its Characteristics

As an smart system of both energy and data, smart grid is the abundant source of information, which covers the data from route of electricity transmission, generation, distribution and consumption. These data comprise the electrical information from distribution stations, distribution switch stations, electricity meters, and non-electrical information like marketing, meteorological as well as local economic data as shown[2] in Fig.1. Collection and analysis of them provide crucial help in scheduling of power plants, operation of subsystems, maintenance for important power equipment and business behavior in marketing.

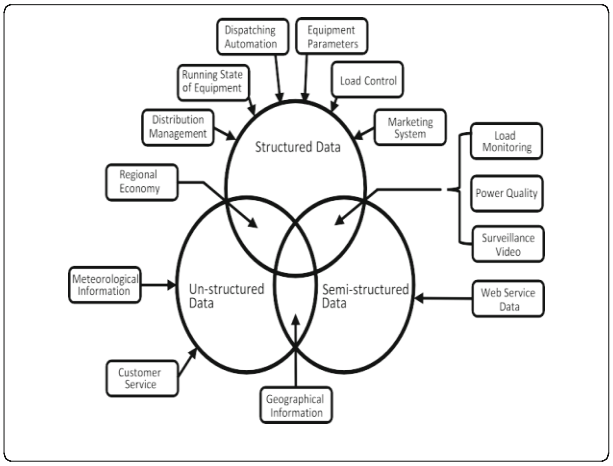

Figure 1. Big data sources inSG

The data sources stated above can be sorted into three categories: business data, measurement data and external data. Most of the operation constraints in power system are measured through installed sensors and smart meters, specifying the system's current and historical status. The weather conditions and social events like festivals are the external data that cannot be measured from smart meters but have an impact on the operation and preparation in power system. The business data mainly includes the marketing plans and behaviors.

ISSN: 2454 - 4248
87 – 91

The characteristics of big data in smart grid are also in accordance with the universal 5 V big data model in many researches as below:

Volume – refers to the massive amount of data generated, which makes data sets too big to store and analyze using traditional database technology. The likely solution to this problem is the distributed systems to store data in different sites, connect them by networks and bring them together by software. In smart grid the extensive application of smart meter and advanced sensor technology provide huge amount of data.

Velocity – refers to the hustle at which new data is generated and the speed at which data moves around. The desires for real-time exchange of data is increasing. With a sampling rate of 4 times per hour, 1 million smart meters installed in the smart grid would effect in 35.04 billion records, equivalent to 2920 Tb data in quantification.

Variety – refers to the types of statistics we can now use. In the past, we focus on planned data that neatly fits into tables or rational databases such as financial or meteorological data. With a big data knowledge, we need to handle different types of shapeless data including messages, social media conversations, digital images, sensor data, video or voice recordings, and bring them together with more outdated, structured data. According to the extensive data sources in smart grid as shown in Fig.1, the formats and dimensions of data are diverse in organization.

Veracity – refers to the griminess or trustworthiness of the data. The quality and accuracy are less trustworthy with such large amount of big data, which task the outcome data analysis. Errors of quantities in smart grid may exist due to the imperfections in devices or mistakes in data transmission. The protected and efficient power system operation relays on the data assessment and state estimation.

Value – refers to our capability to extract valuable information from the vast amount of data and derive a clear understanding of the value it brings. Thesuperior the amount of data is, the lower the density of valuable information will be. With the improvement of smart devices adopted in smart grid, moreand more value of big data analytics is revealed according to the various applications.

IV. Concept Of DataAnalytics,Machine Learning & Artificial Intiligence

Data Analytics: It is the discovery and communication of significant patterns in data. Data analytics is a (sometimes automated) process used to discover new, valid, useful and potentially motivating knowledge from large data sources which is otherwise difficult to uncover. If statistics is to be considered a branch of mathematics, data analytics is inclined towards performing the same functionality for computer knowledge. Visual tools and techniques are the preferred means of communicating the results of performing data analytics.

