2013 IREP Symposium-Bulk Power System Dynamics and Control –IX (IREP), August 25-30, 2013, Rethymnon, Greece

The Role of Big Data in Improving Power System Operation and Protection

Mladen Kezunovic, Le Xie
Texas A&M University

Santiago Grijalva
Georgia Institute of Technology

Abstract

This paper focuses on the use of extremely large data sets in power system operation, control, and protection, which are difficult to process with traditional database tools and often termed *big data*. We will discuss three aspects of using such data sets: feature extraction, systematic integration for power system applications, and examples of typical applications in the utility industry. The following analytics tasks based on big data methodology are elaborated upon: corrective, predictive, distributed and adaptive. The paper also outlines several research topics related to asset management, operation planning, real-time monitoring and fault detection/protection that present new opportunities but require further investigation.

Introduction

By virtue of deploying a smart grid infrastructure the utility industry is facing new challenges in dealing with extremely large data sets, often called *big data*, and using them to improve decision-making. Work in [1] points out that big data in the electric power industry can be of large volume, high velocity, increasing variety, or all three. The volume of data has been growing significantly with introduction of new metering devices. Velocity refers to the temporal constraints on collecting, processing and analyzing data, which is the case with synchrophasor data. For efficient condition-based asset management and outage prevention real-time operation is necessary, often requiring fast processing of large volumes of data. Variety refers to data coming from many different sources that are not necessarily part of the traditional electric utility data.

As mentioned, typical examples of big data encountered by the utility industry include streaming synchrophasor measurement system (SMS) data used for situational awareness, and similarly extensive data sets collected by poling automated revenue metering (ARM) systems for billing purposes. Another type of big data stems from asset management and may consist of condition-based measurements acquired by intelligent electronic devices (IED) as well as nameplate and maintenance data entered off-line. Other very large data sets, not directly obtained through the utility field measurement infrastructure but widely used in decision making, such as weather data, data from the National Lightning Detection Network (NLDN), Geographic Information System (GIS) data, and electricity market data with different planning spans, are becoming more readily available. Several recent papers deal with the integration of GIS data [2-4], US NLDN data [5], and electricity market data [6], for power system applications. The new Hierarchically Coordinated Protection (HCP) approach is proposed in [7, 8]. This approach utilizes local and wide area measurements organizing them in three HPC levels: fault anticipation and prediction, adaptive fault detection, and relay operation correction in case of unwanted tripping. Big data plays an important role in the integration of variable renewable energy sources [9].

Data is typically not handled by utilities using a single, consistent, data management framework making ad-hoc utilization by novel decision making applications needlessly complicated. While data analytics may need to reach across diverse data sets, if the data is not correlated in time and space, if it does not have common data syntax and semantics to assure ease of use, and if it is not correlated to a unified and generalized power system model, such data analytics may not be easily implementable. Past experience, particularly from large blackouts that have affected the grid in recent years, has shown the need for better situational awareness about network disturbances such as faults and dynamic events, sudden changes of intermittent power from renewable resources such as wind generation, outage management tasks such as fault location and restoration, and monitoring of system operating conditions such as voltage stability. All of the mentioned tasks, and others, are handled reasonably well by existing solutions but improvements in decision-making are highly desirable in order to produce more cost-effective and timely decisions, facilitating more efficient and secure grid operation. This paper focuses on utilization of integrated data sets and systematic data processing to improve decision-making.

The paper first explains what the typical large data sets of interest in power system operation and protection are, then it discusses how they may be processed and managed, and finally examples of future applications are provided.

Types of Big Data Used in Power Systems

Big data sets and their relationship to sample applications are depicted in Fig. 1.

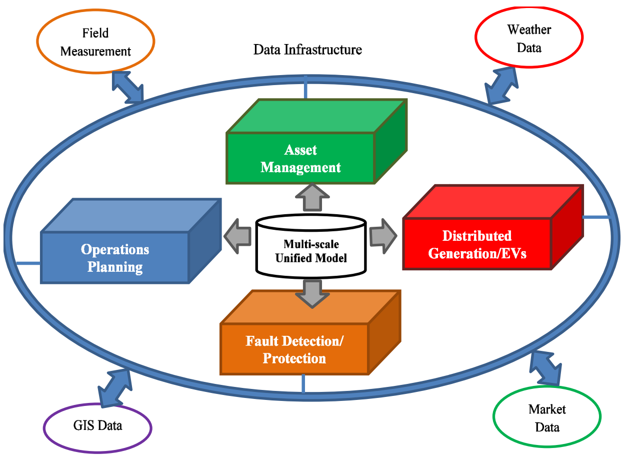

Figure 1. Big Data and applications

The deployment of phasor measurement units (PMUs) has provided operators the ability to measure instantaneous values of voltages and currents with accurate timestamps [10]. Since PMUs provide both magnitude and angle measurements they can be imported to state estimators to provide direct state measurement [11]. This is especially relevant when system topology such as automated breaker operation changes rapidly. PMU data can also be used for protective relaying applications [12]. The use of PMU data can aid in faster and more accurate tripping response. Improved communication and visualization would allow engineers to better analyze wide area events. Many power system oscillations may not be detected by the traditional slow-response SCADA system. Wide-area information can be utilized to implement new tools for better situational awareness and help operators determine the best response to an event, or allow systems to act autonomously in cases where human response time would be exceedingly slow [13, 14]. Such automated applications are increasingly significant in light of variable renewable sources which are likely to create more unpredictable stress on the existing long distance transmission lines connecting renewable sources from remote areas to the load centers.

