Identification of Electrica
Learning: A Study in Col

Konrad

Introduction and Problem Statement

Managing energy consumption in smart homes is challenging due to the increasing number of devices. Homeowners lack detailed information about their energy use, resulting in inefficiencies and higher costs. Non-Intrusive Load Monitoring (NILM) aims to solve this by analyzing total energy consumption to identify and estimate the usage of individual devices. However, each household has unique patterns, making it complex to accurately disaggregate this data.

Purpose and Importance

This research aims to develop and refine NILM methods to improve accuracy in identifying individual device usage within households. By providing detailed insights into energy consumption patterns, NILM can help optimize energy usage, reduce costs, and promote sustainable living. Enhanced NILM technologies will empower homeowners with the information needed to manage their energy effectively.

Background

Smart homes have evolved to allow remote control and automation of various systems, enhancing convenience. However, the proliferation of smart devices complicates energy management. NILM technologies analyze aggregated energy data to pinpoint device-specific usage without separate meters. Each household's unique energy patterns add complexity to this process. This research seeks to advance NILM methods, providing critical insights for better energy management and sustainability.

NILM Methods

1. Data Collection and Preparation

- Gather Data: Collect comprehensive datasets capturing both aggregated power usage and individual device consumption.
- Signal Processing: Preprocess data to filter out noise and extract relevant features.

2. Problem Formulation

- Analysis of Devices: Researching what devices to use in NILM
- Data Clustering: Use k-means clustering to simplify power measurements by predicting device operating states.
- Feature Selection: Extract additional features like hour of the day, day of the week, and weekend indicators to enhance model performance.

3. Disaggregation Methods

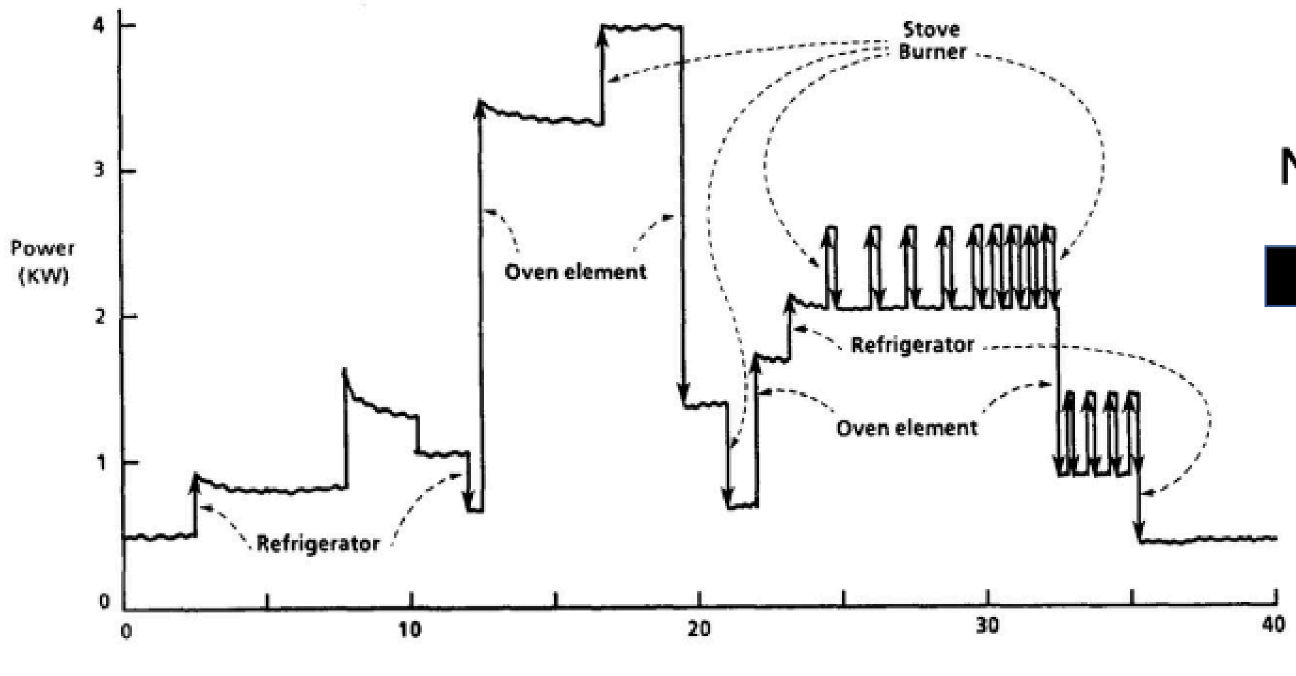
- Random Forest: Implement a Random Forest for initial disaggregation, focusing on robustness and feature importance.
- RNN/LSTM: Utilize Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks to capture temporal dependencies in sequential data.
- Process Mining: Apply process mining techniques to predict device usage patterns and provide insights into sequential device operations.

