Big Data Analytics in Electric Power Distribution Systems

Dr. Nanpeng Yu

Department of Electrical and Computer Engineering

WCH 428

nyu@ece.ucr.edu

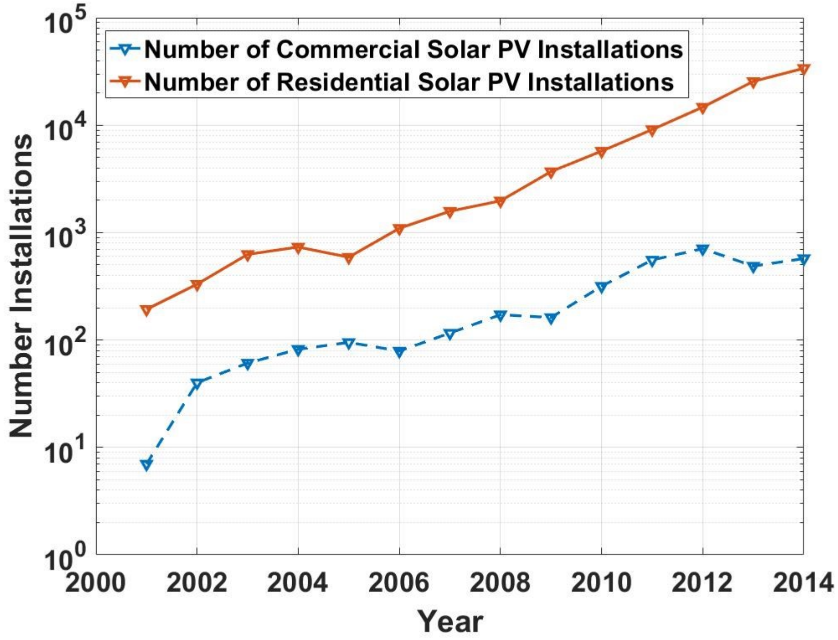
951.827.3688

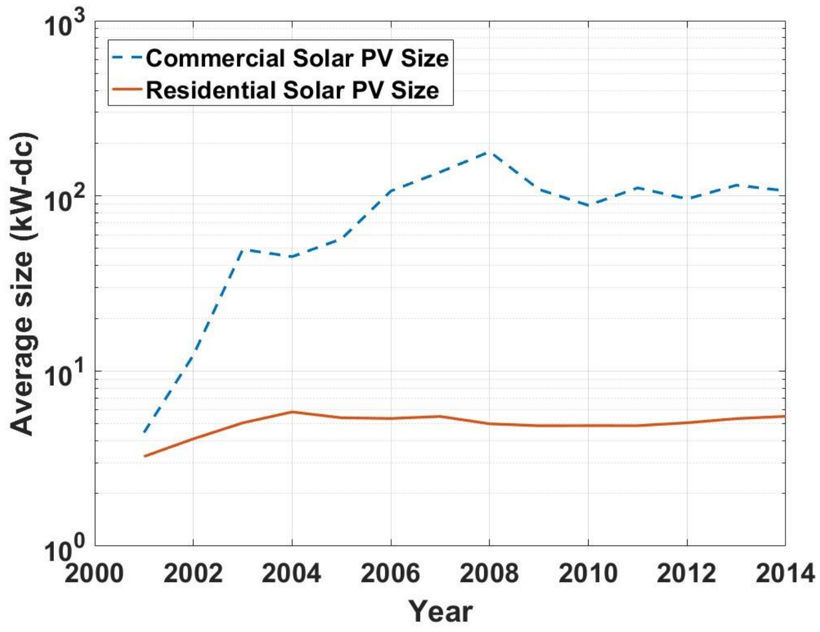
Outline

- Why do we focus on electric power distribution systems?
- Big data in power distribution systems
  - Volume, Variety, Velocity, and Value
- Big data applications in distribution systems
  - Electricity Theft Detection, Detection of Electric Vehicle
  - Phase Connectivity Identification, Transformer to Customer Association
  - Granular Load Forecast, Solar Adoption Forecast
  - Predictive Maintenance

Why distribution systems?

- Increasing penetrations of distributed energy resources (DERs) in electric power distribution systems
  - E.g. California’s transition to local renewable energy, 12,000 MW by 2020 (peak load 50,000 MW)
- DERs
  - Rooftop solar PV systems (1.84 GW of installed capacity by June 2017)





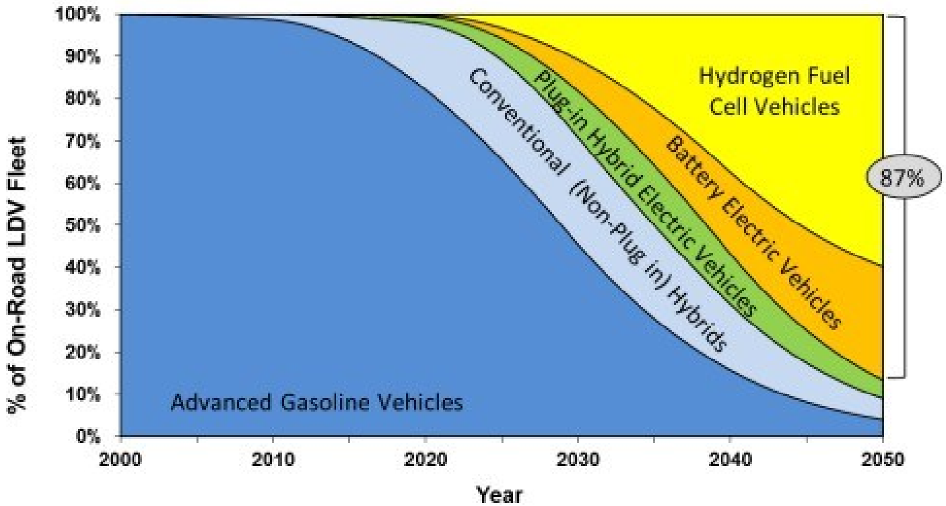
Why distribution systems?

- Increasing penetrations of distributed energy resources (DERs) in electric power distribution systems
  - E.g. California's transition to local renewable energy, 12,000 MW by 2020 (peak load 50,000 MW)
- DERs
  - Energy storage systems
    - In California 1,325 MW of energy storage will be integrated into the power system by 2020.

System Level	Applications & Revenue Streams	Technical Requirements	
		Typical Cycles / Year	Typical Discharge Duration
Distribution	13. Distribution Peak Shaving	20 to 50	1 to 4 hours
	14. Distribution Voltage Support	50 to 100	1 to 4 hours
	15. Distribution Power Quality	50 to 100	1 to 4 hours
	16. Retail Energy Time-Shift	20 to 50	15 min to 1 hour
Microgrid / Consumer	17. Energy Cost Minimization	N/A	N/A
	18. Microgrid Voltage Support	50 to 100	1 to 4 hours
	19. Microgrid Power Quality	50 to 100	1 to 4 hours
	20. Demand Charge Management	50 to 100	1 to 4 hours

Why distribution systems?

- Increasing penetrations of distributed energy resources (DERs) in electric power distribution systems
  - E.g. California's transition to local renewable energy, 12,000 MW by 2020 (peak load 50,000 MW)
- DERs
  - Electric vehicle
    - In Nov 2016, the cumulative sales of battery electric and plug-in hybrid sales in California hits 250,000 which accounts for 20% of global cumulative sales.



The need for advanced modeling, monitoring, & control of distribution systems!