The management of large data at the distribution level is also important in order to facilitate applications such as demand response, integration of distributed energy resources (DERs), and electric vehicles [15]. Turning loads into dispatchable resources and allowing DERs to participate in electricity markets will require improved and secure two way communications as well as better management of the associated big data.

Data Processing

Data processing steps that will be discussed in the paper are illustrated in Fig. 2 and are based on three important innovations: representation of multiscale models to allow the big data to be matched at an appropriate resolution, utilization of advanced data mining techniques to select the most relevant data features, and specification of the multi-domain graphics analytics to enhance decision-making. The technical approach discussion may be illustrated by the flow of data and associated data analytics as shown in Fig 2. The presented concept indicates that as one moves past the measurement collection and data conversion layers, the focus is on three key components of new data analytics: extraction of information features, the use of features to expand knowledge, and the use of knowledge to define the best visualization for facilitated decision-making.

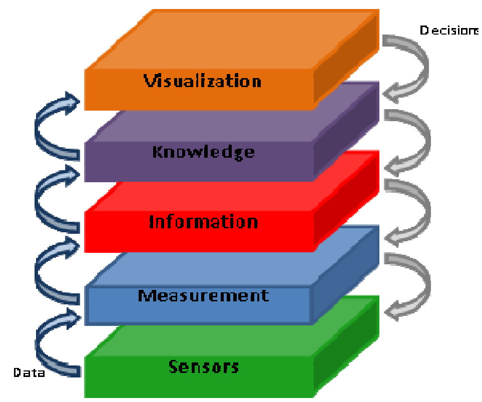

Figure 2. Data management steps

Data Correlation

To ensure that the relevant data may be correlated for application purposes, the paper will now discusses how such a correlation may be established in time and space. The considerations included in this discussion are shown in Fig. 3. For each relevant data set the discussion focuses on how the data may be correlated to the power systems or data model, and what the features of interest that need to be correlated are.

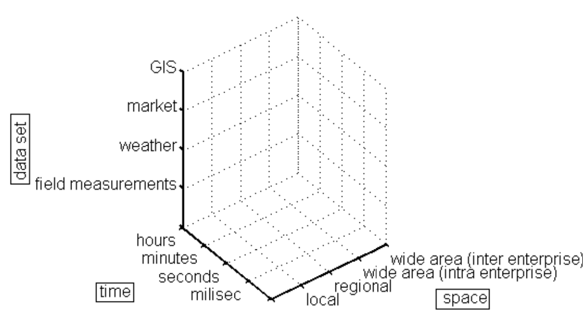

Figure 3. Spatiotemporal correlation of big data

With the use of a Geographical Information System (GIS) the collected data can be viewed in a geographic context. An important source of information for GIS is the Global Positioning System (GPS) which provides location and time information thus giving the data a spatial and temporal component.

In order to reduce the number of interruptions in service the power industry is interested in catastrophic weather conditions such as lightning, severe storms and winds, which can lead to outages in power system grid. This information is collected as a set of measurements from weather radar systems located at various locations across a region and it is commonly integrated and spatially indexed with GIS. Depending on methodology used to locate lightning strikes the temporal resolution ranges from 0.1 to 5 μsec and the source position is uncertain to within a few hundred meters [5].

Corrective, Predictive, Distributed and Adaptive Decision Making

The main focus of the decision-making framework is on the exploration of innovative computational concepts to allow novel applications:

a) Corrective: This concept facilitates just-in time actions aimed at rectifying undesirable conditions as soon as they develop. An example of such an application is risk-based asset management and maintenance optimization.

b) Predictive: This concept provides a very detailed forecasting of system behavior allowing for enhanced operations planning. An example of such an application is operations-planning convergence and interactions of renewable generation and mobile loads (EVs).

c) Distributed: This concept demonstrates an ability to assess system state based on a distributed processing so that fast control decisions can be executed locally. An example application is online assessment of voltage stability.

d) Adaptive: This concept enables operators to monitor unfolding events very closely allowing adjustments in operating strategy. An example of such an application is enhanced disturbance detection and on-line outage management.

The following sections illustrate research topics that may be pursued to get full use of big data in power system applications in the future.