4. Model Evaluation

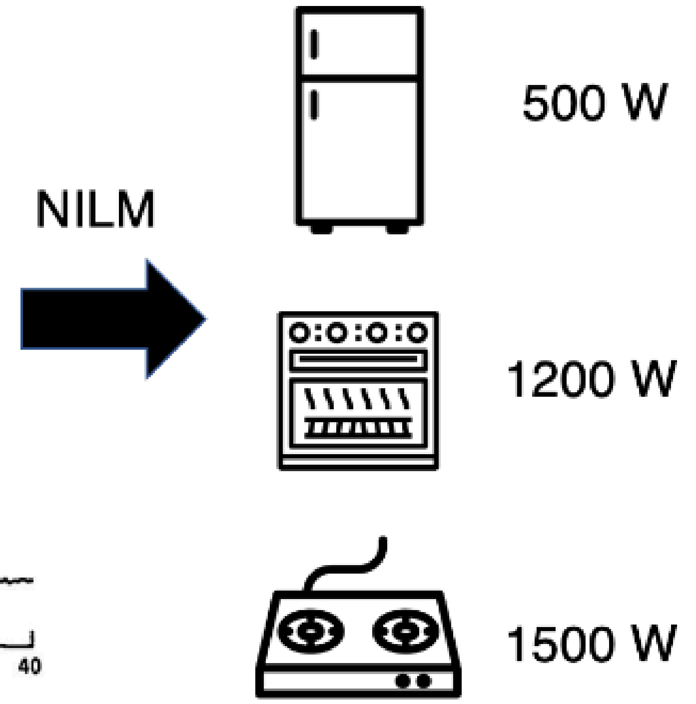
- Split data into training (70%) and testing (30%) sets to ensure robust model evaluation.
- Evaluate models based on accuracy, precision to determine effectiveness in identifying device-specific energy usage.

al Devices Using Machine laboration with Origin AS

. Jarczyk



* Hart (1992)



3. Disaggregation Methods

The problem of device identification is a classification problem. Therefore, it is used algorithms that excel in classification for disaggregation methods. The DataFrame with the chosen devices is divided into training and testing subsets.

- **Random Forest**: We start with Random Forest classification, employing the MultiOutputClassifier. This method is straightforward and effective, especially with large datasets, providing robust performance and feature importance insights.
- **RNN with LSTM Layers**: For more complex temporal dependencies, we use RNN with LSTM layers. The dataset is divided into sequences of 10,400 records (equivalent to one day). By structuring the data into separate columns for each device and predicting all devices together, the model captures inter-device dependencies, enhancing prediction accuracy.

