

誰識KOL?

2020台灣總統大選在Facebook上的社群網絡分析(II)

- 游騰林
- 2020-11-28



游騰林 數據分析師



東華大學 社會學研究所



國泰世華 數據部



TLYu0419@gmail.com



<https://github.com/TLYu0419>



國泰金控

Cathay Financial Holdings

國泰世華
國泰人壽
國泰產險
國泰證券
國泰創投
國泰投信

By collaborating with Cathay's subsidiaries and consolidating the group resources, DDT provides innovative digital products and services driven by data with brilliant customer experiences, aiming to become the top in Asia Pacific Fintech ecology.

時間安排

前情提要

情感分析介紹

情感分析實務

發現

前情提要

- 為什麼需要 SNA
- SNA 簡介
- 評估重要性
- 視覺化

情感分析介紹

- 簡介情感分析
- 應用方式
- 常見的情感分析方式

情感分析實務

- 實務上的情感分析
- 小技巧分享

發現

- 發現
- 總結

前情提要

Previously on PyConTW20

為什麼需要 SNA ?

前情提要

情感分析介紹

情感分析實務

發現



IN SIDE 5G AI 新創 評論 焦點 活動 Jobs 好工作 繁 / 簡 訂閱 :

趨勢

抓到網軍！Twitter 永久刪除來自中、俄、土耳其的 3.2 萬個帳號

2020/06/12 · Heemie · Twitter、網軍、言論、仇恨言論

除了刪掉中國核心網軍的大量帳號，Twitter 也發現有額

軍帳號所發表內容的正當性。

網軍帶風向價碼 網站全

04:10 2020/06/26 | 中國時報 | 吳家豪

▲ Photo Credit: Shutterstock/達志



蝦皮優選

蝦皮超便宜

facebook

粉絲專頁100(讚+追蹤) 40元

貼文100讚 40元

台灣真人50讚 50元

台灣真人留言讚 2.8元

直播 LIVE 80元

全館最便宜

全店免運

facebook 按讚 / 直播 / 真人 / 留言讚 / 粉絲專頁讚 / 直...

\$10 - \$50

Free

月銷量 1萬

台南市佳里區

為什麼需要 SNA ?

前情提要

情感分析介紹

情感分析實務

發現



資料萃取邏輯

前情提要

情感分析介紹

情感分析實務

發現

蔡英文 Tsai Ing-wen 6月30日

中國曾經承諾香港五十年不變，今天港版國安法的強行通過，不僅代表中國罔顧香港民意，背棄承諾，更證明了「一國兩制」不可行。

我們希望，香港人民可以持續堅持他們所珍惜的自由和人權。台灣也已經啟動「香港人道援助關懷行動專案」，對於需要協助的香港人民，將提供便捷的服務及基本照顧。

我們援港專案的「台港服務交流辦公室」，將在明天正式營運。因應變局，台灣將不分政府民間，共同協力，給予香港人民最堅定的支持。

#自由的台灣撐香港的自由

1 陳井泰和其他11萬人 7,424則留言 5,061次分享

讚 留言 分享

Johnson Ho 我相信這一刻大部分的香港人都跟我一樣充滿著焦慮、不安、擔憂和憤怒的感覺。從去年反送中開始到現在強行推出國安法，就連現任的特首也不知道有什麼條例下也麻木地叫所有人支持，簡直是荒謬到一個點了吧！因此我在此希望蔡總統和廣大的台灣朋友可以在未來的日子優先協助一些因政治而受到逼害的人，而我也希望可以透過更多的工作經驗再次融入到台灣生活中（好懷念大學時期台灣生活的日子喔！）

明天7.1香港人難覓見！那麼快7.047已經到了 😢😢😢

讚 · 回覆 · 2週 · 已編輯

Yu-Ching Hsieh Johnson Ho
我只想問一句
應該保護中國香港孩子的中國香港大人們，在幹嘛？
台灣幹嘛要幫你們帶小孩啊？
你們是有付錢？還是怎樣？

讚 · 回覆 · 2週

Enjolras La Victoire Yu-Ching Hsieh 如果對台灣方面有財政負擔，請你向總統提案，我們留在香港的大人負責處理

讚 · 回覆 · 23

Facebook 可以萃取出 **互動** 與 **留言** 的資訊

Python 套件

前情提要

情感分析介紹

情感分析實務

發現



Ref:

<https://tlyu0419.github.io/2020/03/17/Crawl-Facebook-Pages>

誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析(II)

Mail: TLYu0419@gmail.com

Python 套件

前情提要

情感分析介紹

情感分析實務

發現

Facebook粉絲專頁爬蟲筆記

發表於 2020-03-17 | 更新於: 2020-04-08 | 閱讀次數: 2760

先前曾發表過 [網路爬蟲_Facebook粉絲團貼文與留言](#) 的文章，可以透過 Selenium 來抓取 Facebook 粉絲專頁的貼文與留言，但由於 Selenium 的效率實在太慢了，因此這篇文章是記錄我嘗試用 request 抓取資料的筆記，並且於文末附上台灣的2020總統大選 3位候選人從 2019 年 1 月至 2020 年 2 月間的貼文與留言資料~

簡報與完整的 Python 程式碼如下
<https://github.com/TLYu0419/FindAndMeetKOLs>

Ref:

<https://tlyu0419.github.io/2020/03/17/Crawl-Facebook-Pages>

誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析(II)

Mail: TLYu0419@gmail.com

資料結構

前情提要

情感分析介紹

情感分析實務

發現

SOURCE_NAME	SOURCE	TARGET_NAME	TARGET	TIME	TEXT
Linda****	1795****	林****	1000****	2020-01-01 19:53	選賢與能唯一支持宋楚瑜
陳****	1000****	蔡英文 Tsai Ing-wen	4625****	2019-11-19 15:24	蔡英文總統 加油 加油 連任 一定要當選臺...
海老****	1000****	蔡英文 Tsai Ing-wen	4625****	2019-11-13 20:37	加油
Hsu****	1000****	蔡英文 Tsai Ing-wen	4625****	2019-11-19 18:59	蔡總統加油
Ingwen****	1000****	蔡英文 Tsai Ing-wen	4625****	2019-11-30 05:31	台灣人為何一直擔心統一大陸十四億人口 如...
Faithé****	1000****	Faithé****	1000****	2020-01-01 22:00	洪福 韓草包沒能力
...