Asset Management

This application combines large data sets from on-line condition based monitoring of power system assets with large data sets representing the asset model. We will focus on the example of using condition based data from a CB control circuit. The relevant measurements are illustrated in Fig. 4, including the trip initiate, trip coil current, and contact voltages, with associated timings τ_1 through τ_6 [16]. The explanations of these timings are included in Table 1.

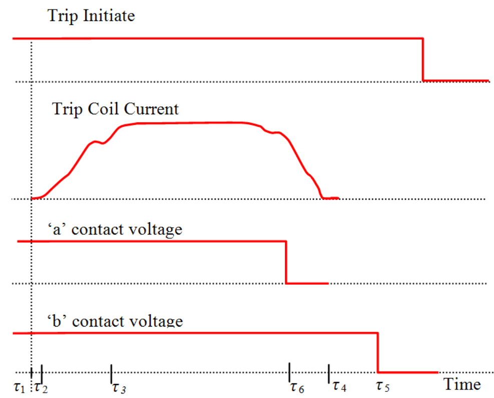

Figure 4. Signal features for CB monitoring signals

Records of timings, across many operations (tripping) of the circuit breaker and across many circuit breakers in the system, may be used to form probability distributions of these events.

Table 1. Explanation of signal features for CB monitoring signals

Event	Event Description	Signal
1	Trip or close operation is initiated (change from LOW to HIGH)	τ_1
2	Coil current picks up	τ_2
3	Coil current dips after saturation	τ_3
4	Coil current drops off	τ_4
5	B contact breaks or makes (changes from LOW to HIGH or vice versa)	τ_5
6	A contact breaks or makes	τ_6

Together with actual conditions of circuit breakers these historical records may be used to build statistical models of circuit breaker operating conditions. Comparing any particular timing to the adequate historical probability distribution related to that timing, as illustrated in Fig. 5, may be used to predict when a circuit breaker is malfunctioning or classify those circuit breakers which are close to malfunctioning (proper operation shown in blue).

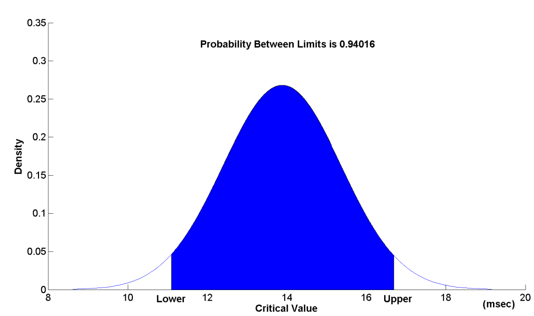

Figure 5. Historical distribution of timings τ_2

As illustrated in Fig. 5 if a statistical model of a timing, in this case τ_2 is well known, and the proper thresholds of failure, in this case the interval [11 s, 16.5 s], are available, it is possible to classify the operation of a circuit breaker according to the recorded timings.

Operations-Planning Convergence

This application refers to the utilization of unified models, facilitating massive data integration and use. The realized actual physical state of the power system is the result of a set of overlapping decisions that take place at different temporal scales, from long-term planning to operations planning to real-time control. This is illustrated in Fig. 6.

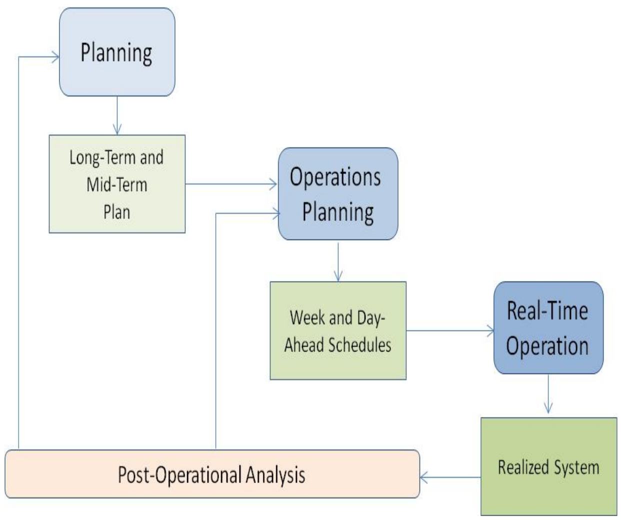

Figure 6. Planning, Operations Planning and Real-Time Operations Process

Planning provides long-term and mid-term plans, which are utilized by operations planning. Operations planning in turn provide week and day-ahead schedules to real-time operations, which are then delivered to the realized system. A post-operational analysis provides feedback in order to affect and enhance plans and schedules.

The term *operations-planning convergence* refers to the ability of a utility enterprise to realize the future conditions of the power system with high probability and high accuracy. This is difficult to achieve without systematic data management and unified models.

