问题3-改进

In [1]:
```
import numpy as np
import pandas as pd
import cufflinks as cf

import scipy
import scipy.cluster.hierarchy as sch

from sklearn.metrics import *
from sklearn.ensemble import IsolationForest

import plotly
import plotly.express as px
import plotly.graph_objects as go
import plotly.figure_factory as ff

import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False

from IPython.display import HTML
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = 'all'

import pylatex
import latexify
```

孤立森林-改进

DMA1 的 2 类都要做孤立森林，DMA2 的 4 个类中的 2 类合并成一类做孤立森林

In [2]:
```
writer_q3 = pd.ExcelWriter('问题3-模型改进-孤立森林数据.xlsx')
```

DMA1

In [3]:
```
path = './模型改进数据.xlsx'
sheet = 'DMA1的用户用水量'
```

```
DMA1_data = pd.read_excel(path, sheet_name=sheet, index_col=0)
DMA1_data.index = DMA1_data.index.strftime("%m-%d")
DMA1_data
```

Out[3]:

	0:0:00	0:15:00	0:30:00	0:45:00	1:0:00	1:15:00	1:30:00	1:45:00	2:0:00	2:15:00	...	21:30:00	21:45:00	22:0:00	22:15:00	22:30:00	22:45:00	23:0:00	23
04-15	25.560	14.450	11.110	10.000	10.000	8.890	8.890	10.000	8.890	11.110	...	37.780	37.780	36.670	41.110	44.450	45.560	43.340	
04-16	26.220	17.330	16.220	15.110	12.890	11.780	9.560	9.560	5.110	5.110	...	35.110	36.220	39.560	42.890	40.670	40.670	36.220	
04-17	27.780	20.000	15.560	15.560	15.560	21.110	20.000	15.560	15.560	10.000	...	44.450	44.450	46.670	45.560	46.670	43.340	43.340	
04-18	24.777	22.557	15.887	14.777	12.557	12.557	12.557	6.997	6.997	4.777	...	43.667	44.777	45.887	44.777	42.557	43.667	48.107	
04-19	31.665	28.335	25.005	18.335	18.335	16.115	10.555	8.335	10.555	9.445	...	36.115	38.335	40.555	46.115	50.555	51.665	52.775	
04-20	28.335	19.445	15.005	16.115	15.005	12.785	10.565	11.675	11.675	8.335	...	36.115	37.225	40.565	42.785	40.565	39.445	37.225	
04-21	26.115	22.775	17.225	15.005	15.005	12.775	11.665	8.335	11.665	11.665	...	36.115	37.225	37.225	41.665	43.895	43.895	41.665	
04-22	28.890	24.450	16.670	17.780	16.670	17.780	18.890	12.230	11.110	11.110	...	37.780	37.780	36.670	40.000	40.000	42.230	43.340	
04-23	22.557	19.217	13.667	9.217	8.107	8.107	6.997	6.997	4.777	5.887	...	35.887	33.667	36.997	35.887	41.447	39.217	39.217	
04-24	21.447	15.887	14.777	12.557	12.557	9.217	5.887	8.107	8.107	9.217	...	34.777	34.777	35.887	41.447	42.557	43.667	44.777	
04-25	28.884	27.774	21.104	16.664	16.664	15.554	12.214	11.104	6.664	5.554	...	39.994	39.994	39.994	42.214	48.884	47.774	45.554	
04-26	30.555	23.895	18.335	19.445	19.445	18.335	16.115	13.895	13.895	10.555	...	37.225	36.115	42.775	42.775	43.895	41.665	40.555	
04-27	21.780	14.000	11.780	14.000	10.670	11.780	7.330	5.110	4.000	5.110	...	35.110	40.670	42.890	46.220	45.110	39.560	34.000	
04-28	23.334	15.554	14.444	14.444	13.334	11.104	9.994	7.774	7.774	5.554	...	41.104	39.994	42.214	46.664	47.774	54.444	55.554	
04-29	29.994	23.334	21.104	18.884	14.444	12.214	13.334	13.334	11.104	7.774	...	37.774	38.884	42.214	44.444	45.554	43.334	44.444	
04-30	33.340	25.560	15.560	5.560	4.450	3.340	2.220	0.000	0.000	0.000	...	31.110	30.000	30.000	31.110	31.110	30.000	31.110	
05-01	18.890	14.450	12.220	6.670	5.560	5.560	5.560	3.330	3.330	6.670	...	36.670	34.450	33.330	33.330	37.780	37.780	36.670	