1620690 rows × 6 columns

註：時間範圍為大選前2個月(2019/11/1 ~ 2020/1/10)

誰識KOL? 2020台灣總統大選在Facebook上的社群網絡分析(II)

Mail: TLYu0419@gmail.com

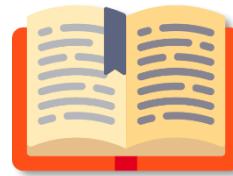
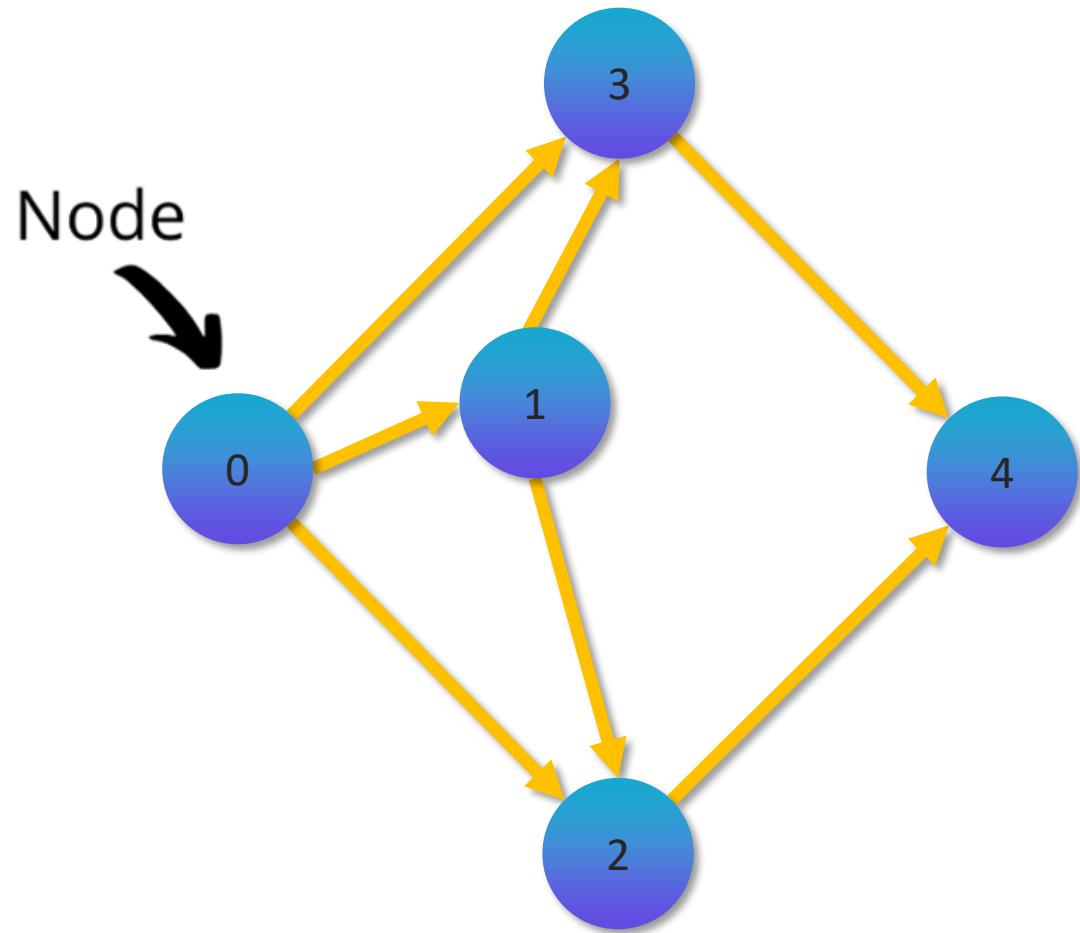
社群網絡分析介紹

前情提要

情感分析介紹

情感分析實務

發現



網絡中的節點(臉友)



大小表示數量
深淺表示重要性



重要性評估方式
找出意見領袖

如何評估節點重要性？

- Degree
 - Indegree
 - Outdegree
- Closeness
- Betweenness
- Pagerank



PageRank 介紹

前情提要

情感分析介紹

情感分析實務

發現



- 網頁的重要性排序
- 想法
 - In-Link越多越重要
 - 來自重要節點的引用較重要

如果是不重要的節點(假帳號)
藉此創造出的互動也不會有價值

Ref : <http://jpndbs.lib.ntu.edu.tw/DB/PageRank.pdf>

誰識KOL? 2020台灣總統大選在Facebook上的社群網絡分析(II)

