

舆论事件热度对新冠肺炎新增确诊数量有影响吗？

苏格拉底大王

2020.02.21 11:16:53 字数 916 阅读 0

编辑文章

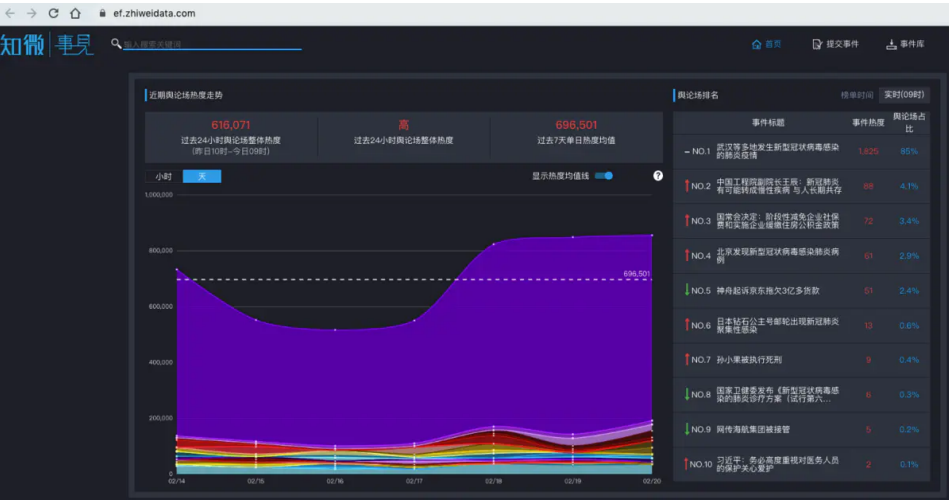
问题

RQ: 舆论事件热度对新冠肺炎新增确诊数量有影响吗？

数据

我们使用知微事见的数据来理解这一研究问题。知微事见,专注于热点事件、企业危机事件、营销事件的研究与分析。当事件符合下列标准之一时，将被收录至知微事见事件库中：1.在短时间内达到高传播量的事件；2.在长期内都保持一定传播量的事件；3.在网络社交媒体中引起热议的事件。

1.什么是事件影响力指数？事件影响力指数（Event Influence Index，EII）是基于全网的社交媒体和网络媒体数据，用来刻画单一事件在互联网上的传播效果的权威指标。指数的计算数据来自全网的社交媒体和网络媒体数据。事件影响力指数是根据事件在社交媒体（以微博、微信为主）和网络媒体上的传播效果进行加和，加和后的事件影响力再通过归一化运算得到范围在0 ~ 100之间的事件影响力指数。



知微事见

可以看到新冠肺炎依然占据舆论热点。

我们使用的数据来自针对新冠肺炎事件脉络的梳理：<http://xgml.zhiweidata.net/?from=floating#/>

苏格拉底大王

总资产 1 (约0.09元)

2019ncov细颗粒度历史数据

阅读 4

微博讨论和2019ncov疫情有多大关系？

阅读 10

2019ncov确诊人数最多的小区是哪一个？

阅读 196

推荐阅读

居民出入证下来了

阅读 1,327

每日三省吾身

阅读 9,253

儿童笑话：“妈，你杂不说你做饭难吃呢，”

阅读 1,270

“现在的女孩也太好睡了吧”

阅读 96,878

邻居

阅读 5,501



数据读取和可视化

```
1 import pylab as plt
2 import pandas as pd
3 import seaborn as sns
4 import json
5
6 j = json.load(open('zhiwei_line.json'))
7 df = pd.DataFrame(j)
8 df.tail()
```

```
j = json.load(open('zhiwei_line.json'))
df = pd.DataFrame(j)
df.tail()
```

executed in 12ms, finished 10:48:33 2020-02-21

	time	voice	heat	case	allCase
47	2020-02-15		434828.9191	2009	68500
48	2020-02-16		414931.4885	2048	70548
49	2020-02-17		438663.6598	1886	72436
50	2020-02-18		651786.8827	1749	74185
51	2020-02-19		706464.2559	394	74576

