

AI4C - Python Programs Information

Henokh Y. Fibrianto

1. Function of each program

- a. AI4C_Viz_Cdat.py
Extract and visualize any hidden information and pattern in the pre-processed training data such as demand pattern at specific location, etc..
- b. AI4C_PreProc.py
Pre-process the training data into a more suitable data format for analysis. The resulting data contains the complete demand at each unique location at each unique time.
- c. AI4C_LocProfile.py
Generate a 'location profile' data containing the statistical properties of the demand and the change of demand at each location, each unique time in a day and unique 'weekday' (monday, ..., sunday).
- d. AI4C_GenTrain12.py
Generate the final training data set complete with the features value for each data. The features are obtained from the location profile data.
- e. AI4C_NN.py
Create or load the designed artificial neural network (ANN) model, and train the ANN using the final training data set.
- f. AI4C_inputIF.py
Convert the input data with the original training data (training.csv) format into the input data that is suitable for the designed forecasting model.
- g. AI4C_NNTest.py
Test the designed forecasting model to predict the future demand.

2. Recommended usage:

Please create a folder named as 'Traffic data' within the project folder, and put the 'training.csv' within the Traffic data folder.

(execution sequence follows the written sequence)

- a. Analysis
 - i. AI4C_PreProc.py
 - ii. AI4C_Viz_Cdat.py
- b. Training
 - i. AI4C_PreProc.py
 - ii. AI4C_LocProfile.py
 - iii. AI4C_GenTrain12.py
 - iv. AI4C_NN.py

- c. Testing
 - i. AI4C_PreProc.py (can be skipped if it has been ran before)
 - ii. AI4C_LocProfile.py (can be skipped if it has been ran before)
 - iii. AI4C_GenTrain12.py (can be skipped if it has been ran before)
 - iv. AI4C_NN.py (can be skipped if it has been ran before)
 - v. AI4C_NNTest.py

3. Requirement:

- a. Python 3
- b. numpy version $\geq 1.16.3$
- c. pandas version $\geq 0.24.2$
- d. Geohash version ≥ 1.0
- e. h5py version $\geq 2.9.0$
- f. scikit-learn version $\geq 0.21.2$
- g. keras version $\geq 2.2.4$

4. Resources consumption considering all training data (except for the Testing)

Hardware used during experiment:

- a. Intel Core i3-3217U CPU @ 1.80GHz
- b. 4 GB of RAM

Hard Disk usage : 1.2 GB (location profiles related data takes only 66~67 MB, and the rest is taken by the demand data used to create the location profiles and train the NN).

Time:

- a. The Analysis takes about 20 minutes (most of the time spent on running PreProc.py).
- b. The Training (including PreProc.py) takes up to 2 and a half hours with the following durations:
 - i. AI4C_PreProc.py : around 18 minutes
 - ii. AI4C_LocProfile.py : around 25 minutes
 - iii. AI4C_GenTrain12.py : around 35 minutes
 - iv. AI4C_NN.py : around 45 minutes
- c. The Testing (AI4C_NNTest.py only) takes about 5 minutes using 1000 training data for $T + 1$ up to $T + 5$ predictions.