Human Occupancy/Activities Prediction

Based on ECO Dataset

Yicheng Wang wych@ucla.edu
Longjia Niu niulongjia@ucla.edu

Our website: https://sites.google.com/view/activities-prediction-202b

Background & Prior work

Human occupancy detection is the core part of building automation systems.

Human activities detection is an enhanced version of human occupancy detection, which allows automation system to adopt different power saving strategies based on specific human activities.

Current building automation systems

Extra cost and installation complexity.

Human Occupancy/Activities Prediction

Save energy not only from the household appliances but also the monitoring system for occupancy detection.

Our goals in this project

Human **occupancy**/**Activities** prediction

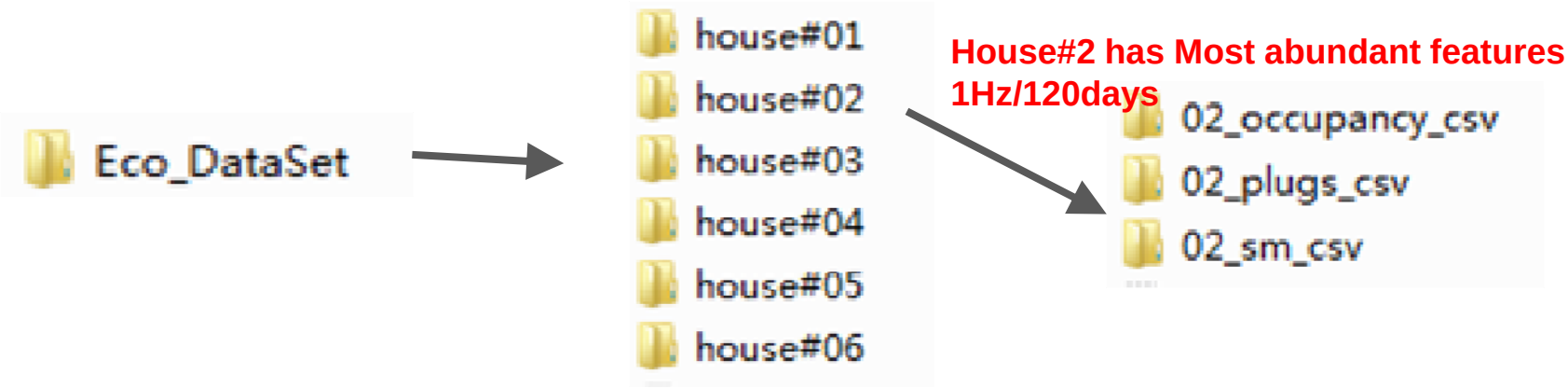
- **Predict** human occupancy/activities states with different classification models.
- **Evaluate** the performance.
- Try to **enhance** the performance.

ECO Dataset

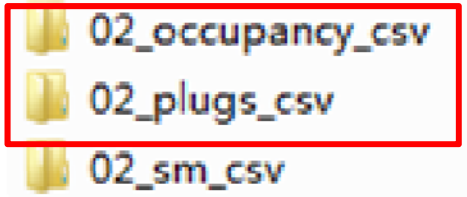
ECO (Electricity Consumption and Occupancy) data set is the first data set available that contains both electricity consumption and ground truth occupancy information of households.

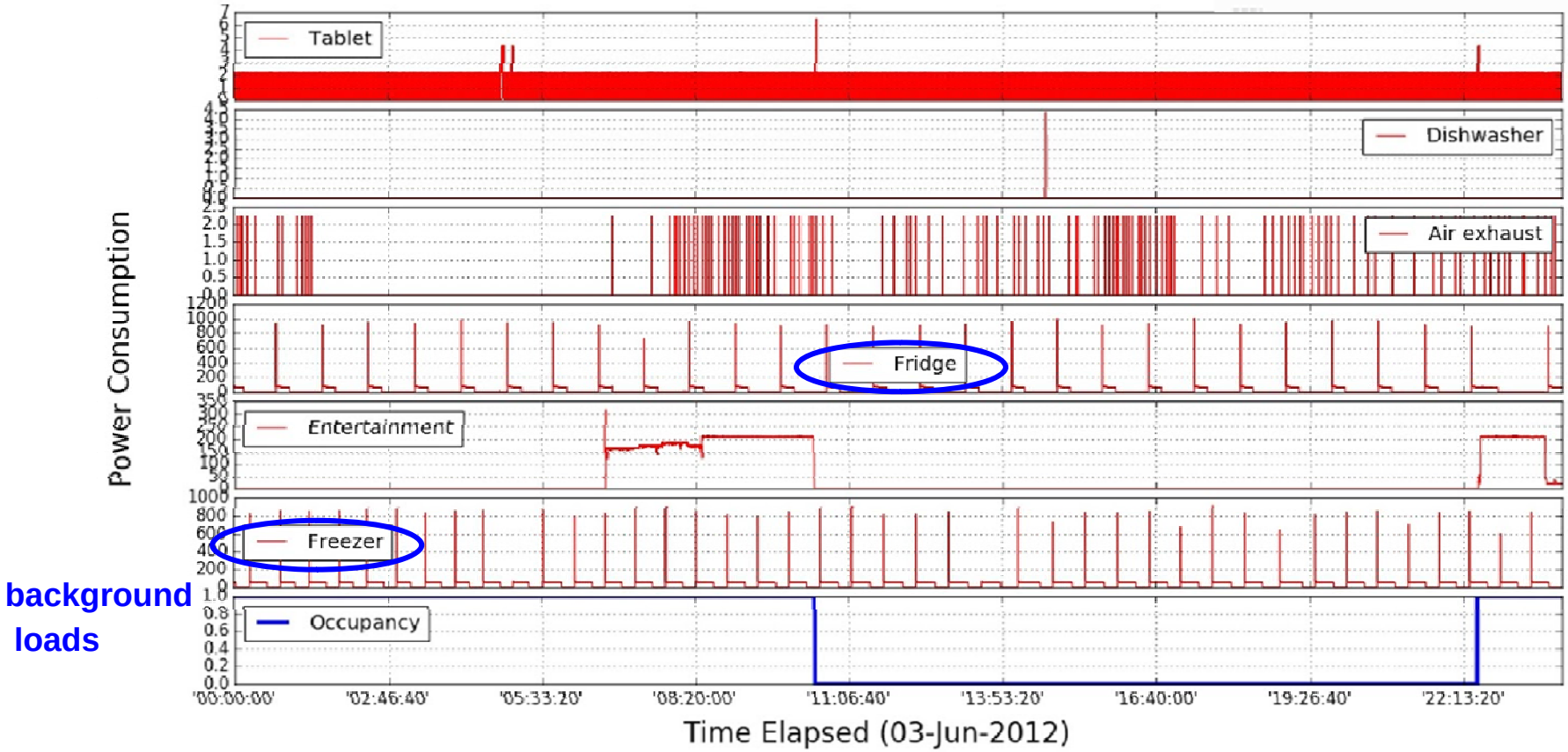
It was collected in 6 Swiss households over a period of 8 months

