Unsupervised Classification of Residential Appliance Energy Usage

David Newcomb
Stanford University
04/22/15

Problem

The multifaceted modern day issues of energy and sustainability can be aprroached from several angles. Addressing these large scale issues requires not only large scale fixes, but also small scale solutions. One particularly interesting and promising area of small scale research involves residential energy use. Reducing energy usage on the residential level can lead to significant reductions in primary energy consumption, since it represents 21% of total primary consumption (United States).

One current problem is that the average homeowner only sees the total electricity usage each month and has no clear idea about which appliances are using how much energy or when. Knowledge of applicance-specific energy use has proven to effectively increase user energy efficiency by 15%. *The primary point of interest in this project, therefore, is to figure out how to use unsupervised machine learning to disaggregate the total electricity usage profile into its components. This is called Non-Intrusive Load Monitoring (NILM).*

Research

A significant part of the research put forth thus far has involved Hidden Markov Models, k-means clustering, and Support Vector Machines. Hidden Markov models help identify the presence of low-power applicances in the presence of high loads, whereas k-means and SVM classifiers have been better at helping identify the more prominent profiles. My goal in this project will be to focus on identifying the high-power applications and doing so with a method that performs as close to real time as possible. Being able to quickly identify the higher loads can help with demand response initiatives involving smart meters. I largely anticipate challenges with the speed of the algorithm and algorithmic accuracy.

Data

Currently, there is one dataset aimed at opening up the question of disaggregation, called the Reference Energy Disaggregation Dataset (REDD). It is an open dataset created by MIT professors and it tracks the energy usage of real homes by delivering a signal representing whole home energy use and up to 24 separate subsignals representing various home appliances. The data, therefore, have both training and testing components. For the moment, this is my sole source of data and I will continue to search for additional datasets.

References

[1] Z. Ghahramani and M. I. Jordan. Hidden Markov Models. *Machine Learning*, 1997.

[2] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011.

[3] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 2012.