Machine Learning: It is the ability of machineries (associated with computers) to learn automatically without being explicitly programmed. It deals with picture and generalization of data and creates a representation of instances and functions which are evaluated on these data. Simplification is the unique property that the machine learning systems will try to yield, that is, the ability of the systems to perform well even on unseen data instances. Artificial Intelligence: It is the intelligence exhibited by machines, as opposed to natural intelligence exhibited by humans or animals. AI encompasses techniques which can endow an object or a program with human-like intelligence. AI also includes smart agents, entities that perceive their environment and take actions based on that perception. Figure 2 shows the interconnection of statistics, data analytics, machine learning and artificial intelligence.

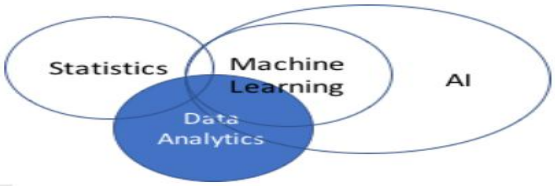

Figure 2. Venn diagram illustrating the interconnectedness of data analytics, ML&AI

V. Big Data Computing Platforms

The following subsections presents commonly used platforms for big data analytics and compared their performance [3] in Table II.

Hadoop: Apache Hadoop is an open source framework for storing and processing large datasets using Map Reduce programming model. The Hadoop consists of storage part (known as hadoop distributed file systems (HDFS)) and processing part (known as Map Reduce programming model). Mainly, Hadoop splits files into large slabs and distributes them across nodes so as to process data in parallel. Due to distributed storage structure, HDFS not only ensures high availability, but also high fault tolerance against hardware failures. OSI-Soft, which is one of the widely used database and data analytics platform in electric utility, uses Hadoop for performing data analytics.

Spark: Spark is a fast, in-memory, open-source big data processing engine which is designed to overcome the disk I/O limitations of Hadoop. Spark can perform in-memory computations and allow the data to be cached in memory, thereby eliminating the Hadoop's disk overhead limitation for iterative tasks. Spark is a general engine for large-scale data processing which is up to 100 times faster than Hadoop Map Reduce when the data can fit in the memory and up to 10 times faster when data resides on the disk.

Storm: Apache Storm is also an open source distributed real-time computation system that can reliably process unbounded streams of data. It is scalable, fault-tolerant, and easy to set up and operate, thereby having several use cases, including real-time analytics, online machine learning, and real-time computation.

ISSN: 2454 - 4248
87 – 91

Apache Drill: Apache Drill is an open framework that supports data-intensive distributed applications for interactive analysis of large-scale datasets. Drills is able to scale 10,000+ servers and process petabytes of data and trillions of records within seconds. In addition, Drill can notice schemes on-the-fly, thereby delivering self-service data explorationcapabilitiesondatastoredinmultipleformatsinfilesordatabases.Drillcan seamlessly integrate with several visualization tools, thereby making big-data platform interactive.

High Performance Computing: HPC is a vertical scale up platform for big data processing which consists of a powerful machine with thousands of cores. Due to high quality hardware implementation, fault tolerance in HPC systems is not problematic as hardware failures are extremely are. Even though HPC system can process terabytes of data, they are not scalable as horizontal processing platforms. Furthermore, initial placement and scaling costs are higher compared to other horizontal scale-out platforms.

TABLE II. COMPARISON OF VARIOUS BIG DATA ANALYTICS PLATFORMS

Platform	Data Scaling	Scalability	Fault Tolerance	I/O Performance	Application
Hadoop	Horizontal	Yes	Yes	Limited	Batch Processing
Spark	Horizontal	Yes	Yes	Moderate	Batch and real-time processing
Storm	Horizontal	Yes	Yes	Moderate	Real-time processing
Drill	Horizontal	Yes	Yes	Good	Interactive analytics
HPC	Vertical	Limited	Yes	Very good	Batch, stream, and interactive

VI. APPLICATION OF BIG DATA IN SMART GRID

In smart grids, the big data coming from several sources carry valuable information, and the cross fertilization of the heterogeneous data sources can unlock several novel applications beneficial to all the stakeholders, i.e., electric utilities, grid operators, customers, etc., for planning and operational decisions. The big data has potential to a) improve reliability and resiliency of power grid, b) deliver optimum asset management and operations, c) improve decision making by sharing information/data, and d) to support rapid analysis of extremely large data sets for performance improvement. However, the current trend in smart grid is that the smart meter big data is primarily used for demand response, load forecasting, baseline estimation, and load clustering type of applications, while the application of PMU big data is focused mainly on transmission grid visualization, state estimation, and dynamic model calibration.Fig. 3 shows [4]some of the potential applications of big data in smart grid useful for various stakeholders.