Online database: http://data-archive.ethz.ch/delivery/DeliveryManagerServlet?dps_pid=IE594964

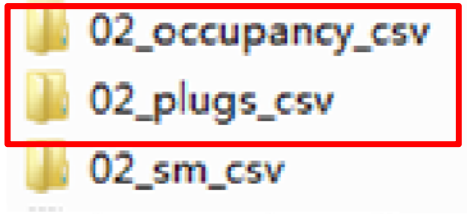


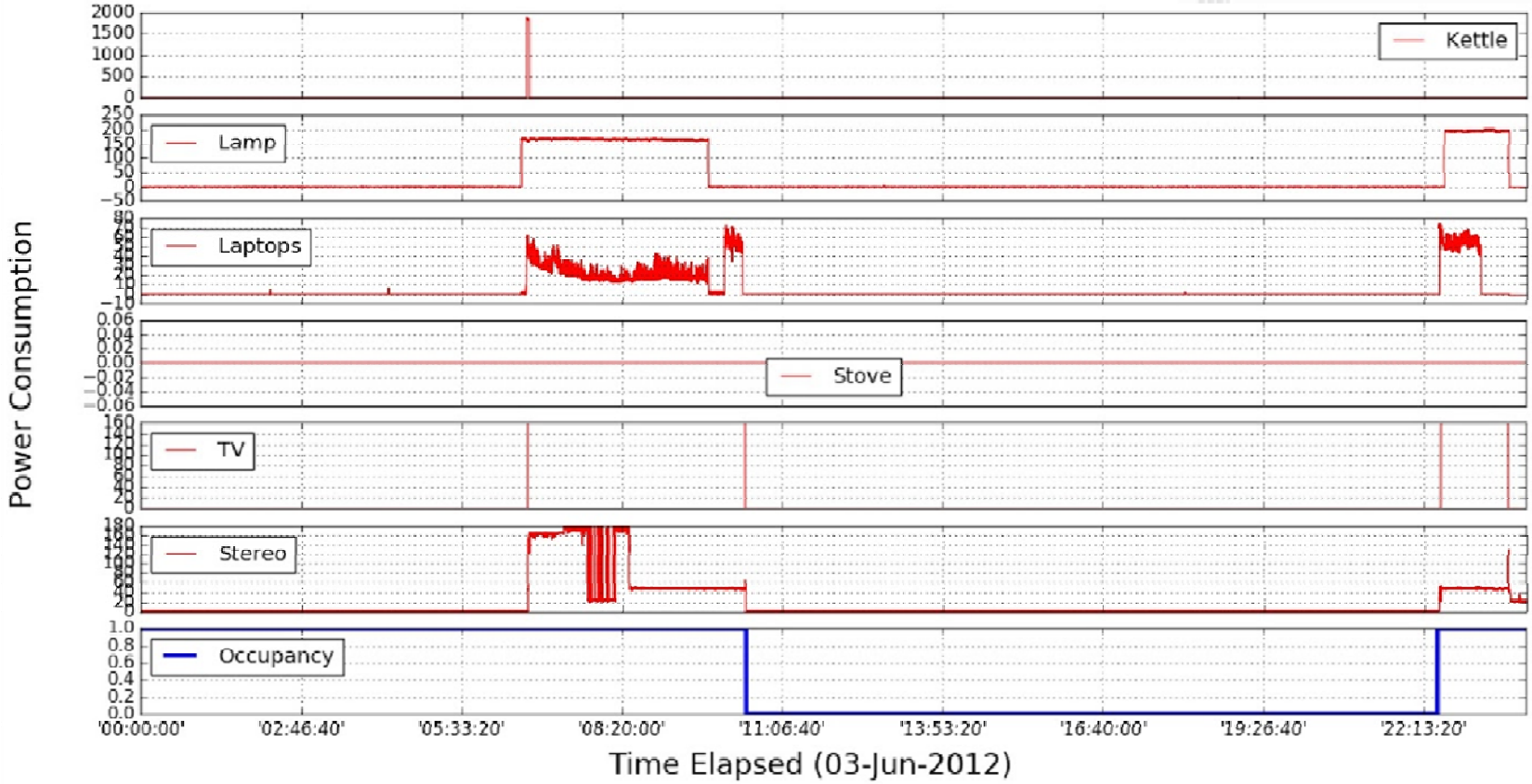
House# 2 Data Visualization --- 1



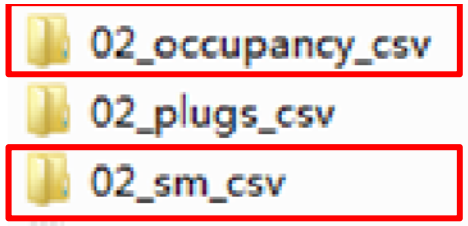


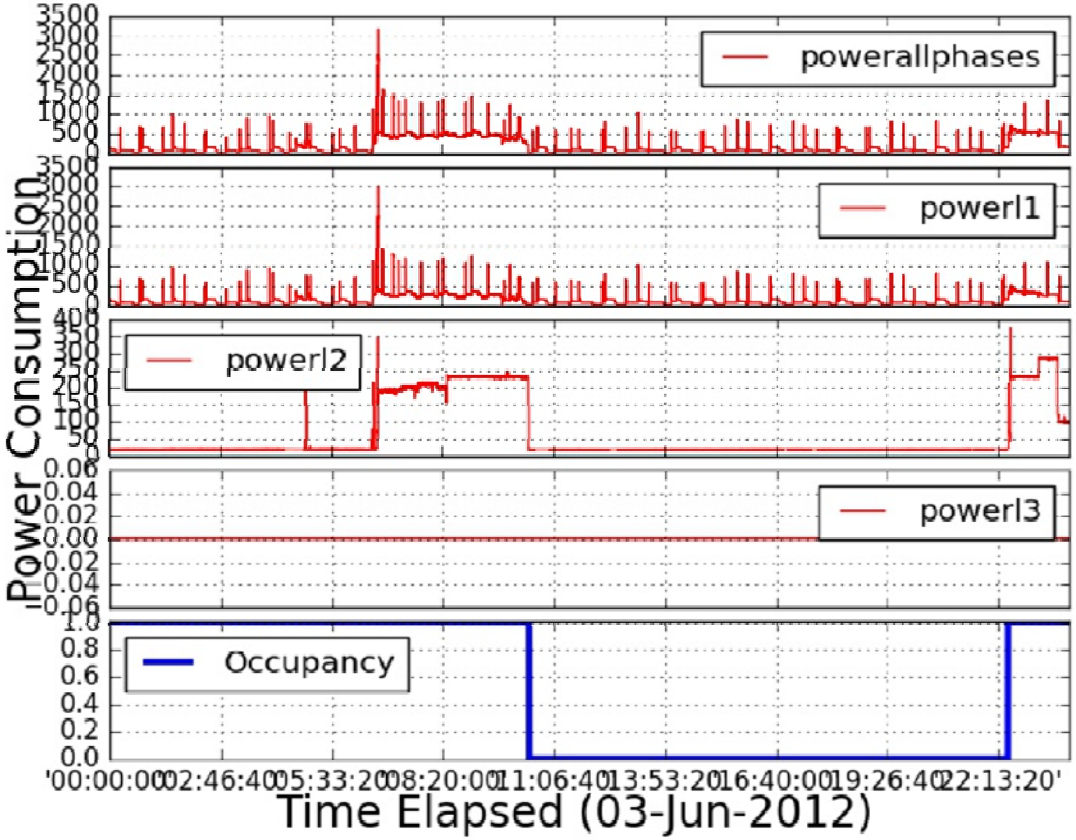