RNN/LSTM Architecture

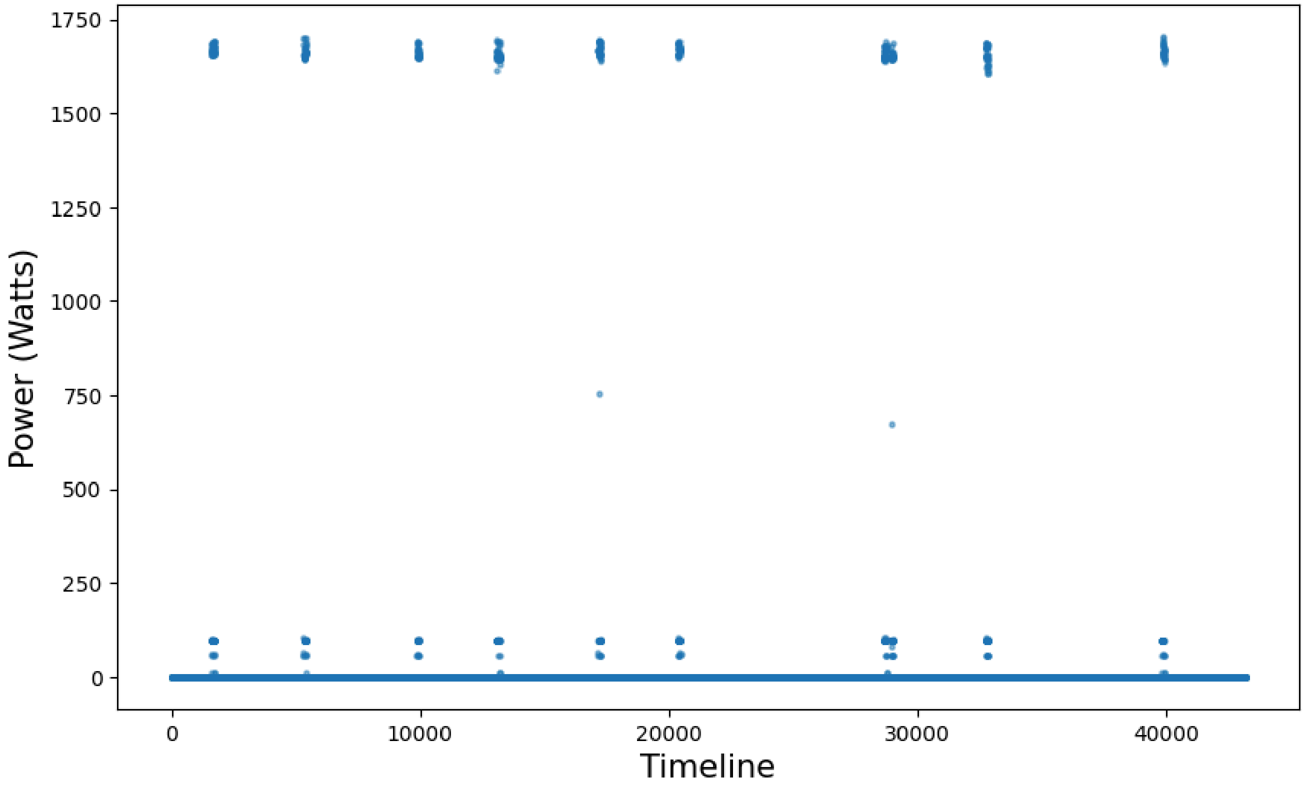
Problem Solution

1. Data Collection and Preparation

The dataset used is UK-DALE, which includes data from five houses, each with unique users and devices. Data from all devices is merged at one-minute intervals along with the aggregated power consumption.

