The Role of Big Data Analytics in Future Control Centers

Mladen Kezunovic, Regents Professor
Texas A&M University
May 15, 2019
Outline

Expectations
Big Data Properties
Data Science & Big Data Processing Infrastructure
Examples: Predicting outages
- Transmission
- Distribution
Takeaways
Expectations
Big Data Properties

Volume
Size of Data

Velocity
The Speed at which Data is Generated

Variety
Different type of Data

Veracity
Data Accuracy

Value
Useful Data

Validity
Data quality, Governance, Master Data Management on Massive

Variability
Dynamic, Evolving Behavior in Data Source

Venue
Distributed Heterogeneous Data from Multiple Platforms

Vocabulary
Data Models, Semantics that describes data Structure

Vagueness
Confusion over Meaning of BigData and Tools used

---

**Big Data Properties**

- **Volume**: Size of Data
- **Velocity**: The Speed at which Data is Generated
- **Variety**: Different type of Data
- **Veracity**: Data Accuracy
- **Value**: Useful Data
- **Validity**: Data quality, Governance, Master Data Management on Massive
- **Variability**: Dynamic, Evolving Behavior in Data Source
- **Venue**: Distributed Heterogeneous Data from Multiple Platforms
- **Vocabulary**: Data Models, Semantics that describes data Structure
- **Vagueness**: Confusion over Meaning of BigData and Tools used

---

**Time Scales**

- **NanoSecond**
- **MicroSecond**
- **MilliSecond**
- **Second**
- **Minute**
- **Hour**
- **Day**
- **Month**
- **Year**

---

**Life Span of an Assets**

- **T&D Planning**
- **Demand response**
- **Service restoration**
- **Wind and solar output variation**
- **Protective relay operations**
- **Dynamic system response (stability)**
- **Weather**
- **Hour-ahead scheduling and resolution of most renewables integration studies**

---

**Units of Measurement**

- **Big Data Properties**
- **Volume**: Byte, Kilobyte, Megabyte, Gigabyte, Terabyte, Petabyte, Exabyte, Zettabyte, Yottabyte
- **Velocity**: Clock accuracy, Absolute accuracy of GPS time stamp, 256 samples per cycle, 1 cycle, 12 cycles

---

**Texas A&M University**

©2019 Mladen Kezunovic, All Rights Reserved
Data Science & Processing Infrastructure
Example:

Big Data Analytics for T&D Outage Prediction
Cause of outages

Major causes of power outages in the U.S.

- Weather/Tree-related: 62%
- Equipment failure: 15%
- Unknown/Other: 10%
- Public or Animal contact: 7%
- Power Grid failure: 5%
- Maintenance: 1%

Number of outages

- Storms and severe weather
- Cold weather and ice storms
- Hurricanes and tropical storms
- Tornadoes
- Extreme heat and wildfires

©2019 Mladen Kezunovic, All Rights Reserved
Impact

International Electricity Grid Reliability

Customer outage minutes per year

= 1000 TWh/yr Consumption

Lithuania
Portugal
United States
Australia
Hungary
Czech Republic
New Zealand
Spain
UK
France
Ireland
Austria
Finland
Netherlands
Japan
Denmark
Urban China

GDP per capita (USD)

5,000 15,000 25,000 35,000 45,000 55,000 65,000

Source: The Brattle Group, Galvin Power Institute, Council of European Energy Regulators, China Southern Power Grid

Annual Business Losses from Grid Problems

Primen Study: $150B annually for power outages and quality issues

SMART GRID CENTER
TEXAS A&M ENGINEERING EXPERIMENT STATION

©2019 Mladen Kezunovic, All Rights Reserved
Weather Driven Risk Analysis

\[ \text{Risk} = \text{Hazard} \times \text{Vulnerability} \times \text{Economic Impact} \]

- Probability of hazardous weather conditions
- Depends on Weather Forecast
- Pick a moment in time (or a period of time) and estimate probability of hazardous conditions

- Probability that hazardous conditions will cause an event in the network
- Depends on Historical Weather and Outage Data
- Learn from the historical data what may happen if hazardous conditions occur

- Expected economic impact in case of an event
- Depends on the type of economic loss that the user wants to consider
- Identify type of economic loss that is of interest for the study and calculate it
Results – Risk Maps

Weather Hazard

Network Vulnerability

Risk Map

©2019 Mladen Kezunovic, All Rights Reserved
Spatiotemporal Predictive Model

TIME

Future

Event 3

Event 2

Event 1

SPACE

High risk

Low risk

Different types of measurements

Prediction

Training Set
Predictive Data Analytics

To reduce the computational complexity of learning and inference, $A$ and $I$ can be constructed as quadratic functions of $y – \text{Gaussian Conditional Random Fields}$

$$P(y|X) = \frac{1}{Z} \exp(- \sum_{i=1}^{N} \sum_{k=1}^{K} \alpha_k (y_i - R_k(X))^2 - \sum_{i,j}^{L} \sum_{l=1}^{L} \beta_{l} e_{ij}^{(l)} S_{ij}^{(l)}(X)(y_i - y_j)^2)$$

