Energy Disaggregation at Greenely

Eric Leijonmarck

Royal Insitute of Technology

Content

Contents of this presentation

- Disaggregation; state of 2015
- Sparse-coding implementation
- Results and Extensions
- Pre-processing data
- Suggestions on how to proceed next

State of 2015

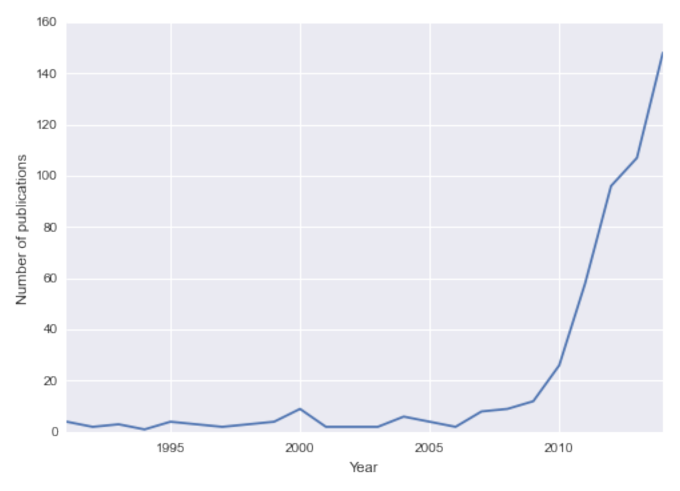
Competitors

- Smart-meter disaggregation
  1. EEme, energy disaggregation
  2. SenseHome, startup in Boston
  3. Bidgely, PlotWatt, ComEd, Neurio, Navetas, Belkin, Intel
- Hardware Competitors
  1. Watty, KTH-based start-up
  2. Smappee, real-time Disaggregation startup
- Overall energy product companies
  1. ITG, signal-processsing company in New York
  2. British Gas, holistic energy provider
  3. Opower, energy efficiency

State of 2015

Research

- Research in Energy Disaggregation

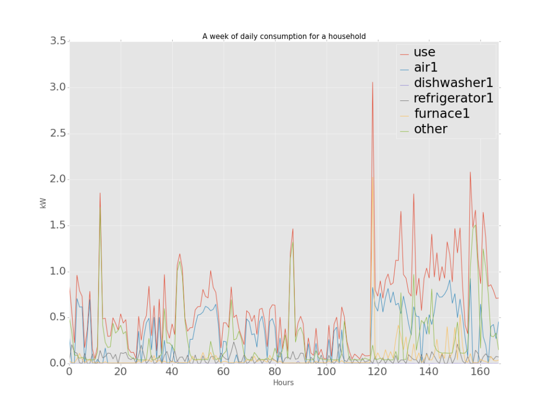


Increase at 2010 from 20 - 140 in 4 years

State of 2015

Challenge

- Challenge



- Lots of methods one common denominator, Data
- Perform unsupervised on test, and supervised on scaled houses

State of 2015

Solution

- Data, Data-preparation, Data-data, Data-data-data
- Ed Freeman, More Data or Better Algorithms?
  1. Having a good problem to work on
  2. Having a good approach to that problem
  3. Having decent data. Quality is much more important than quantity
  4. Handling the data well – good transforms, good missing value handling, making sure that all approaches make sense for the problem.

Input data

We have $1:k$ appliances

- We define one class (e.g. heater) $\mathbf{X}_i \leftarrow 1, \ldots, k$
- Where $\mathbf{X}_i \in \mathbb{R}^{T \times m}$, T hourly week data for m houses
- **One** aggregated household $\bar{\mathbf{X}} \leftarrow \sum_{i:k} \mathbf{X}_i$
- Assuming we have individual energy readings $\mathbf{X}_1, \ldots, \mathbf{X}_k$
- Goal: test with new data $\bar{\mathbf{X}}'$ to components $\mathbf{X}'_1, \ldots, \mathbf{X}'_k$

Input data

We have $1:k$ appliances

$$\underbrace{\mathbf{X}_i \in \mathbb{R}^{T \times m}}$$

$$\begin{bmatrix} \text{App} & \mathbf{x}_1^{(j)} & \cdots & \mathbf{x}_1^{(m)} \\ 1h & 0.8kWh & \cdot & \cdot \\ 2h & 0.7kWh & \cdot & \cdot \\ \vdots & \vdots & \cdot & \cdot \\ 168h & 0.1kWh & \cdot & \cdot \end{bmatrix}$$

How do we get this format?

Pre-processing raw-data

- Number of missing values in data
- When missing - what to do?
- These points need to be considered before dealing with any algorithm
- Demonstation of processing

Sparse coding pre-training

pre-train activations and basis vectors

Algorithm 1: Dicriminative disaggregation sparse coding

input: data points for each individual source $\mathbf{X}_i \in \mathbb{R}^{T \times m}, i = 1 : k$, regularization $\lambda \in \mathbb{R}_+$, with gradient step size $\alpha \in \mathbb{R}_+$.

Sparse coding pre-training:

1. Initalize $\mathbf{B}_i$, $\mathbf{A}_i; \geq 0$, scale columns $\mathbf{B}_i$ s.t. $\left\| \mathbf{b}_i^{(j)} \right\|_2 = 1$
2. For each $i = 1, \ldots, k$, iterate until convergence:

$$\mathbf{A_i} \leftarrow \operatorname{argmin}_{A \geq 0} \|\mathbf{X}_i - \mathbf{B}_i \mathbf{A}\|_F^2 + \lambda \textstyle\sum_{p,q} \mathbf{A}_{pq}$$

$$\mathbf{B}_i \leftarrow \operatorname{argmin}_{B \geq 0, \left\|\mathbf{b}^{(j)}\right\|_2 \leq 1} \|\mathbf{X}_i - \mathbf{B}\mathbf{A}_i\|_F^2$$

Discriminiative disaggregation training

perceptron algorithm

Algorithm 2: Dicriminative disaggregation sparse coding

Discriminative disaggregation training:

3. Set $\mathbf{A}^*_{1:k} \leftarrow \mathbf{A}_{1:k}, \hat{\mathbf{B}}_{1:k} \leftarrow \mathbf{B}_{1:k}$.
4. Iterate until convergence:

   $\hat{\mathbf{A}_{\mathbf{1:k}}} \leftarrow \operatorname{argmin}_{A_{1:k} \geq 0} F\left(\bar{\mathbf{X}}, \tilde{\mathbf{B}}_{1:k}, \mathbf{A}_{1:k}\right)$

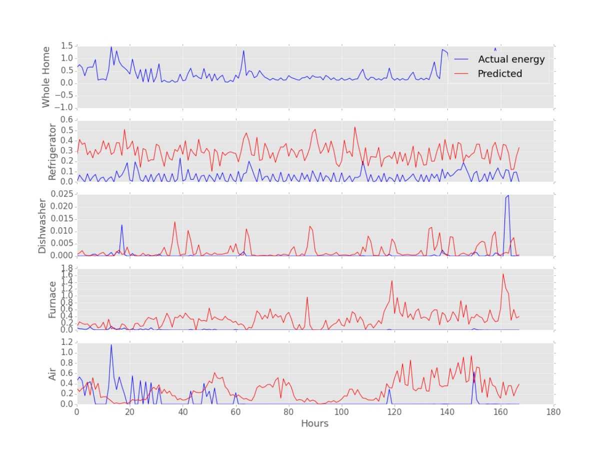
   $\tilde{\mathbf{B}} \leftarrow \left[\tilde{\mathbf{B}} - \alpha\left((\bar{\mathbf{X}} - \tilde{\mathbf{B}}\hat{\mathbf{A}})\hat{\mathbf{A}}^T - (\bar{\mathbf{X}} - \tilde{\mathbf{B}}\mathbf{A}^*)(\mathbf{A}^*)^T\right)\right]_+$

   $\forall \quad i, j, \mathbf{b}_i^{(j)} \leftarrow \mathbf{b}_i^{(j)} / \left\|\mathbf{b}_i^{(j)}\right\|_2$

Given aggregated test examples $\bar{\mathbf{X}}'$

5. $\hat{\mathbf{A}}'_{1:k} \leftarrow \operatorname{argmin}_{\mathbf{A}_{1:k} \geq 0} F(\bar{\mathbf{X}}', \tilde{\mathbf{B}}_{1:k}, \mathbf{A}_{1:k})$
6. Predict $\hat{\mathbf{X}}'_i = \mathbf{B}_i \hat{\mathbf{A}}'_i$

Trained models
DDSC



Finds patterns within the data

Tests

DDSC

$$\text{Accuracy of Week} \equiv \frac{\sum_{i,q} \min\left\{\sum_p (\mathbf{X}_i)_{pq}, \sum_p (\mathbf{B}_i, \hat{\mathbf{A}_i})_{pq}\right\}}{\sum_{p,q} \bar{\mathbf{X}}_{i\,p,q}}$$

The overlap of the predicted appliance use and the actual appliance usage.

Extensions

DDSC

- Gridsearch across hyperparameters
- Clustering - label data
- Adding extensions to the model, through Andrew

Total energy priors:

$$F_{TEP}(\bar{\mathbf{X}}, \mathbf{B}_{1:k}, \mathbf{A}_{1:k}) =$$
$$F(\bar{\mathbf{X}}, \mathbf{B}_{1:k}\mathbf{A}_{1:k}) + \lambda_{TEP} \sum_{i=1}^{k} \left\| \mu_i \mathbf{1}^T - \mathbf{1}^T \mathbf{B}_i \mathbf{A}_i \right\|_2^2$$

Group Lasso:

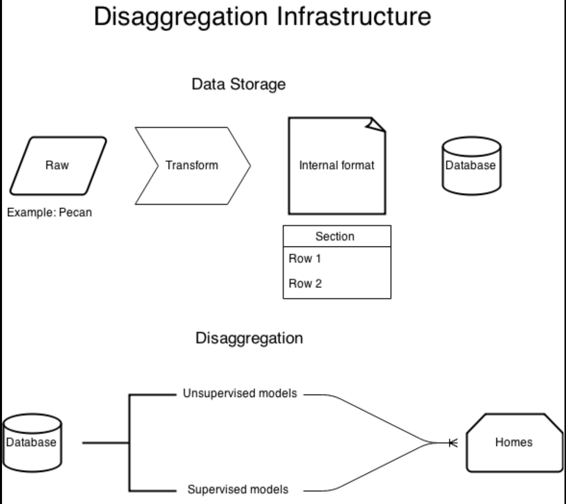
$$F_{GL}(\bar{\mathbf{X}}, \mathbf{B}_{1:k}, \mathbf{A}_{1:k}) = F(\bar{\mathbf{X}}, \mathbf{B}_{1:k}\mathbf{A}_{1:k}) + \lambda_{GL} \sum_{i=1}^{k} \sum_{j=1}^{m} \left\| \mathbf{a}_i^{(j)} \right\|_2$$

Suggestions
DDSC

- Data, Data, DATA

- Quality data == better models. Period
- Model using fewest assumptions is most likely to be correct.
- Simple algorithm quality data
- All models are wrong, but some are useful. (George E. P. Box)

Suggestions
DDSC



Thank you!

Questions?