AI/ML for 5G-Energy Consumption Competition Solution

Ning Jia

https://www.linkedin.com/in/ning-jia-04b55b241/

Contents

Challenges

Unseen configurations

There are unseen configurations in the test dataset. 'B_854' and 'B_835' are exceptional stations in the train data. They have only one observation: ' B_854' is Type12, and 'B_835' is Type11.

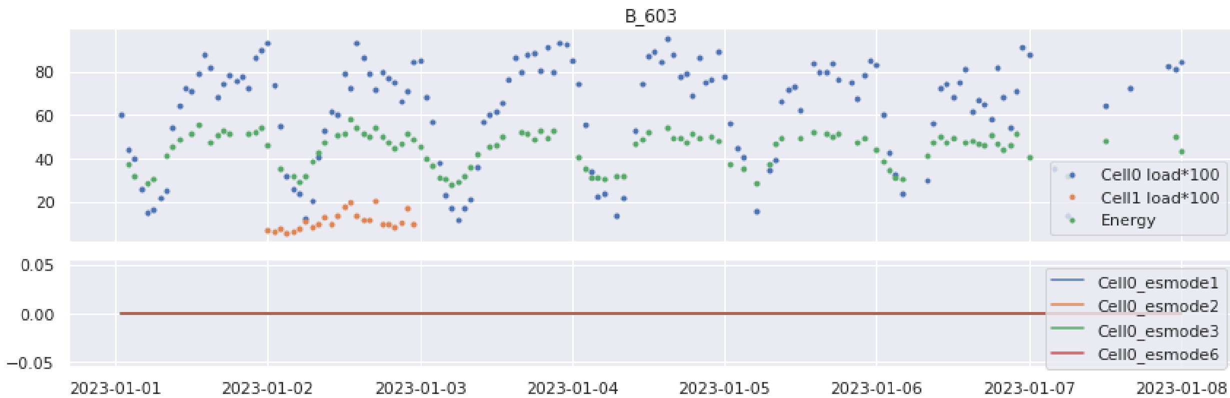
Type12 and Type11 have new settings not shown in the train. Type12 has brand new settings: frequency:979.9980, bandwidth:8, antennas:64. For Type11, only B_837 and B_838 have new settings: frequency: 715.9980, bandwidth: 8.

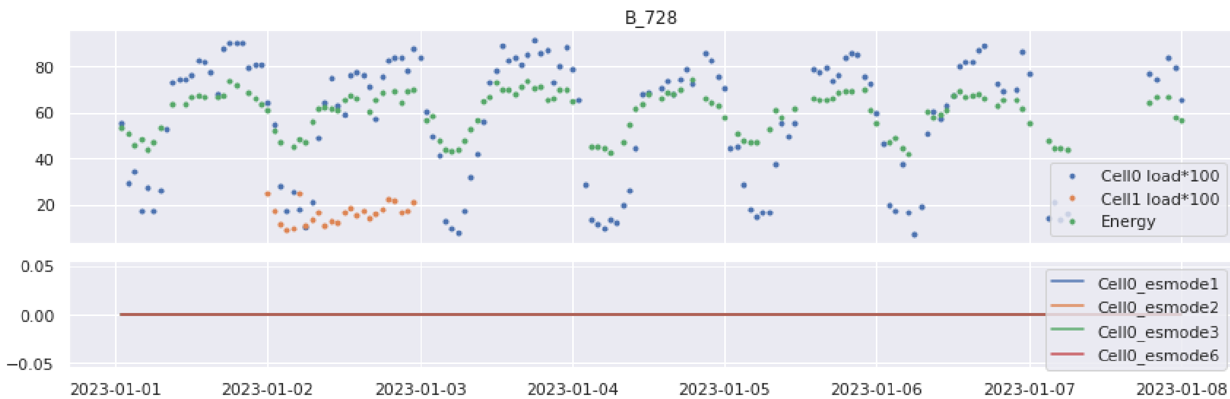
Hidden factors depend on stations.