Mail: TLYu0419@gmail.com

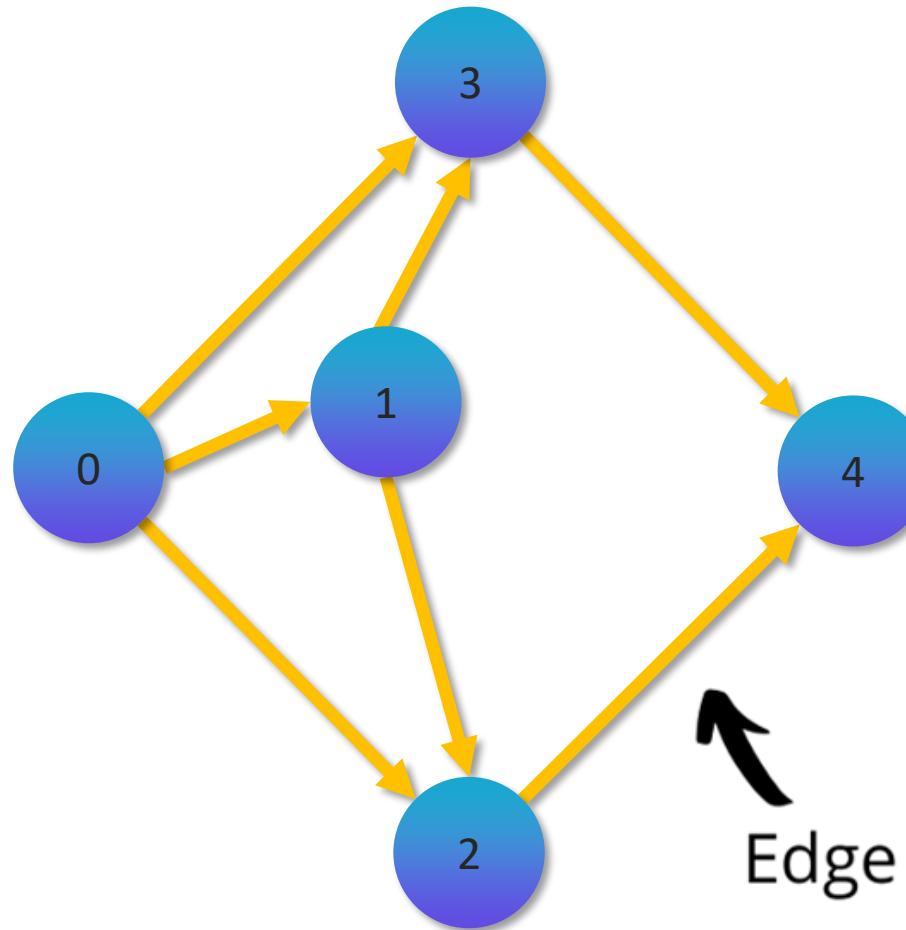
社群網絡分析介紹

前情提要

情感分析介紹

情感分析實務

發現



互動的方向與頻率



箭頭表示方向
粗細表示數量



觀察互動方式
剖析網絡社群

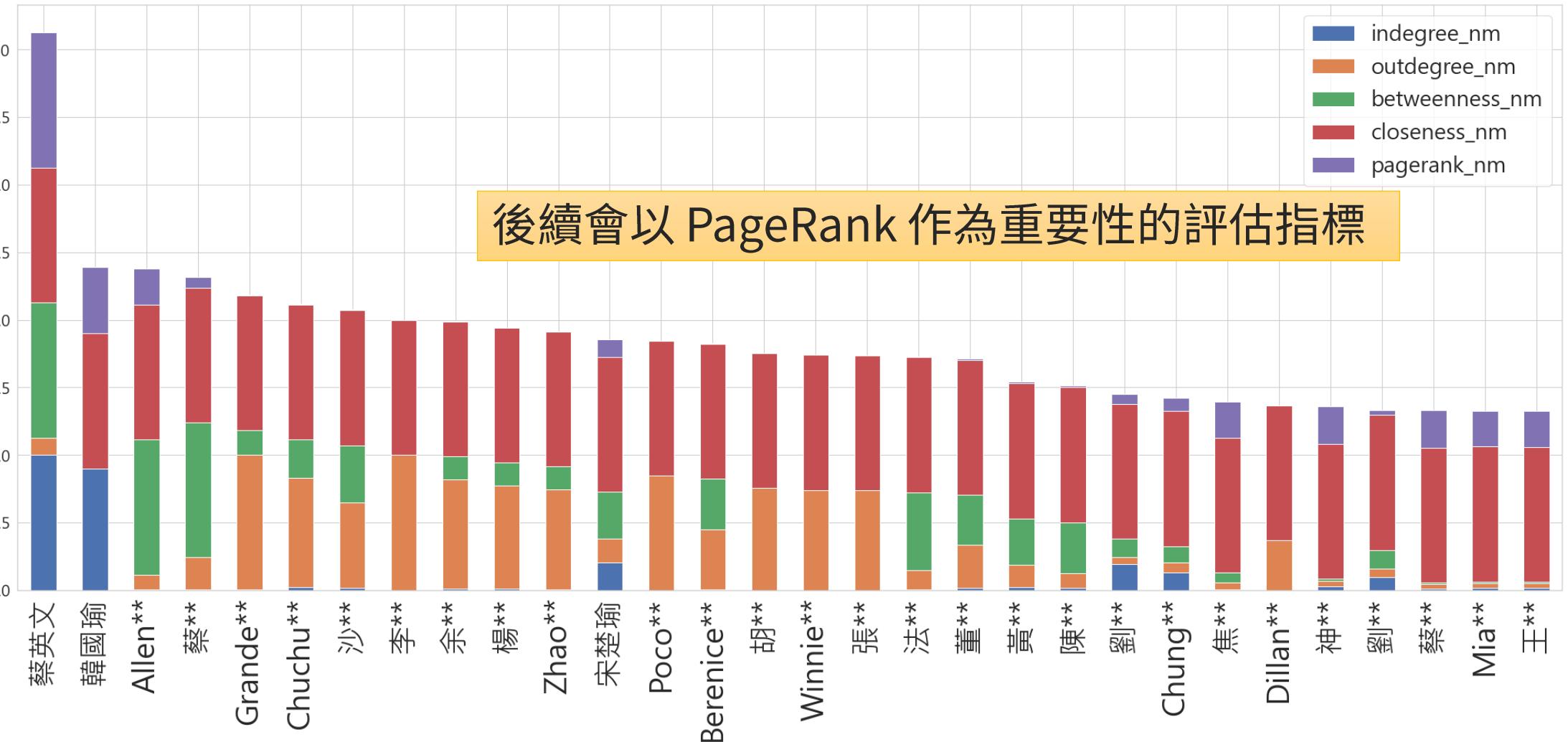
誰是網絡中最重要的人？

前情提要

情感分析介紹

情感分析實務

發現





情感分析简介

Sentiment Analysis

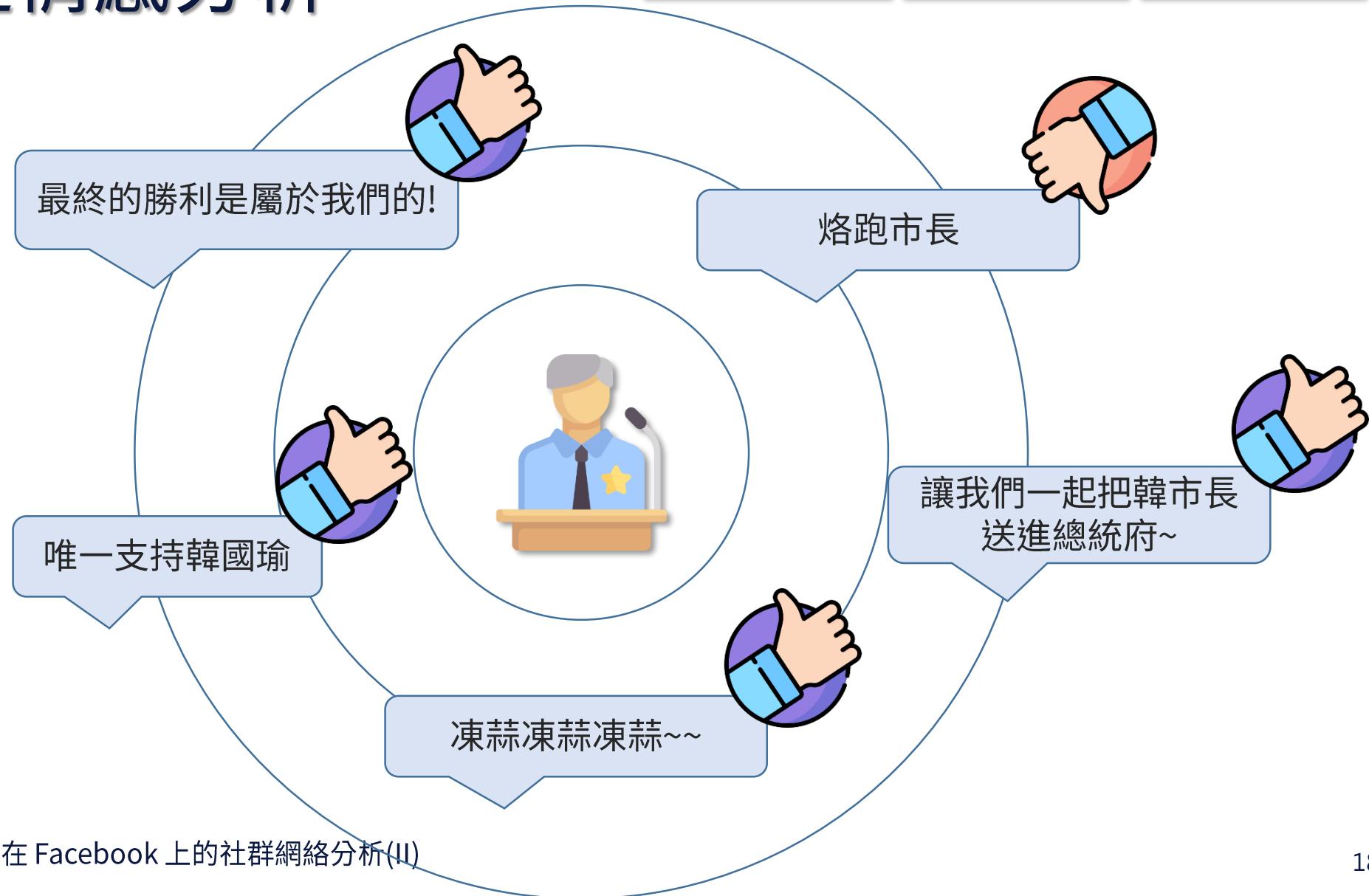
什麼是情感分析

前情提要

情感分析介紹

情感分析實務

發現



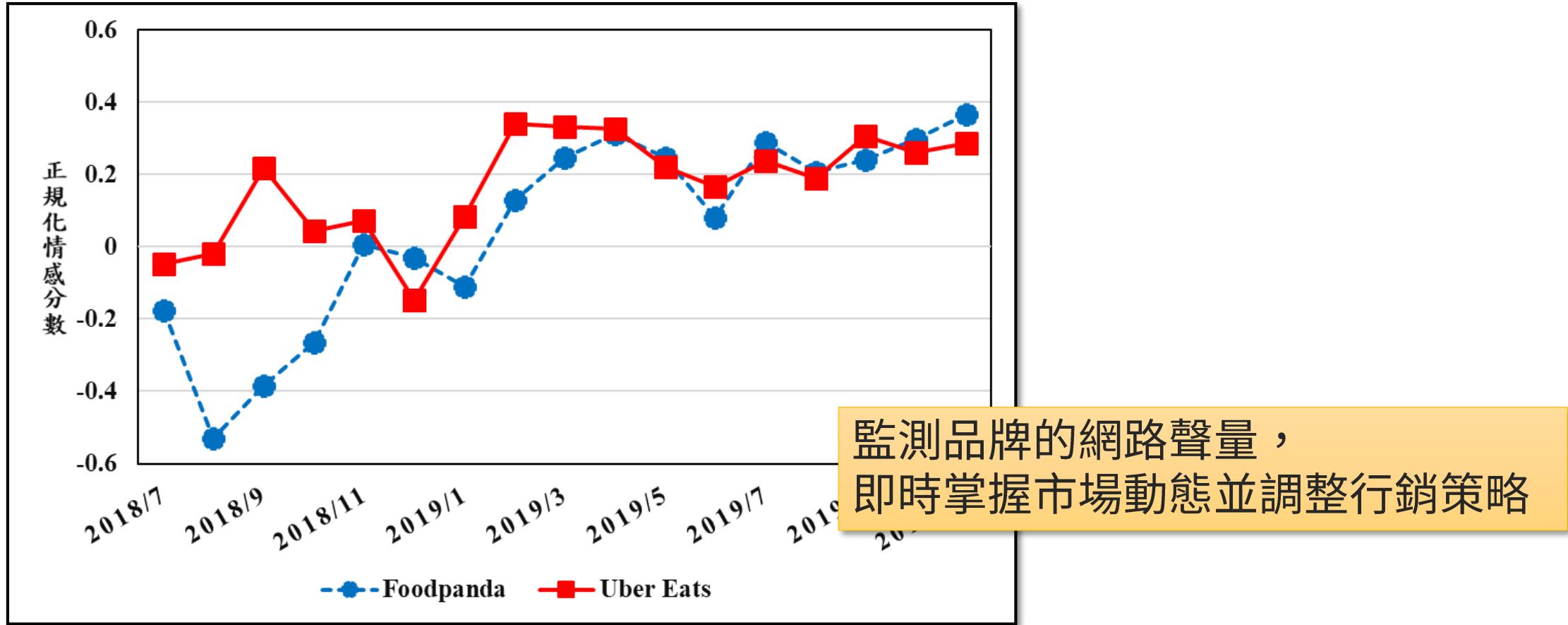
情感分析應用案例

前情提要

情感分析介紹

情感分析實務

發現



情感分析應用案例

前情提要

情感分析介紹

情感分析實務

發現

