South Carolina

A Dataset and Baseline Approach for Identifying Usage States from Non-intrusive Power Sensing with MiDAS IoT-Based Sensors

Bharath Muppasani[1], Cheyyur Jaya Anand[2], Chinmayi Appajigowda[2], Biplav Srivastava[1], Lokesh Johri[2]

[1] AI Institute, University of South Carolina, Columbia, South Carolina, USA
[2] Tantiv4, San Jose, California, USA

AI INSTITUTE #AIISC UNIVERSITY OF SOUTH CAROLINA

TANTIV4

Accepted for **IAAI - 2023**

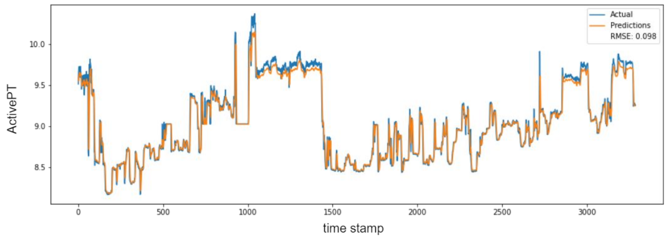
Outline

- Motivation
- Dataset
- State Identification
- Baseline Results
- Contributions
- References

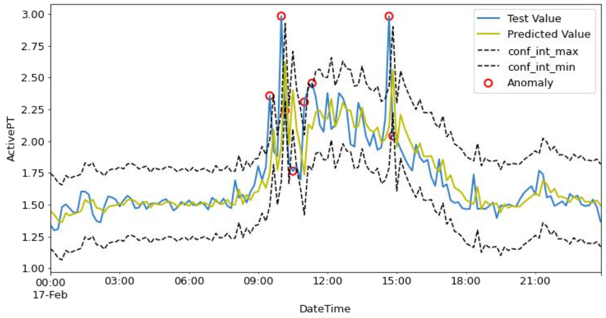
Motivation

- Major focus of literature work
  - Power Forecasting, Anomaly Detection, Non-Intrusive Load Monitoring.
- Traditionally the malfunctions in a system like a machine beginning to fail due to a fault are analysed as vibration data in a technique called as *Condition Based Monitoring.*
  - But this is expensive and intrusive.
- In an electrical system running on AC frequency, any perturbation in the system manifests as an energy spectrum. These spectrum patterns, or signatures, reflects the *state* in which the system could be. Few examples of a state -
  - A factory floor with all machines running normally or a machine beginning to fail due to a fault;
  - In a datacenter or office, the system under heavy load;
  - In any industry, an unusual load that the system has not seen till now and needs operator attention.

Power Forecasting



Anomaly Detection

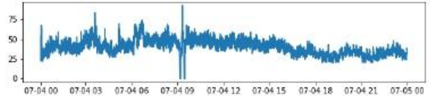
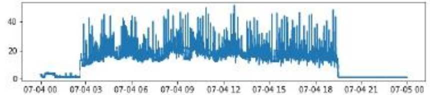
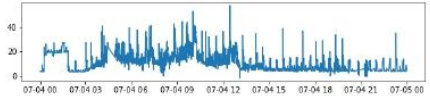
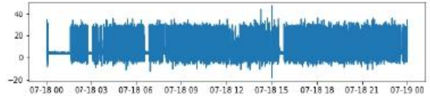
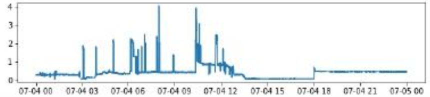
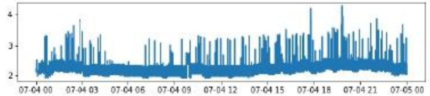
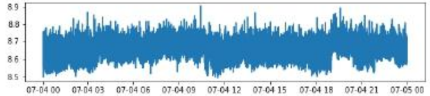


Data Collection

- The MiDAS IoT sensor from Tantiv4 measures all the crucial electrical parameters like current, voltage, power and other values every 300ms.
- The device is also capable of collecting three-phases of current and voltage harmonics data from 2 to 32 harmonic levels along with total harmonic distortion for each phase of current and voltage every 500ms.