Data exploration shows that although the stations have identical configurations and similar load patterns, the energy consumption could be in different ranges. Please see the example below. Station

'B_603' and 'B_728' have the same configurations, the same enabled saving mode and similar load patterns, but the energy consumptions are in the range of 20 to 60 and 40 to 80, respectively.

	bs	cellname	rutype	mode	frequency	bandwidth	antennas	txpower
603	B_603	Cell0	Type1	Mode2	365.0000	20	2	6.8759
1156	B_603	Cell1	Type1	Mode2	426.9800	2	2	7.3259
728	B_728	Cell0	Type1	Mode2	365.0000	20	2	6.8759
1186	B_728	Cell1	Type1	Mode2	426.9800	2	2	7.3259





This is especially true for Type1. I assume this is an old type; it may have some unique configurations or settings not available in the dataset to define the **basic consumption** for each station.

Evaluation metrics can be cheated easily by using future data.

The prediction becomes interpolating when future data is referenced, and the evaluation metric will be improved. Future energy consumption is not allowed, and other future features cannot be used. For example, the load is a strong indicator of energy consumption; if future load is leaked, that will have a similar impact when future energy consumption is known.

Solutions

Algorithm.

Tree-based algorithms generally perform well for small tabular datasets. Here, the popular LightGBM with 10 KFold folds is used.

Separate models for seen and unseen base stations

We will train separate models for seen and unseen base stations.

We improve the training dataset for unseen base stations by pseudo-labelling 3 unseen test stations which not only increase the sample size but also provide the model with basic guidelines on the unseen configurations.

Preprocess

Remove esmode4 and esmode5 because esmode4 is never enabled, and esmode5 is enabled for 3 data points.

Remove the load of cell2 and cell3 because they are only available for stations 'B_105' and 'B_745'.

Remove 'B_854' and 'B_835' when modelling for the seen stations.

Easy and fewer input features.

The models will use less than 30 features that are easy to implement and interpret. Complicated features may have the potential to include future data. For instance, aggregating the daily load for each station is actually peeking into the future.

The 27 features can be grouped into four categories:

Station load and energy-saving modes	Cell0_load,Cell0_esmode1,Cell0_esmode2,Cell0_esmode3,Cell0_esmode6 Cell1_load,Cell1_esmode1,Cell1_esmode2,Cell1_esmode3,Cell1_esmode6
Station configuration	rutype, cell_count, antenna-txpower, freq-band, mode Cell0_antennas,Cell0_frequency,Cell0_txpower,Cell0_bandwidth Cell1_antennas,Cell1_frequency,Cell1_txpower,Cell1_bandwidth
Special configuration	config_group (see below configuration group feature) bs_encode (see below station Id encoded feature)
Time	hour, day

Configuration group feature

We group the stations based on the configuration. If two stations have the same configurations, they will be clustered into the same group. In the above example, stations 'B_603' and 'B_728' belong to a group with 119 stations. Some groups only have two stations. After clustering, there are 84 unique groups. This group feature is the most crucial feature for seen stations.

Station Id encoded feature

As pointed out in the above Challenge section, stations may have hidden configurations unavailable in the dataset. The only clue is the station ID. We may encode the station with encodings, such as the average daily energy consumption, the frequency of load above certain thresholds, etc. However, they all introduce using future data. The safe encoding is to cast the station identities as categories. It turns out that these categories are the second most important feature for seen data.

Predict unseen stations in the test.

First, because the stations in the test are unseen and their configuration combinations have never been used in the train, we will train separate models without using the features of the configuration group and station ID. Therefore, the models will use 25 features.