The screenshot shows a news article from the anue 鉅亨 website. The headline reads: 「潤寅傳詐貸」13家踩雷銀行現形 台企銀成最大受災戶. The article was written by 陳蕙綾 and published on 2019/07/02 at 18:23. The text discusses a major loan scam involving Ruyin Industrial, which has affected 13 banks, with Tai企銀 (Tai企 Bank) being the largest victim. The article also mentions other banks like Shang Dao, Taiwan企銀, Chia Feng, Hwa Kun, Yiheng, Hua Nan, Yuan Da, Guotai Shihua, An Tai, Yu Shan, Sin Ming, Yang Xing, and Kaohsiung. On the right side of the page, there is a sidebar with a 'Main Editor's Selection' section featuring an article about LIBOR withdrawal and market risks.

anue 鉅亨 新聞 搜尋新聞、代碼或名稱

〈潤寅傳詐貸〉13家踩雷銀行現形 台企銀成最大受災戶

鉅亨網記者陳蕙綾台北 2019/07/02 18:23

銀行界爆發大規模詐貸案！老牌貿易商潤寅實業以虛假應收帳款詐貸，傳出有 13 家銀行踩雷，傳為苦主的 13 家銀行現形，包括：王道、台企銀、兆豐、合庫、一銀、華南銀、元大銀、國泰世華銀、安泰銀、玉山銀、星展銀、陽信銀行、高雄銀行等。

潤寅實業成立於 1982 年，資本額 3.29 億元，代表人為楊文虎，主要從事工業化纖原料與塑橡膠產品進出口，是老牌進出口貿易公司，往來企業客戶幾乎都是上市櫃公司，包括台化、福懋、中纖與南亞等。