	0:0:00	0:15:00	0:30:00	0:45:00	1:0:00	1:15:00	1:30:00	1:45:00	2:0:00	2:15:00	...	21:30:00	21:45:00	22:0:00	22:15:00	22:30:00	22:45:00	23:0:00	23
05-02	24.440	14.440	14.440	6.670	5.560	3.330	1.110	1.110	0.000	1.110	...	36.670	36.670	38.890	41.110	41.110	42.220	38.890	
05-03	25.560	20.000	16.670	10.000	7.780	5.560	3.330	1.110	2.220	1.110	...	41.110	40.000	38.890	42.220	41.110	42.220	36.670	
05-04	26.670	18.890	11.110	6.670	5.550	4.440	2.220	3.330	1.110	2.220	...	43.330	42.220	41.110	41.110	40.000	41.110	37.780	
05-05	22.220	17.780	8.890	6.660	3.330	3.330	1.110	3.330	3.330	2.220	...	33.330	33.330	31.110	34.440	37.780	35.550	35.550	
05-06	28.890	18.890	14.440	11.110	6.660	5.550	4.440	2.220	1.110	0.000	...	41.110	36.660	37.780	38.890	41.110	38.890	36.660	
05-07	22.220	14.450	5.560	2.220	2.220	2.220	1.110	2.220	2.220	1.110	...	33.330	36.670	35.560	35.560	38.890	37.780	36.670	
05-08	30.000	21.110	14.450	12.220	6.670	3.330	4.450	6.670	4.450	4.450	...	31.110	32.220	33.330	34.450	37.780	36.670	35.560	
05-09	26.670	12.220	8.890	6.670	6.670	4.440	1.110	1.110	0.000	2.220	...	31.110	33.330	34.440	42.220	43.330	38.890	38.890	
05-10	23.330	11.110	8.890	5.560	4.440	3.330	1.110	3.330	0.000	0.000	...	34.440	33.330	35.560	37.780	42.220	43.330	40.000	
05-11	30.000	18.890	12.220	8.890	4.440	3.330	4.440	1.110	4.440	7.780	...	37.780	38.890	38.890	37.780	37.780	35.560	32.220	
05-12	16.670	11.110	6.670	2.220	2.220	1.110	1.110	2.220	2.220	0.000	...	41.110	46.670	46.670	50.000	52.220	52.220	48.890	
05-13	32.220	18.890	18.890	12.220	5.550	4.440	3.330	2.220	0.000	2.220	...	45.550	44.440	45.550	50.000	50.000	48.890	45.550	
05-14	32.220	16.670	13.330	7.780	4.450	3.330	4.450	5.560	4.450	5.560	...	43.330	43.330	44.450	48.890	51.110	50.000	52.220	
05-15	31.110	20.000	20.000	12.220	7.780	6.670	4.450	2.220	2.220	0.000	...	42.220	44.450	47.780	48.890	51.110	48.890	48.890	
05-16	34.440	27.780	13.330	10.000	7.780	4.440	3.330	1.110	1.110	2.220	...	32.220	34.440	34.440	38.890	38.890	41.110	45.560	
05-17	31.110	27.780	21.110	13.330	7.780	6.670	7.780	4.440	4.440	4.440	...	34.440	36.670	34.440	35.550	38.890	41.110	41.110	
05-18	23.330	14.440	8.890	7.780	6.660	3.330	3.330	3.330	2.220	1.110	...	42.220	42.220	42.220	52.220	51.110	50.000	41.110	