	Electricity Consumption Data	Harmonics Data
Number of features	28	193
Frequency	3.33 Hz	2 Hz

Dataset

Location	Industry	Load Illustration	Size of Released Dataset (15-days)	Load Figures (ActivePT vs datetime)*
India-1	Manufacturing	Laser Cutting Machine	ECD - 212MB Harmonics - 3.2GB 5.34% missing data	
India-2	Hospital	Main Supply	ECD - 221MB Harmonics - 3.1GB 1.51% missing data	
India-3	Manufacturing	Lathe Machine	ECD - 225MB Harmonics - 3.3GB 0.58% missing data	
India-4	Manufacturing	Main Supply	ECD - 439MB Harmonics - 3.8GB 10.29% missing data	
India-5	Manufacturing	CNC Machine	ECD - 266MB Harmonics - 3.8GB 0.13% missing data	
India-6	Manufacturing	Main Supply	ECD - 175MB Harmonics - 2.3GB 0.41% missing data	
USA-1	Education	AI/ML Lab	ECD - 393MB Harmonics - 3.4GB 1.67% missing data	
USA-2	Education	Data center	ECD - 467MB Harmonics - 3.9GB 1.63% missing data	

Location data used for our experimentation

The data is already being used !!!

State Identification Problem

The objective of state identification problem (SIP) is to identify power usage patterns of any system, like buildings or factories, of interest.

Let $\boldsymbol{F} = \{f_1, \dots, f_m\}$ be a list of features that the sensor is able to capture

$\boldsymbol{x}$, where $\boldsymbol{x} \in \mathbf{R}^n$, be a collection of power usage observations for a location at a known interval Δ

SIP is to produce a function $\boldsymbol{f}: \mathbf{R}^n \rightarrow \{1, \dots, k\}$

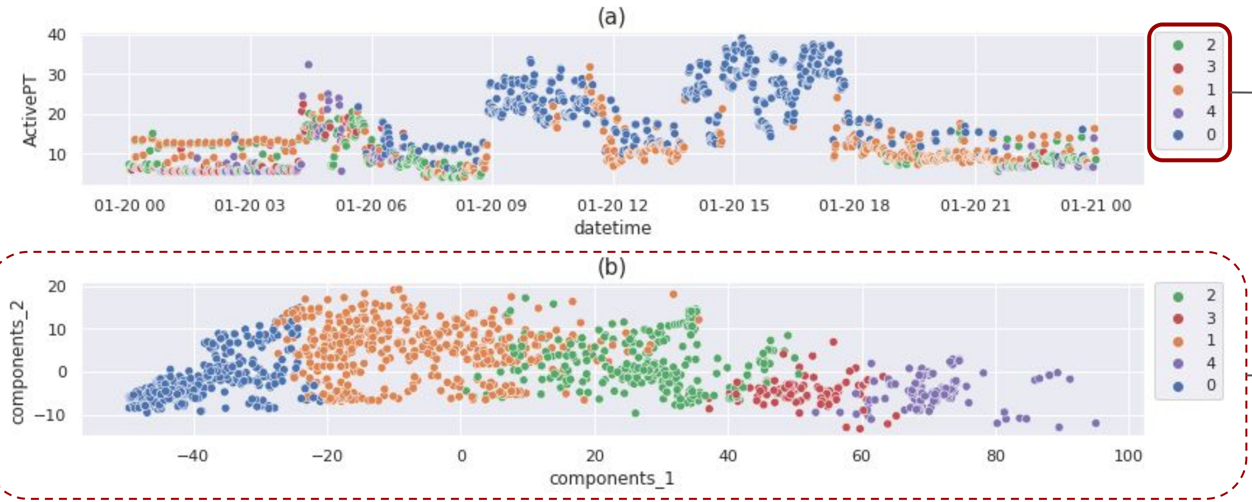
When $\boldsymbol{s} = \boldsymbol{f}(\boldsymbol{x})$, the state corresponding to $\boldsymbol{x}$ is a number capturing the category assigned to the observation

We are interested in the *unsupervised* SIP problem where k is unknown, i.e., the number of states has to be also learnt.