① 主編精選

（觀察）LIBOR退場牽太廣 盤點各項風險成金

新聞能夠更即時的偵測企業的犯罪風險

向出口市場恐雪上加霜
20:20

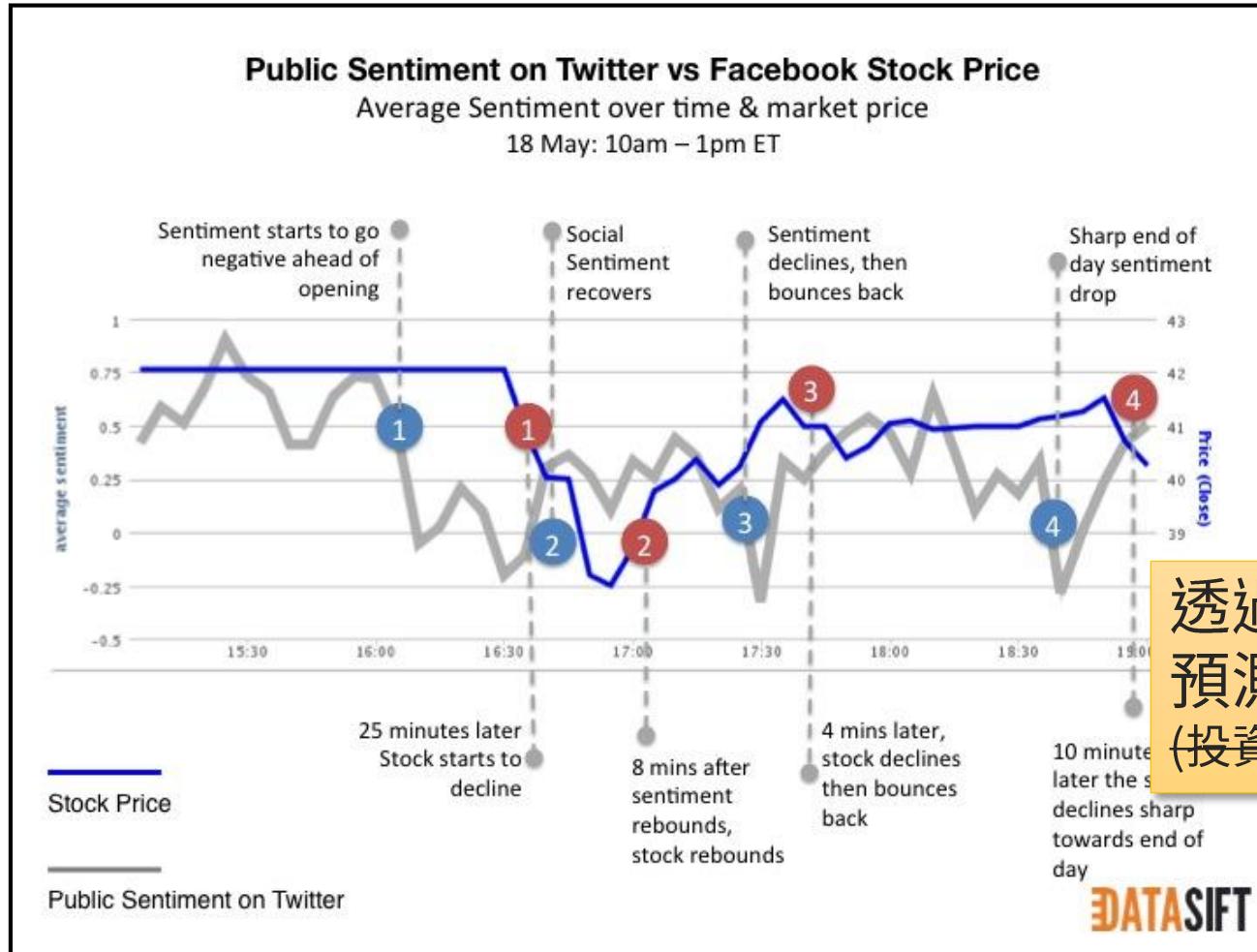
情感分析應用案例

前情提要

情感分析介紹

情感分析實務

發現



透過社群網絡上的正負向情感資訊
預測短時間後的股價走勢
(投資理財有賺有賠，使用前請詳閱說明書)

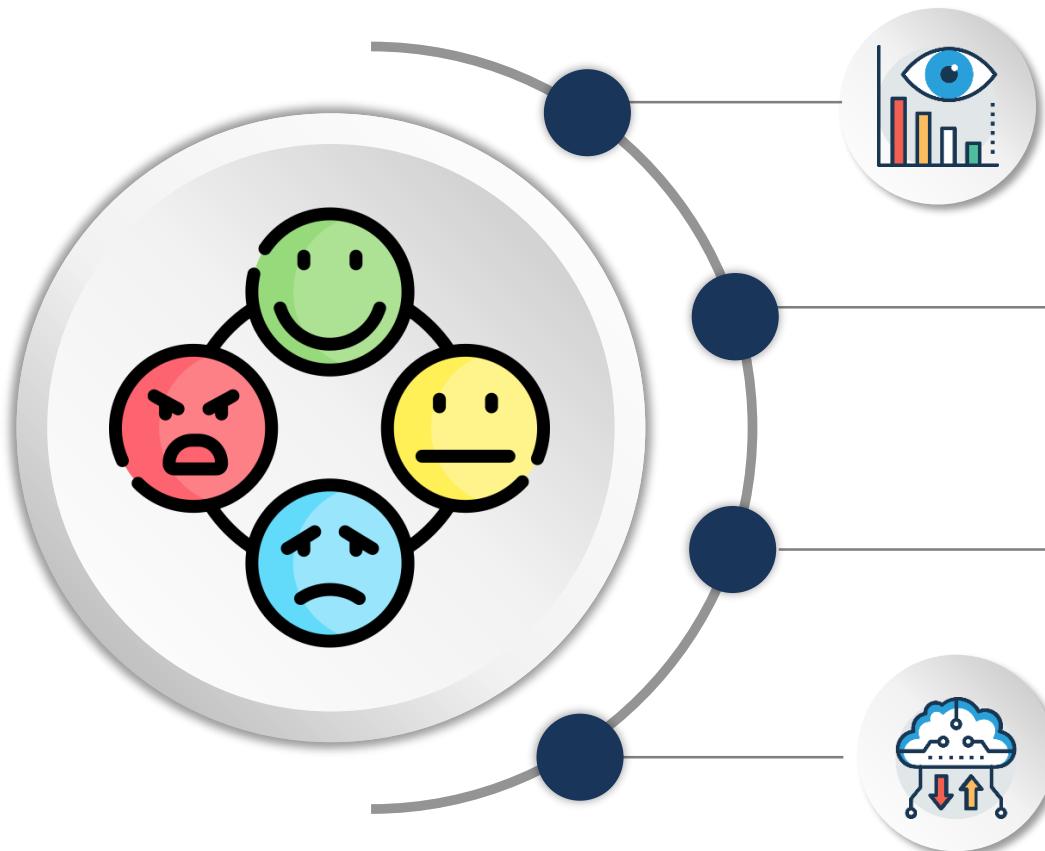
常見的情感分析方法

前情提要

情感分析介紹

情感分析實務

發現



統計值法

藉由資料源的評分來統計情感分數

詞典法

建立情感詞典，統計文本中正負向詞的出現次數

情感分析套件

藉由其他資料源來建立情感評分模型，並套用在新資料源上

訓練模型

將資料標註評分後再建置ML/DL模型

統計值法

前情提要

情感分析介紹

情感分析實務

發現

The screenshot shows the Google Play page for the Uber Eats app. The app icon is a black square with 'Uber' in white and 'Eats' in green. The title is 'Uber Eats: 美食外送'. Below it, it says 'Uber Technologies, Inc. 美食佳飲' and has a '3+' rating. It includes a note about ads and compatibility. The main area shows a rating of 4.4 stars from 3,120,640 reviews, with a bar chart showing the distribution of ratings from 1 to 5. Two reviews are visible: one from 'Scorpio Vickie' dated 2020/11/23, and another from 'Aries Wei'.