- The cold, hard facts about modern power distribution systems
  - Modeling
    - Incomplete topology information in the secondary systems (phase connection, transformer-to-customer mapping)
    - Even the three-phase load flow results are unreliable.
  - Monitoring
    - Most utilities do not have online three-phase state estimation for entire distribution network
  - Control
    - Focus on system restoration
    - Limited predicative and preventive control
      - Volt-VAR control, network reconfiguration

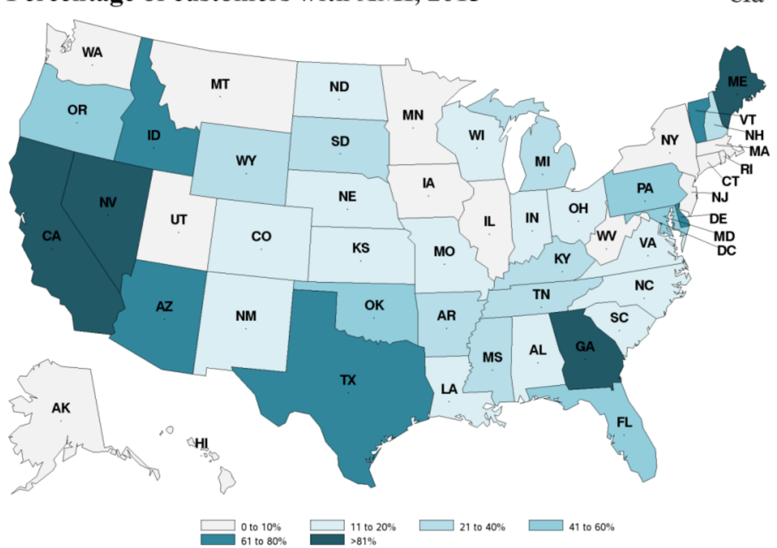
Outline

- Why do we focus on electric power distribution systems?
- Big data in power distribution systems
  - Volume, Variety, Velocity, and Value
- Big data applications in distribution systems
  - Electricity Theft Detection, Detection of Electric Vehicle
  - Phase Connectivity Identification, Transformer to Customer Association
  - Granular Load Forecast, Solar Adoption Forecast
  - Predictive Maintenance

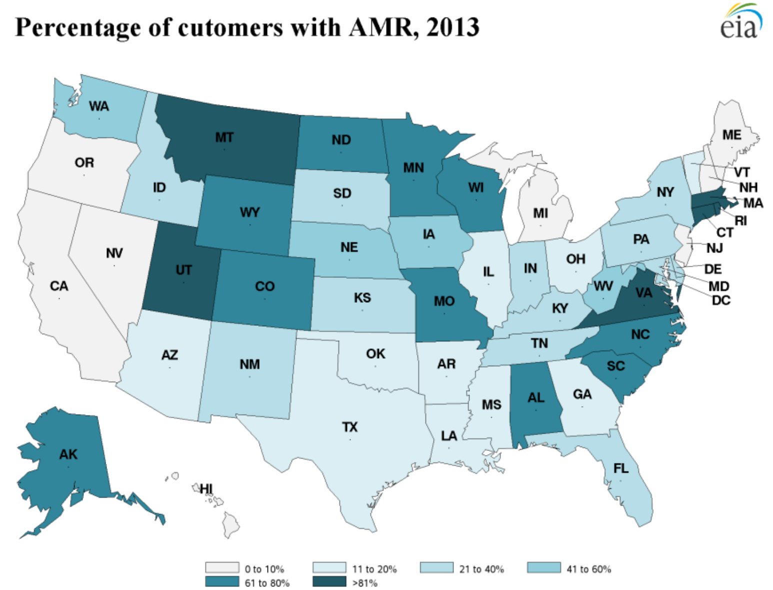
Big Data in Distribution Systems: Volume

- In 2015, the U.S. electric utilities had about 64.7 million AMI installations.
- By the end of 2016, almost 50% of the residential customers in the U.S. have AMI infrastructure.
- The smart meter installation worldwide will surpass 1.1 billion by 2022.
- In 2012, the AMI data collected in the U.S. alone amounted to well above 100 terabytes. More than 2 petabytes of meter data in 2022.

Percentage of customers with AMI, 2013

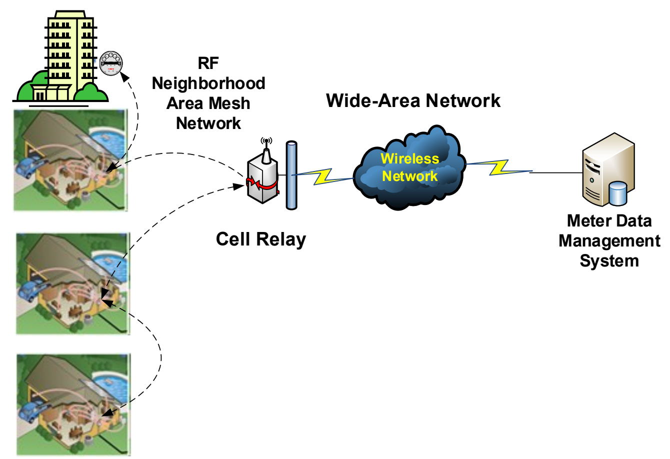


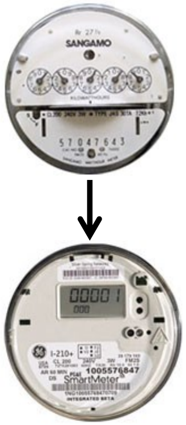
Percentage of cutomers with AMR, 2013



Big Data in Distribution Systems: Variety

- Advanced Metering Infrastructure
  - Electricity usage (15-minute, hourly)
  - Voltage magnitude
- Geographical Information System
- Equipment Monitors
  - Asset health
- Census Data
  - Household variables: ownership, appliance, # of rooms
  - Person variables: age, sex, race, income, education
- SCADA Information
- Micro-PMU
  - Time synchronized measurements with phase angles





Big Data in Distribution Systems: Velocity

- Sampling Frequency
  - AMI’s data recording frequency increases from **once a month** to one reading **every 15 minutes to one hour**.
  - Micro-PMU hundreds (512) of samples per cycle at 50/60 Hz
- Bottleneck in Communication Systems
  - Limited bandwidth for zigbee network
  - Most of the utilities in the US receives smart meter data with ~24 hour delay
- Edge Computing Trend
  - Itron and Landis+Gyr extend edge computing capability of smart meters
  - Increasing data transmission range and computing capabilities of smart meters
  - Centralized → distributed / decentralized

Big Data in Distribution Systems: Value

- The big data collected in the power distribution system had utterly swamped the traditional software tools used for processing them.
- Lack of innovative use cases and applications to unleash the full value of the big data sets in power distribution systems.*
- Insufficient research on big data analytics system architecture design and advanced mathematics for petascale data
- It is estimated that the electric utilities around the world will spend $10.1 billion on automated metering infrastructure (AMI) data analytics solutions through 2021.
- Start-up Companies
  - C3-IOT, Opower/Oracle, Autogrid
- Risk of failing to adhere to data privacy and data protection standards.

* Nanpeng Yu, Sunil Shah, Raymond Johnson, Robert Sherick, Mingguo Hong and Kenneth Loparo, "Big Data Analytics in Power Distribution Systems" *IEEE PES ISGT*, Washington DC, Feb. 2015..

Outline

- Why do we focus on electric power distribution systems?
- Big data in power distribution systems
  - Volume, Variety, Velocity, and Value
- Big data applications in distribution systems
  - Electricity Theft Detection, Detection of Electric Vehicle
  - Phase Connectivity Identification, Transformer to Customer Association
  - Granular Load Forecast, Solar Adoption Forecast
  - Predictive Maintenance

Big Data Applications in Power Distribution Systems

Spatio-temporal Forecasting

Electric Load / DERs – Short-Term / Long-Term

Anomaly Detection

Electricity Theft, Integration of EV

System Monitoring

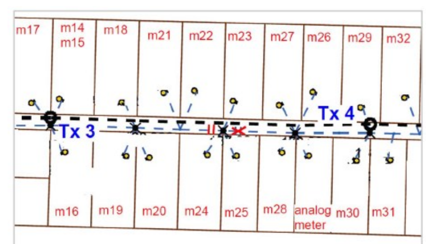
State Estimation & Visualization

Equipment Monitoring

Predictive Maintenance
Online Diagnosis

Customer Behavior Analysis

Customer segmentation, nonintrusive load monitoring, demand response



Network Topology and Parameter Identification

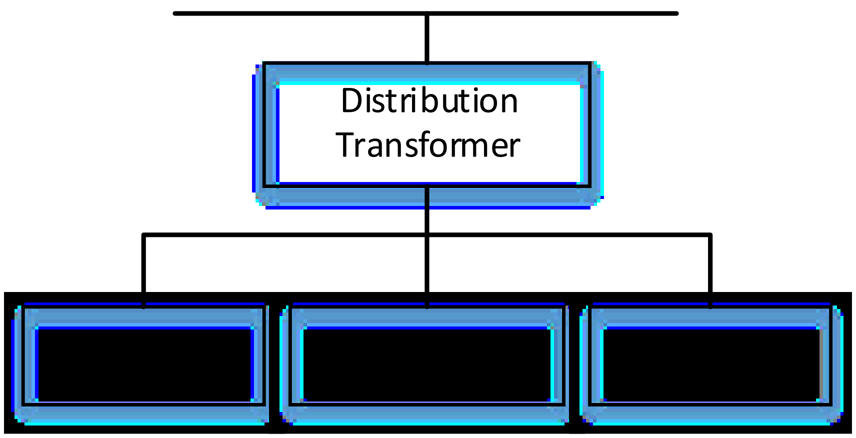
Transformer-to-customer, Phase connectivity, Impedance estimation

Electricity Theft Detection

- Problem Definition
  - Energy Theft: The activity of reducing electricity bill by altering the electricity consumption (physical / cyber)
    - Physical: Bypassing the smart meter, tamper electricity meters
    - Cyber: Hack into meters, communication network to change kWh readings
- Why is it important? (Business Value)
  - According to Northeast Group, LLC, the world loses $89.3 billion annually to electricity theft in 2015 (India $16.2 billion).
  - In the North America energy theft costs between 0.5% and 3.5% of annual gross revenue.
  - B.C. Hydro estimates up to 3% of energy theft with 1500 ‘electrical diversions’ caught in 3 years.
  - Center Point estimates energy theft is 1% to 2%.