Inevitably, external conditions such as outages, errors in demand forecasting, and variability of renewable sources will result in a realized system that is different from the planned system. Besides external and intrinsic system deviations, there are several other reasons why the realized system is different from the planned system. Let us consider the historical records of the system for a certain time range [T_{start}, T_{end}]. We select this range to be large enough so that any decision-making uses information contained within the range. Within this large time range, let us consider the time range of one day:

$$d_{t0} = \{t : T_{start} << t_0 < t < t_0+24 << T_{end}\} \qquad (1)$$

In order to plan that day, there were a set of operational planning procedures that took place before hand. Let us denote the scheduled state of the system for a given time t by $s(t)$, and the realized state of the system by $r(t)$. The difference $r(t) - s(t)$ is the *convergence gap*. Let us denote by $S_{t0} = \{ s(t): t_0 < t < t_0+24\}$ the sequence of scheduled system states for the day ahead that starts at t_0, and by R_{t0} the corresponding sequence of realized states. The day-ahead schedule S_{t0} is a function of past realized system information and some forecasted information available at the time when the schedule was generated.

Assuming a deterministic scheduling procedure, we can utilize the realized information in the range [T_{start}, T_{end}] to reproduce the scheduling process that took place to generate a day-ahead schedule S_{t0}. If the scheduling process was perfect, the scheduled states would be identical to the realized states, e.g., $S_{t0} = R_{t0.}$ However, records for most utilities suggest that there may be significant differences between what was planned and the realization, even under assumptions of deterministic scheduling and certainty of demand and output from renewable sources. There are various reasons for this convergence gap, which can be overcome with the proposed unified methods and systematic data management approach:

a) Models: Traditionally, it has been difficult to compare operational and planning cases even within the same utility, due to the use of diverse modeling approaches by different applications. Thus it has been difficult to determine the extent of the convergence gap, except when discrepancies in individual elements, such as transmission lines flow, become obvious.

b) Format: The formats used by different applications may be incompatible, making comparisons on large data sets very difficult.

c) Data Management: Polling the data from the various subsystems necessary to reproduce planning procedures has been cumbersome. New systematic methods are required.

d) Computation: It has been difficult to set a simulation planning environment for convergence assessment.

e) Application Setting: Different applications that are fundamentally realizations of the same algorithm for operations and planning use different settings. For instance a power flow application may have different options for variable sharing in power plants for the automatic setting of LTC transformers.

The Role of Big Data in Convergence

Under a unified model approach the decision-making setting is illustrated in Fig. 7, where all applications access a unified model (or its database replica).

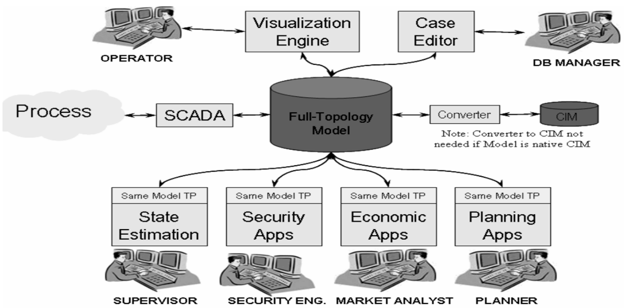

Figure 7. Proposed Unified Topology Model

Convergence of operation and planning means successfully managing the electricity infrastructure: the system was planned correctly, and the system operation took place as planned. Electric utility big data, large amounts of data from emerging sensors about the operation of the power infrastructure that is managed in a systematic and strategic manner, means new opportunities to provide a richer feedback loop for operations. These opportunities are identified at three levels:

a) Being able to provide an otherwise non-existent feedback loop.

b) Being able to take actions to correct and enhance planning or operational processes.

c) Being able to realize the planned system state.

Some of the features of big data that enable convergence are:

a) Availability of more data: more data enables more robust statistical or data mining analysis that allows increased process accuracy and enhanced control. More data requires better storage processes, virtualized storage systems and possibly distributed data centers.

b) New data, which allows creating new feedback loops for planning and operations. For instance, new PMU data can be used to enhance an existing state estimator or help determine model errors. New data must be managed using an integrated and flexible platform; new data is generated as raw facts from sensors, but also as by-products of more and enhanced decision-making functions. For example smart meter data may support enhanced load forecasting. Enhanced load forecasting may enable forecasted prices, etc.

c) Better management of data, which provides clear, actionable information. Traditional utility data has been managed in a server-centric architecture. Big data requires an information-centric architecture, where the storage and other IT resources are optimally shared. An information centric approach allows more flexible and adaptive management including emerging unstructured data.

d) Advanced analytics, to support complex decision-making.

Distributed Online Monitoring of Voltage Instability

Big data could be used for improving the situational awareness in large-scale system operations. As an example, the monitoring of quasi-static voltage collapse can be performed using SCADA data in a fully distributed manner [17]. Theoretically, the quasi-static voltage collapse can be associated with the near singularity of the power flow Jacobian matrix or its reduced order submatrices. There has been several centralized performance indices proposed in the literature for voltage collapse analysis and control, such as minimization of load voltage deviation, maximum loading point, and least damping ratio. There have also been discussions more recently on using second-order approximation of the saddle-node bifurcation surface for voltage instability detection.