	0:0:00	0:15:00	0:30:00	0:45:00	1:0:00	1:15:00	1:30:00	1:45:00	2:0:00	2:15:00	...	21:30:00	21:45:00	22:0:00	22:15:00	22:30:00	22:45:00	23:0:00	23
05-19	23.330	14.440	11.110	7.770	4.440	3.330	2.220	1.110	1.110	0.000	...	36.660	35.550	38.890	37.770	40.000	38.890	36.660	
05-20	23.330	17.770	11.110	5.550	3.330	5.550	1.110	2.220	2.220	0.000	...	35.550	34.440	40.000	46.660	48.880	50.000	48.880	
05-21	31.110	21.110	16.660	10.000	7.770	5.550	2.220	2.220	1.110	2.220	...	43.330	43.330	46.660	47.770	51.110	51.110	45.550	
05-22	24.440	20.000	11.110	8.890	5.550	5.550	2.220	2.220	1.110	3.330	...	35.550	37.780	40.000	43.330	44.440	44.440	38.890	
05-23	34.440	22.220	16.660	8.880	2.220	1.110	2.220	1.110	1.110	2.220	...	41.110	42.220	35.550	42.220	40.000	41.110	41.110	
05-24	25.550	14.440	11.110	7.770	4.440	4.440	2.220	2.220	1.110	2.220	...	37.770	42.220	41.110	44.440	47.770	45.550	43.330	
05-25	31.110	22.220	13.330	8.890	8.890	4.440	5.550	4.440	3.330	1.110	...	40.000	41.110	40.000	43.330	46.660	0.000	0.000	
05-26	52.220	36.670	27.780	18.890	10.000	6.670	6.670	3.330	2.220	2.220	...	44.440	45.560	48.890	52.220	52.220	52.220	47.780	
05-27	33.330	24.440	15.550	10.000	8.890	4.440	2.220	3.330	2.220	6.670	...	36.670	36.670	37.780	45.550	50.000	24.440	1.110	
05-28	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	63.330	64.440	64.440	71.110	71.110	67.780	67.780	
05-29	34.440	23.330	12.220	10.000	8.890	6.660	4.440	4.440	4.440	3.330	...	40.000	42.220	45.550	51.110	56.660	53.330	51.110	
05-30	34.440	23.330	13.330	10.000	7.780	6.670	6.670	3.330	1.110	1.110	...	42.220	44.440	46.670	50.000	53.330	56.670	52.220	
05-31	44.440	32.220	26.670	17.780	10.000	5.550	3.330	6.670	3.330	1.110	...	40.000	38.890	38.890	44.440	44.440	50.000	47.780	
06-01	41.110	33.330	25.560	14.440	8.890	7.780	5.560	7.780	4.440	3.330	...	42.220	42.220	44.440	42.220	44.440	50.000	53.330	
06-02	36.670	22.220	16.670	12.220	8.890	7.780	4.450	5.560	4.450	3.340	...	44.450	42.220	43.340	47.780	53.340	50.000	41.110	
06-03	35.550	27.780	23.330	15.550	8.890	5.550	4.440	3.330	1.110	2.220	...	36.660	37.780	38.890	42.220	44.440	44.440	42.220	
06-04	32.220	20.000	11.110	11.110	6.660	4.440	5.550	4.440	1.110	1.110	...	46.660	46.660	43.330	46.660	50.000	50.000	50.000	

	0:0:00	0:15:00	0:30:00	0:45:00	1:0:00	1:15:00	1:30:00	1:45:00	2:0:00	2:15:00	...	21:30:00	21:45:00	22:0:00	22:15:00	22:30:00	22:45:00	23:0:00	23
06-05	35.550	24.440	16.660	11.110	10.000	8.890	6.660	4.440	2.220	2.220	...	41.110	43.330	46.660	54.440	55.550	57.770	52.220	
06-06	31.110	16.670	12.220	11.110	6.670	6.670	3.330	2.220	3.330	3.330	...	34.440	36.670	36.670	41.110	46.670	48.890	45.560	
06-07	25.560	16.670	12.220	7.780	10.000	3.330	2.220	3.330	0.000	2.220	...	35.560	36.670	36.670	38.890	45.560	45.560	44.440	
06-08	26.660	15.550	17.780	17.780	14.440	11.110	8.890	5.550	2.220	1.110	...	42.220	41.110	43.330	46.660	50.000	56.660	53.330	
06-09	26.660	18.890	13.330	12.220	7.780	4.440	3.330	3.330	2.220	2.220	...	38.890	42.220	44.440	50.000	53.330	52.220	50.000	
06-10	31.110	18.890	11.110	7.780	6.670	4.440	0.000	3.330	2.220	0.000	...	38.890	41.110	43.330	46.670	51.110	53.330	41.110	
06-11	23.330	22.220	17.780	14.440	11.110	7.780	4.440	3.330	3.330	1.110	...	41.110	42.220	44.440	47.780	50.000	48.890	47.780	

