OutbreakPAD

Healthcare-associated infection outbreaks (HAIOs), a sudden increase in the incidence of a disease in a healthcare setting, is usually caused by bacterial or fungal pathogens and results in substantial morbidity and mortality and increase healthcare costs. Outbreak Predictor and Detector (OutbreakPAD) is a Python 3 library that aims to capture the unusual signal of incidence rise and therefore assist in the early recognition of HAIOs. The online service version of OutbreakPAD is freely available at https://github.com/pandafengye/OutbreakPAD.

Algorithm:
The daily (or weekly) case number in a hospital (or a ward, an inpatient building) constitute a time series data set, based on which prediction and detection of HAIOs are carried out.

- Prediction: a combined model of autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN).
- Detection: seven algorithms, including Mann-Kendall, Pettitt, Buishand U Test and Standard Normal Homogeneity Test (SNHT), CUSUM, EWMA and P value-CUSUM.

In theory, OutbreakPAD can also applies to other outbreaks based on time series data.

Download and install

in Linux

```
git clone https://github.com/pandafengye/OutbreakPAD.1.1.git
cd OutbreakPAD.1.1
python setup.py install
```

Usage

```
from OutbreakPAD import *
data=read_data("/PATH/example.csv")
PAD(data,p=2,d=0,q=1,a="ARIMA-GRNN",pvalue_cusum_k=1.5)
```

Input format

Example input:

```
2014-01-01,3
2014-01-02,1
2014-01-03,3
2014-01-04,1
2014-01-05,3
...
2014-08-08,6
2014-08-09,28
2014-08-10,30
2014-08-11,28
2014-08-12,30
2014-08-13,28
2014-08-14,30
```

Note:

- A two columns csv file is required (comma separated).
- Column 1: Dates in the YYYY-MM-DD format (Example: 2014-01-01).
- Column 2: Number of cases.
- Title line is not required.
- To ensure the accuracy of prediction, at least 200 days are required; the longer the better.

Output format

In the output folder there are five result files as follows:

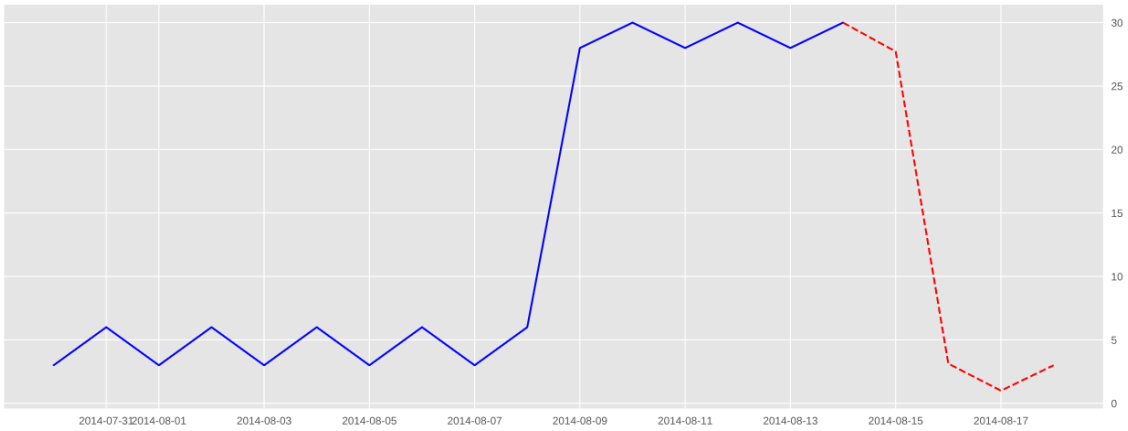
1, Predicted_case_number.txt: The predicted case numbers in the next four future days.

Example:

```
2014-08-15      27.724795837146747
2014-08-16      3.130990343022262
2014-08-17      1.0000000521540648
2014-08-18      3.0000104050647804
```

2, Lineplot_case_number_recent_20_day.svg: Line plot of the case number in recent 20 days, including the four predicted days.

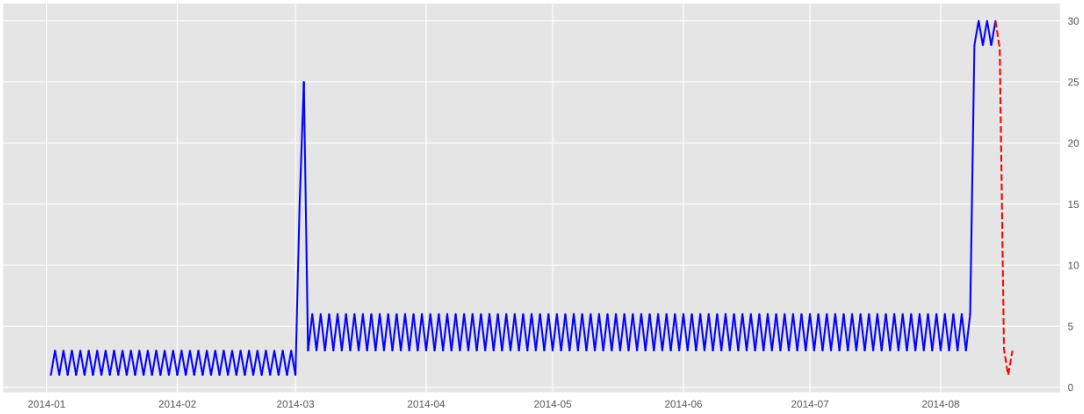
Example:



Actual data, blue; predicted data, red.

3, Lineplot_case_nubmber.svg: Line plot of the all case numbers in the input data as well as the predicted ones.

Example:



Actual data, blue; predicted data, red.

4, Detected_recent_outbreak_signal.txt: The detected outbreak signals in the last day as well as the next four future days.

Example:

```
Buishand_U_Test         ['2014/08/07']
CUSUM   ['2014/03/03-2014/03/13', '2014/08/09-2014/08/18']
EWMA    ['2014/03/03-2014/03/04', '2014/08/09-2014/08/18']
Mann-Kendall    []
P value-CUSUM   ['2014/08/09-2014/08/18']
Pettitt ['2014/03/01']
Standard Normal Homogeinity Test        ['2014/08/08']
```

Column 1, detection method; Column 2, the detected outbreak date.

5, Detected_all_outbreak_signal.txt: All detected outbreak signals in the input data as well as the predicted data.

The format is the same as Detected_recent_outbreak_signal.txt.