Electricity Theft Detection

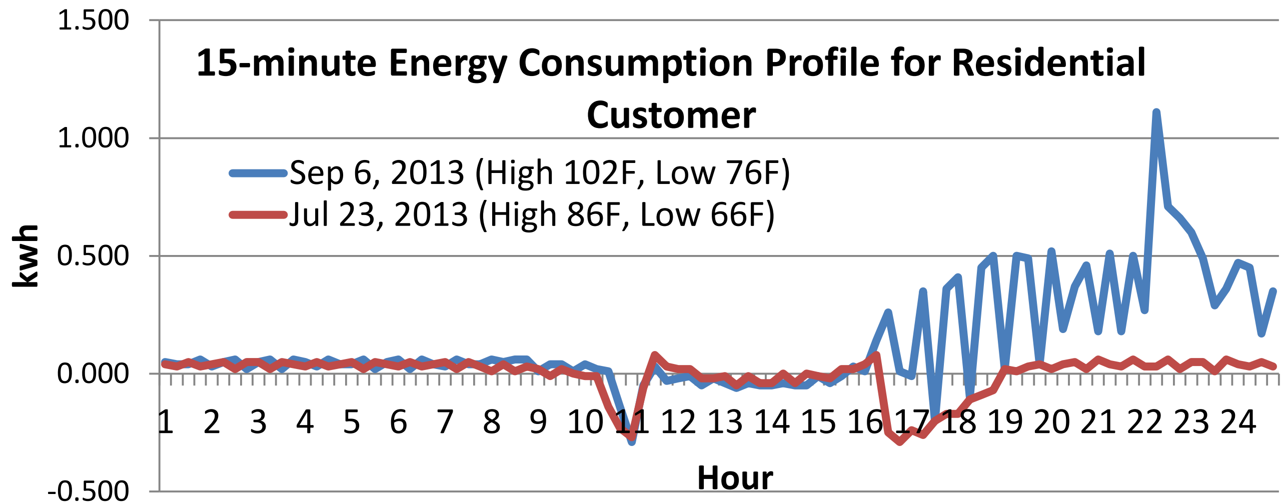
- Primary Data Set
  - Advanced Metering Infrastructure, SCADA, GIS
  - Training data (energy theft cases)
- Solution Methods
  - Physical approach
    - Technical loss model based method, state estimation based method
    - Drawback: assume all distribution network topology and parameters are known or can be estimated accurately. Meter readings are required for transformers as well.



Technical Loss + Non-technical Loss =
Total Loss =
Transformer kWh Reading - $\sum$ Customer kWh Readings

Electricity Theft Detection

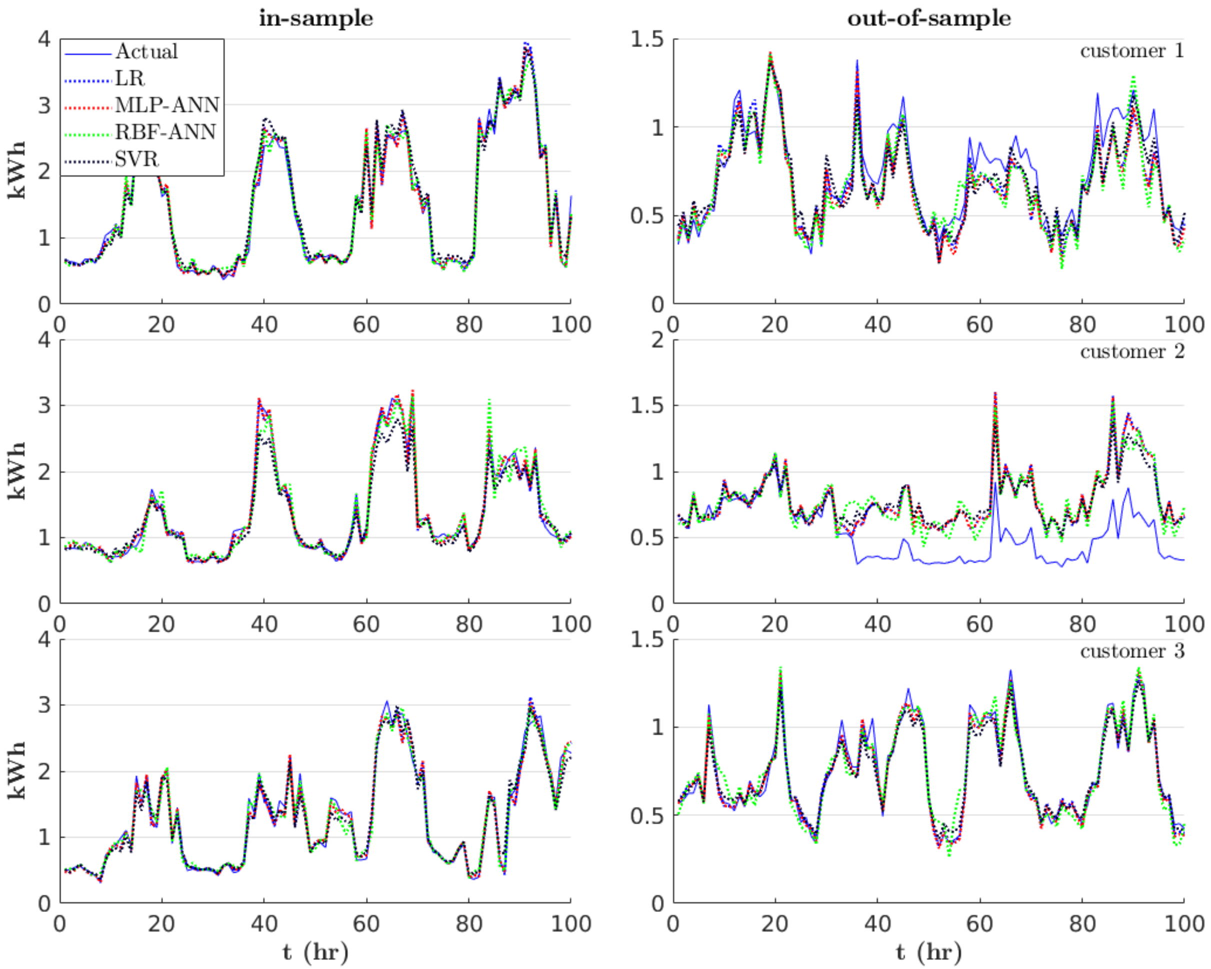
- Solution Methods
  - Machine learning approach
    - Unsupervised: anomaly detection on a single time series, supervised: classification
    - Drawback: Many other factors lead to anomaly in usage pattern, biased training set



  - Hybrid approach
    - The voltage magnitude and electricity consumption of the customers under the same transformer must be in sync. (Kirchhoff Law). Consumption can be fitted with voltage data.
    - Large difference between estimated and metered electricity consumption indicated potential energy theft.

Case Study

- Three customers are connected to the same center tapped transformer
- Realistic customer smart meter data with energy theft introduced

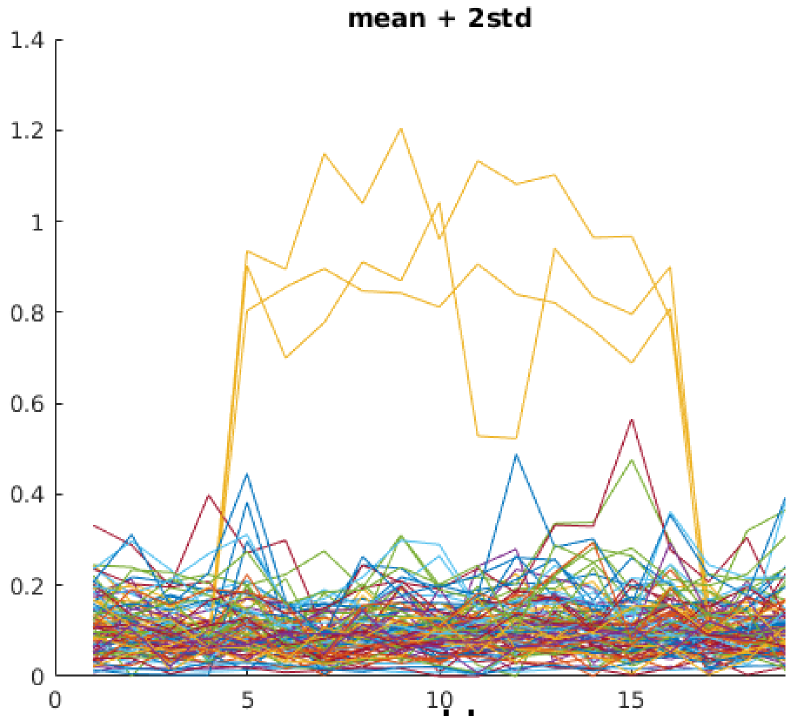


- Which customer is stealing power?
- Answer: Customer 2!

