A Sustainable Approach for Demand Prediction in Smart Grids using a Distributed Local Asynchronous Algorithm

Rajarshi Mallik, and Hillol Kargupta

University of Maryland Baltimore County & Agnik
E-mail: hillol@cs.umbc.edu
Web: http://www.cs.umbc.edu/~hillol

Acknowledgement: NASA
Road Map

• Distributed Data Mining for Smart Grid

• Distributed linear regression for energy demand prediction

• PRIANS

• Conclusions
Distributed Data Mining for Smart Grids: Two Examples

- Distributed Power Consumption/Production Prediction for Smart Grids (*What my student wants me to talk about*)

- Smart Grid and Social Media: P2P computing and social networking (*What I really want to talk about*)
• Distributed Energy Production, Storage, Consumption

• Social interaction in Smart Grid
Data Sources

- Smart Metering data
  - Energy, power load characteristics
- Usage data
  - Smart appliances, detailed energy consumption logs
- Power generation/storage data
  - Alternate power generation (e.g. solar, wind, PV) data
  - Energy storage data
- Customer Relationship Management (CRM) data.
  - Customer contact, complaint, billing data
  - Contractual data
- Network Security data
Additional Data Sources

- Distribution network SCADA (Supervisory Control And Data Acquisition) data

- External data
  - Weather data
  - Emissions data
  - Electricity pricing data
  - Transportation data
  - Other utilities data: gas, water distribution, garbage collection

- Social networks, media data
Computing Architecture & Data Ownership

• Multi-party:
  – Different energy companies
  – Government organizations
  – Individual entities

• Who owns the data?
• Who pays for the infrastructure?
Data Analysis in Smart Grid

• Operation and control of demand response
  • Predicting demand
  • Learning multi-variate models from data

• Clustering users based on consumption behavior characteristics

• Outlier detection for threat monitoring attempting to bring down the grid

• Social media data analysis
• Data Mining: Scalable analysis of data by paying careful attention to the resources:
  – computing
  – communication
  – storage
  – human-computer interaction.

• Distributed Data Mining (DDM): Mining data using distributed resources.
DDM Algorithm Design: Methodology

- Distributed environment $G = (V, E)$
- Each node contains some data $O_k$
  - Same schema
  - Different schemas
- Compute function $f(V)$
- Construct a decomposed representation of $f(V)$ where $f(V)$ can be computed from locally computed functions $p(O_k)$
- Correctness and Scalability
Distributed Linear Regression

- **Linear Regression:**
  Modem structure $A.w = b$

  Learn coefficient vector $w$.
  Cholesky Factorization: $(A^T \cdot A) \cdot w = A^T b$

- Covariance matrix that is additively decomposable
- Local computation of covariance matrix
- Iterative computation of intermediate results
- Use distributed asynchronous techniques to come to a global solution
- Converges quickly to the globally correct solution
Experiments with Grid Data
Power Consumption Data

• Residential Buildings Energy Consumption Survey (12th RECS 2005)*
• Obtained from residential energy suppliers and questionnaire survey
• Data features include
  – Housing Unit characteristics
    • Mobile homes, Single family detached house, apartment buildings with 5 units or more etc.
    • How many people lives in those house unit

*http://www.eia.doe.gov/emeu/recs/recspubuse05/pubuse05.html
Data Features

- **Average energy consumption per house by**
  - Dryer, dishwasher, freezer, Refrigerator, AC, water heater, space heater, and other electric appliances

- **Data collected all over US at specified census regions**
  - Consumption statistics for the state of New York used in experiments
  - For all 62 counties
  - On an hourly basis
Power Production Data

• Photovoltaic (PV) system
  – Data simulated from System Advisor Model (SAM)\(^1\)
  – NREL’s PVWatts\(^2\) calculator
  – Default PV system size of 4 kW approximates to PV array of size 35m\(^2\) (377ft\(^2\))
  – Other weather conditions like temperature, wind speed, shading factored in for net output
  – Hourly production data per household

\(^1\)https://www.nrel.gov/analysis/sam/
\(^2\)http://www.nrel.gov/rredc/pvwatts/
Demand Simulation

- Demand per hour = Consumption per hour - Production per hour

- Consumption and Production based on a typical day in the summer of New York state (June – Sept)

- Weather data from SAM* factored in during simulation for production

  Assumed heavy use of AC during summer and space heater during winter

* https://www.nrel.gov/analysis/sam/
Per household type average Consumption and Production statistics during summer at Albany.
Grid Demand
Actual Demand vs. Centralized Prediction

![Graph showing actual demand vs. centralized predicted demand over time. The x-axis represents time of day, and the y-axis represents power in kWhr. The graph shows peaks and valleys throughout the day, with the actual demand lines generally lower than the predicted demand lines.]
Centralized vs Distributed Demand Prediction

The graph illustrates the comparison between centralized and distributed predicted demand for power in kWhr over a 24-hour period. The green line represents distributed predicted demand, while the blue line represents centralized predicted demand. The x-axis represents the time of day, and the y-axis represents power in kWhr. The data shows fluctuations in demand throughout the day, with peaks occurring during certain times, highlighting the differences in prediction methods.
Communication cost

- Measured as the number of bytes of data transferred
- Against the problem size, here total number of residences
Conclusions

- Grid is fundamentally Distributed

- Analyzing data in Smart Grid needs some fundamental thinking

- DDM offers fundamentally different, commercially proven scalable solutions