In practice, online voltage stability monitoring is a critical function in modern control centers of interconnected power systems such as PJM, NYISO and ERCOT. System operators in control rooms typically use Transfer Limit Calculator (TLC), which simulates the MW transfer capability based on the latest system snapshot obtained by the state estimator, to assess the voltage stability. The TLC calculates the voltage collapse points based on a set of pre-defined interfaces, which include multiple high-voltage transmission lines. Typical scenarios are simulated to assess the power transfer from one side of the interfaces to the other, typically from generation surplus regions to load centers. The choices of interfaces are typically based on the operators' experience and lengthy off-line studies. However, such a centralized scheme of voltage instability monitoring may not be effective under a much broader set of operating conditions for the following reasons:

1) A reliable state estimation is the prerequisite of the TLC calculation. However, in practice global state estimator may not converge reliably when the system is close to stressed conditions/voltage collapse.

2) The pre-defined reactive interfaces may not necessarily be the actual weakest transfer interfaces in the system due to the change of operating conditions, such as the transmission outages and/or stochastic generation from renewable sources.

3) In order to obtain a reliable transfer limit, an accurate equivalence of external systems would be required. However, such accurate equivalence is either unavailable or results in a heavy computational burden.

In our preliminary work, with the decentralized Jacobian matrix J_k, we propose a performance index (PI) for decentralized monitoring of voltage collapse

$$PI_{dk} = \left\| \frac{\partial(1/\sigma_m)}{\partial S} \right\|_\infty \tag{2}$$

This PI represents the highest sensitivity of the smallest singular value (SSV) σ_m of J_k with respect to the regional apparent load vector $S = [S_1, S_2, \ldots, S_n]^T$. Theoretical justification of PI is provided in [17]. By overcoming the small value of widely used SSV of the Jacobian matrix, the proposed PI holds a large value near system singularity, which makes it easier for system operators to detect the voltage collapse.

Based on the theoretical justification of such a distributed PI, large-scale power system operators could develop online distributed monitoring of quasi-static voltage collapse by use of SCADA data. Simulation results on a 118 bus system show a promising opportunity of distributed analytics for voltage monitoring [17]. An entire interconnection can be decomposed into control areas based on their electrical couplings, which preserves the most critical information of system-wide power flow Jacobian matrix and includes only tie-line buses in both neighboring control areas. Then, each control area will solve a decomposed power flow Jacobian matrix, which significantly reduces the computational burden and complexity while maintaining the monitoring accuracy. Such data and online methods allow a fully distributed online monitoring of system-wide instability.

Stability Margin Prediction Using PMU Measurements

Several data mining tools such as Multi-Linear Regression, Neural Networks, Support Vector Machine and Regression Tree have been used to evaluate the system stability status [18-21]. We will focus on a regression tree-based approach to predicting the power system stability margin and detection of impending system events.

Voltage and oscillatory stability are targeted for monitoring. In the case of oscillatory stability the damping ratio (DR) of the critical oscillation mode is used as the stability margin indicator. The voltage stability margin (VSmargin) may be defined using the continuation power flow (CPF) technique. Instability occurs when the load attempts to step beyond the capability of the combined transmission and generation system.

The framework for RT-based stability margin prediction and event detection is shown in Fig. 8. After the PMU measurements are collected and time-aligned by the Phasor Data Concentrator (PDC), they are delivered to the Wide Area Measurement System (WAMS) server located at the central control facility. The regression trees for monitoring oscillatory and voltage stability are trained and updated periodically. Whenever the checking of corresponding thresholds indicates insufficient stability margin operators are alerted with the possible event and preventive control strategies can be initiated. The commercial software CART [22] is used to develop regression trees used for evaluating oscillatory and voltage stability margins respectively.

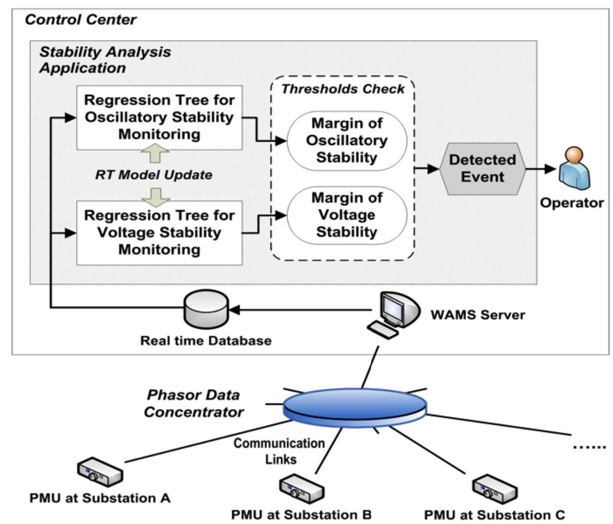

Figure 8. Framework of the RT-based stability margin prediction and event detection

Fault detection/protection

One example of problem associated with traditional protection schemes comes from the fact that the transient signal during a fault is non-stationary, containing fundamental frequency components, DC offset with damping, harmonics, etc. In some extreme situations, it will result in an inaccurate estimation of phasor representation of the faulted signal, which may cause misoperation in relays [23]. Under such conditions the thresholds may have to be recalculated, which is not feasible in real-time.

