

Research on Application of Big Data Technique in Smart Distribution Grids

Da-Jiang REN

Power Supply Bureau, Zhangjiakou, Hebei, China

14592318@163.com

Keywords: Smart Distribution Grids, Big Data, Energy Management System.

Abstract. Smart grid is a complete automation system, where large pool of sensors is embedded in the existing power grids system for controlling and monitoring it by utilizing modern information technologies. The data collected from these sensors are huge and have all the characteristics to be called as Big Data. The Smart-grid can be made more intelligent by processing and deriving new information from these data in real time. This paper presents Apache spark as an energy management system which is suitable for storing and performing Big Data analytics on smart grid data for applications like automatic demand response and real time pricing.

Introduction

Human energy consumption has gradually increased greenhouse gas concentrations in the atmosphere and is considered the main source of global warming. With the power of emerging technologies in information and communication systems, the present day power grid scenario is evolving into a sensor-embedded network (Smart Grid), which has the ability to control and automate the entire processes. Sensors of various types are deployed across the length and breadth of the smart grid. All these sensors produce different types of data (Heterogeneous), which are then collected at the utility data-centers. The variability, variety and velocity of the data make it into a very special category called “Big Data”. The data collected can be analyzed to reveal the knowledge of unseen patterns which are hidden in large datasets and utilized for making strategic, tactical and operational level decisions. In data science, the term data analysis, data mining and textmining refers to the same technique of deriving hidden information using various machine learning algorithms from the data acquired. Among the data mining methods the most widely used are:

1) Pattern Matching and Associative rule: It involves learning the frequently occurring trends in the data to define rules for future decision making. Pattern matching algorithms includes Eclat, FP-Tree, so on.

2) Classification: A supervised machine learning method in which the data is divided into training and testing sets. Then a classifier model is trained using training set in-order to predict the class labels for the given test data. Some of the most widely used classification algorithms are Decision tree, K-Nearest Neighbors (K-NN), Support Vector Machine (SVM) and Naive-Bayes.

3) Clustering: It is an unsupervised learning process. The goal of clustering is to group data points in the dataset together into a number of groups, depending upon its distribution in higher dimensional space. The choice for number of clusters depends on data and problem definition. Some of the most commonly used clustering algorithms are k-means, Expectation Maximization and Hierarchical clustering.

A smart grid is the next-generation power system able to manage electricity demand in a sustainable, reliable and economic manner, by employing advanced digital information and communications technologies.

Smart Grid Data Flow

Apart from the electrical aspect, Smart grid is now becoming an interesting research area for Data scientists. Smart grid data are broadly classified into Generation data, Transmission/Distribution data

and Consumer data. Data-centric sensor network view and Distribution of data in smart grid is visualized in the Fig.1. A lot of data analysis can be done over this data to make the Grid more intelligent and smart.