BD Data Aggregation

Utility measurements

Weather Forecast

Animals Data

Network Assets Data

Lightning Data

Vegetation Indices

Market data

GIS

UAS
### BD Data Properties

<table>
<thead>
<tr>
<th>Source</th>
<th>Data Type</th>
<th>Temporal Resolution</th>
<th>Spatial Resolution</th>
<th>Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated Surface Observing System</td>
<td>Land-Based</td>
<td>1 min</td>
<td>900 stations</td>
<td>Air Temperature, Dew Point, Relative Humidity, Wind Direction, Speed and Gust, Altimeter, San Level Pressure, Precipitation, Visibility…</td>
</tr>
<tr>
<td>Level-2 Next Generation Weather Radar</td>
<td>Radar Data</td>
<td>5 min</td>
<td>160 high-resolution Doppler radar sites</td>
<td>Precipitation and Atmospheric Movement</td>
</tr>
<tr>
<td>NOAA Satellite Database</td>
<td>Satellite Data</td>
<td>Hourly, daily, monthly</td>
<td>4 km</td>
<td>cloud coverage, hydrological observations (precipitation, cloud liquid water, total precipitable water, snow cover…), pollution monitoring</td>
</tr>
<tr>
<td>Vaisala U.S. National Lightning Detection Network</td>
<td>Lightning Data</td>
<td>Instantaneous</td>
<td>Median Location Accuracy 200-500m</td>
<td>Date and Time, Latitude and Longitude, Peak amplitude, Polarity, Type of event: Cloud to Cloud to Ground</td>
</tr>
<tr>
<td>National Digital Forecast Database</td>
<td>Weather Forecast Data</td>
<td>3 hours</td>
<td>5 km</td>
<td>Wind Speed, Direction, and Gust, Relative Humidity, Convective Hazard Outlook, Tornado Probability, Probability of Thunderstorms…</td>
</tr>
<tr>
<td>Texas Parks &amp; Wildlife Department</td>
<td>Texas Ecological Mapping Systems Data</td>
<td>static</td>
<td>10 m</td>
<td>Distribution of different tree species</td>
</tr>
<tr>
<td>Texas Natural Resources Information System</td>
<td>NAIP</td>
<td>year</td>
<td>50 cm – 1 m</td>
<td>High Resolution Imagery</td>
</tr>
<tr>
<td>National Aeronautics and Space Administration</td>
<td>3D Global Vegetation Map</td>
<td>static</td>
<td>1 km</td>
<td>Canopy height data</td>
</tr>
<tr>
<td>National Cooperative Soil Survey</td>
<td>gSURGO</td>
<td>static</td>
<td>10 m</td>
<td>Soil type</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Class</th>
<th>Data Source (Measurements)</th>
<th>VOLUME (Data file size)</th>
<th>VELOCITY (Rate of use)</th>
<th>VERACITY (Accuracy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM</td>
<td>120GB per day/device</td>
<td>Every 5-15 min</td>
<td>err &lt; 2.5%</td>
<td></td>
</tr>
<tr>
<td>PMU</td>
<td>10GB per day/device</td>
<td>240 samples/sec</td>
<td>err &lt; 2%</td>
<td></td>
</tr>
<tr>
<td>ICM</td>
<td>5GB per day/device</td>
<td>250 samples/sec</td>
<td>err &lt; 1%</td>
<td></td>
</tr>
<tr>
<td>DFR</td>
<td>10GB per field/device</td>
<td>1600 samples/sec</td>
<td>err &lt; 0.2%</td>
<td></td>
</tr>
<tr>
<td>Radar [27]</td>
<td>612 MB per radar</td>
<td>Every 4-10 min</td>
<td>1-2 kHz; in 1 s³</td>
<td></td>
</tr>
<tr>
<td>Satellite [28]</td>
<td>At least: 10 GB per day</td>
<td>Every 1-15 min</td>
<td>VIS/2%; HH=7K</td>
<td></td>
</tr>
<tr>
<td>ASOS [29]</td>
<td>10 MB/day per station</td>
<td>Every 1 min</td>
<td>T&lt;18°F, P&lt;1%, Wind speed&lt; 5%, R&lt;4%</td>
<td></td>
</tr>
<tr>
<td>NLIN [30]</td>
<td>40 MB/day</td>
<td>During lightning</td>
<td>SE &lt; 200 ma, PCE&lt;4%</td>
<td></td>
</tr>
<tr>
<td>NOFD [31]</td>
<td>5-10 GB/day per model</td>
<td>1 - 12 hours</td>
<td>Varies by parameter</td>
<td></td>
</tr>
<tr>
<td>TPWD EMIST [32]</td>
<td>2.7 GB for Texas</td>
<td>static</td>
<td>SE &lt; 10 m</td>
<td></td>
</tr>
<tr>
<td>TNRS [33]</td>
<td>300 GB for Texas</td>
<td>static</td>
<td>SE &lt; 1 m</td>
<td></td>
</tr>
<tr>
<td>LIDAR [34]</td>
<td>7 GB for Harris Co.</td>
<td>static</td>
<td>HE &lt; 1 m, VE &lt; 150 cm</td>
<td></td>
</tr>
</tbody>
</table>
BD Analytics Outcomes

Probabilities of outages for no outage

Probabilities of outages for vegetation

Probabilities of outages for lightning

Probabilities of outages for ice
Takeaways

• Extensive research is needed to bring BD Analytics into utility practice:
  - Data analytics has been used in the power system domain for over 50 years, but Big Data Analytics is in its infancy
  - The Big Data Applications require intensive and costly effort to prepare the data (ingestion, cleansing, curation)
  - The gap between the Big Data platforms and utility legacy software (EMS, DMS, MMS) uses is huge, and costly
  - Utility predictive methods do not explore data sciences advances (Deep learning, spatiotemporal scaling, etc.)

• Lessons learned:
  - Assessment of risk is not meaningful without clear mitigation steps (design, component health, operating steps)
  - Big Data Predictive Analytics is cost effective and feasible if Big Data is readily available
  - Acceptance of Big Data Analytics depends on whether it is able to solve problems that otherwise are not solved
  - The target need to be great challenges with high returns if solved to justify the cost of implementation
  - The solutions are not necessarily intuitive, so extensive training and mind set change may be needed


http://smartgridcenter.tamu.edu/resume/long_resume/Html/index.html#publ