Visualization of Energy Theft

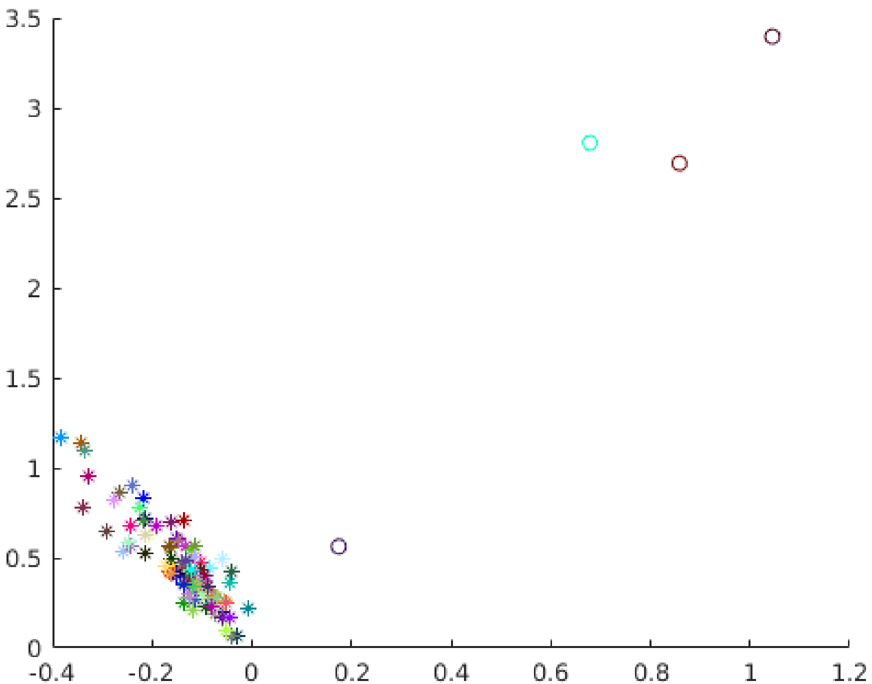
- Residual (Electricity Consumption Estimation)

Normalized
Residual



Hour

- Dimension Reduction



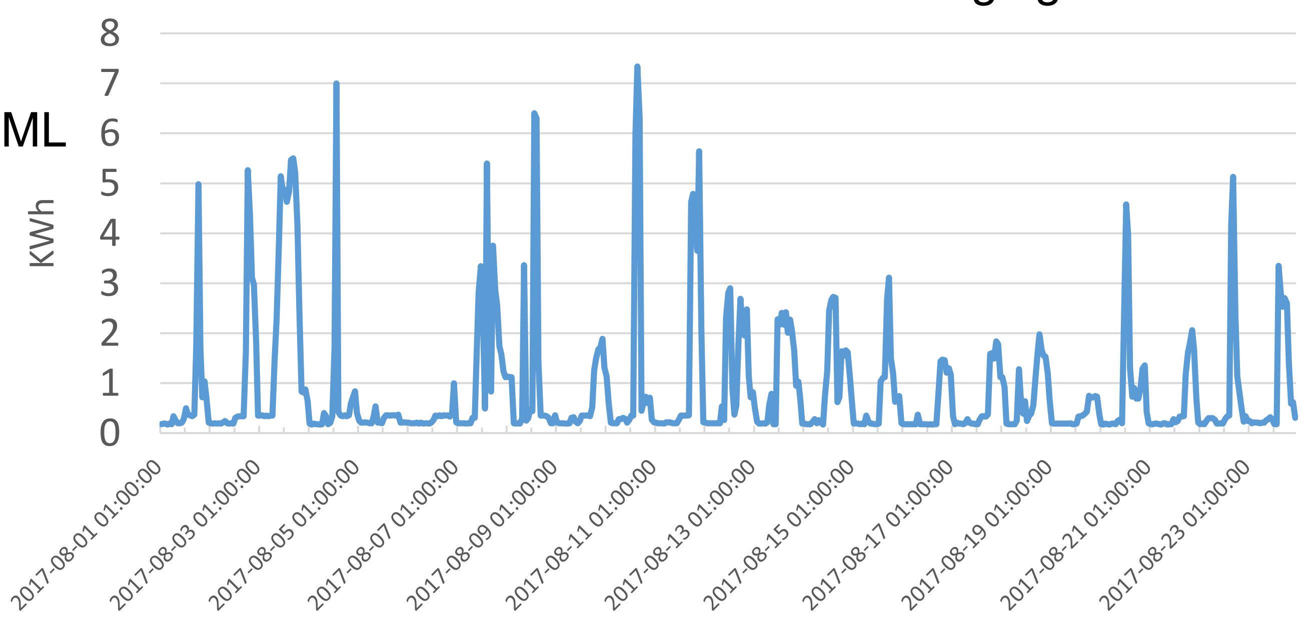
Detection of Electric Vehicle

- Problem Definition
  - Identify which customer(s) have adopted electric vehicle
  - Detect charging of electric vehicle and estimate the power consumption from charging activities.
- Why is it important? (Business Value)
  - When buying a plug-in electric vehicle (PEV), or a plug-in hybrid electric vehicle (PHEV), the consumer has no obligation to inform the electric utility.
  - On average, a typical household draws 0.7 kW of load from their local power utility. EV draws up to 3.7 kW per hour. This presents a problem because one EV owner alone can indirectly add 4 household worth of load to a transformer.
  - Unexpected charging of electric vehicles can lead to overloaded assets in a distribution system and premature equipment failure.
  - Targeted demand response / EV charging program info can be distributed to the right group of customers.

Detection of Electric Vehicle

- Primary Data Set
  - Advanced Metering Infrastructure, Customer Information System
  - Census Data at Block Group Level (Income, age, vegetation level, No. Rooms.)
  - Training data (customers who informed the utility about EV purchase)
- Solution Methods
  - Supervised ML
  - Semi-Supervised ML

Difficult to detect level-1 charging



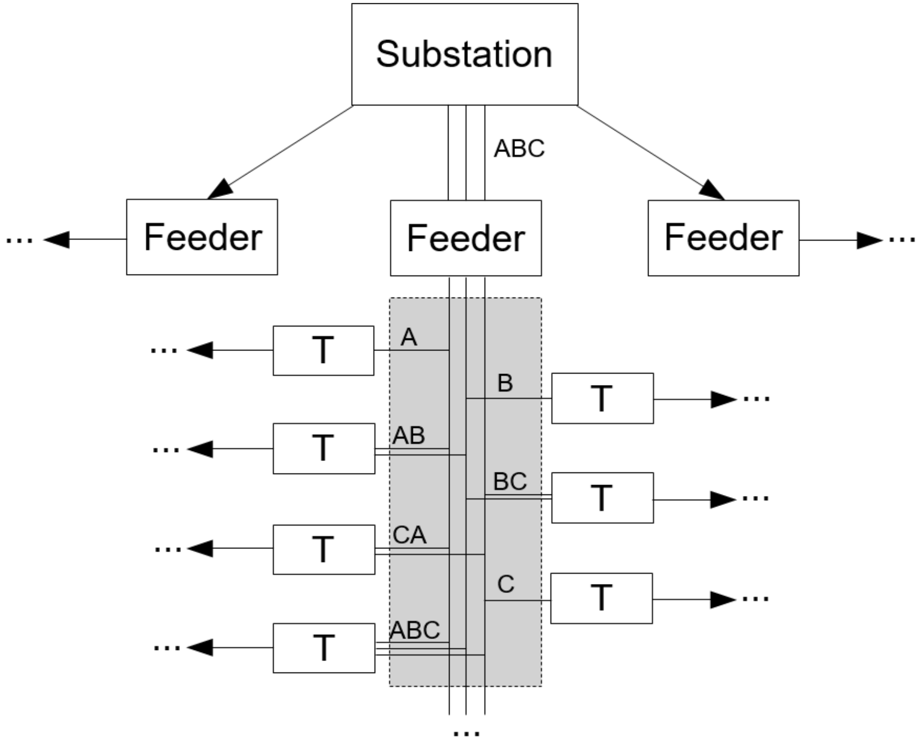
Does this household has an EV?