58 rows × 96 columns

DMA1的类1

In [4]:
```
DMA1_class1_April = [30, 27, 24, 16, 23, 15, 28]
DMA1_class1_May = [16, 22, 5, 8, 7, 9, 11, 23, 10, 19, 2, 17, 3, 4, 6, 1]
DMA1_class1_June = [3, 7]
DMA1_class1 = [f'04-{i}' for i in DMA1_class1_April] + [f'05-{i}' for i in DMA1_class1_May] + [f'06-{i}' for i in DMA1_class1_June]

DMA1_data_class1_mask = DMA1_data.apply(lambda x: x.name in DMA1_class1, axis=1)
DMA1_data_class1 = DMA1_data[DMA1_data_class1_mask]
DMA1_data_class1.to_excel(writer_q3, 'DMA1-class1-all')

rng = np.random.RandomState(24)
n, m = DMA1_data_class1.shape

xtrain = DMA1_data_class1
clf = IsolationForest(n_estimators=n, random_state=rng)
clf.fit(xtrain)
ypred = clf.predict(xtrain)
# print(ypred)
# print(ypred < 0)
```

```
DMA1_data_class1[ypred > 0].to_excel(writer_q3, 'DMA1-class1-normal')
DMA1_data_class1[ypred < 0].to_excel(writer_q3, 'DMA1-class1-abnormal')
```

Out[4]:
```
IsolationForest(n_estimators=14,
                random_state=RandomState(MT19937) at 0x173D201D360)
```

DMA1的类2

In [5]:
```
DMA1_class2_ = DMA1_class1 + ['05-25', '05-27', '05-21', '05-28', '06-06', '05-15']

DMA1_data_class2_mask = DMA1_data.apply(lambda x: x.name not in DMA1_class2_ , axis=1)
DMA1_data_class2 = DMA1_data[DMA1_data_class2_mask]
DMA1_data_class2.to_excel(writer_q3, 'DMA1-class2-all')

rng = np.random.RandomState(24)
n, m = DMA1_data_class2.shape

xtrain = DMA1_data_class2
clf = IsolationForest(n_estimators=n, random_state=rng)
clf.fit(xtrain)
ypred = clf.predict(xtrain)
# print(ypred)
# print(ypred < 0)
DMA1_data_class2[ypred > 0].to_excel(writer_q3, 'DMA1-class2-normal')
DMA1_data_class2[ypred < 0].to_excel(writer_q3, 'DMA1-class2-abnormal')
```

Out[5]:
```
IsolationForest(n_estimators=38,
                random_state=RandomState(MT19937) at 0x173D201D570)
```

DMA2

In [6]:
```
path = './模型改进数据.xlsx'
sheet = 'DMA2的用户用水量'
DMA2_data = pd.read_excel(path, sheet_name=sheet, index_col=0)
DMA2_data.index = DMA2_data.index.strftime("%m-%d")
DMA2_data
```

Out[6]:

