Subtitle

Deep Learning-based Electric Load Prediction with Partial Information for Smart Grids

Author



# Introduction

Estimating the future electric load of server company in US by using historical data. The predicted values are used for strategies decisions making by smart grids operator in order to maximize the possible return. Then, the actual values will be updated/verified by the companies/users. The *load estimator* uses the past loads for predicting the future electric demand.

However, only a portion of the data is verified by users which means the historical data used by the *load estimator* containing two types of data: *verified* data and *predicted* data. This situation leads to the decrease in the prediction accuracy. Therefore, in this research, we aim to solve this problem by leveraging the advances of deep learning techniques.

# Problem description

## Dataset

Hourly Load Estimated dataset: This data contains estimated integrated hourly loads that are calculated from meter data. Estimated loads reflect revenue-quality meter information but have not yet been verified by the electric distribution companies and are subject to later adjustment. This information is provided for informational purposes only and should not be relied upon by any party for the actual billing values.

#### Data Fields

|  |  |  |
| --- | --- | --- |
| Field name | Data type | Description |
| Datetime Beginning UTC | DateTime | Datetime Beginning according to Coordinated Universal Time |
| Datetime Beginning EPT | DateTime | Datetime Beginning according to Eastern Prevailing Time |
| Datetime Ending UTC | DateTime | Datetime Ending according to Coordinated Universal Time |
| Datetime Ending EPT | DateTime | Datetime Ending according to Eastern Prevailing Time |
| NERC Region | String | The regions of North American Electric Reliability Corporation. Research on RFC, SERC, RTO |
| Market Region | String | The market region of the NERC regions focus on |
| Transmission Zone | String | Transmission Zone Location |
| Load Area | String | Fully Metered Electric Distribution Company |
| MW | Number | Load Power in MW |
| Verified | Boolean | Indicates whether the metered load has been verified by the Electric Distribution Company |

In this study, we consider to estimate the load for each “load area” hourly using the historical data from 2017 to 2019. Since there are some unverified values in the dataset, for evaluating the performance of the prediction model, we consider all the values are verified.

## Problem formulation

Assume that there are “load areas” in the grids. Let be the load value of area () at time-step which is consider as current time-step. The problem is formulated as follows.

##### Input:

where:

* : the predicted value of area at time-step .
* : the ground-truth (i.e. verified) value of area at time-step
* : the binary variable indicates that the value will be updated/verified by user.
* : the number of historical time-steps used for prediction

##### Output:

where:

* : the number of time-steps need to be predicted

## Proposed Approach

#### Diffusion Convolutional Recurrent Neural Network with static graph

## Evaluation

Dataset

The hourly load estimated dataset is divided into three subsets for training, validating and testing. The divided ratio as following: 60% for training set, 20% for validating set and 20% for testing set.

Experiments setting

|  |  |  |
| --- | --- | --- |
|  | Determined latter | The number of time-steps used for each prediction |
|  | **Determined latter** | The number of time-steps need to be predicted |
|  | **Determined latter** | The percentage of verified data in the period of time-steps |

After each prediction, a portion of predicted values are updated by the ground-truth randomly based on

Model training

Preparing training data for LSTM encoder-decoder

Random create binary matrix

for

for

1. Data need to be transformed as format: where is the input with shape (), is the target with shape ().
2. If

Preparing training data for DCRNN

In the training phase, training data need to be prepared as follows.

Random create binary matrix

for # is the number of time-steps in training set

1. Data need to be transformed as format: where is the input with shape (), is the target with shape ()
2. Change the value whose as , where is the stdev of the training set.

Evaluation’s Metrices

1. MAE
2. RMSE
3. MAPE

Compared approaches

1. ARIMA
2. LSTM encoder-decoder