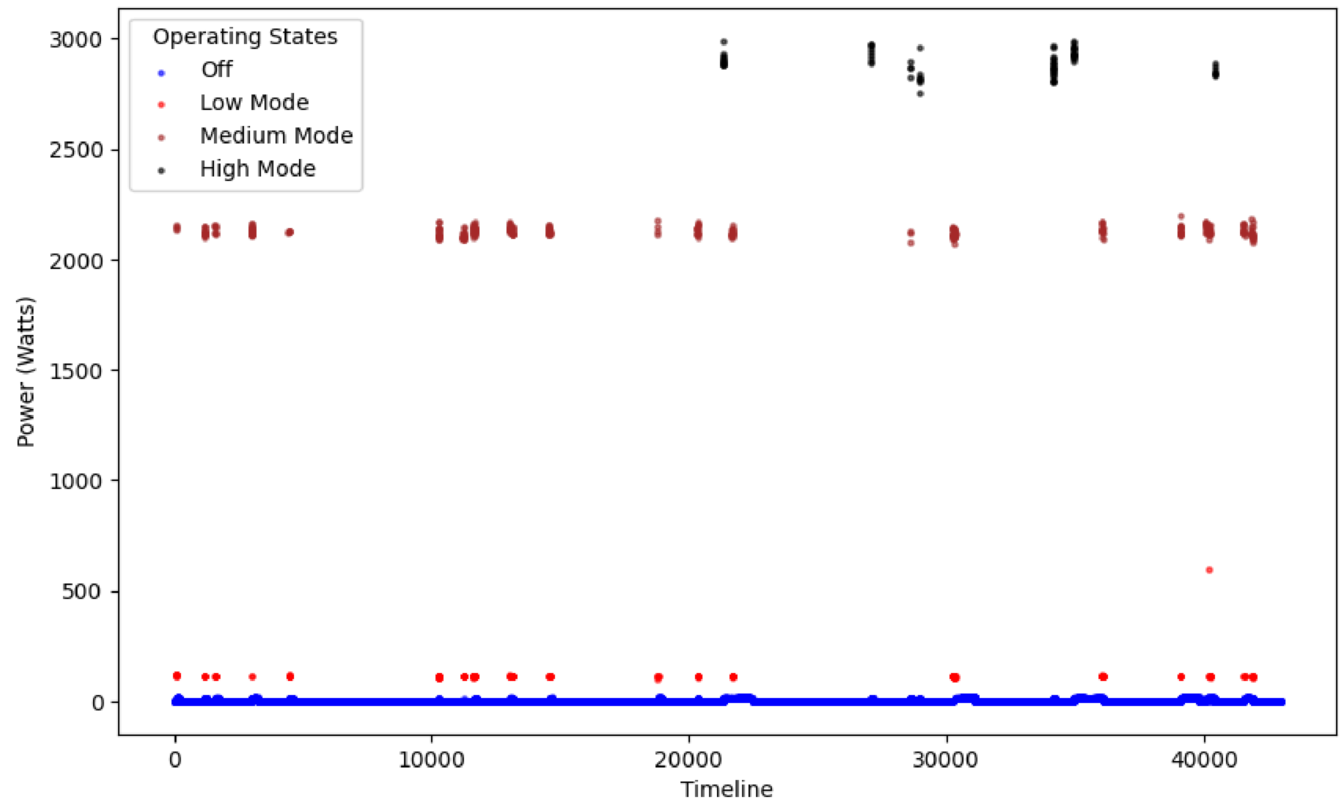
2. Problem Formulation

There are various types of devices in a household, some of which are turned on irregularly, while others have consistent operating patterns. For NILM analysis, only devices that operate in repeatable cycles and have a steady number of operating states are chosen. Devices with variable energy consumption, such as dimmable lights or TVs with custom settings, are excluded from this analysis.



Plot of dishwasher energy consumption over a month period.

Using the k-means algorithm, the selected devices with a fixed number of operating states are clustered into their respective states.



Clustered usage of oven over a month.

Machine learning algorithms require specific features to accurately identify devices. To enhance the predictive power, additional features are extracted from the timestamp and aggregated power load data.

	timestamp	what_hour	what_day	is_weekend	aggregate	agg_clustered
0	2014-06-29 16:23:00	16	6	1	769.000000	2
1	2014-06-29 16:24:00	16	6	1	1077.700000	3
2	2014-06-29 16:25:00	16	6	1	2264.444444	5
3	2014-06-29 16:26:00	16	6	1	763.500000	1
4	2014-06-29 16:27:00	16	6	1	1980.400000	4

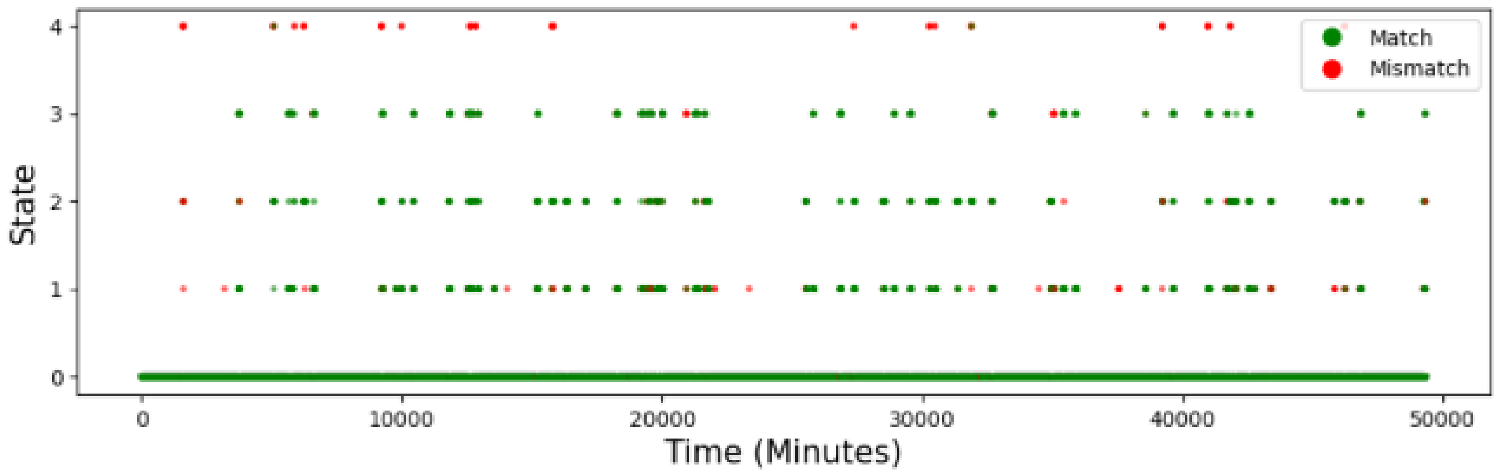
Snippet of DataFrame with selected features