	0:0:00	0:15:00	0:30:00	0:45:00	1:0:00	1:15:00	1:30:00	1:45:00	2:0:00	2:15:00	...	21:30:00	21:45:00	22:0:00	22:15:00	22:30:00	22:45:00	23:0:00	23
04-15	11.775	4.675	3.835	4.755	5.755	6.835	7.775	8.615	9.305	9.065	...	6.555	5.985	8.315	8.325	7.975	8.805	9.725	
04-16	9.990	5.710	6.370	7.920	9.040	10.390	11.120	12.150	8.740	8.310	...	9.350	9.700	9.350	8.820	9.010	9.330	7.910	
04-17	12.630	11.140	7.960	10.230	11.230	12.110	12.800	11.730	8.930	9.730	...	8.590	10.590	10.110	9.220	9.450	10.250	10.500	
04-18	13.015	11.975	8.135	9.705	10.165	11.425	12.845	9.555	9.115	9.725	...	11.805	11.825	11.275	10.365	10.635	10.935	11.415	
04-19	11.205	10.145	10.875	9.545	11.185	12.225	9.405	8.445	9.315	9.945	...	11.535	11.785	10.855	9.925	9.945	10.055	10.945	
04-20	7.285	4.465	2.935	4.805	6.285	7.155	8.035	8.755	8.465	7.745	...	5.965	6.425	6.425	6.255	6.895	7.445	7.665	
04-21	7.935	5.115	3.775	5.625	6.765	7.295	8.205	7.235	9.445	9.615	...	3.695	3.685	3.345	3.455	3.455	4.335	5.175	
04-22	10.175	11.455	6.565	8.115	9.315	10.805	11.925	10.265	8.755	9.075	...	10.145	11.025	10.295	9.515	9.545	10.155	10.715	
04-23	10.490	8.230	7.820	5.620	6.850	8.000	8.340	9.280	9.760	8.850	...	9.780	10.360	10.630	10.100	10.100	11.500	12.490	
04-24	7.375	3.485	4.635	6.805	7.975	7.665	6.085	6.635	7.915	8.305	...	4.095	4.455	4.375	5.055	5.295	6.025	7.635	
04-25	10.370	11.990	11.060	8.340	10.020	10.970	12.020	11.140	8.480	9.040	...	7.080	7.280	6.120	5.630	7.900	8.920	9.730	
04-26	8.570	6.100	5.210	7.030	9.010	10.100	10.760	10.230	9.770	8.140	...	7.630	7.760	6.940	6.070	5.280	5.560	6.440	
04-27	11.280	5.680	6.170	7.550	9.320	11.010	10.970	7.700	8.710	8.960	...	11.050	10.750	10.490	10.230	11.390	10.530	11.130	
04-28	8.935	4.805	5.215	6.785	8.485	8.845	6.895	6.585	7.485	8.035	...	4.785	5.895	5.265	4.255	4.125	4.905	5.885	
04-29	9.450	7.500	8.680	9.910	8.090	8.490	9.810	10.520	10.880	8.890	...	7.390	9.000	8.460	7.280	7.400	7.930	9.120	
04-30	18.188	14.518	9.718	7.908	7.668	7.648	7.528	7.538	7.618	7.568	...	17.908	17.948	18.138	17.428	17.818	18.118	18.688	
05-01	14.680	12.790	10.850	7.960	7.610	7.160	7.330	7.440	7.440	7.500	...	15.500	15.900	15.470	14.810	14.760	15.040	15.910	

