Between Big Data and Analytics

IEEE PES Webinar
November 26, 2018
For presentation but not distribution
PES edited version 10/1/18
Updated 11/25/18

Objectives

Inform Big Data and Analytics professionals working with electric power systems of some of the typical problems in implementation. The technical management of these projects is often as important as the math

Potential Audiences

Electric Power Utility algorithm developers, and suppliers to that industry, who would benefit from an informal view of the non-mathematical aspects.

Lessons Learned

Issues that may be encountered by a team of people, the I&C and SCADA engineers, the power engineers, the data scientists, and the IT support groups, on the journey to create successful big data projects and the subsequent analyses

New Contributions

A condensed survey from the electronic, numerical analysis, information technology, and business impact perspectives of topics related to successful big data manipulation and analytics

Premise:

First we had SCADA.

Followed by BIG DATA.

Then we had analytics.

Next, the arrival of Data Scientists

Successful projects need the whole picture

Hypotheses

Smart Grid is Internet of Things 101 for the electric power grid

What your analytic sees in the IoT world may not be what you think

SCADA data is just the beginning

Add the velocity of PMUs, the volume of unstructured data, the veracity of the information

Garbage In, Garbage Out, Goes Global

- A common theme is Analytics and Optimization
  - This involves data from the real world
  - Not all people are familiar with the risks with equipment
- The questions to be asked include
  - What is the provenance of this data ?
  - What does it really measure ?
  - What can affect the measurement, other than the one particular physical parameter I think the data represents ?
- Once I'm convinced the data is useful, and worth analyzing, and then even good enough for making optimization decisions based upon it
  - How can I do any validation of what I think I know, to what is really happening ?
  - *"Reality…what a concept"*
    - *Title of a Robin Williams 1978 comedy album*

Lessons Learned

- The data the utility RFP writers thought was available may not be at implementation time
  - Or not at least easily
  - Taxonomy differences between systems holding the data
  - Availability of data from systems no longer on-line but ‘available somewhere’
    - Meta data disparities from these previous systems
    - Physical ability to read historical data

Lessons Learned

- Cognitive systems can recognize unstructured data
  - From written field service and maintenance reports
  - From visual
    - Drone inspections
    - Thermal imaging cameras, satellite, LIDAR, helicopter
- Thousands of utility engineering hours may go in to system models that are stored in proprietary applications
  - Costing more to use that information in external analytics projects

Freedom to the Data

- Analytics assume you know everything to be analyzed
- There should be a discovery mode
  - Read-only data for all on the intranet
    - You should not assume members of just one group will have that “ah hah” moment
      - ODBC/JDBC
      - Corporate standard tools
        - » Excel
        - » Mat Lab
        - » SPSS
        - » SAS
        - » Cognos
  - Peoples’ visual pattern recognition skills are not to be underestimated
    - Thus, the User Interface is critical

IEEE PES Power & Energy Society® IEEE

Total Data Cost of an Application

- Installation of new major application
  - Software and services from vendor
  - System Integration costs
  - Security costs – for new integrations; for overall project
    - “Throw it over the wall” isolation
  - Consider effort by vendor to help utility organize its data before vendor can import it
    - Plus effort by utility to support that data organization / cleansing
  - Performance of application if it becomes an unwilling data server
    - May not be possible if crossing IT / OT security boundaries

Total Benefit of Data Organization

- Utility engineers do it once
  - Not hidden on a per-project basis
  - Change propagation and review
    - Forward consistency
- Utility owns tooling and data and format
  - Better if something CIM related
  - Consistency checker; data dictionary; taxonomy; mapping
  - Automate (Cognitive, AI) most of the unstructured data prep
- Utility puts forth tooling to application vendors as part of RFP process
  - Reduce vendor's cost and schedule risks
    - Vendors may be adding their typical extra cost to their response

Data Origination

- Meta data for all points/tags should include a pointer to a person's name or group
  - Who "owns" the data in the sense that they are responsible for its accuracy
  - Who knows the data to validate sensor readings
  - Who can dissect a calculated value and determine a suspicious input

Value Your Engineers

- By the time a corporate initiative has started, some people may already doing it
  - Find the Mat Lab, SAS, SPSS, Cognos, …, users, and ask them what they are doing on the O&M side
    - You may already have your first valuable analytics defined
      - Time to take them to the enterprise level
      - The basic equations are often in college textbooks
        - » It is the constants in the equations that make the analytic relevant to your utility
        - » It is the rules of the exceptions that make the analytic reliable and robust

Listen to the Customer (Meter)

- AMI provides information on the product you are delivering: quality electrons
  – Take all of the smart meter data, look for anomalies
    - Relative harmonic amplitude
    - Momentary outages
    - Voltage profiles
  – Then ask why
- Of course, you are also looking for frequent outage and maintenance patterns

Further Considerations Pre-Project

- If you believe there may be value in the data, even if you don't see it now, save it!
- Cloud storage may be your friend. Security? Encode, rename, obfuscate, Hybrid Cloud
- Organize, time stamp, add known meta-data now
  - Time compression may not be optimal for real-time data
- Find older on-line systems and give them as much storage as possible
  - Work out how to extract data and save from loss due to limited storage capacity devices
- Everything doesn't have to come back in real time
  - Store it on the edge as long as it can be commanded back