House# 2 Data Visualization --- 2



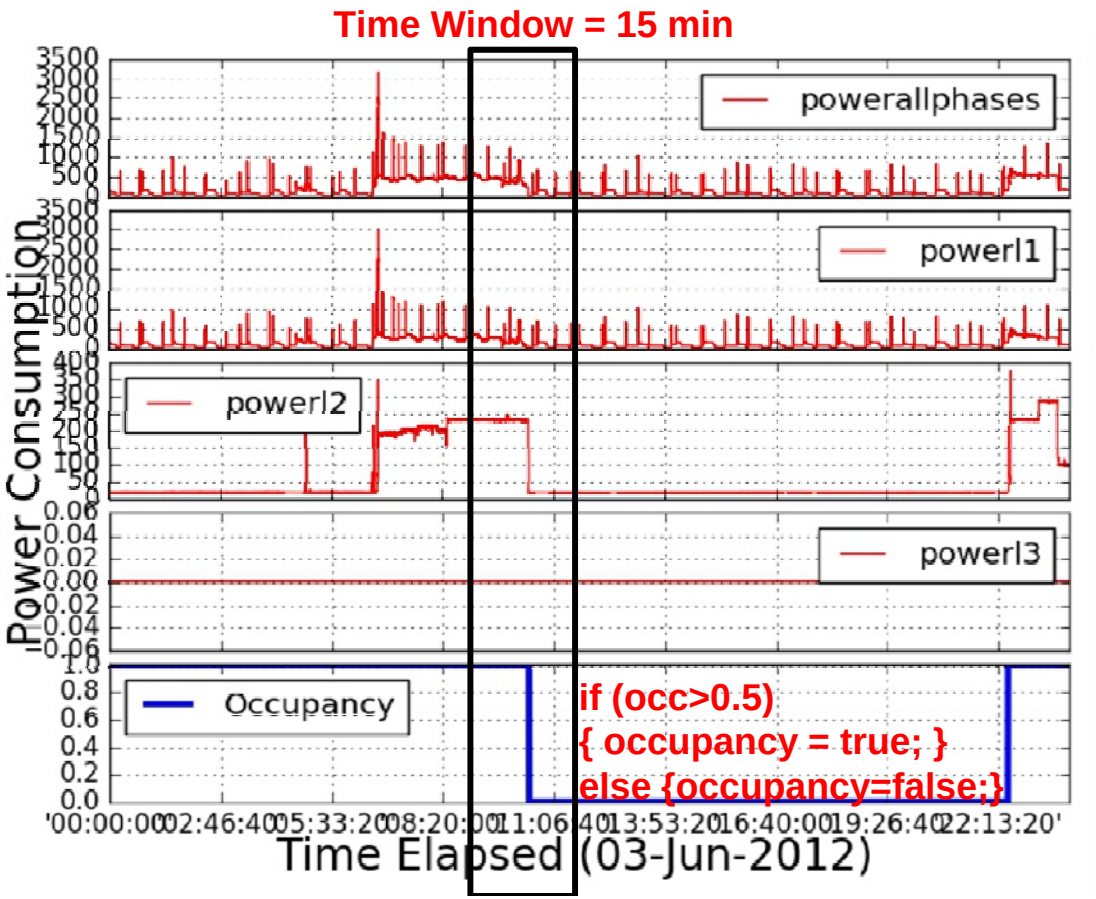


House# 2 Data Visualization --- 3





Human Occupancy Prediction --- Data Preprocessing



Human Occupancy Prediction --- Features Selection

Using **plug-level** power consumption to predict human occupancy.