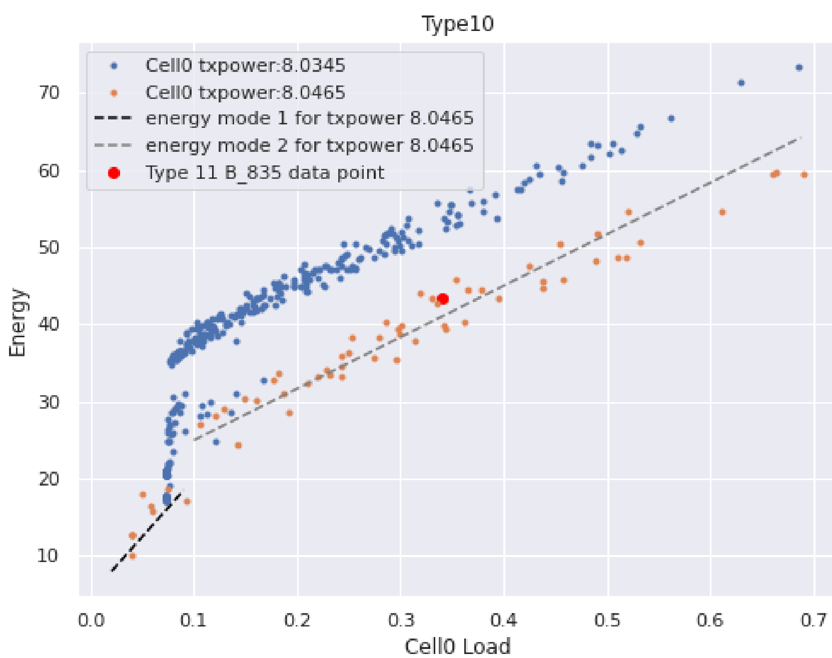
Second, we pseudo-labelled three stations with Type11 and Type12 in the test. This will incorporate the insights learned from the analytic models and increase the sample size for unseen rutype and configurations.

Pseudo-label Type11 with analytic models

Type11 is an unseen rutype. Exploration shows the majority of stations of Typ11 have similar configurations as Type10 in terms of frequency, bandwidth and txpower. Most stations of type10 and type11 have one cell. Hence, we may get some clues from Type 10.

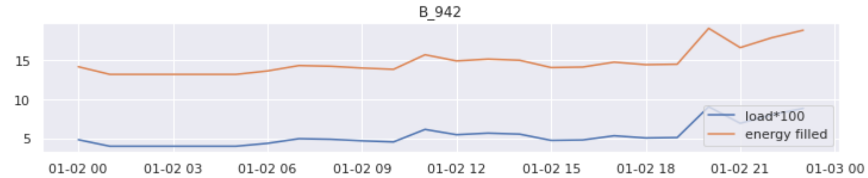
The below plot demonstrates the main insights from Type10:

1) Energy consumption is almost linear relative to the Cell0 load.
2) The Cell0 txpower determines the intercept of the linear relationship.
3) Cell0 load of 0.1 divides the linear relationship into two regimes (shown as energy mode 1 and mode 2). One is below 0.1, the station could be in a dormancy state. The other is above 0.1, the station could be active. These two regimes have different slopes.

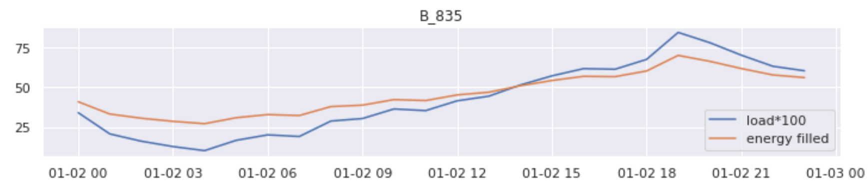


B_835 is the only type 11 station in the train dataset. Its txpower is 8.0465, and its only observation is the red data point in the plot. This single data point suggests that most type 11 may have the same energy mode patterns as Type 10 txpower 8.0465.

Therefore, I picked two stations in tests and filled the energy values with the linear analytic models. Station 942 is in the dormancy state (load less than 0.1) and the energy values are labelled with the linear regression output of mode 1.