Energy Management Related Applications

Two-way flow of power and information in smart grid provides opportunities to small scale consumers, energy producers, and distribution system operators to take active part in grid management and ancillary services. In order to support energy management in real- time, we have to efficiently and intelligently process large volumes of data in smart grids. Improved forecasting tools for energy resources and loads, improved demand response (DR) methods, efficient data management framework, and data analytics are critical to enable the energy management for the optimized operation of power grids. It requires methods for dimensionality reduction of data (e.g., Random Projection method), algorithms that can extract load patterns from large-scale data set (e.g., K-means and ANNs), design of machine learning algorithms for improved forecasting, design of data compression for low memory requirements, development of scalable and distributed computing architecture for real-time performance, and so on.

Through large amount of data obtained from smart meter and home devices, utilities not only can get near real-time information of consumption, but also can develop proper incentives and operational strategies to better utilize behind the meter energy resources. Big data analytics can dynamically classify and categorize consumer consumption behaviors and electrical characteristics that can help utilities to make better operational decisions. Similarly, energy consumption pattern in big cities are identified using big data techniques.

Improvement of Smart Grid Reliability and Stability

Data collected from social media is used to identify and locate the power outage, which could help enhance power system reliability. This is an interesting application of big data techniques applied to smart grids based on data collected from social media (non-electrical data).Similarly, PMU big data could be used for stability margin prediction and real-time asset health monitoring.The big data helps to improve reliability and stability of power grids. Also event detection is possible through µPMUs.

Visualization

Advanced visualization is one of the key application area of big data analytics that can improve the overall assessment of smart grids. Big data analytics with the visualization technologies is used for monitoring real-time power system status as well as accurate grid connectivity information. Conventionally, various visualization techniques such as single line diagram, 2D, and 3D charts/plots were used for grid visualization. However, due to increased number of variables and their interdependencies, advanced visualization techniques are often required for the big data visualization in smart grid. Commercial tools, such as Real Time Dynamics Monitoring System (RTDMS), are avail- able for visualization using PMU big data. RTDMS provides several visualization options including dashboard display for situational awareness, voltage angle contour plots, voltage magnitude plot, frequency plot, oscillatory mode plot, etc.

ISSN: 2454 - 4248
87 – 91

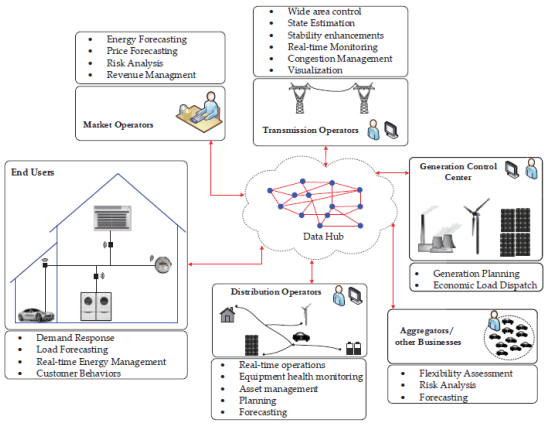

Figure 3. Applications of Big data analytics in SG

Parameter/State Estimation

Parameter and state estimations are essential for power system planning, operation, and control. Estimations are used for several applications including operational resource planning, real-time system monitoring, and resilient control design against cyber and physical-attacks. The availability of huge amount of data within the smart grid framework provides challenges as well as opportunities for state estimation. Due to availability of large dataset from various sensors and intelligent devices across the grid, system will be more visible, thereby having better and more accurate state estimation. However, due to introduction of large number of active nodes, power system optimization problems become mix-integer,nonlinear, and non-convex, thereby making the system computationally challenging. For instance, the trend in Volt/Var regulation is to utilize a large mix of voltage regulation resources (e.g. smart inverters, solid state transformers, on-load tap changers, voltage regulators, STATCOMs) on the feeder. The coordination of these resources will require real time monitoring and predictive tools to optimize the utilization of these resources and lead to reduced operational costs and increase the power quality and reliability of the system.