4. Model Evaluation

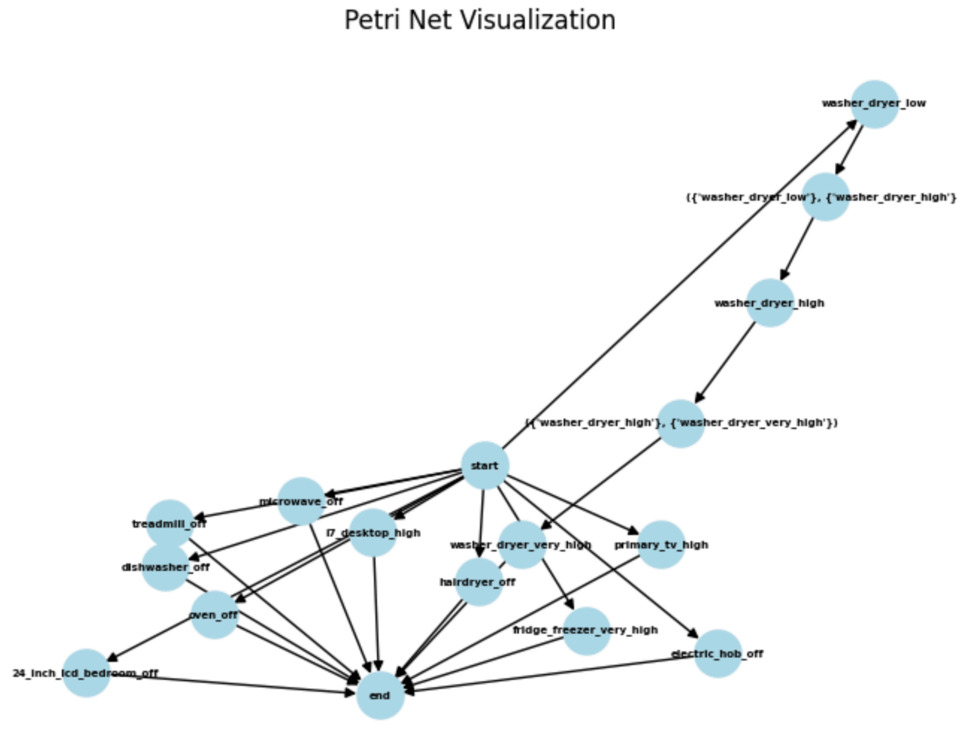
The results indicate that the Random Forest method performs slightly better than the RNN/LSTM method. Both methods show similar hamming accuracy, meaning each value for every device is correctly identified. However, general accuracy, which measures the correctness of all values in a row at a given timestamp, shows a more significant difference between the models, favoring Random Forest.

	ACCURACY
Random Forest	99.21%
RNN / LSTM	97.65%
CNN	88.45%

Model Results comparison



Predicted vs. Actual States of Dishwasher



Process Mining

Conclusion

The results demonstrate that our NILM algorithms can accurately identify individual device usage within a household. Further work could explore integrating process mining techniques to enhance device identification. Additionally, efforts could focus on remodeling machine learning algorithms and refining feature engineering to further boost performance and adaptability in diverse household settings.