# Energy Profiling Using NILMTK

Ziyou Hu

05/18/2024

## 1 Summary

Energy datasets tend to be very large and different energy datasets vary significantly in how they measure and format. It is difficult to get an overview of what the dataset looks like and compare different datasets in a short period of time.

NILMTK provides converters for several well-known dataset into HDF format [1]. NILMTK also provides various methods that will run on the HDF format. Utilizing the converters and the methods of HDF from NILMTK, I developed a preliminary prpogram that can provide an overview of datasets.

There are 2 main files in the repository: **generate_ampds_profile.ipynb** and **generate_redd_profile.ipynb**. They generate energy profile for their corresponding dataset.

The 2 profiles are essentially identical except modification of file path names and one skip over in **generate_ampds_profile.ipynb** which I will explain later.

I converted REDD to redd.h5 by myself, and used AMPds.h5 from the offical website of AMPds [2] [3].

## 2 Effects of the Energy Profile

I also modified line 719 metergroup.py of NILMTK: deleting ”inclusive=’left’” from the code because Pandas I am running does not accept ”inclusive” as an argument for *pd.date_range()*.

I also tried to run the energy profile on GREEND, however, the kernel dies everytime shortly after building 1 is converted.

## 3 Energy Profile

1. Overview of the activation periods:
   (a) See the longest and shortest activation period: **longerst_period()**, **shortest_period()**
   (b) See several random activation periods: **random_ap()**

(c) See if there are significant differences between the average energy usage among activation periods: **plot_avg()**. For example, a space heater may have different modes, with some modes consume more energy than other ones.

(d) See how stable are the energy usage during activation period. For example, if the variances in the entire plot are small, then then appliance is stable: **plot_var()**

If some variances are high and some are low, this suggests that some activation periods are very non-stable while others are very stable. This would suggest that there might be different modes in how the appliance can be used.

(e) Through those function calls, users can see activation periods in both extremes in terms of duration and anything in between

2. Overview of the duration of activation periods

   (a) Find the distribution of duration of activation periods for a single meter: **plot_ap_durations()**

   (b) Users can see whether the durations of the activation periods are generally similar or different, what is a typical duration, and so on.

   (c) **duration_stats()**: Produce statistics about the durations of active periods. Saved in each appliance folder.

      i. avg: the average duration of an activation period.

      ii. var: the variance of activation period. i.e. do the activation periods last similar in time

   (d) Histogram for average duration of activation periods: see if there exist typical length for activation period. For exmaple, if there is a single peak in the histogram, then the activation periods tend to be of a specific length.

3. Overview of the gaps between activation period

   (a) Gap between 2 activation a and b is the time between the end of a and the start of the b. In other words, it is length of the inactivation period in between: **plot_ap_gaps()**

   (b) Users can see if the appliance is on and off periodically. For example, a person may use a toaster every morning (about once per 24 hours).

4. Overview of Energy Usage **usage_stats** plots the following information:

   (a) a_a: the average energy level when the appliance is active. It is the average of the average list produced by usage_list()

   (b) a_v: the average variance of energy level during active periods. Assess the stability of energy usage while active. i.e. does energy usage vary a lot within each active periods.

(c) s_a: the standard deviation of the average list. It assesses the variances among active periods. i.e. are different active periods similar in energy level

(d) Histogram for averge wattage during activation periods: see if there exists a typical intensity during activatio period. For example, a light may have 3 peaks if it has 3 very distinct levels of brightness.

5. Compare all appliances in a building: Within each building folder, there exists a txt file called **variance_ranking.txt**. This ranks the appliances in a building based on the variance of its energy use across different activation periods from low to high. Variance ranking is based of a_v (average variance): the average variance of energy level during active periods. Assess the stability of energy usage while active. i.e. does energy usage vary a lot within each active periods.

The use of histograms is to replace clustering to find out the typical activation periods of an appliance if there exists one. If the data are all spread out (no significant peaks) in both histograms, the appliance is unlikely to have a typical activation period.

# 4 Example of Using the Energy Profile

We can take a look at in AMPds building 1 as an example of what the energy profile can reveal.

Take the heat pump of building 1 from AMPds as an example.

In its "denoised ap_gaps" plot, there seems to be several values that the duration cluster upon. Some of the values are: 1200, 1080, 1020, 960, and so on. The indices of the activation periods seem to have little correlation with the gaps between activation periods. This suggests some high regularity in terms of when the heat pump is active. "ap_gaps_distribution" supports this observation. In addition, "ap_gaps_distribution" suggest that there exists a normal distribution that almost evenly spreads our from 1000.

From the "denoised_ap_durations", we can tell that an activation periods last from 0 to 20000, with majority clustering at 2500 to 10000. I think the units should be in seconds. The "durations_distributions" reflects the same information.

The plots of the random activation periods are puzzling. The longest period seemed to have lasted 9 days of values around 30-40 wattage with high fluctuation. The wattages are quite low for a heat pump, but the pump is turned on.

The random activation periods last about 1-3 h. The first half of the activation period, energy usage is only slightly above 0. Then, the energy level jumps to 1000 or 2000. This might be a particular behaviour of the heat pump, or a bug in determining activation periods in NILMTK.

From "usage_stats", we know that the avergae of denoised standard deviations is 759. The jump of wattage during activation periods might explain the scale of the standard deviation.

Lastly, we can take a look at the variance ranking of AMPds building 1. ['fridge', 'sockets', 'sockets_1', 'sockets_2', 'unknown', 'unknown_1', 'unknown_2', 'unknown_5', 'unknown_6', 'unknown_7', 'light', 'light_1', 'light_2', 'light_3', 'heat pump', 'television', 'na']. Since lights are considered high in variance, the appliances in this data set is more stable during activatiion period than REDD appliances of building 1.

# 5 Known bugs and Other Comments

Users and future developers of this energy profile should be aware of the following information.

1. The energy profiles I generated are: *mnt/research/sbarker/mhu/AMPds_profile* and *mnt/research/sbarker/mhu/REDD_profile*.
2. The paths in the *.ipynb* use paths specific to the directories I used. Please adapt them if you want to run on a different directory.
3. You can adapt the jupyter notebooks into a single notebook that takes in a dataset and outputs the energy profile. The current notebooks are dataset specific.
4. Bug: The y-axis of the "longest activation period" should be wattage.

# References


[1] "NILMTK." [Online]. Available: https://github.com/nilmtk/nilmtk

[2] S. Makonin, B. Ellert, I. V. Bajic, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Scientific Data*, vol. 3, no. 160037, pp. 1–12, 2016.

[3] J. Z. Kolter and M. J. Johnson, "REDD." [Online]. Available: https://tokhub.github.io/dbecd/links/redd.html