

[Code] Split Topical Phases

Code Snippet

```
import numpy as np
import pandas as pd
import scipy.signal as signal
import matplotlib.pyplot as plt

df = pd.read_pickle('./vietnam_data.pkl', compression='gzip')

daily = df.groupby('date')
daily_count = daily.count()['id']

N = 1 # Filter order
Wn = 0.2 # Cutoff frequency
B, A = signal.butter(N, Wn, output='ba')

smooth_data = signal.filtfilt(B,A, daily_count.tolist())

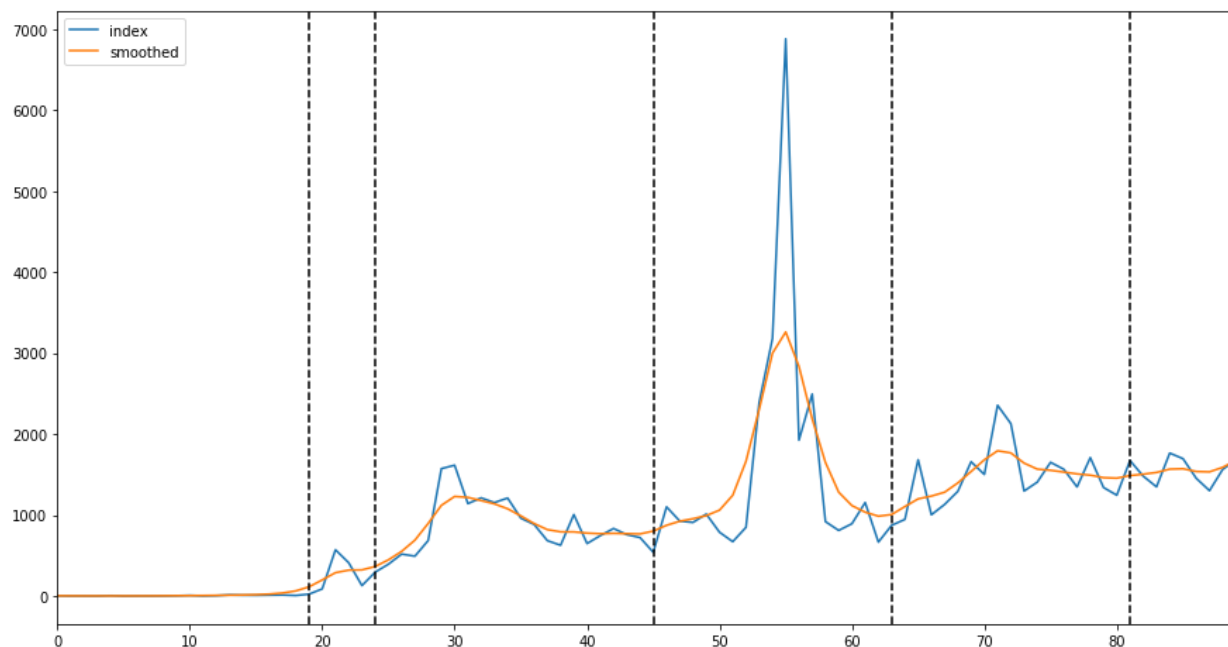
daily = daily_count.reset_index()
daily['smoothed'] = smooth_data

# Calculate diff and double_diff
diff = np.append([0], np.diff(smooth_data)).tolist()
daily['diff'] = diff
double_diff = np.append([0], np.diff(diff)).tolist()
daily['double_diff'] = double_diff

# Extract the id_list and plot
idx_list = daily[(daily['diff'] < 50) & (daily['diff'] > -10) & (daily['double_diff'] > 15)].index
daily[['id', 'smoothed']].plot(figsize = (15, 8))

for idx in idx_list:
    plt.axvline(x=idx, color='k', linestyle='--')
    print(daily['date'][idx])
```

Output Example



```
## [Iran] Extract the id_list and plot => updated by Shaun

total_df = pd.read_pickle('./data/Farsi/total_tokens_farsi.pkl', compression='gzip')
daily = total_df.groupby('date')
daily_count = daily.count()['id']

N = 1 # Filter order
Wn = 0.2 # Cutoff frequency
B, A = signal.butter(N, Wn, output='ba')

smooth_data = signal.filtfilt(B, A, daily_count.tolist())

daily = daily_count.reset_index()
daily['smoothed'] = smooth_data

# Calculate diff and double_diff
diff = np.append([0], np.diff(smooth_data)).tolist()
daily['diff'] = diff
double_diff = np.append([0], np.diff(diff)).tolist()
daily['double_diff'] = double_diff

print("Total # of the target days:", daily['date'].count(), '\n')
daily[['diff', 'double_diff']].plot(figsize = (15, 8))

## On 19 February, two people tested positive for SARS-CoV-2 in the city of Qom.

first_date = '2020-02-19'
first_date_idx = daily.loc[daily.date==first_date].index.tolist()[0]

first_smoothed = daily['smoothed'][first_date_idx]
first_diff = math.floor(daily['diff'][first_date_idx]) + 1
first_double_diff = math.floor(daily['double_diff'][first_date_idx])

idx_list = daily[(daily['diff'] > 0) & (daily['diff'] < first_diff) & (daily['double_diff'] > first_double_diff)].index
daily[['smoothed']].plot(figsize = (15, 8))
plt.title("Iran")

for idx in idx_list:
    plt.axvline(x=idx, color='k', linestyle='--')
    print("index:", idx)
    print("date:", daily['date'][idx])
```

```
print("smoothed:", round(daily['smoothed'][idx], 2))
print("diff:", round(daily['diff'][idx], 2))
print("double_diff:", round(daily['double_diff'][idx], 2))
print('\n')
```