Right Data, Wrong Analytics

- Do the equations in the analytics apply to the whole range of possible measured values?
  - State change, chemical kinetics, combustion, material deformation, …
- If differential equations are involved

$$\frac{d^2 i(t)}{dt^2} + \frac{R}{L}\frac{di(t)}{dt} + \frac{1}{LC}i(t) = 0$$

  - How sensitive is it to (possibly erroneous) sudden erroneous changes in sensor value ?
    - The sensor system susceptibility to problems is usually time invariant, so the time step in the solver can make a big difference
      - Glitch results in voltage data set {500, 497.8, 501.3, 999, 503.4, 498.3, …}
      - If Δt in the solver is 2 seconds, then dV/dt is 248.85
      - If Δt in the solver is 100 ms, then dV/dt is 4977

Computational Delinquency

- When deploying an iterative solver, can the solution be done in all cases before next data set comes in ?
  - If Δt is two seconds, then solver should be done in all cases within two seconds
  - Limits on iterative solvers re: convergence

Data Time Skew

- If three phase voltage and current going in to and out of a transformer are measured, then any classic calculation on them assumes all 12 measurements are taken simultaneously
- If the values are not time aligned, one doesn't have much hope for the underlying physics
- Steady state happens mostly in the lab

To Image Process or Not:

Is the water boiling?

Do you see bubbles ?

Energy input: Current through the burner over time (is the pot in place?)

Measure water temperature (consider altitude ?)

“Rube Goldberg” vs. Dilbert / KISS / Sensor cost versus CPU cost

To Image Process or Not

Some visual recognition areas of current interest:

Some people like problems solved with a large sparse matrix of partial differential equations, because that's what they did in grad school

However, the multitude of simpler, passive devices cause much of the maintenance spend

Therefore could have the highest RoI

What if you could determine pole tilt, trees close to distribution lines, insulator condition, storm damage AHEAD OF TIME?

Cognitive vision can help

Coupled with clues from SCADA data

How do your customers like a truck roll just for inspection during a major outage, then two days later after a damage assessment is done, the trucks are back to do the repairs?

Whether to use Weather

Weather prediction is quite advanced today

- Impacts loads forecast
- Affects equipment damage, especially above ground distribution lines. But don't dismiss underground flooding in urban areas
- Determines wind turbine and solar panel output
- Wildfires

Weather prediction is sensor and simulation based

- It is an IoT input

Non-Technical Concerns

- Installation cost of sensors, their wiring, and connectivity, are often the largest economic deterrent to more sensing
- Outside of a plant environment, wired sensing often isn't practical
- Sometimes the economical answer to classic manual gauges is remote vision

- Data Privacy
- Security - hack the sensors and mislead the operators
  - The field systems are more "out of sight, out of mind" than the central ones
- Third party sensors
  - Certification / trust

Non-Technical Concerns

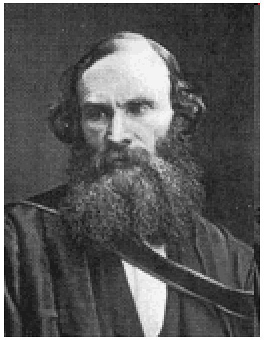
(Not legal statements, just questions to consider)

- Potential Liability
  - Perhaps you might have known it could explode, since you monitored the system. Could someone say you were negligent in not having an analytic to detect it?
- Regulatory
  - Might someone ask for your data that shows your plant operated in violation of license or statutes?
- Supply Chain
  - Our plant data is proof that XYZ pumps don’t meet spec
- Plant Operations
  - The data from a remote monitoring system could show you ran the plant out of manufacturer’s guidelines, so your warranty claim is rejected

Short Version of Sensing

Data Scientists Didn't Invent Data

Lord Kelvin



"In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of *Science*, whatever the matter may be."

Popular Lectures and Addresses (1891-1894, 3 volumes) , vol. 1, "Electrical Units of Measurement", 1883-05-03

IEEE PES Power & Energy Society®

IEEE

Sensors

- In general
  - Find a phenomena that changes an electrical property when the physical property changes
  - Convert that electrical parameter in to a digital form
    - Analog to Digital Converter
  - Transmit that digital data to a computer
  - Apply an equation in the computer to transform the number back in to the physical quantity you measured
    - Engineering units conversion
  - Use the engineering datum appropriately (range, scan rate, error propagation)

Classic Sensors

- Measure a physical property
  - Temperature
    - RTD, Thermocouple
  - Pressure
    - Direct
    - Strain gauges
  - Flow
    - Often as $\sqrt{\Delta P}$
  - Electrical Voltage and Current
    - Easy transition

More Interesting Sensors

- Light
  - Light beam weighs very little, thus little impact on measured item
  - Often needs more local preprocessing
  - Examples
    - Embedded sapphire fiber optic in metal
      - Optical properties changed by heat, stress
    - Electricity passing through optical fiber loop