To overcome such problems intelligent methods such as neural networks and fuzzy logic should be explored. The neural network learning process consists of the procedure for converting the power system field data to information, which then can be processed to form the knowledge, which is based on the inter-neuron connection strengths, or synaptic weights. In order to integrate artificial neural networks with power system field data the following questions should be answered: what data is required for neural network based fault detection and classification; what kind of model should be used to convert the data to information; and how the translation from information to knowledge leads to improved decision-making.

Neural network training and testing environment

The learning techniques for neural networks can be classified into two broad categories: supervised learning and unsupervised learning [24]. In supervised learning, each input signal is associated with the labeled output. The task of learning a correct mapping of inputs to outputs is that of adequately adjusting the synaptic weights to minimize the overall error between the output patterns and their corresponding input patterns, across the entire set of data designated for training. In unsupervised learning, the categories of the outputs are not known in advance. The network is used to self-organize data by clustering to identify concentrations of input patterns maximizing mutual similarity within a cluster and mutual differences between clusters. A neural network based on a combined unsupervised/supervised training scheme has proven to be more capable of handling large data sets of random fault scenarios than solely using supervised training schemes [25]. The input data vectors of samples from voltage and/or current signals are mapped into clusters that contain information about fault existence and type. Fig. 9 shows the training environment process.

During the testing procedure, distances between each test pattern and established clusters are calculated. The outcomes of testing are class labels assigned to test patterns according to the most common value among the *K* nearest prototypes.

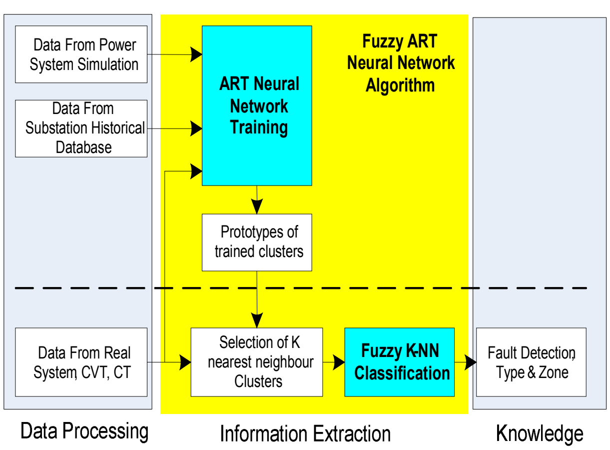

Figure 9. Fuzzy ART neural network algorithm for fault diagnosis

The input into the neural network is in the form of a moving data window containing samples of voltage and current signals from local measurements and simulations. Patterns are extracted from these data and placed together in one row to form feature vector components, as shown in Fig. 10. Then the pattern is arranged using the post-fault samples of three phase voltage and current signals. The zero sequence values of voltage $3v_0 = v_a + v_b + v_c$ and current $3i_0 = i_a + i_b + i_c$ are also included to precisely detect ground faults. In this case, all fault types can be differentiated very well [24]. Thousands of such patterns obtained from power system simulation or substation databases of field recordings are used to train the neural network offline, and then the recovered pattern prototypes are used to analyze faults on-line by using the Fuzzy *K*-NN classifier.

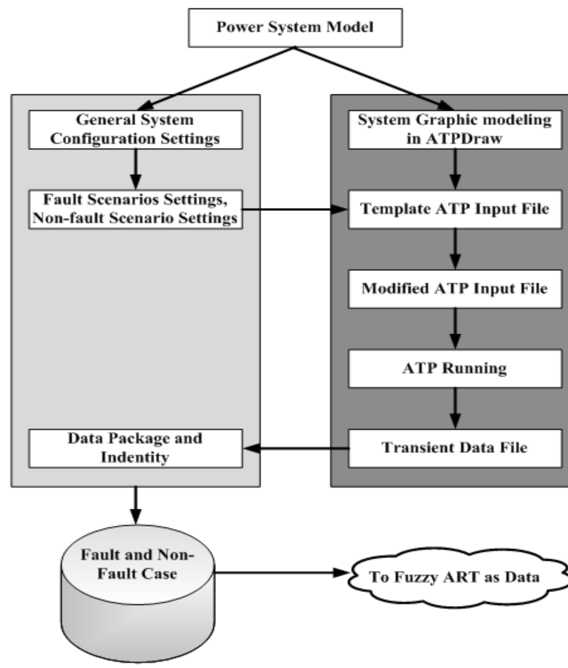

Figure 10. Simulations for fault and non-fault scenarios

Data Processing

The neural network algorithm requires a large number of fault and non-fault cases to complete the process of training and testing for neural network tuning. Those training and testing cases are quite different for various transmission lines due to the selection of different simulation parameters and settings. To perform comprehensive tests, two categories of data can be acquired: field signal measurements and data from simulated cases. As it is not possible to acquire enough fault cases from the field, the needed data can be generated by simulation of the fault and non-fault scenarios, based on ATP/ATPDraw [26] and MATLAB [27]. The power system of interest is first modeled in ATP/ATPDraw. A user can define the desired fault or non-fault cases by initializing the simulation setting parameters in MATLAB. The measured three-phase voltage and current samples, which may also include a time stamp, are extracted in the data format files defined by a user.