	0:0:00	0:15:00	0:30:00	0:45:00	1:0:00	1:15:00	1:30:00	1:45:00	2:0:00	2:15:00	...	21:30:00	21:45:00	22:0:00	22:15:00	22:30:00	22:45:00	23:0:00	23
05-02	17.860	11.630	11.770	7.840	7.570	7.430	7.280	7.300	7.330	7.290	...	15.170	15.030	16.600	15.330	16.130	16.670	17.260	
05-03	9.590	5.330	4.190	1.270	0.360	0.250	0.080	0.030	0.050	0.120	...	10.290	10.450	9.610	9.080	9.080	9.800	10.100	
05-04	11.620	7.110	1.940	0.460	0.450	0.130	0.000	0.000	0.010	0.000	...	10.410	10.380	10.870	11.160	11.380	12.350	12.070	
05-05	16.141	14.191	9.941	7.761	7.411	7.411	7.461	7.471	7.541	7.511	...	15.821	15.591	16.251	16.641	17.101	18.281	19.521	
05-06	12.550	4.740	1.060	0.520	0.190	0.140	0.110	0.230	0.050	0.160	...	9.570	9.930	9.500	8.980	9.170	9.960	10.090	
05-07	11.680	6.330	0.580	0.440	0.230	0.050	0.040	0.090	0.120	0.080	...	9.790	9.890	9.420	9.700	9.580	10.140	10.370	
05-08	11.150	5.750	1.330	0.840	0.470	0.110	0.060	0.190	0.190	0.210	...	10.840	10.900	9.770	9.440	10.040	11.130	11.330	
05-09	11.750	3.140	2.160	0.520	0.560	0.280	0.240	0.000	0.000	0.090	...	8.370	8.140	8.550	9.860	10.220	10.650	11.860	
05-10	12.030	3.930	1.100	0.540	0.370	0.200	0.200	0.170	0.380	0.000	...	9.870	9.430	9.180	8.580	8.130	8.030	8.330	
05-11	11.600	5.900	2.030	0.800	0.520	0.170	0.100	0.030	0.000	0.100	...	8.890	9.700	9.790	9.870	10.670	11.430	12.890	
05-12	10.640	4.460	0.920	0.340	0.060	0.110	0.000	0.010	0.020	0.040	...	7.330	10.040	9.930	9.000	9.460	10.200	11.070	
05-13	11.060	4.430	5.200	2.650	0.410	0.080	0.040	0.030	0.000	0.080	...	7.980	8.260	8.460	7.980	7.890	8.480	8.370	
05-14	10.750	3.520	3.400	1.600	0.530	0.390	0.410	0.460	0.450	0.460	...	8.430	8.450	7.900	7.030	7.140	7.810	8.910	
05-15	9.210	3.920	6.210	1.930	0.920	0.610	0.650	0.430	0.510	0.180	...	7.960	8.190	7.830	6.880	6.610	7.460	8.200	
05-16	10.920	8.750	2.030	1.300	0.850	0.470	0.230	0.230	0.350	0.330	...	8.550	8.230	6.840	5.890	5.680	5.730	7.950	
05-17	7.600	5.970	4.970	1.720	0.870	0.570	0.790	0.390	0.560	0.510	...	10.590	11.110	11.090	10.240	10.530	9.320	9.480	
05-18	11.140	4.430	1.120	0.690	0.560	0.420	0.450	0.420	0.450	0.290	...	10.610	10.730	11.270	10.470	10.890	9.410	10.530	

	0:0:00	0:15:00	0:30:00	0:45:00	1:0:00	1:15:00	1:30:00	1:45:00	2:0:00	2:15:00	...	21:30:00	21:45:00	22:0:00	22:15:00	22:30:00	22:45:00	23:0:00	23
05-19	11.030	4.950	2.310	0.570	0.200	0.000	0.040	0.050	0.150	0.180	...	8.680	8.900	8.340	7.660	8.080	8.520	9.390	
05-20	10.750	4.990	1.460	0.320	0.270	0.380	0.520	0.270	0.440	0.200	...	8.860	8.550	7.630	9.260	9.840	10.660	11.840	
05-21	11.070	3.060	1.790	0.570	0.370	0.100	0.000	0.000	0.040	0.000	...	9.680	9.890	9.600	8.870	8.810	9.580	10.700	
05-22	8.420	7.020	1.620	0.630	0.200	0.060	0.000	0.120	0.200	0.110	...	11.040	11.510	11.150	10.270	10.110	10.460	9.770	
05-23	12.060	5.920	2.890	0.600	0.240	0.340	0.460	0.460	0.480	0.000	...	9.630	10.110	9.930	9.040	9.180	9.680	10.320	
05-24	11.030	3.750	1.110	0.700	0.310	0.210	0.120	0.170	0.240	0.010	...	9.220	9.610	9.210	8.370	7.920	7.820	8.750	
05-25	10.860	6.590	1.330	1.030	0.990	0.550	0.510	0.610	0.550	0.440	...	7.470	7.220	7.050	6.610	7.040	0.000	0.000	
05-26	7.340	4.120	4.280	2.900	0.820	0.570	0.550	0.410	0.320	0.000	...	8.370	8.590	7.970	7.380	7.360	8.050	9.450	
05-27	10.930	6.170	2.890	1.040	0.650	0.380	0.300	0.340	0.340	0.490	...	4.110	3.770	3.140	3.550	3.800	0.000	0.000	
05-28	4.650	4.410	3.980	2.800	1.220	1.140	1.260	0.960	0.800	0.710	...	30.700	30.470	30.240	29.750	29.610	30.040	30.730	
05-29	10.920	5.750	0.840	0.210	0.310	0.380	0.250	0.340	0.240	0.090	...	8.510	8.400	7.540	8.290	7.590	8.490	8.700	
05-30	11.590	5.100	1.620	0.970	0.630	0.610	0.330	0.060	0.010	0.000	...	9.020	8.620	7.870	8.360	7.800	8.100	8.730	
05-31	10.120	6.060	4.750	2.420	0.930	0.490	0.180	0.310	0.260	0.110	...	9.120	8.780	8.390	7.370	6.810	7.010	7.700	
06-01	10.020	7.780	3.110	1.250	0.860	0.670	0.550	0.610	0.450	0.430	...	9.830	9.790	9.630	9.750	9.340	9.180	9.000	
06-02	10.840	4.510	1.660	0.940	0.610	0.460	0.450	0.450	0.420	0.370	...	7.230	7.320	7.460	6.880	6.700	8.120	7.710	
06-03	10.590	6.000	3.520	1.310	1.070	0.410	0.390	0.030	0.150	0.110	...	8.680	8.780	8.650	7.720	7.960	9.140	10.340	
06-04	10.200	4.700	0.750	0.380	0.670	0.070	0.120	0.130	0.100	0.220	...	6.170	6.140	5.700	4.890	4.810	5.250	6.880	