读取数据

查看一下数据基本类型：

```
1 df.info()
```

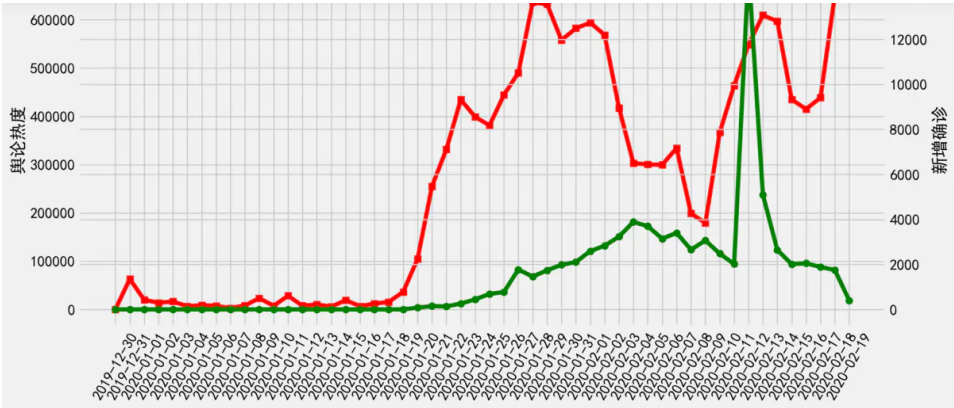
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52 entries, 0 to 51
Data columns (total 5 columns):

Column Non-Null Count Dtype

0 time 52 non-null object
1 voice 52 non-null object
2 heat 52 non-null object
3 case 52 non-null object
4 allCase 52 non-null object
dtypes: object(5)
memory usage: 2.2+ KB

结果发现都是object，未能被正确识别。

```
1 # 将object转化为数值型的类型
```



可视化

双y坐标轴可视化

```
1 # 双y坐标轴可视化
2 fig = plt.figure(figsize=(12,6),dpi = 200)
3 plt.style.use('fivethirtyeight')
4
5 ax1=fig.add_subplot(111)
6 ax1.plot(df['time'], df['heat'], 'r-s')
7 ax1.set_ylabel('舆论热度', fontsize = 16)
8 ax1.tick_params(axis='x', rotation=60)
9 ax1.legend(('舆论热度',),loc='upper left')
10
11 ax2=ax1.twinx()
12 ax2.plot(df['time'], df['case'], 'g-o')
13 ax2.set_ylabel('新增确诊', fontsize = 16)
14 ax2.legend(('新增确诊',),loc='upper right')
15
16 plt.show()
```

格兰杰检验

接下来进行格兰杰因果检验。

The Null hypothesis for [grangercausalitytests](#).

H0: the time series in the second column, x2, does NOT Granger cause the time series in the first column, x1.

Grange causality means that past values of x2 have a statistically significant effect on the current value of x1, taking past values of x1 into account as regressors. We reject the null hypothesis that x2 does not Granger cause x1 if the pvalues are below a desired size of the test.

```
1 import statsmodels.api as sm
2 from statsmodels.tsa.stattools import grangercausalitytests
3 import numpy as np
4
5 data = df[21:][['case','heat']].pct_change().dropna()
6 data.plot();
```

写下你的评论...

评论0 赞 ...

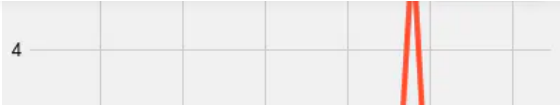


image.png

```
1 | gc_res = grangercausalitytests(data,4)
```

Granger Causality
number of lags (no zero) 1
ssr based F test: F=0.0800 , p=0.7796 , df_denom=26, df_num=1
ssr based chi2 test: chi2=0.0892 , p=0.7652 , df=1
likelihood ratio test: chi2=0.0891 , p=0.7653 , df=1
parameter F test: F=0.0800 , p=0.7796 , df_denom=26, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test: F=7.5340 , p=0.0030 , df_denom=23, df_num=2
ssr based chi2 test: chi2=18.3436 , p=0.0001 , df=2
likelihood ratio test: chi2=14.1086 , p=0.0009 , df=2
parameter F test: F=7.5340 , p=0.0030 , df_denom=23, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test: F=13.4029 , p=0.0001 , df_denom=20, df_num=3
ssr based chi2 test: chi2=54.2818 , p=0.0000 , df=3
likelihood ratio test: chi2=29.7563 , p=0.0000 , df=3
parameter F test: F=13.4029 , p=0.0001 , df_denom=20, df_num=3

Granger Causality
number of lags (no zero) 4
ssr based F test: F=8.8419 , p=0.0005 , df_denom=17, df_num=4
ssr based chi2 test: chi2=54.0915 , p=0.0000 , df=4
likelihood ratio test: chi2=29.2519 , p=0.0000 , df=4
parameter F test: F=8.8419 , p=0.0005 , df_denom=17, df_num=4

结论

舆论热度可以显著地预测新冠肺炎的确诊数量。

练习

请使用格兰杰检验看一下反过来的问题：RQ2: 新冠肺炎确诊数量是否可以预测舆论热度？

👍

0人点赞 >

👎

📖

日记本

⋮

"小礼物走一走，来简书关注我"

赞赏支持

写下你的评论...

全部评论 0

只看作者

关闭评论

按时间倒序

按时间正序

被以下专题收入，发现更多相似内容

投稿管理

+ 收入我的专题