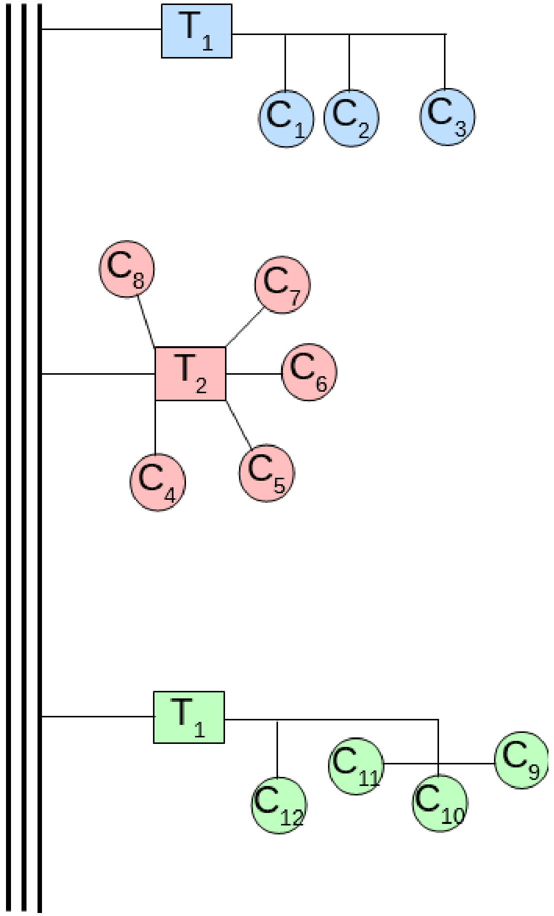
Outline

- Why do we focus on electric power distribution systems?
- Big data in power distribution systems
  - Volume, Variety, Velocity, and Value
- Big data applications in distribution systems
  - Electricity Theft Detection, Detection of Electric Vehicle
  - Phase Connectivity Identification, Transformer to Customer Association
  - Granular Load Forecast, Solar Adoption Forecast
  - Predictive Maintenance

Distribution System Topology Identification

- The distribution system topology identification problem can be broken down into two sub-problems
  - The phase connectivity identification problem
  - The customer to transformer association problem



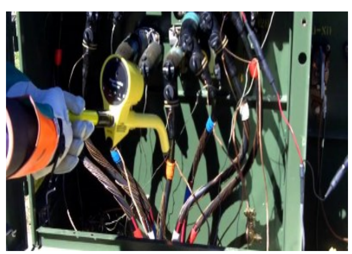


Phase Connectivity Identification

- Problem Definition
  - Identify the phase connectivity of each customer & structure in the power distribution network.
  - Very few electric utility companies have completely accurate phase connectivity information in GIS!
- Why is it important? (Business Value)
  - Phase connectivity is crucial to an array of distribution system analysis & operation tools including
    - 3-phase Power flow
    - Load balancing
    - Distribution network state estimation
    - 3-phase optimal power flow
    - Volt-VAR control
    - Distribution network reconfiguration and restoration

Phase Connectivity Identification

- Primary Data Set
  - Advanced Metering Infrastructure, SCADA, GIS, OMS
  - Training data (field validated phase connectivity)
- Solution Methods
  - Physical approach with Special Sensors
    - Micro-synchrophasors, Phase Meters
    - Drawback: expensive equipment, labor intensive ($2,000 per feeder), 3,000 feeders for a regional electric utility company ($6 million)

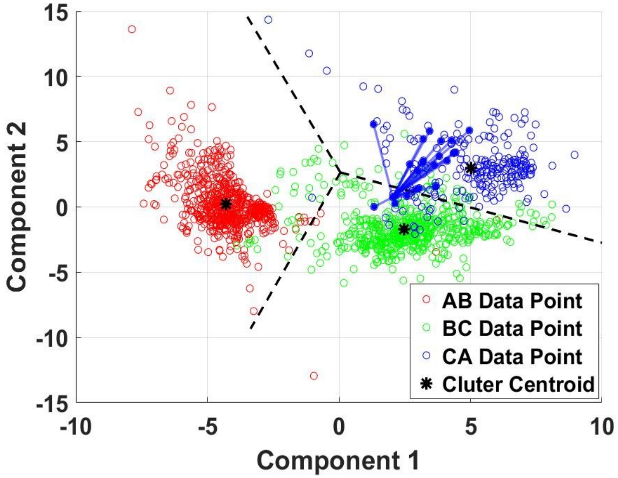



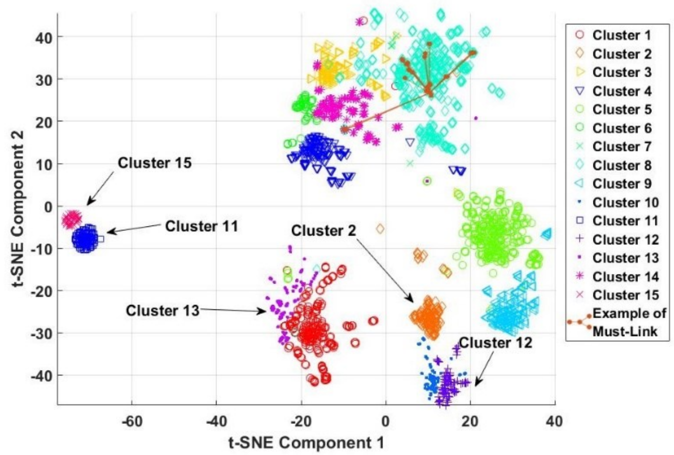
Phase Connectivity Identification

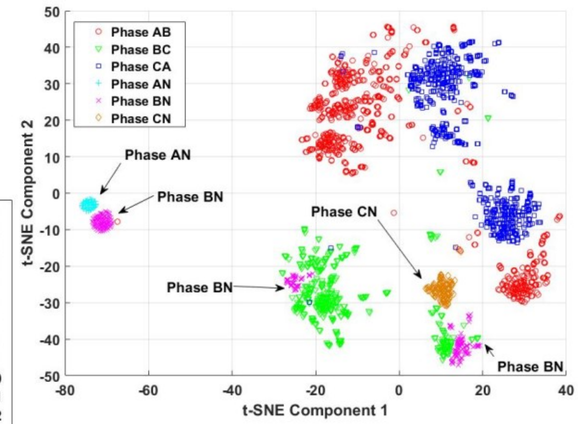
- Solution Methods
  - Integer Optimization, Regression and Correlation based Approach
    - 0-1 integer linear programming (IBM)
    - Correlation/Regression based methods (EPRI)
    - Drawback: cannot handle delta connected Secondaries, low tolerance for erroneous or missing data, low accuracy and high computational cost
  - Data-driven phase identification technology
    - Synergistically combine machine learning techniques and physical understanding of electric power distribution networks.
    - Unsupervised, supervised, and semi-supervised machine learning algorithm
    - High accuracy on all types of distribution circuits, (overhead, underground, phase-to-neutral, phase-to-phase)