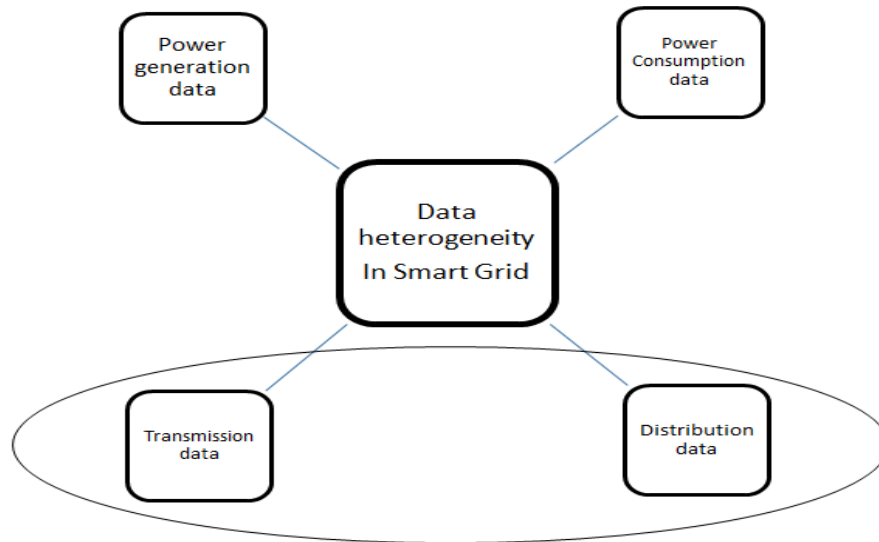


Fig.1 Heterogeneity in Smart Grid Data

Power Transmission Distribution Big Data

Once the power has been produced, it is feed to step-up transformer and intensified into high voltage, and is transmitted to multiple substations. At each substation, the high voltage electricity is reconstructed to a low state, which is best suited for real time consumption. This is done with a step-down transformer. Further it is distributed to the consumers for utilization. In General, distributed control system (DCS) and supervisory control and data acquisition (SCADA) are propriety control and monitoring systems used in power transmission and distribution. Kaplan et al. [6] provides a list of such monitoring sub systems, which can be used for monitoring the grid activities. In recent years a number of grid monitoring technologies have been developed. Some of them are listed below.

1) Grid monitoring system: Phasor measurement unit (PMU) enabled with GPS (global positioning system) measures the spontaneous magnitude of voltage and current from selected grid locations. It is then transmitted to the servers with timestamps. This opens up a global as well as dynamic perspective of the power system.

2) Backscatter radio: contributes improved data. Cautions about of transmission and distribution component failing.

The data collected from these sources can be utilized to analyze power system state estimation, a real-time stability determination and commence significant sequence of action. Power system state estimation is used to ensure the stableness of the grid and prevent blackouts. Likewise Chen et al. [5].Discuss calculation of power system state prediction via weighted least-square method. When some part of the network are detached from the grid Islanding will occurs and such events can leads to stability issues in grid. In [7] Naive-Bayes classifier is used to for islanding detection. A real time power quality assessment is much needed for power systems in which power disturbances like harmonics, swell and sag can affect the performance. Hongke and Linhai developed a data analysis framework for power quality, using SQL Server and OLAP, an Online Analytical Processing. In power systems grid fault identification, failure cause identification is other major issues. Mori made a survey over the smart power grid oriented papers, which deals with various applications over data mining on the power systems, like failure classification; examine the transient faults, etc.

Summary

The development of the Smart Grid concept is the pathway for assuring high reliability, control and management requirements in future electric power distribution systems. The Smart Grid can be defined as an electricity network supported by an intelligent infrastructure, both hardware and software, capable of accommodating high shares of Distributed Energy Resources. Within this line, a Smart Grid laboratorial infrastructure was developed, being dedicated to advanced research and demonstration activities. The adopted laboratorial architecture was developed according to the Microgram concept, where Electric Vehicles are regarded as active and flexible players. Following the laboratory implementation, this paper provides a detailed description of its infrastructure and experimental capabilities, presenting and discussing different experimental set-ups and associated results.

References

- [1] Swartz R.A, Lynch J.P, Zerbst S, Sweetman B, Rolfes R. Structural monitoring of wind turbines using wireless sensor networks. *Smart Structures and Systems* 6, 114, 2010.
- [2] Li J q, Wang S l, Niu C l, Liu J z. Research and application of data mining technique in power plant. In *Proceedings of the 2008 International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2, p. 250253, 2008.
- [3] Hong T. Short term electric load forecasting. Ph.D. thesis, North Carolina State University, USA, 2010.
- [4] Zhang H T, Xu F Y, Zhou L. Artificial neural network for load forecasting in smart grid. In *Proceedings of the 2010 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 6, p. 32003205, 2010.
- [5] Chen Y, Huang Z, Liu Y, Rice M J, Jin S. Computational challenges for power system operation. *Proceedings of the 2012 Hawaii International Conference on System Sciences (HICSS)*, p. 21412150, 2012.
- [6] Najy W, Zeineldin H, Alaboudy A K, Woon W L. A Bayesian passive islanding detection method for inverter-based distributed generation using ESPRIT, *IEEE Transactions on Power Delivery* 26, 26872696, 2011.
- [7] Hongke H, Linhai Q. Application and research of multidimensional data analysis in power quality. In *Proceedings of the 2010 International Conference on Computer Design and Applications (ICCD)*, vol. 1, p. 390393, 2010.