美味佳餚，輕鬆送達。

肚子餓了嗎？從您最愛的餐廳訂購想吃的美食，由 Uber 為您快速外送。

隨時隨地，盡情享用自己喜愛的美食。只要輕輕點選按鈕，就能享用您想吃的當地佳餚。

閱讀完整內容

★★★★★ 3,120,640 ⚡ 編輯精選

4.4

★★★★★

共 3,120,640 則 評分

Scorpio Vickie ★★★★★ 2020年11月23日

Aries Wei

有些資料源會同時蒐集滿意度分數與意見回饋
這時候可以直接使用滿意度分數監測網路輿情

Ref:[如何爬取 GooglePlay 的APP留言與評分](#)

誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析(II)

Mail: TLYu0419@gmail.com

辭典法

前情提要

情感分析介紹

情感分析實務

發現



評分詞典



改良
出色
努力

卓越
順利
…



不足
不法
中止

不明
代價
…

銀行界爆發大規模**詐貸**案！老牌貿易商潤寅實業以**虛假**應收帳款**詐貸**…**損失**最大的苦主為台企銀，曝險金額高達 13.38 億元，因放款多屬無擔保品，**損失**比重高達 9 成，為此案的**重災戶**。

建立標註正/負向詞彙的詞典，
並藉此統計文章中出現正/負向詞彙的數量

Ref: [NTU Sentiment Dictionary User Form](#)

誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析(II)

Mail: TLYu0419@gmail.com

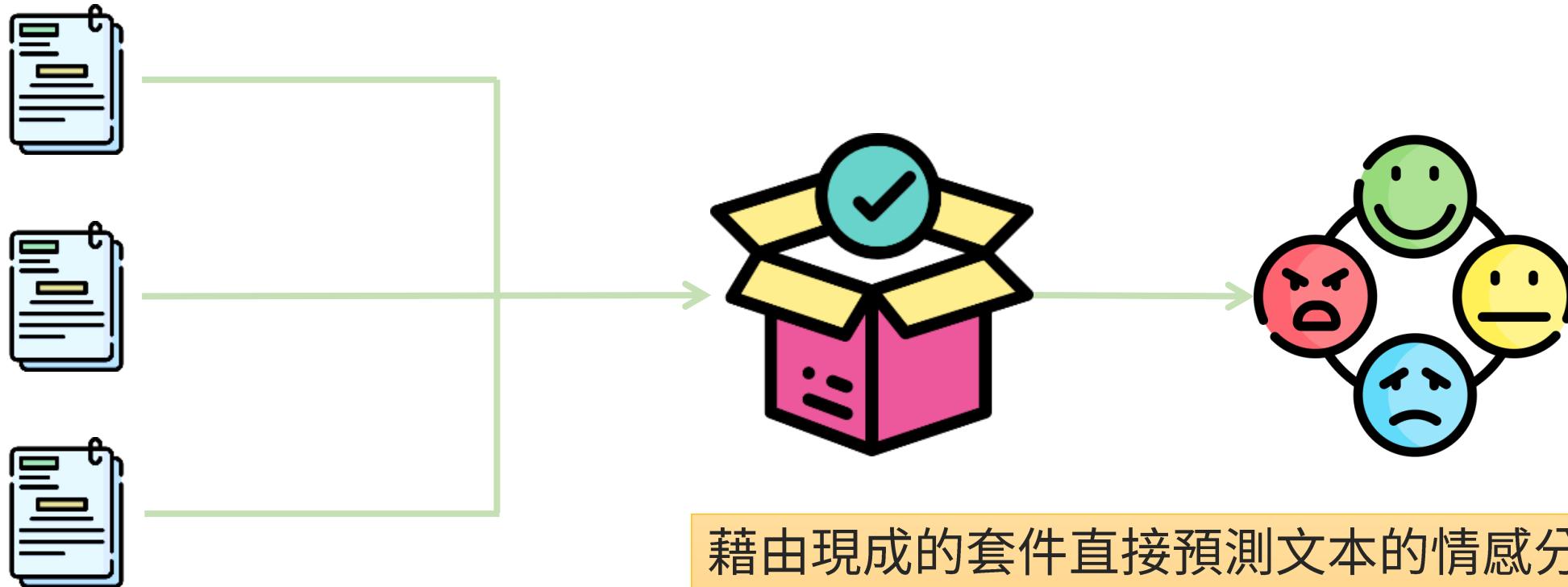
情感評分套件法

前情提要

情感分析介紹

情感分析實務

發現



藉由現成的套件直接預測文本的情感分數
中文：snowNLP(2015)
英文：NLTK(2020)

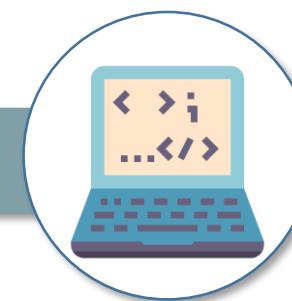
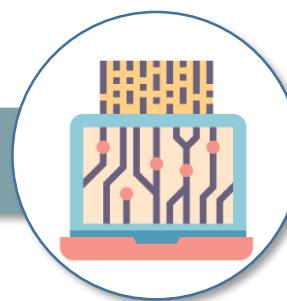
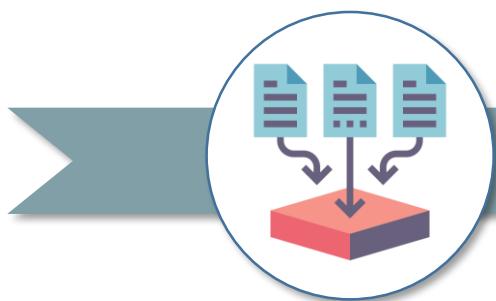
訓練模型法

前情提要

情感分析介紹

情感分析實務

發現



收集語料

- 社群網站
- 新聞文本
- 商品評論
- ...

標註資料

- 自行標註
- 外包

特徵工程

- 詞頻矩陣
- TF-IDF
- 詞向量
- ...