Case Study

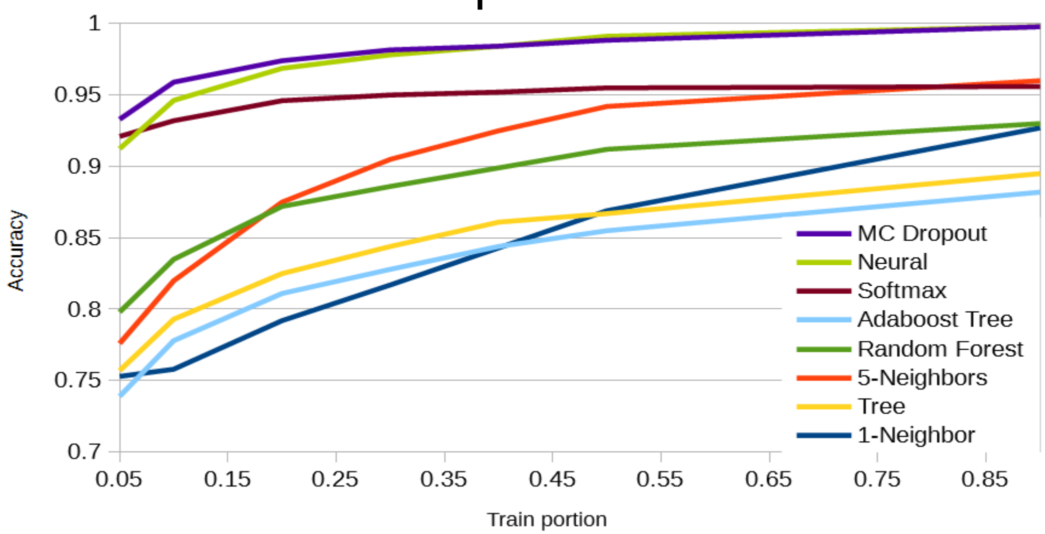
Unsupervised



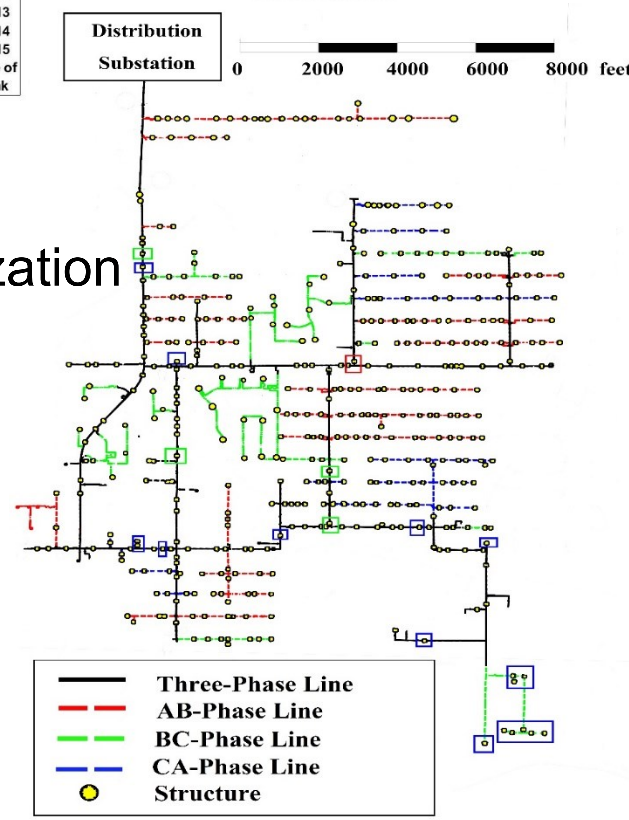




Supervised



Visualization



Transformer to Customer Association

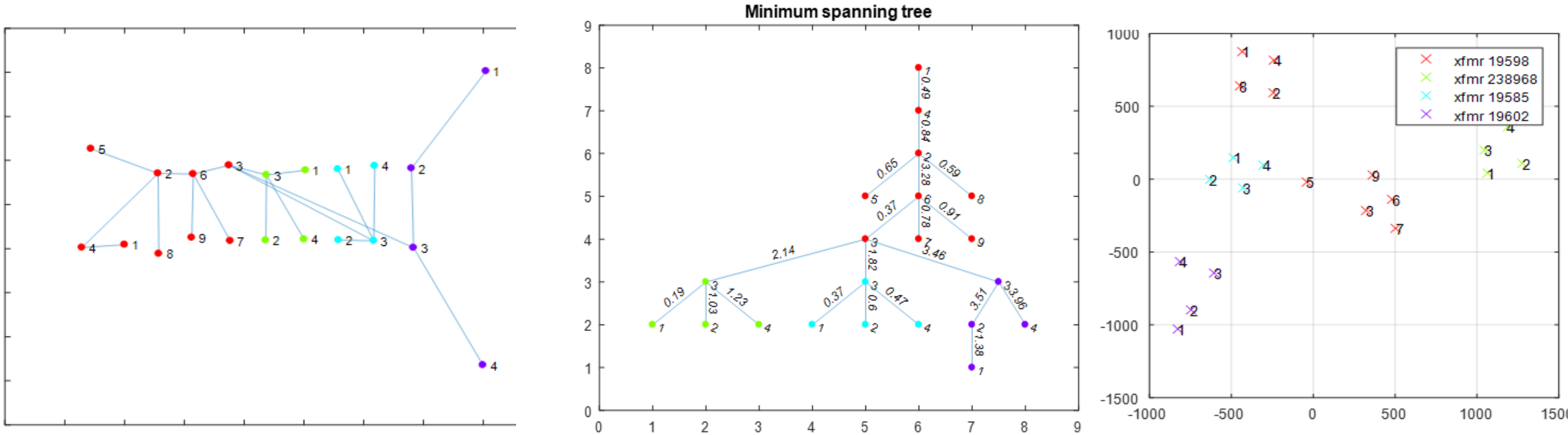
- Problem Definition
  - Correct the connection info between smart meters and transformers in GIS.
  - The current transformer to customer association data is 40% - 90% accurate in U.S. electric utilities' GIS.
- Why is it important? (Business Value)
  - Outage reporting
  - Identify a potential source of a transformer issue
  - Sizing transformers
  - Preventive maintenance of transformers
  - Electric vehicle hosting capacity estimation

Transformer to Customer Association

- Primary Data Set
  - Advanced Metering Infrastructure, SCADA, GIS, OMS
  - Customer Information System, Asset Management System
  - Training data (field validated transformer to customer mapping)
- Solution Methods
  - Physical Approach
    - Field validation (visual inspection for overhead configuration)
    - Drawback: time consuming, labor intensive, distribution network topology undergoes constant change
  - Pure Data Driven Approach
    - Linear regression, logistic regression, correlation based method
    - Voltage magnitude and GIS information are inputs

Transformer to Customer Association

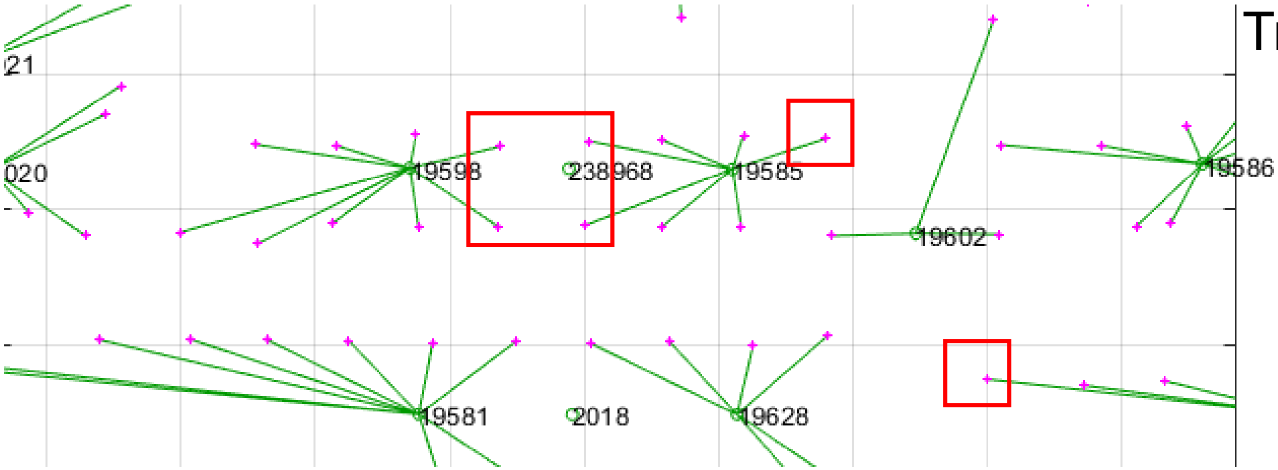
- Hybrid Solution
  - Nonlinear dimension reduction with density based clustering
    - Real-world data reside on a lower-dimensional space, customers connected to the same transformer are close to each other on the lower-dimensional feature space.