Results

Baseline method:

Electrical Current Harmonics Data → Find optimal number of clusters (*K-Means* and *Elbow method*) → Train the Classifier (*Random Forest Classifier*)



*Power usage observations (**x**) classified into different states (**s**) by the classifier*

Higher dimensional feature dataset is reduced to 2 dimensions using *Principal component analysis.*

Figure: Classifier output for location India-4, January 20, 2022 (Thursday) (a) Clusters (states) viewed with Active Power on Y-axis (b) PCA for 2-dimensional visualization of the clusters (states).

Walkthrough of GitHub

Location: https://github.com/ai4society/PowerIoT-State-Identification

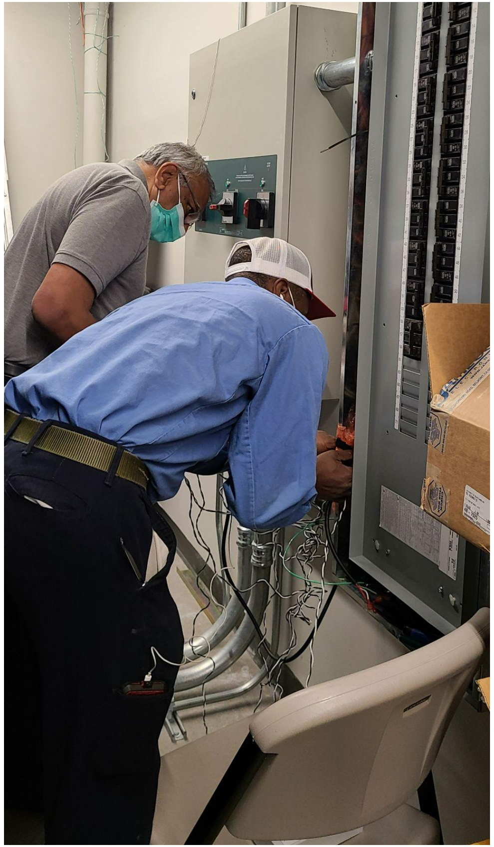
- Documentation
- Datasets
- Baseline results
- Code to replicate results
- Validation details
  - Summary of states identified, using our baseline method, along with the centroid details of each state for the released datasets are provided.
  - *location_states.json* file contains the details about the states identified for each location in the released dataset and the respective centers for each state.

Contribution

- We make a large electrical consumption and harmonics dataset available that is collected using MiDAS sensors from Tantiv4 from 8 institutions in manufacturing, education and medical institutions from the US and India for 15 days.
  - We release the datasets following the FAIR (findability, accessibility, interoperability, and reusability) principles. Follow our GitHub documentation to obtain the link to the released datasets hosted on Zenodo - https://github.com/ai4society/PowerIoT-State-Identification
  - Additional data for these locations can be obtained upon request.
- We introduce and describe a generic state identification problem that is of much interest to the industry.
- We describe an initial un-supervised machine learning based solution to extract different operating states of the power load for a location using current harmonics that can serve as a baseline to the community to accelerate research in this area.

References

1. Batra, N., Kelly, J., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., Singh, A. and Srivastava, M., 2014, June. NILMTK: An open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems* (pp. 265-276).
2. Hart, G.W., 1992. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, *80*(12), pp.1870-1891.
3. Han, Y. and Song, Y.H., 2003. Condition monitoring techniques for electrical equipment-a literature survey. *IEEE Transactions on Power delivery*, *18*(1), pp.4-13.
4. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.
5. Rajapaksha, D. and Bergmeir, C., 2022. LIMREF: Local Interpretable Model Agnostic Rule-based Explanations for Forecasting, with an Application to Electricity Smart Meter Data. *arXiv preprint arXiv:2202.07766*.
6. Muppasani, B., Anand, C.J., Appajigowda, C., Srivastava, B. and Johri, L., 2022. Power Forecasting and Anomaly Detection with MIDAS IoT-based Sensor.
7. Song, F., Guo, Z. and Mei, D., 2010, November. Feature selection using principal component analysis. In *2010 international conference on system science, engineering design and manufacturing informatization* (Vol. 1, pp. 27-30). IEEE.
8. Huang, S.J. and Shih, K.R., 2003. Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. *IEEE Transactions on power systems*, *18*(2), pp.673-679.



Thank You !