	0:0:00	0:15:00	0:30:00	0:45:00	1:0:00	1:15:00	1:30:00	1:45:00	2:0:00	2:15:00	...	21:30:00	21:45:00	22:0:00	22:15:00	22:30:00	22:45:00	23:0:00	23
06-05	11.030	5.160	2.440	0.800	0.510	0.500	0.390	0.340	0.180	0.230	...	8.570	8.690	8.120	7.690	7.880	9.000	10.020	
06-06	8.450	3.120	0.660	0.930	0.740	0.760	0.610	0.290	0.310	0.340	...	7.660	7.730	7.450	7.690	8.770	7.750	8.300	
06-07	7.840	3.300	0.980	0.000	0.530	0.370	0.100	0.120	0.030	0.080	...	7.590	7.590	7.330	7.590	7.310	8.020	8.700	
06-08	8.100	3.650	1.060	1.360	0.920	0.610	0.670	0.560	0.460	0.220	...	7.090	7.420	7.360	7.110	7.120	7.120	8.360	
06-09	8.070	2.680	1.390	0.870	0.640	0.580	0.420	0.480	0.200	0.220	...	7.460	7.580	7.500	8.390	8.510	9.300	9.890	
06-10	8.370	3.620	1.020	0.410	0.630	0.620	0.290	0.260	0.300	0.090	...	7.660	7.770	8.130	7.000	8.070	8.550	3.770	
06-11	4.520	4.740	3.210	3.330	1.270	0.550	0.330	0.150	0.040	0.140	...	7.200	8.480	8.360	8.300	8.320	7.880	8.140	

58 rows × 96 columns

DMA2的类1

In [7]:
```
DMA2_class1_April = [1, 2, 5, 28]
DMA2_class1_ = [f'05-{i}' for i in DMA2_class1_April] + ['04-30']

DMA2_data_class1_mask = DMA2_data.apply(lambda x: x.name not in DMA2_class1_, axis=1)
DMA2_data_class1 = DMA2_data[DMA1_data_class1_mask]
DMA2_data_class1.to_excel(writer_q3, 'DMA2-class1-all')

rng = np.random.RandomState(24)
n, m = DMA2_data_class1.shape

xtrain = DMA2_data_class1
clf = IsolationForest(n_estimators=n, random_state=rng)
clf.fit(xtrain)
ypred = clf.predict(xtrain)
# print(ypred)
# print(ypred < 0)
DMA2_data_class1[ypred > 0].to_excel(writer_q3, 'DMA2-class1-normal')
DMA2_data_class1[ypred < 0].to_excel(writer_q3, 'DMA2-class1-abnormal')
```

Out[7]:
```
IsolationForest(n_estimators=14,
                random_state=RandomState(MT19937) at 0x173D201D678)
```

In [8]:
```
writer_q3.save()
```

In [ ]: