Discriminative Sparse Coding for Energy Disaggregation

J.Zico Kotler, Siddarth Batra and Andrew Y. Ng

Eric Leijonmarck

Royal Insitute of Technology

Content

Contents of this presentation

- Input data
- Sparse-coding pre-training
- Discriminitive disaggregation training and prediction
- Extensions
- Suggestions

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

$$\overbrace{\begin{bmatrix} \text{App} & \mathbf{x}_1^{(j)} & \cdots & \mathbf{x}_1^{(m)} \\ 1h & 0.8kWh & \cdot & \cdot \\ 2h & 0.7kWh & \cdot & \cdot \\ \vdots & \vdots & \cdot & \cdot \\ 168h & 0.1kWh & \cdot & \cdot \end{bmatrix}}^{\mathbf{X}_i \in \mathbb{R}^{T\times m}}$$

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:

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

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

Discrimininiative 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 $\hat{\mathbf{X}}'_i = \mathbf{B}_i \hat{\mathbf{A}}'_i$

Extensions

DDSC

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:

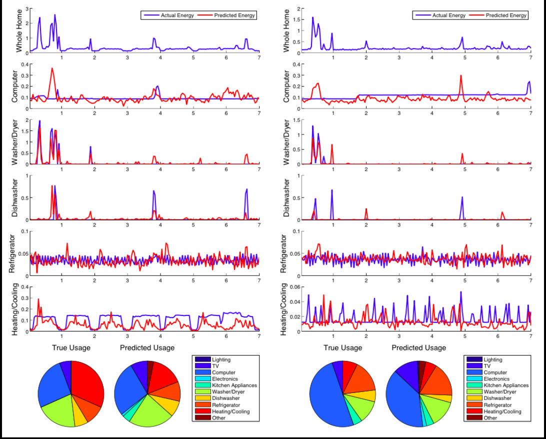
$$F_{GL}(\bar{\mathbf{X}}, \mathbf{B}_{1:k}, \mathbf{A}_{1:k}) = F(\bar{\mathbf{X}}, \mathbf{B}_{1:k}\mathbf{A}_{1:k}) + \lambda_{GL} \sum_{i=1}^{k} \sum_{j=1}^{m} \left\| \mathbf{a}_i^{(j)} \right\|_2$$

Shift Invariant Sparse Coding:

Convolutional sparse coding; each basis is convolved over the input data, with a separate activation for each shift position. Performed poorly due to the dimensional space of the data, $7 \times 24 = 168$.

Trained models

DDSC



Clear problem with heating/cooling.

Tests

DDSC

Method	Training Set Disagg. Err.	Training Set Acc.	Test Accuracy Disagg. Err.	Test Accuracy Acc.
Predict Mean Energy	20.98	45.78%	21.72	47.41%
SISC	20.84	41.87%	24.08	41.79%
Sparse Coding	10.54	56.96%	18.69	48.00%
Sparse Coding + TEP	11.27	55.52%	16.86	50.62%
Sparse Coding + GL	10.55	54.98%	17.18	46.46%
Sparse Coding + TEP + GL	9.24	58.03%	14.05	52.52%
DDSC	7.20	64.42%	15.59	53.70%
DDSC + TEP	8.99	59.61%	15.61	53.23%
DDSC + GL	7.59	63.09%	14.58	52.20%
DDSC + TEP + GL	7.92	61.64%	**13.20**	**55.05%**

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

Suggestions
DDSC

- Combining HMM
- Evolutionary NN
- Heating/Cooling - solve using Poisson processes
- The sparse coding
  - "Feature map" - Kernel to map input training samples
  - Separate activation appliances versus constantly on
  - "Usage" kernel - discuss whiteboard
- Inverse Structured Prediction Model, largest margin approach

Thank you!

Questions?