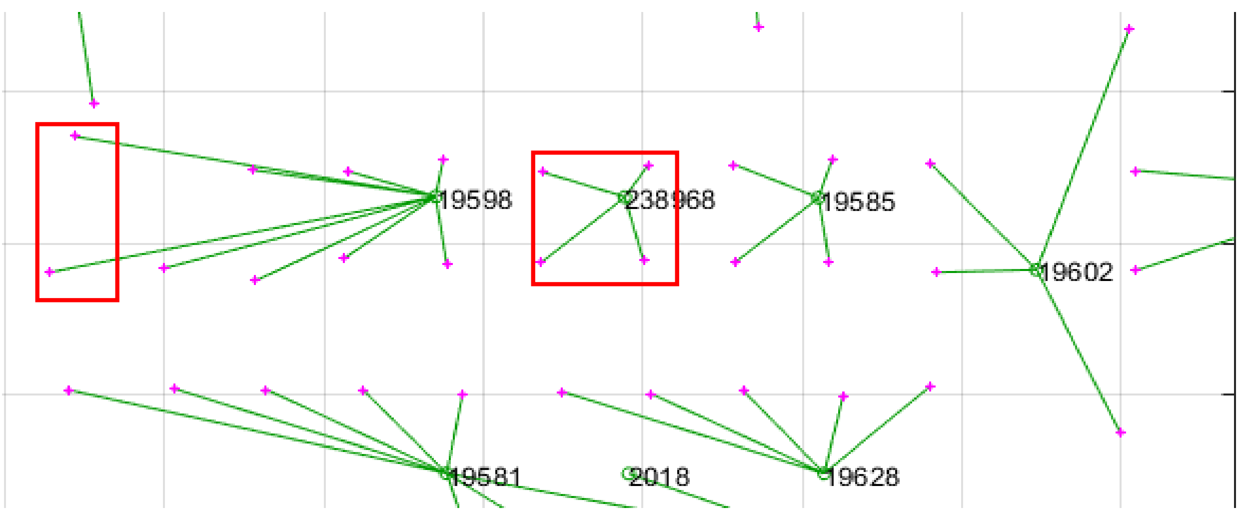
  - Physically inspired method
    - Minimum weight spanning tree based method, find the minimum weight spanning tree with second order voltage covariance as the edge weight.
    - (A subset of edges of a connected undirected graph that connects all vertices, without any cycles and with the minimum possible total edge weights.)

Case Study

Incorrect Customer to Transformer Association



Field Validated Customer to Transformer Association



Outline

- Why do we focus on electric power distribution systems?
- Big data in power distribution systems
  - Volume, Variety, Velocity, and Value
- Big data applications in distribution systems
  - Electricity Theft Detection, Detection of Electric Vehicle
  - Phase Connectivity Identification, Transformer to Customer Association
  - Granular Load Forecast, Solar Adoption Forecast
  - Predictive Maintenance

Granular Load Forecasting

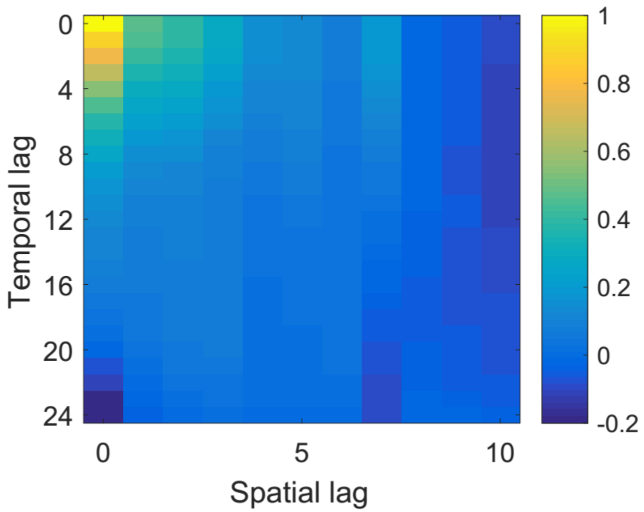
- Problem Definition
  - Electric utilities have been doing territory wide load forecasting with regression / time series / neural network models.
  - However, load forecasting at feeder / lateral / service transformer level has been done in an ad hoc manner.
- Why is it important? (Business Value)
  - Transformer sizing
  - Distribution circuit upgrades
  - Distribution system planning
  - Resource dispatch in power distribution systems
  - Distribution network reconfiguration and restoration

Granular Load Forecasting

- Primary Data Set
  - Advanced Metering Infrastructure, GIS
  - Customer Information System
  - Census Data at Block Group Level (Income, age, vegetation level, No. Rooms.)
- Solution Methods
  - Off-the-shelf Time Series Models
    - Vector Autoregressive Moving Average Model (VARMA)
    - Drawback: Curse of dimensionality, the number of parameters explode
  - Spatio-temporal Models
    - Extended Dynamic Spatial-temporal model
    - Exploit the spatial correlations of the electric load data

Granular Load Forecasting

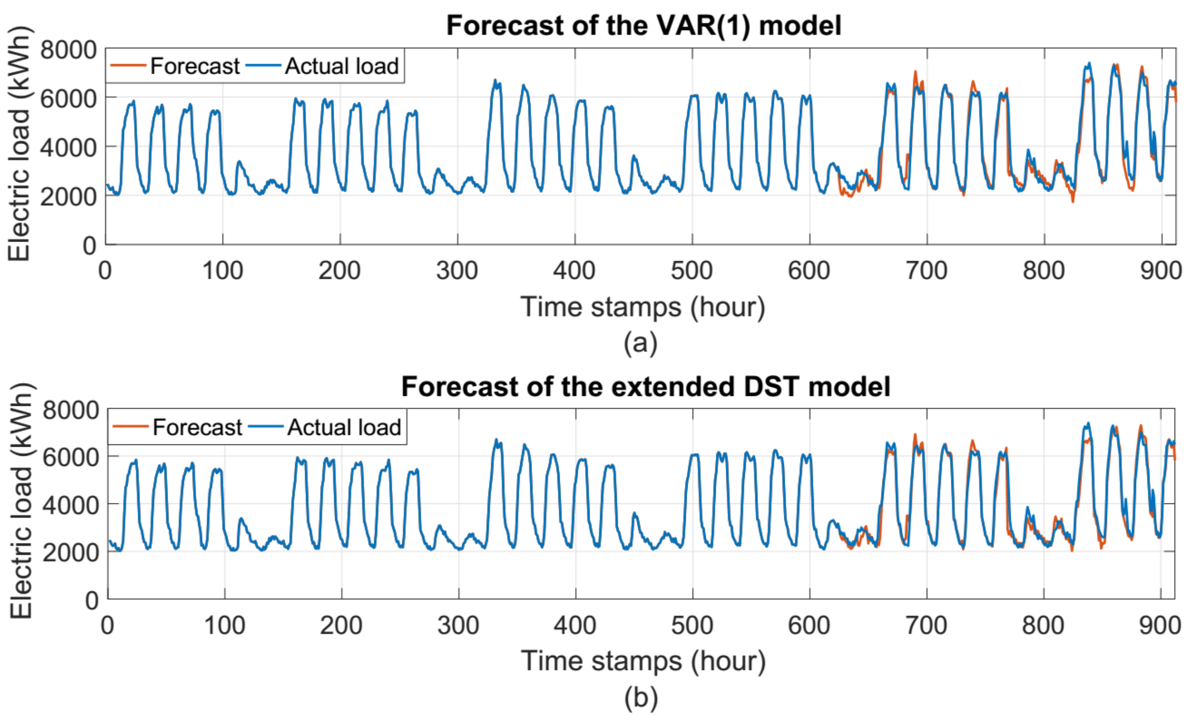
- Extended Dynamic Spatial-temporal model (DST)
  - The correlation between feeder loads is stronger when both spatial lag and time lag are small
  - The correlation drops quickly when the spatial lag increases with time lag fixed at 0.
  - DST model: $y(t) = v + (\Lambda + \Gamma W \Theta) y(t-1) + n(t)$
  - Spatial weight matrix W characterizes the spatial correlation among different feeders.
  - Exponential distance weights: $w_{ij} = \exp(-\alpha h_{ij})$



- Neural Network Model
  - Feedforward neural network
  - Recurrent and Recursive Nets
    - Specialized for processing sequential data

Case Study

Load forecasting results of a sample feeder



(a) VAR(1) model

(b) Extended DST model

Average forecasting performance of VAR(1) and extended DST models

Model	VAR(1) model	Extended DST model
Average RMSE [kWh]	554.64	490.39
Average MAPE	12.26%	10.63%

Solar PV Adoption Forecast

- Problem Definition
  - Perform spatial-temporal solar Photovoltaic system adoption forecast at the customer / feeder level.
  - Separate commercial and residential solar PV adoption forecast models.
- Why is it important? (Business Value)
  - More accurate forecast of distributed solar PV adoption will greatly facilitate distribution system planning.
  - Spatial-temporal solar PV adoption forecast serves as an important input to hosting capacity analysis.
  - A solar PV adoption model is a useful tool for policy evaluation (federal and state incentive programs, CSI, ITC).
  - Understanding the drivers behind the solar PV adoption could help policy makers / utilities improve design of future renewable energy incentive programs.

