Toward a Semi-Supervised Non-Intrusive Load Monitoring System for Event-based Energy Disaggregation

GlobalSIP 2015, 14th Dec.
Karim Said Barsim and Bin Yang
Institute of Signal Processing and System Theory
University of Stuttgart, Stuttgart, Germany
{karim.barsim,bin.yang}@iss.uni-Stuttgart.de
Introduction

- Semi-Supervised Learning (SSL): leverage both labeled and unlabeled data.

Motivations
- Labeled data are: scarce ... costly... time-consuming
- Unlabeled data are: plentiful ... cheap ... rapidly growing

Advantages
- Improved performance with reduced labeling efforts.
- NILM systems learning over time.

Requirements / Limitations:
- Cluster assumption: decision boundaries through sparse regions only!
- Manifold assumption: same labels are close in geometry!
- Lazy training (transductive SSL models).
Introduction

Labeled and unlabeled data

Task & tools

Self-training 1
Self-training 2
Transductive learning 1
Transductive learning 2
Inductive learning 1
Inductive learning 2

Labeled data

- Sub-metered loads (eventless)
- Labeled events (event-based)

Unlabeled data

- Aggregate signals (eventless NILM)
- Segmented signals (event-based NILM)
Is semi-supervised learning suitable for NILM systems?

An SSL model: self-training

- Advantages
  - Simple SSL model
  - A wrapper model
  - Does not require unsupervised components

- Requirements
  - A learning algorithm $h$ or a seed classifier $f^0$
  - Confidence-rated predictions

- Limitations
  - Separable data/classes

NILM test dataset: BLUED\(^1\)

- Suitable for event-based NILM

Self-training 1

- Labeled dataset: \( \mathcal{L} = \{(x_n, y_n^{(n)})\}_{n=1}^{N_L} \)
- Unlabeled dataset: \( \mathcal{U} = \{x_n\}_{n=N_L+1}^{N_U} \)
- Training: \( \hat{f}(t) = h(\mathcal{L} \cup \hat{\mathcal{L}}) \)
- Prediction: \( \hat{y}(n) = \hat{f}(t)(x_n), \quad x_n \in \mathcal{U} \)
- Selection: \( \hat{\mathcal{L}} = \text{Sel}\left(\{(x_n, \hat{y}(n))\}_{n=1}^{N_{a}+N_{b}}\right) \)
Self-training 2: the double crescent problem

- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.

- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i - x_j\|^2}$
  - Minimal labelling (1 sample/class)

- Selection:
  - Farthest from boundary.

Labeled and unlabeled data

Task & tools

Self-training 1

Self-training 2

Transductive learning 1

Transductive learning 2

Inductive learning 1

Inductive learning 2

Classification problem:

- Iteration = 0, Accuracy = 62%
- Iteration = 0, Accuracy = 42%
- Iteration = 0, Accuracy = 71%

- Iteration = 0, Accuracy = 63%
- Iteration = 0, Accuracy = 100%
- Iteration = 0, Accuracy = 82%

- Iteration = 0, Accuracy = 78%
- Iteration = 0, Accuracy = 88%
- Iteration = 0, Accuracy = 71%

K. S. Barsim
Stuttgart University

(6/15) 14th Dec. 2015
Self-training 2: the double crescent problem

- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.

- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel \( e^{-\|x_i - x_j\|^2} \)
  - Minimal labelling (1 sample/class)

- Selection:
  - Farthest from boundary.

- 300 Iterations: > 70%
Self-training 2: the double crescent problem

- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.

- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i-x_j\|^2}$
  - Minimal labelling (1 sample/class)

- Selection:
  - Farthest from boundary.

- 300 Iterations: > 70%
- 500 Iterations: > 80%
Self-training 2: the double crescent problem

- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.
- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i-x_j\|^2}$
  - Minimal labelling (1 sample/class)
- Selection:
  - Farthest from boundary.
- 300 Iterations: > 70%
- 500 Iterations: > 80%
- 700 Iterations: > 90%
Self-training 2: the double crescent problem

- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.

- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i-x_j\|^2}$
  - Minimal labelling (1 sample/class)

- Selection:
  - Farthest from boundary.

- 300 Iterations: > 70%
- 500 Iterations: > 80%
- 700 Iterations: > 90%
- 1000 Iterations: optimal!
- Reduced labeling efforts: how much labeling is needed for near-optimal performance?
- When should SSL replace purely supervised models?
- **Object of classification:** Event-based features ($[dP, dQ]^T$ feature vectors)
- **Classifier:** Support Vector Machine (SVM) with a linear kernel
- **Selection:** nearest to class mean (based on the labeled samples)
- $1$ sample per class per iteration, $3$ iterations

**Dataset:** *BLUED dataset (refined)*
- Phase A: $749$ samples, $23$ classes
- Phase B: $1284$ samples, $45$ classes
Transductive learning: how much labeling?

- Reduced labeling efforts: how much labeling is needed for near-optimal performance?
- When should SSL replace purely supervised models?

![Graph showing performance comparison between semi-supervised and supervised models.](image-url)

**Phase A**
- Semi-supervised model
- Supervised model

**Phase B**
- SSL is no longer required! (~12% of labeling)
- Semi-supervised model
- Supervised model

Manifold assumption violation
Inductive learning: learning over time?

- Effect of increasing unlabeled dataset.
- Test dataset is fixed and includes inductive and transductive inference tests.
- **Object of classification:** Event-based features ($[dP, dQ]^T$ feature vectors)
- **Classifier:** Support Vector Machine (SVM) with a linear kernel
- **Selection:** nearest to class mean (based on the labeled samples)
- 1 sample per class per iteration, 3 iterations
- **Dataset:** BLUED dataset (refined)
  - Phase A: 749 samples, 23 classes
  - Phase B: 1284 samples, 45 classes
Inductive learning: learning over time?

- Effect of increasing unlabeled dataset.
- Test dataset is fixed and includes inductive and transductive inference tests.

<Diagram>

- Semi-supervised model
- Supervised model

Learning phases

Effects of cluster-assumption violation

Irrelevant labeling (samples behind support vectors)

Unlabeled samples have no effect on purely supervised models

4.67% of labeling in both phases

Mean F1-score

Time [day]
Discussion

Thank you for your attention