Applications to Cyber-Physical Systems

Since smart grid is a critical infrastructure, any cyber or physical vulnerabilities could lead to widespread impacts. Conventionally power system planner used to perform contingency analysis to provide resiliency under sudden disturbances against system faults and/or natural disasters. Due to close interdependencies between power and communication infrastructure, the future grids subject to increased risk of cruel attacks. However, most of the existing power system were not designed by accounting cyber-security. Unlike random nature of equipment failure probability distribution, cyber-attacks are normally coordinated and deliberately targeted to most critical components of the energy system. Such structured attacks can lead to cascading failures in the system. Therefore, tight cyber-physical coupling is necessary to extend power system security into both cyber and physical attacks. Integration of big data analytics provides an excellent opportunity to timely identify such malicious attacks and prevent the system from huge damages.

VII. Conclusion

This paper presented a overall review of big data analytics for smart grids.This paper provides detailed information and items to consider for utilities looking to apply big data analytics to, and details on how utilities can utilize big data analytics to develop new business models and revenue streams. The data which may contain valuable information are collected from smart meters installed in the power system, electricity market, GIS, meteorological information system, social media, and so on. The purpose of advanced ICT technology in power system is to associate the traditional physical parameters in power system to the external variables to develop potential regulations and scientific problems. Applications of data analytics mentioned in the paper are nearly involved in every aspect of smart grids, including the operation, maintenance, load forecasting, protection as well as fault detection and location.

References

[1] S. Sagiroglu, R. Terzi, Y. Canbay and I. Colak, “Big data issues in smart grid systems,” IEEEInternational Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, 2016, pp. 1007-1012.

[2] Yang Zhang,TaoHuang and Ettore FrancescoBompard, “Big data analytics in smart grids: areview,” Zhang *et al. Energy Informatics* (2018).

[3] Bishnu P. Bhattarai , Sumit Paudyal , Yusheng Luo , Manish Mohanpurkar , Kwok Cheung , Reinaldo Tonkoski , Rob Hovsapian, Kurt S. Myers , Rui Zhang , Power Zhao , Milos Manic, Song Zhang , Xiaping Zhang, “Big data analytics in smart grid:State-of-the-Art,Challenges ,Opportunities and Future directions,” IET Generation, Transmission and Distribution · February 2019.

[4] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar and S. Mishra, “Decision Tree and SVM- based Data Analytics for Theft Detection in Smart Grid,”IEEE Transactions on Industrial Informatics, vol. 12, no. 3, pp. 1005-1016, 2016.

[5] Shyam R, Bharathi Ganesh HB, Sachin Kumar S, Prabaharan Poornachandran, Soman K P, “Apache Spark a Big Data Analytics Platform for Smart Grid,” Procedia Technology 21 (2015) 171 – 178.

[6] Panagiotis D. Diamantoulakis, Vasileios M. Kapinas, George K. Karagiannidis, “Big Data Analytics for Dynamic Energy Management in Smart Grids,” Preprint submitted to Elsevier Big Data Research, Published in vol. 2, no. 3, pp. 94-101, Sep. 2015.

[7] Y. Sun, H. Song, A. J. Jara, and R. Bie, “Internet of things and big data analytics forsmart and connected communities,”*IEEEAccess*,vol.4,pp.766–773,2016.

[8] Alam Mollah Rezaul, Kashem M. Muttaqi, Abdesselam Bouzerdoum. Evaluating the effectiveness of a machine learning approach based on response time and reliability for islanding detection of distributed generation. IET renewable power generation (volume: 11, Issue: 11, 2017)

[9] J. HuandA.V. Vasilakos, “Energy big data analytics and security: challenges and opportunities,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2423–2436, Sep.2016.