Feature Extraction

One of the tasks that is often time consuming, to an extreme extent when dealing with big data, is that of distinguishing between important and unimportant attributes. It is often desirable to remove from consideration measurements that are not helpful in the decision-making process. Automated reasoning tools, in support of the decision-making process, often suffer severe penalties in both accuracy and time complexity when applied to data with many unnecessary attributes. However the methodology of distinguishing, within an often large amount of measurements, between those which are useful and should be included in decision-making, and those which are not, is difficult to objectively perform in ad-hoc fashion.

Let us describe one powerful method of removing from consideration measurements that are not useful to decision-making. Starting from the full set of measurements we remove those for which a good Market Blanket approximation can be made using other features. The intuition behind this method is to keep in consideration only those measurements which either cannot be derived from other measurements or which are not used to derive other measurements, while pertaining to the task at hand. The Hilbert-Schmidt Independence Criterion may be used to determine dependence between measurements in kernel space for this purpose. In [28] this method, solidly grounded in theory, has been successfully applied to select among 850,000 measurements those which are useful.

Detecting and Classifying Faults by Mapping Data to Labeled Clusters

The purpose of the information processing is to form the knowledge, which could allocate the training patterns into homogeneous clusters by a grouping technique. Then the clusters are assigned to classes, which are in our case the expected fault events in the power system, such as fault type, etc. The number of clusters is increased and their positions are updated automatically during learning, and there is no need to define them in advance. Fig. 11 shows the raw training patterns as information on the left and the clusters after the training processing, as knowledge on the right. The typical types of classification are based on detecting the fault type and fault zone. The classification of mapping data to labeled clusters is performed by using the *K*-nearest neighbor rule (*K*-NN) [25].

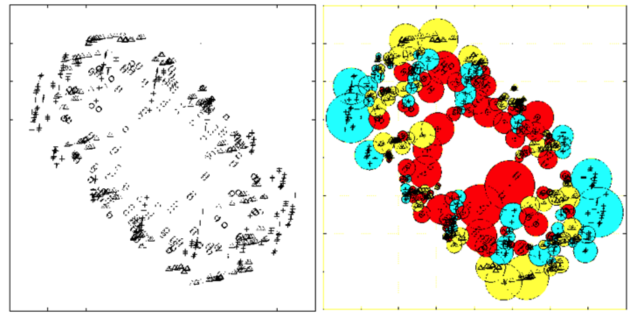

Figure 11. The raw training patterns (information) and the patterns allocated to the clusters after a training processing (knowledge)

Conclusions

This paper provides several contributions:

- It introduces the relevance of the big data, the potential uses, and expected benefits
- It outlines the critical steps in big data processing and establishes a decision-making framework
- It illustrates the proposed concepts with several new applications that have tangible benefits
- It offers a pathway to problem formulation and research methodologies for Big Data utilization

References

[1] IBM Software, “Managing big data for smart grids and smart meters,” May 2012.

[2] J. Syllignakis, C. Adamakis, T. M. Papazoglou, “A GIS web - application for power system of Crete,” *Universities Power Engineering Conference, 42nd International*, pp. 414-418, Sept. 2007.

[3] X. Q. Li, Z. Y. Zeng, Y. C. Zhang, X. J. Xu, “A Study of Distribution Load Transfer Operation Based on GIS,” *International Conference on Machine Learning and Cybernetics*, pp. 1428-1433, Aug. 2007.

[4] A. N. Sekhar, K. S. Rajan, A. Jain, “Spatial Informatics and Geographical Information Systems: Tools to Transform Electric Power and Energy Systems,”, *TENCON 2008 - 2008 IEEE Region 10 Conference*, pp. 1-5, Nov. 2008.

[5] K. L. Cummins, E. P. Krider, M. D. Malone, “The US National Lightning Detection Network™ and applications of cloud-to-ground lightning data by electric power utilities,” *IEEE Trans on Electromagnetic Compatibility*, Vol. 40 , No. 4, pp. 465-480, Nov. 1998.

[6] T. Niimura, M. Dhaliwal, K. Ozawa, “Fuzzy regression models to represent electricity market data in deregulated power industry,” *IFSA World Congress and 20th NAFIPS International Conference, Joint 9th*, pp. 2556-2561, July 2001.