Features	Description
Min	Minimum power value in one time slot for 12 appliances
Max	Maximum power value in one time slot for 12 appliances
Mean	Arithmetic power average in one time slot for 12 appliances
Std	Standard deviation of power in one time slot for 12 appliances
Range	Range of power in one time slot for 12 appliances
Cor	Autocorrelation at lag 1 computed over one time slot for 12 appliances
Time	Time of the day (morning/afternoon/evening) for 12 appliance

Using **smart meters** power consumption to predict human occupancy.

Features	Description
Min	Minimum power value in one time slot for all power phases
Max	Maximum power value in one time slot for all power phases
Mean	Arithmetic power average in one time slot for all power phases
Std	Standard deviation of power in one time slot for all power phases
Sad	Sum of absolute differences of power in one time slot for all power phases
Range	Range of power in one time slot for all power phases
Cor	Autocorrelation at lag 1 computed over one time slot for all power phases
Time	Time of the day (morning/afternoon/evening) for all power phases
onoff	Number of on/off state switch in one time slot for all power phases

Human Occupancy Prediction --- Features Selection

Absolute value of the power consumption

The min, max and mean features are used to capture the absolute value of power consumption at each time slot. A high value of power consumption may provide an indicator of human occupancy.

Variability of power consumption

Significant power changes are often the result of human actions. We use std, range, cor, and onoff features to indicate the variability.

Dependency of occupancy

Time. The occupancy heavily depends on the time of the day, for example, the house is much more possible to occupied in the evening than in the afternoon.

Human Occupancy Prediction --- Classification Models

K nearest neighbors: A simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).

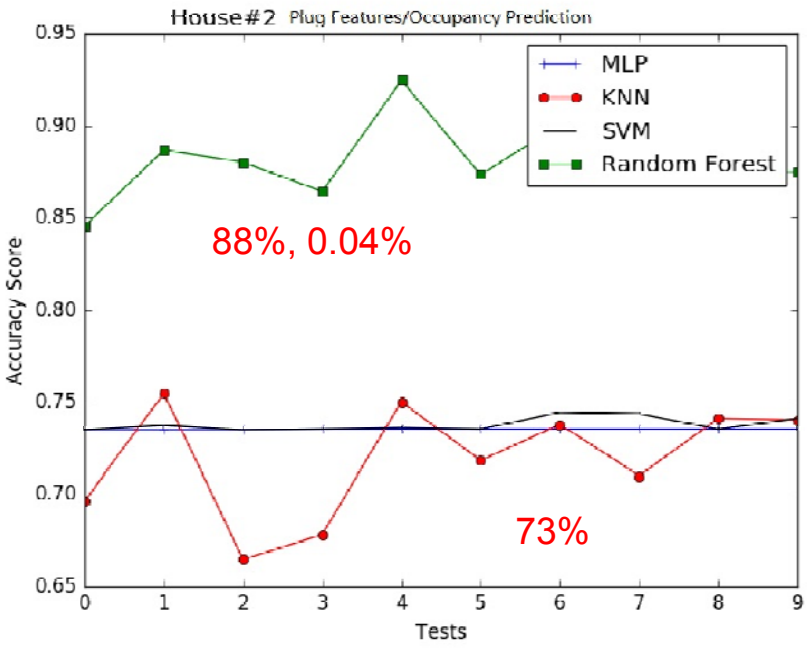
Support vector machine: A method of constructing a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks.

Multilayer perceptron (MLP): A feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs.

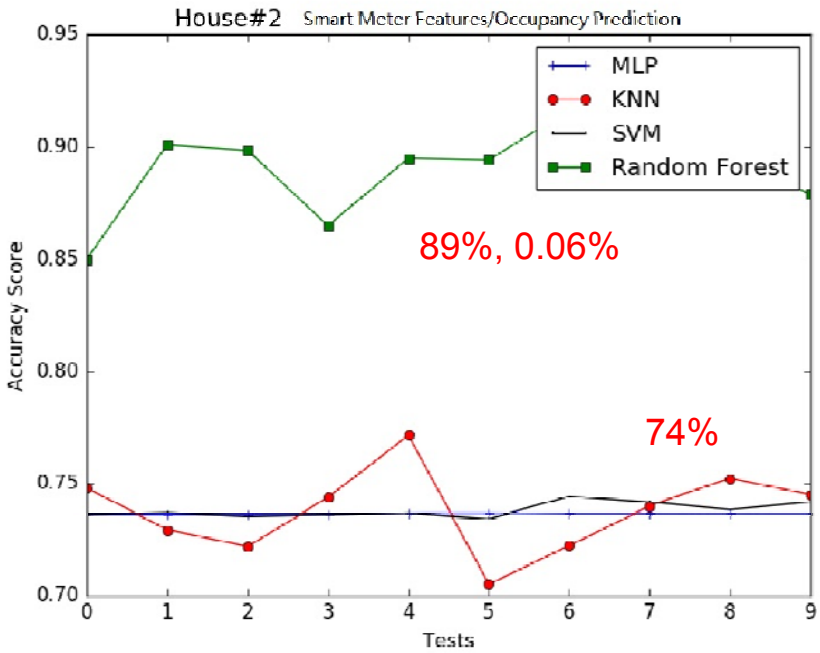
Random forests: An ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees.

Human Occupancy Prediction --- Results & Evaluation

Using **plug-level**
power consumption
to predict human occupancy.



Using **smart meters**
power consumption
to predict human occupancy.



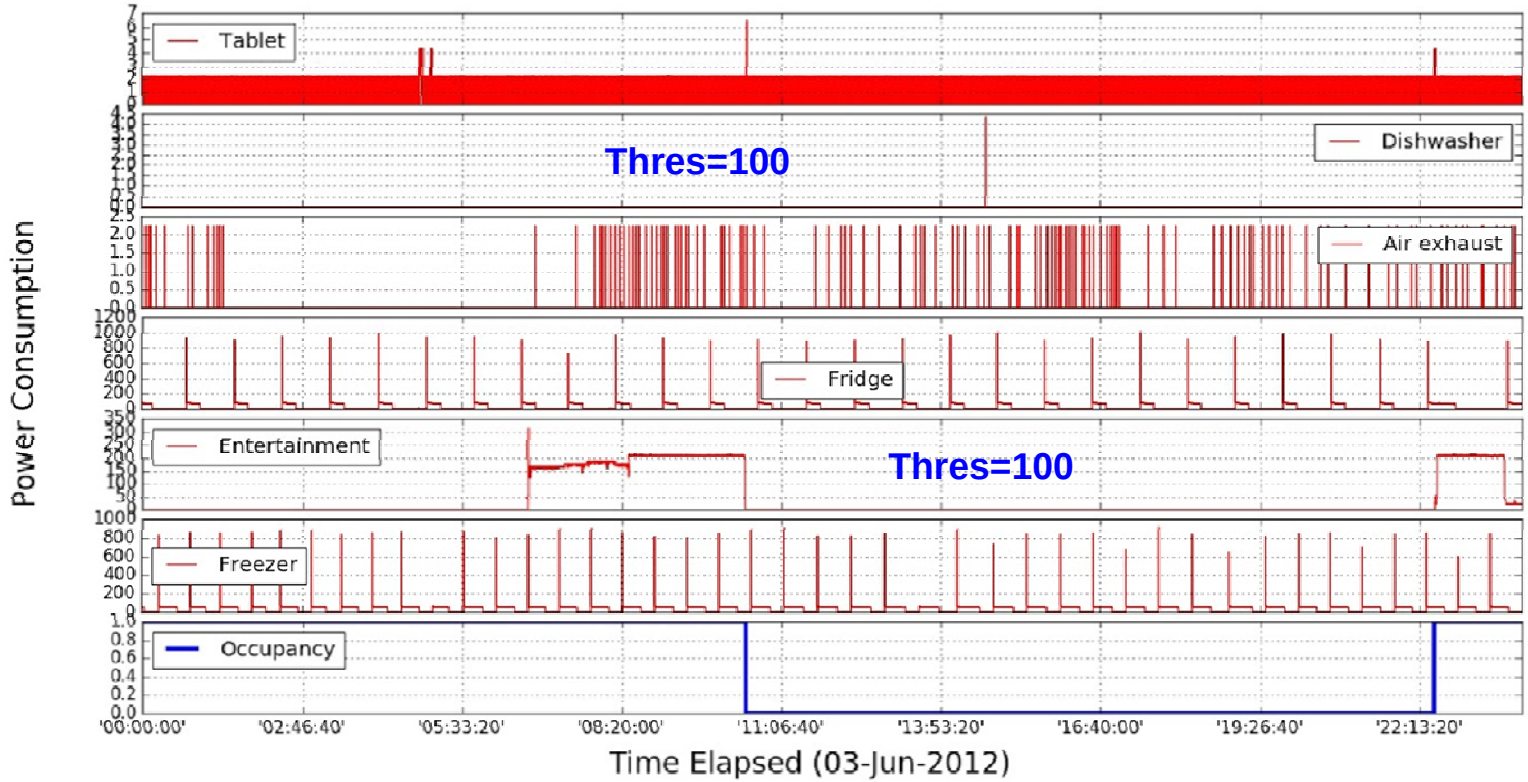
Human Occupancy Prediction --- Results & Evaluation

Prediction result analysis

Data type	plug-level power consumption data	smart meters power consumption data
dataset size	12000	12000
Accuracy	88% (Random Forest)	89% (Random Forest)
Variation	0.0004	0.0006
Effective features	**Min,max,...,cor** (below 30% for MLP, 60-70% accuracy for other 3) **Onoff** (increase 20% for MLP, 5% for other 3) **Time** (increase around 15% accuracy)	**Min,max,...,cor** (below 30% for MLP, 60-70% accuracy for other 3) **Onoff** (increase 40% for MLP, 5% for other 3) **Time** (increase 17% accuracy)

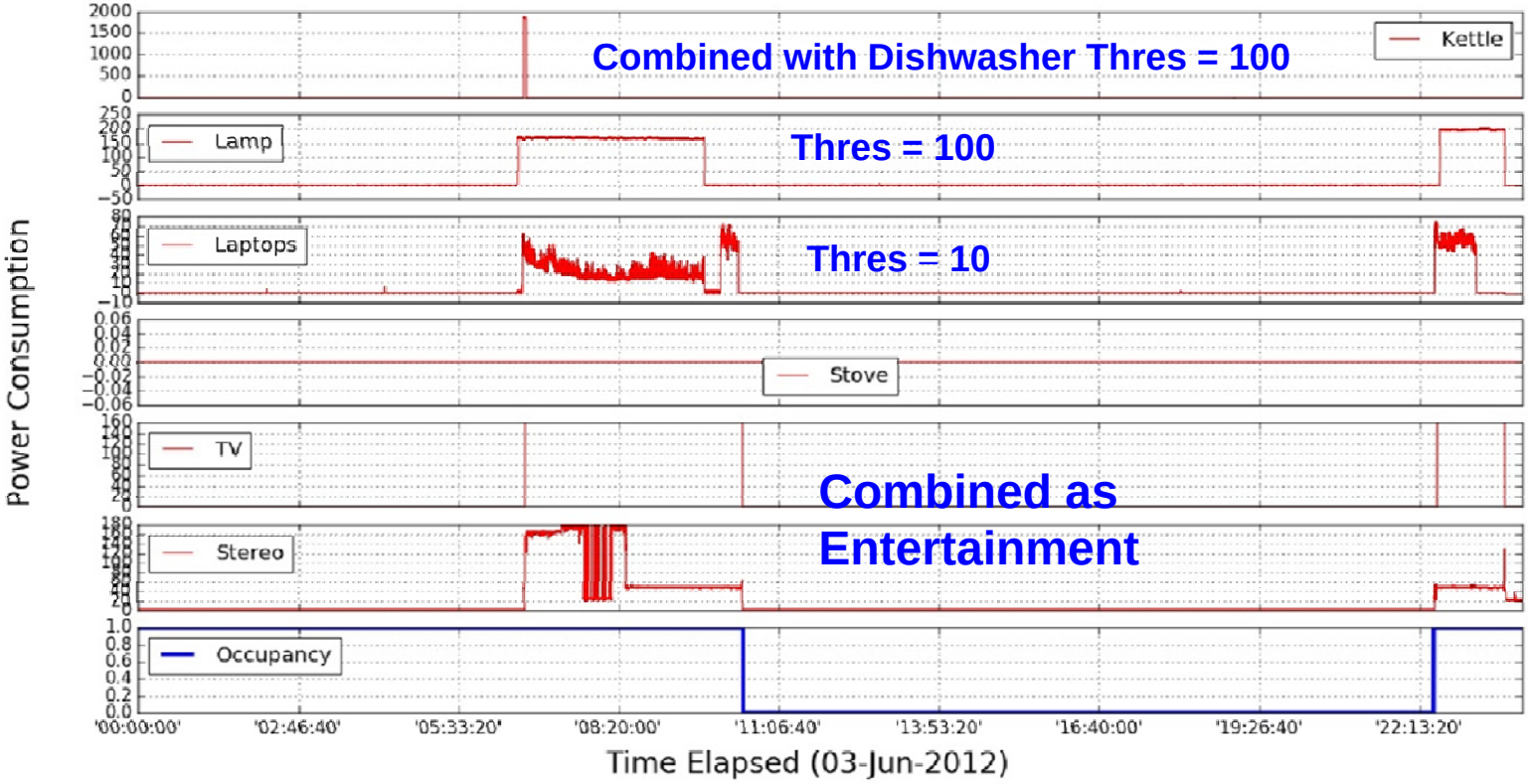
Human Activities Prediction --- Data Preprocessing#1

Ground truth human activities are not provided, our first step is to make reasonable assumptions about human activities from plug-level power consumption data.

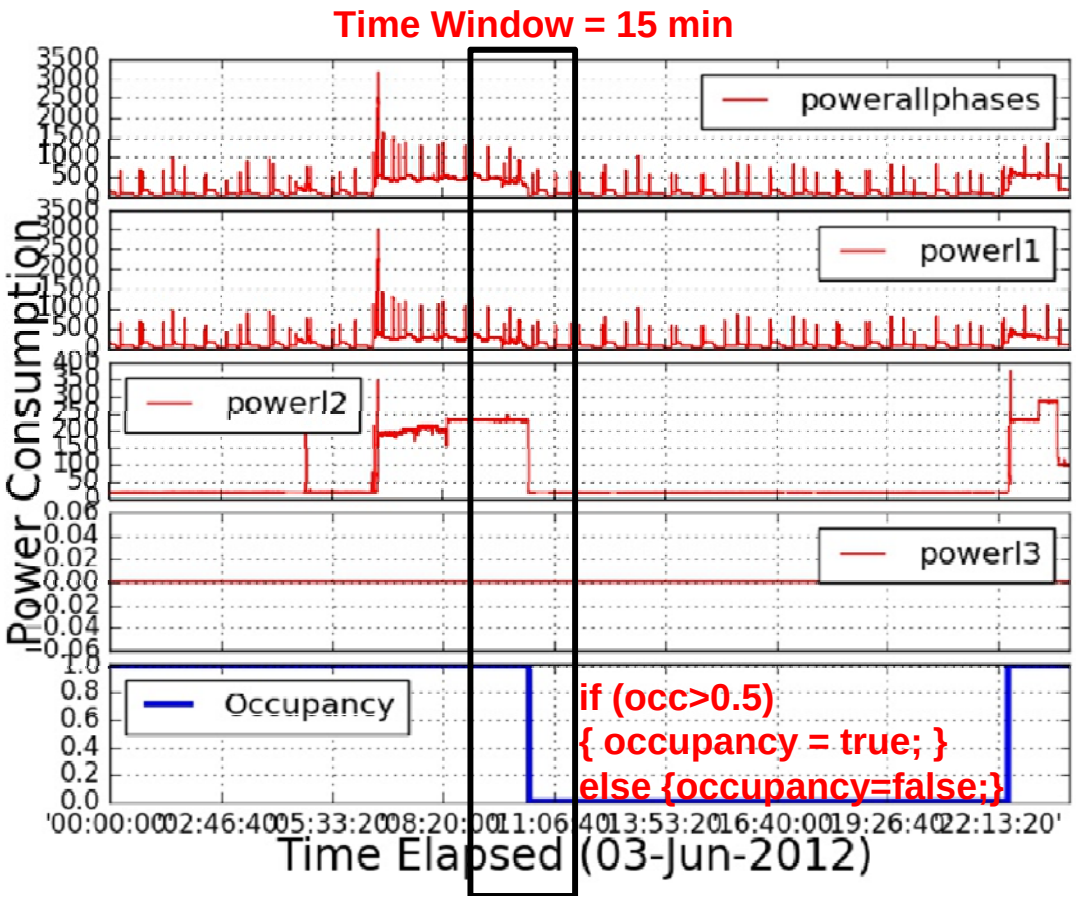


Human Activities Prediction --- Data Preprocessing#2

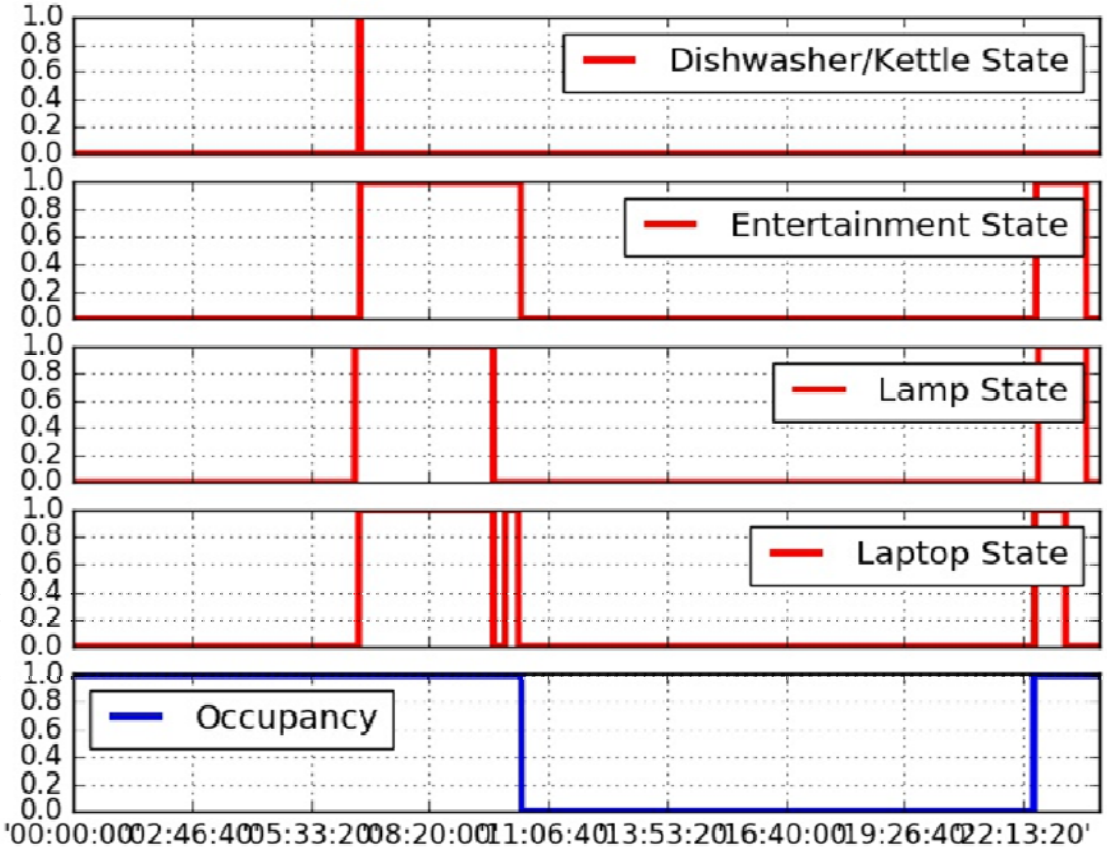
Ground truth human activities are not provided, our first step is to make reasonable assumptions about human activities from plug-level power consumption data.



Human Activities Prediction --- Data Preprocessing#3



Human Activities Prediction --- After Data Preprocessing



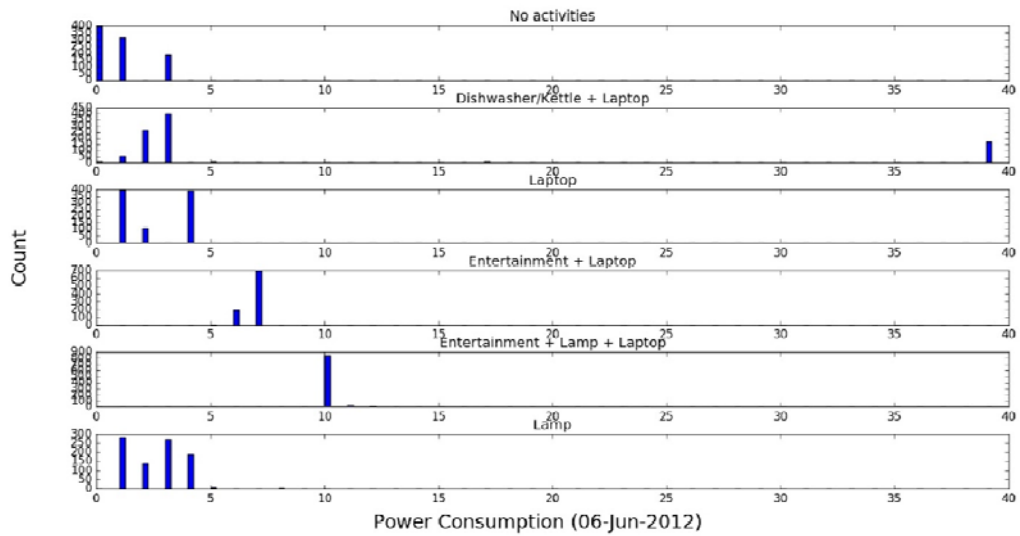
Human Activities Prediction --- Features Selection

Same features

as human occupancy prediction

Features	Description
Min	Minimum power value in one time slot for all power phases
Max	Maximum power value in one time slot for all power phases
Mean	Arithmetic power average in one time slot for all power phases
Std	Standard deviation of power in one time slot for all power phases
Sad	Sum of absolute differences of power in one time slot for all power phases
Range	Range of power in one time slot for all power phases
Cor	Autocorrelation at lag 1 computed over one time slot for all power phases
onoff	Number of on/off state switch in one time slot for all power phases

Power consumption distribution to predict human activities



Human Activities Prediction --- Target Generation

Target:

0101 -> 5

Dishwasher/Kettle state is **OFF**

Entertainment state is **ON**

Lamp state is **OFF**

Laptop state is **ON**

Human Activities Prediction --- Classification Models

K nearest neighbors: A simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).