Don’t Forget the Person

- Human Machine Interaction
  - The user does not understand the correct data you have presented
    - http://www.edwardtufte.com/tufte/courses

- The facts about the process or the instrumentation are wrong / out of date / mislabeled but the utility people ‘just know’ what is right
  - Essential to have in your database a person who can validate the data, if your analytic seems to be misbehaving. Find your I&C group
- Maintenance operations
  - While fixing one part of the plant, something else may become disconnected, damaged, or reconfigured wrong
    - The analytics “tag out”

Conclusion

Analytics Require Process, Instrumentation, and Numerical Analysis Design, Software Development, GUI, User Composition

- Just sensing a simple physical measurement isn't so simple
- Reliable and secure communication of the data is important
- Recognize data problems before processing in to the results from an analytic

Some IBM Resources

https://www.ibmbigdatahub.com/blog/shorten-your-path-ai-watson-knowledge-catalog

https://www.ibm.com/developerworks/library/ba-data-becomes-knowledge-1/index.html

https://www.ibm.com/developerworks/library/ba-intro-data-science-1/index.html

https://www.ibm.com/developerworks/library/ba-cleanse-process-visualize-data-set-1/index.html

https://www.ibm.com/us-en/marketplace/geospatial-big-data-analytics

https://developer.ibm.com/dwblog/2017/overview-graph-database-query-languages/

https://www.ibm.com/us-en/marketplace/infosphere-info-server-for-datamgmt

https://www.ibm.com/us-en/marketplace/text-analytics-for-data-science

https://www.ibm.com/products/watson-explorer

https://www.ibm.com/us-en/marketplace/decision-optimization-for-watson-studio

Some IBM Resources Suggested by IEEE PES Big Data Subcommittee

IBM Big Data & Analytics Hub: Energy & Utilities

https://www.ibmbigdatahub.com/category/925/whitepapers

White Papers & Reports

https://www.ibmbigdatahub.com/whitepaper/managing-big-data-smart-grids-and-smart-meters

Use Cases

There are industry products designed to enable IT to leverage big data in a variety of ways that can contribute to the success of energy companies. Capabilities include time-series data flow, streaming data analysis, data security, data warehousing, archiving, data mining and reporting. Each of the following power and utility industry use cases presents technical challenges:

- Managing smart meter data
- Monitoring the distribution grid
- Optimizing unit commitment
- Optimizing energy trading
- Forecasting and scheduling loads

Industry Big Data Solutions

- High Performance Time Series databases to capture and load smart meter data as part of a meter data management system
- Deploy Data Management solutions for archiving meter data to comply with mandated retention periods
- Use layered security software on top of third party data bases to help ensure customer information privacy
- Employ Complex Event Processing to process and analyze data in motion
- Leverage data warehousing appliances and solutions for insights as the stores for deep analytics and customer behavior analysis

A Use case: quantized results

Table below shows some of the performance and storage efficiency improvements the utility realized with a specialized Time Series database, versus an MDMS, including dramatically reducing the amount of time required to perform critical data analysis and storage tasks

Challenges	Other relational database	Specialized Time Series
Load time for 1 million meters	7 hours	18 minutes
Time to run required reports	2 to 7 hours	25 seconds to 6 minutes
Storage required for 1 million meters	1.3 TB	350 GB

Source: http://www-935.ibm.com/services/multimedia/Managing_big_data_for_smart_grids_and_smart_meters.pdf

How does big data change utility daily work?

Using predictive analytics on their data, companies can make a wide range of forecasts such as:

- How much excess energy will be available, when to sell it and whether the grid can transmit it
- When and where equipment downtime and power failures are most likely to occur
- Which customers are most likely to feed their generated energy back to the grid, and under what circumstances
- Which customers are most likely to respond to energy conservation and demand reduction incentives
- How to manage the commitment of larger, traditional plants in a scenario where peaks from distributed generation are becoming relevant

How does big data change utility daily work?

Integrated data management solutions and analytics engines enable energy and utility companies to:

Analyze a variety of information

- Apply novel analytics to a broad set of mixed information
- Analyze mixed (structured and unstructured) data sets that could not be analyzed before

Analyze information in motion

- Perform streaming data analysis
- Easily handle large volume data bursts

How does big data change utility daily work?

Analyze high volumes of information

- Cost-efficiently process and analyze petabytes of information
- Manage and analyze structured, relational data to extract value, discover, and experiment
- Perform ad hoc analysis

How does big data change utility daily work?

Discover and experiment

- Perform ad hoc analysis
- Enable data discovery and experimentation
- Gain insight and value from large volumes of low-economic-value data

Manage and plan

- Enforce data structure, integrity and control
- Ensure consistency for repeatable queries

To Go Further

- IEEE PES Utility Big Data Workshop
  - September, 2017
  - http://2017isap.tamu.edu/ieee-utility-big-data-workshop/
- Big Data Application in Power Systems
  - https://www.elsevier.com/books/big-data-application-in-power-systems/arghandeh/978-0-12-811968-6