[7] B. Matic-Cuka, M. Kezunovic, "Improving Smart Grid Operation with New Hierarchically Coordinated Protection Approach," *The 8th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2012)*, Cagliari, Italy, October 2012.

[8] M. Kezunovic, B. Matic Cuka, "Hierarchical Coordinated Protection with High Penetration of Smart Grid Renewable Resources (2.3)" *PSERC Future Grid Initiative*, 2013.

[9] D. E. Bakken, A. Bose, C. H. Hauser, D. E. Whitehead, and G. C. Zweigle, "Smart Generation and Transmission With Coherent Real-Time Data," *Proceedings of the IEEE* , vol. 99, no. 6, pp. 928 - 951, June 2011.

[10] A. Bose, "Smart Transmission Grid Applications and Their Supporting Infrastructure," *Smart Grid, IEEE Transactions on*, vol. 1, no. 1, pp. 11 - 19, June 2010

[11] A. Abur and F. Galvan, “Synchro-phasor assisted state estimation (SPASE),” *Proc. IEEE PES Innovative Smart Grid Technologies (ISGT)*, pp. 1 – 2, Jan. 2012.

[12] J. O’Brien et al., “Use of Synchrophasor measurements in Protective Relaying Applications,” Power System Relaying Committee, Report of Working Group C-14 of the System Protection Subcommittee, 2012. [Online] Available: http://www.pes-psrc.org/c/C14/Use of Synchrophasor Measurements in Protective Relaying Applications_draft 6_3.pdf

[13] Y. Zhang et al., “Wide-Area Frequency Monitoring Network (FNET) Architecture and Applications,” *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 159 - 167, Sep. 2010.

[14] A. Silverstein, J. E. Dagle, "Successes and Challenges for Synchrophasor Technology: An Update from the North American SynchroPhasor Initiative," *45th Hawaii Int. Conf. on System Science (HICSS)*, pp. 2091 - 2095, Jan. 2012.

[15] F. Rahimi, and A. Ipakchi, "Demand Response as a Market Resource under the Smart Grid Paradigm," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 82 - 88, June 2010.

[16] M. Kezunovic, Z Ren, G. Latisko, D. R. Sevcik, J. Lucey, W. Cook, E. Koch, “Automated Monitoring and Analysis of Circuit Breaker Operation,” *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp 1910 – 1918, July 2005.

[17] L. Xie, Y. Chen, and H. Liao, “Distributed Online Monitoring of Quasi-static Voltage Collapse in Multi-area Power Systems,” *IEEE Transactions on Power Systems*, Vol. 27, No. 4, pp. 2271-2280, Nov. 2012.

[18] L. Bruno and V. Ajjarapu, “Development of multi-linear regression models for online voltage stability margin estimation,” *IEEE Trans. Power Syst.*, Vol. 26, No. 1, pp. 374-383, Feb. 2011.

[19] D. Q. Zhou, U. D. Annakkage, and A. D. Rajapakse, “Online monitoring of voltage stability margin using an artificial neural network,” *IEEE Trans. Power Systems*, Vol. 25, No. 3, pp. 1566-1574, Aug. 2010.

[20] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, “Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements,” *IEEE Trans. Power Systems*, Vol. 26, No. 3, pp. 1474-1483, Aug. 2011.

[21] C. Zheng, V. Malbasa, M. Kezunovic, “Regression tree for stability margin prediction using synchrophasor measurements,” *IEEE Trans. Power Syst.*, Vol. 28, No. 2, pp. 1978-1987, May 2013.

[22] Steinberg, Dan and Mikhail Golovnya, CART 6.0 User’s Manual. San Diego, CA: Salford Systems, 2006.

[23] J. S. Thorp, A. G. Phadke, S. H. Horowitz, J. E. Beehler, “Limits to impedance relaying,” *IEEE Trans. Power Apparatus Syst.* ,vol. 98, no. 1, pp. 246–260, Jan./Feb. 1979.

[24] Zhang, Tong, and Rie Ando. "Analysis of spectral kernel design based semi-supervised learning." *Advances in neural information processing systems*, Vol. 18, 2006

[25] S. Vasilic, M. Kezunovic, “Fuzzy ART Neural Network Algorithm for Classifying the Power System Faults,” *IEEE Transactions on Power Delivery*, Vol. 20, No. 2, pp. 1306-1314, April 2005.

[26] Alternative Transients Program, ATP-EMTP, 2010. [Online]. Available: http://www.emtp.org/.

[27] Mathworks Inc. MATLAB R2012b User’s Guide. [Online]. Available: http://www.mathworks.com.

[28] Q. Lou, H. P. Parkman, M. R. Jacobs, E. Krynetskiy, Z. Obradovic, “Exploring Genetic Variability in Drug Therapy by Selecting a Minimum Subset of the Most Informative Single Nucleotide Polymorphisms through Approximation of a Markov Blanket in a Kernel-induced Space,” *Proc. IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, San Diego, CA, pp. 156-163, May 2012.