Support vector machine: A method of constructing a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks.

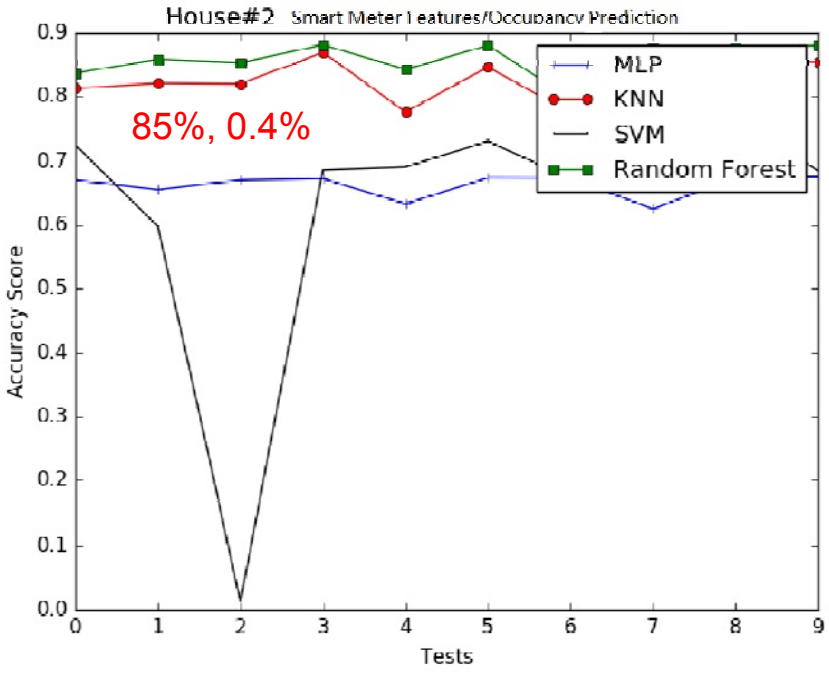
Multilayer perceptron (MLP): A feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs.

Random forests: An ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees.

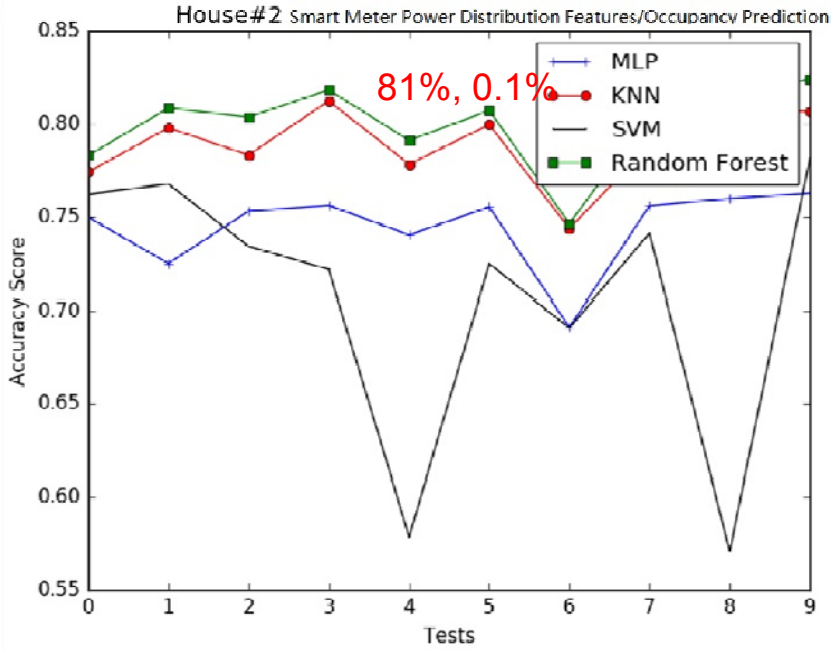
Human Activities Prediction --- Results & Evaluation

Same features

as human occupancy prediction



- **Power consumption distribution** to predict human activities



Human Activities Prediction --- Results & Evaluation

Prediction result analysis

Feature type	Min, max, mean, Range, std, cor, onoff	Power consumption distribution
Data type	Smart meter power consumption data	Smart meter power consumption data
dataset size	23000+	23000+
Accuracy	85.8% (Random Forest)	81% (Random Forest)
Variation	higher variation (0.04 for svm)	lower variation (below 0.001)

Summary

Visualize the ECO dataset and process the power data for feature extraction and prediction.

Implement different features for the prediction and compare the performance.

Build and evaluate different classification models to predict **human occupancy** with house energy consumption information.

Build and evaluate different classification models to predict **human activities** with house energy consumption information.

Future Work

ECO Dataset does not have ground truth data on sleeping, the electricity consumption is usually less indicative of occupancy during sleeping, thus more assumptions/data collection/analysis should be done.

Optimize the prediction by finding more features, changing the parameters in current classification models and using other classification models.

Predict how many people in the house based on data from smart meter and plugs if we can get related ground truth datasets.

References

Tang Y, Ono C. Detecting Activities of Daily Living from Low Frequency Power Consumption Data[C]//Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. ACM, 2016: 38-46.

Kleiminger W, Beckel C, Santini S. Household occupancy monitoring using electricity meters[C]//Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2015: 975-986.

Kleiminger W, Beckel C, Staake T, et al. Occupancy detection from electricity consumption data[C]//Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings. ACM, 2013: 1-8.

Thanks for watching!