An Active Learning Framework for Non-Intrusive Load Monitoring

3rd International Workshop on NILM
May 15, 2016

Xin Jin, Ph.D.
National Renewable Energy Laboratory
Golden, Colorado, USA
xin.jin@nrel.gov
National Renewable Energy Laboratory (NREL)

NREL is a national lab of the United States Department of Energy
• Dedicated solely to energy efficiency and renewable energy technologies
• Highest number of patents, copyrights, and technology transfers among all labs
• Focusing on advancing science and technologies to address the nation’s energy and environmental goals
Energy Systems Integration Facility at NREL

• ESIF has been named 2014 Laboratory of the Year by the R&D Magazines
• A first-of-its-kind research user facility: an ultra-energy efficient workplace, one of the most energy efficient HPC data centers, and sophisticated high-bay lab spaces
• Partners include major players in high tech, home automation, and utility industries
Work with NREL

Why work with NREL?
• World-class testing facilities and researchers
• Strong partnership with clean technology industry
• Recognized neutral third-party for technology validation

NREL is the Technical Lead for DOE’s Building America program
• Develop cost-effective efficiency solutions for residential buildings
• Demonstrated in over 45,000 homes over past 20 years across the U.S.
• Partnership with industry (including many of the top U.S. home builders) to bring cutting-edge innovations and resources to the market

Ways to work with NREL
• ESIF User Facility Program
• Strategic partnership
• Cooperative Research and Development Agreement (CRADA)
• Wells Fargo Innovation Incubator Program (for clean tech startups)
• Funding opportunity collaboration (DOE, non-DOE, SBIR/STTR, etc.)
Key Barriers to Wide Adoption of NILM

NILM technology has existed for more than two decades but still hasn’t been widely adopted in most homes. Key barriers include:

• **Non-trivial initial setup is required**
  - Need to provide excessive information to train the NILM algorithms
  - Need to install additional hardware, such as current sensors to the service mains – which should only be done by a qualified electrician

• **Performance of NILM systems can be improved**
  - Pattern libraries cannot be shared among different suites of appliances
  - Difficult to achieve high disaggregation accuracy with limited user inputs

• **Cost of NILM systems are still higher than desirable**
  - The additional hardware (current/voltage sensors, displays, etc.) keeps NILM systems expensive
  - Software-based NILM systems are less expensive but they offer less granular disaggregation results
A Modular Approach to NILM

- Hundreds of NILM methods have been developed but most of those methods work on their own and do not interoperate.

- Every NILM method has its own strength in solving subproblems of NILM:
  - Appliance signature extraction
  - Event detection and/or classification
  - Interaction with users
  - ...

- **Can we make NILM methods modular** to help build up more robust systems using separate state-of-the-art approaches for each subproblems?
Active Learning-based NILM

What is active learning?
- A machine learning method that interactively queries the user for the class label information
- Widely adopted in scenarios where labeled samples are difficult, time-consuming or expensive to obtain
- A perfect fit for the NILM problems where data are abundant but labels are expensive to obtain

Why does AL work for NILM?
- Not all samples need to be labeled by the user; some samples are more informative
- We can use as few labeled samples as possible while still achieving high accuracy
- AL overcomes the key barriers of NILM by minimizing the cost of obtaining class labels
Pool-based Active Learning with $L^2$ Norm

Pool-Based Sampling:
Large collections of unlabeled samples can be collected at once and queries are selectively drawn from the pool.

Problem Formulation:
Active learning aims to find an unlabeled sample $x_q$ such that

$$x_q = \arg\max_{x \in X_U} \frac{1}{n} \sum_{i=1}^{n} d(x, x_i)$$

where $x_i \in X_K$, $X_K$ is a set of labeled samples, $X_U$ is a pool of unlabeled samples, $d(a, b)$ is the $L^2$ Norm, and $\frac{1}{n} \sum_{i=1}^{n} d(x, x_i)$ is the informativeness measure.

Procedures for generating a query:
- Compute the informative measure
- Identify the most informative sample
- Check the stopping criterion
- Obtain the class label from the user
- Store the label and remove the sample
Experiments on BLUED Dataset

BLUED Dataset
• BLUED is a fully labeled public dataset for event-based NILM research
• Voltage and current measurements were sampled at 12 kHz for a whole week
• Every state transition of each appliance was labeled and time-stamped

Feature Extraction
• The raw voltage and current data were converted to 60 Hz power data
• 150 data points were taken around an event, simulating an ideal event detector
• $\Delta P$, $\Delta Q$ and $P_{\text{overshoot}}$ (transient real power overshoot) were extracted as features
Results

- A four-fold cross validation is implemented
- A $k$-NN classifier is included in the learning loop to perform event classification every time a new label is provided by the user
- A random selection algorithm is implemented for comparative evaluation
- Both algorithms are implemented for 50 iterations

The active learning method reaches its peak (or near peak) performance by automatically selecting only using 10% of the total user-labeled samples.
Conclusion

• The proposed active learning method provides a modular solution for augmenting existing NILM systems.

• The proposed method only needs the minimally sufficient information to build a compact yet highly representative load signature library by querying the most informative samples first.

• The proposed method is able to reduce the user inputs by up to 90% while still achieving similar disaggregation performance compared to a heuristic method.

• The proposed method has the potential to overcome the key barriers to the wide adoption of NILM technologies:
  o **User Inputs**: substantially reducing the burden on the user
  o **Performance**: improving the performance of a NILM system with very limited user inputs
  o **Cost**: minimizing the incremental cost over existing systems by implementation in the cloud or on mobile platforms
Future Research

- Extend active learning to **proactive learning** that does not assume the user is *infallible* or *indefatigable* to bridge the gaps.

- Combine active learning with **semi-supervised learning** to further reduce the labeling effort:
  - Active learning: exploit the least confident samples
  - Semi-supervised learning: exploit the most confident samples
  - Combined learning scheme: only highly uncertain samples are labeled by a user and all others are automatically labeled.

- Explore other **informativeness measures** other than the Euclidean distance, such as information-theoretic measures.

- Develop a **labeling assist** module that uses time/frequency-of-use to infer the appliance information to reduce the labeling effort.

- Integrate active learning-based NILM techniques to address all of the energy sensing needs in a home: electricity, water & gas.
Thank You!