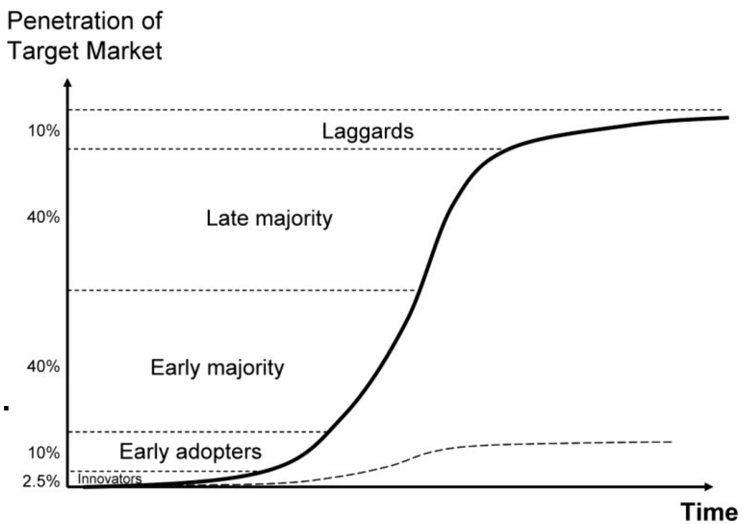
Solar PV Adoption Forecast

- Primary Data Set
  - Advanced Metering Infrastructure, GIS, Customer Information System
  - Census Data, Historical PV adopter information, Financing
  - Retail Rate, Historical Installed PV Cost, Incentive Program Data, Roof Info
- Solution Methods
  - Discrete Choice Experiment with Surveys (EPRI)
    - Identify attributes that influence consumers' decisions.
    - The attributes were tested with a focus group and then used to develop questions for surveys administered to more than 2,500 customers.
    - The "choice" mode is built for determining the combinations of attributed likely to drive customer preference.

Solar PV Adoption Forecast

Bass Diffusion Model

- The basic Bass model is well-established to model the innovation and technology adoption in any market.
- The probability of adoption of a new product at time T given that it has not yet been adopted would depend linearly on two forces, innovation p and imitation q.
- $\frac{f(t)}{[1-F(t)]} = p + qF(t)$



Generalized Bass Diffusion Model

- To include "marketing effort", the GBM is introduced.
- $\frac{f(t)}{[1-F(t)]} = [p + qF(t)]x(t)$
- We shall call $x(t)$ "current marketing effort", reflecting the influence of market factors on the adoption rate at time t.
- $x(t)$ could represent energy savings, government incentives, etc.

Outline

- Why do we focus on electric power distribution systems?
- Big data in power distribution systems
  - Volume, Variety, Velocity, and Value
- Big data applications in distribution systems
  - Electricity Theft Detection, Detection of Electric Vehicle
  - Phase Connectivity Identification, Transformer to Customer Association
  - Granular Load Forecast, Solar Adoption Forecast
  - Predictive Maintenance

Predictive Maintenance - Transformer

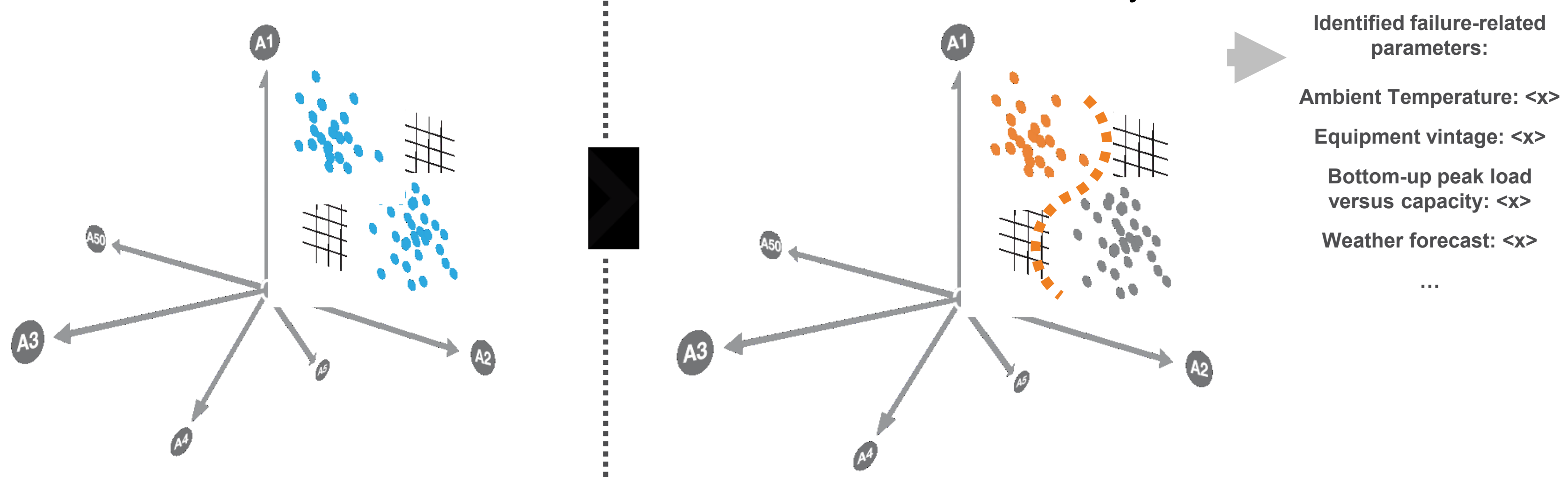
- Problem Definition
  - The lack of data and knowledge about the transformers prevents utilities from performing predictive maintenance until the transformer fails, resulting in interruption in service to one or more customers.
  - Assign and maintain a health index score, and predict remaining life for transformers.
- Why is it important? (Business Value)
  - Reduce System Average Interruption Duration Index (SAIDI) index and enhance system reliability.
  - $SAIDI = \frac{Total\ Duration\ of\ Interruptions\ for\ a\ Group\ of\ Customers}{Number\ of\ all\ Customers}$
  - Utilities can maintain the same level of overall maintenance activity on the system while shifting from reactive maintenance to preventive maintenance.

Predictive Maintenance - Transformer

- Primary Data Set
  - GIS, Weather, Climate Zone, OMS, AMI, SAP (Manufacture, Age, etc.)
  - SCADA, Historical Failures, Lightning Data
- Solution Methods
  - Basic aging algorithms and thermal models developed by IEEE and International Electrotechnical Commission (IEC)
  - Supervised, Semi-supervised ML

Historical recorded asset information

Known asset failure / reliability cases



Thank You

- Contact information
  - Dr. Nanpeng Yu
  - Department of Electrical and Computer Engineering, UC Riverside, United States
  - Phone: 951.827.3688
  - Email: nyu@ece.ucr.edu
  - Website: http://www.ece.ucr.edu/~nyu/

Big Data Applications in Short-term Operations

- Short-term Spatio-temporal Forecasting
  - Load forecast*
  - Solar PV forecast
  - Demand Response forecast☼
- Anomaly Detection
  - Energy theft detection
- Estimation
  - AMI data-driven Three-phase State Estimation†
- Distribution System Visualization

* Jie Shi and Nanpeng Yu, "Spatio-temporal modeling of electric loads" to appear in 49th North American Power Symposium, pp.1-6, Morgantown, WV, 2017.

☼ Xiaoyang Zhou, Nanpeng Yu, Weixin Yao and Raymond Johnson, "Forecast load impact from demand response resources" IEEE Proceedings, Power and Energy Society General Meeting, pp. 1-5, Boston, USA, 2016.

† Yuanqi Gao and Nanpeng Yu, "State estimation for unbalanced electric power distribution systems using AMI data" The Eighth Conference on Innovative Smart Grid Technologies (ISGT 2017), pp. 1-5, Arlington, VA.

Big Data Applications in Long-term Planning

- Distribution Network Topology Identification
  - Transformer-to-customer association
  - Phase connectivity identification* ☼
- Customer Segmentation
- Nonintrusive Load Monitoring
- Long-term Spatio-temporal Forecasting
  - Solar PV Adoption Forecast†
  - EV Penetration Forecast
- Equipment Preventive Maintenance

* W. Wang, N. Yu, B. Foggo, and J. Davis, “Phase identification in electric power distribution systems by clustering of smart meter data” *15th IEEE International Conference on Machine Learning and Applications* (ICMLA), pp. 1-7, Anaheim, CA, 2016.

☼ W. Wang and N. Yu, "AMI Data Driven Phase Identification in Smart Grid," the Second International Conference on Green Communications, Computing and Technologies, pp. 1-8, Rome, Italy, Sep. 2017.

† W. Wang, N. Yu, and R. Johnson “A model for commercial adoption of photovoltaic systems in California” *Journal of Renewable and Sustainable Energy*, Vol. 9, Issue, 2, pp.1-15, 2017.