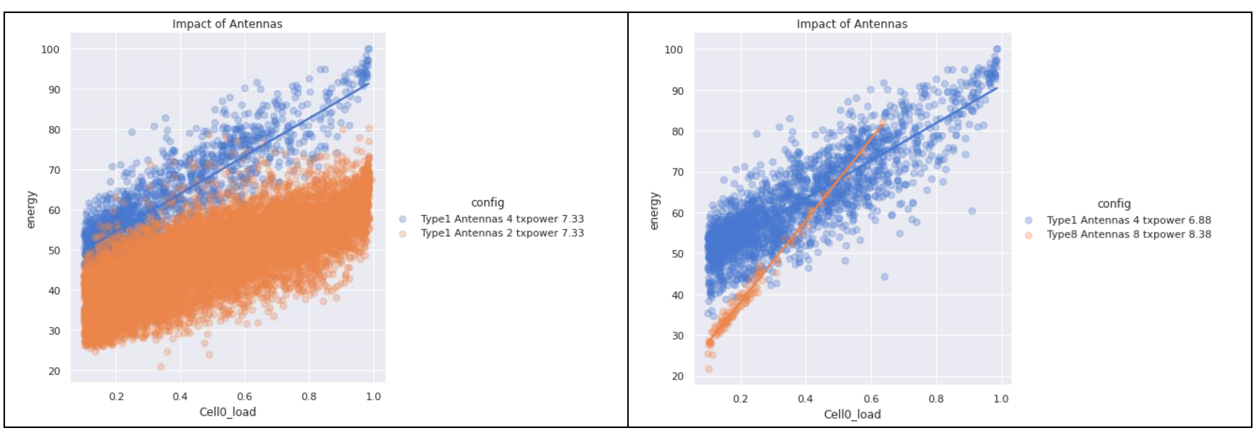
Station 835 is in the active state (load is more significant than 0.1) and the energy values are labelled with linear regression of mode 2.



Pseudo-label Type12 with analytic models

Type 12 is trickier because its configurations are all new.   First, from below SHAP analysis without using special configuration features, we can see that antennas and txpower are the most configuration settings. The higher the two values, the higher the energy consumption. This makes sense as more antennas and transition powers consume more energy.

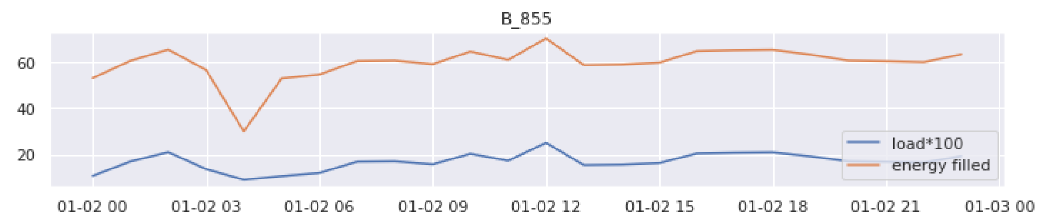
From an antennas and txpower perspective, the rutype that is most like Type12 is Type10. Most Type12 stations have 64 antennas with 8.036 txpower. Most of the Type10 stations have 32 antennas with 8.035 txpower. Now let's estimate the impact of antennas. For example, if the number of antennas of Type10 is double, what will the energy and cell0 load relation be?



These two pictures show examples of the study on the antennas. For instance, the left compares Type 1 Antennas 4 vs Type 1 Antennas 2. If we fit Huber Regression models, the slope of the 4 Antennas is 0.47, and the slope of the 2 Antennas is 0.26. Similarly, the Type 8 Antennas 8 slope is 0.99 on the right picture, and Type1 Antennas 4 is 0.46.

Studies suggest that doubling the number of antennas may double the energy consumption, given other conditions are identical.

Now, I double the slope of the Type 10 txpower 8.035 linear response and use the new line to fill one of the Type12 stations, which is B_855.



Final Model Performance:

Public Score: 0.081038064

Private Score: 0.081328210

Best Private Score: 0.069686581