訓練模型

- 機器學習
- 深度學習

訓練模型的精度最高，
將資料標註後即可訓練機器/深度學習模型



情感分析實務

Sentiment Analysis in Practice

情感分析的現實面

分享一些五花八門的小技巧XD

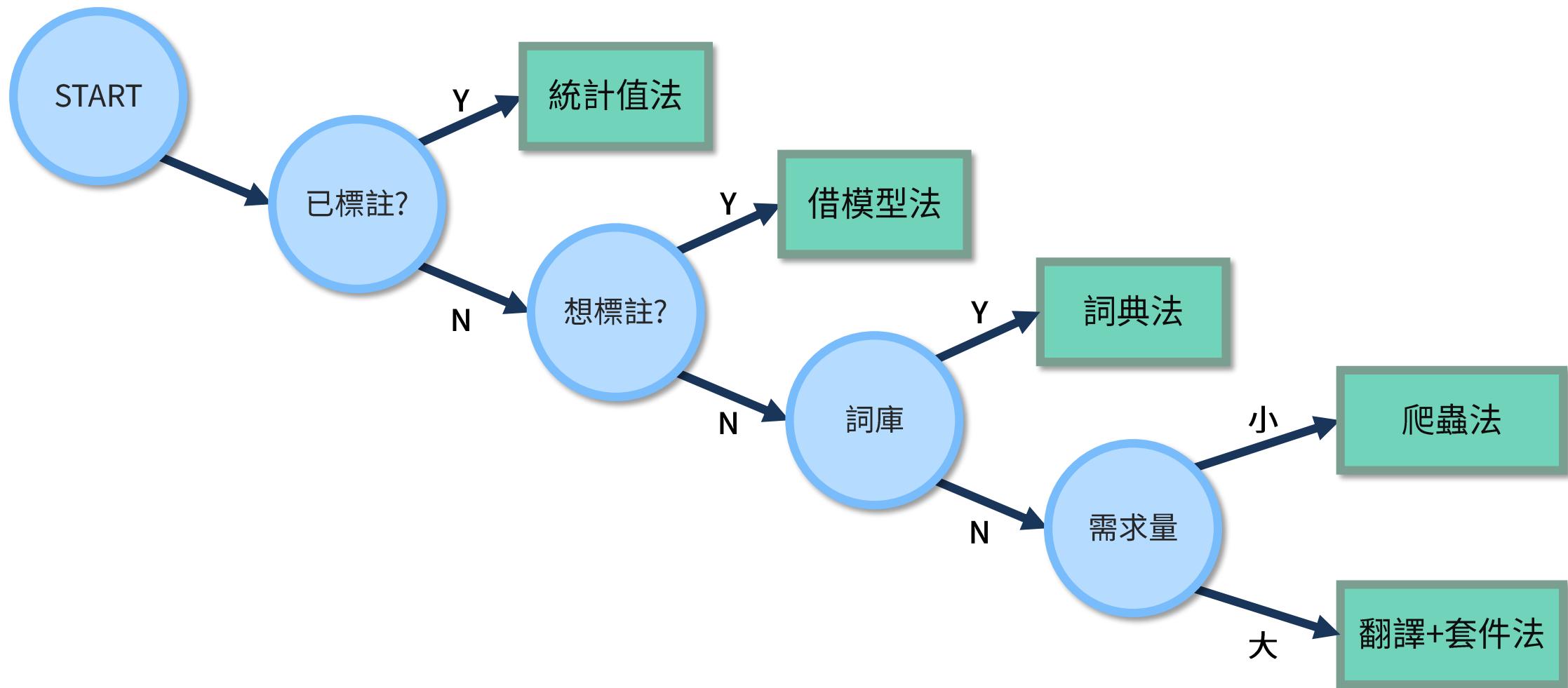
情感分析策略

前情提要

情感分析介紹

情感分析實務

發現



借模型法

前情提要

情感分析介紹

情感分析實務

發現

The screenshot shows the Baidu AI Open Platform interface. At the top, there's a navigation bar with links: 产品服务, 解决方案, 合作伙伴, 开发资源, AI加速器, AI市场, 资讯, 社区, and 控制台 (Control Panel). The Control Panel link is highlighted in blue. Below the navigation bar, there's a large banner for '情感倾向分析' (Sentiment Analysis) with a background image of a person's face. The banner text includes: '自动对包含主观信息的文本进行情感倾向性判断, 为口碑分析、话题监控、舆情应用提供基础技术支持, 同时支持用户自行定制模型效果调优'. Below the banner are two buttons: '立即使用' (Use Now) and '技术文档' (Technical Documentation). A yellow callout box is overlaid on the bottom left of the banner, containing the text: '借用 百度/ Google 的情感傾向分析來幫我們的資料標註情感分數 (註冊帳號後就能免費體驗50萬次)'. To the right of the banner is a table showing pricing for the sentiment analysis service:

规格	有效期	单价	数量
10万次	12个月	¥ 230.00	- 0 +
30万次	12个月	¥ 630.00	- 0 +
	12个月	¥ 950.00	- 0 +
	12个月	¥ 1,700.00	- 0 +
	12个月	¥ 4,500.00	- 0 +

Ref: [百度 / AI 開放平台 / 情感傾向分析](#)

誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析(II)

Mail: TLYu0419@gmail.com

借模型法

前情提要

情感分析介紹

情感分析實務

發現

```
# !pip install baidu-aip
from aip import AipNlp

# 輸入 APP 的資料
APP_ID = '17584307'
API_KEY = 'tvqsXpPaX7lDq1xxYZAinwQm'
SECRET_KEY = 'X4XksoWpBXXrKMG5Sk3bPmNvPplgckHy'
client = AipNlp(APP_ID, API_KEY, SECRET_KEY)

# 想做情感分析的文本
text = "希望从情感的角度來看服务平台谁可以成为未來的标准"
client.sentimentClassify(text)
client.emotion(text)
```



在網站上申請後
取得的資料

借模型法

前情提要

情感分析介紹

情感分析實務

發現

```
# !pip install baidu-aip  
from aip import AipNlp
```

輸入 APP 的資料

```
APP_ID = '17584307'  
API_KEY = 'tvqsXpPaX7lDq1xxYZAinwQm'  
SECRET_KEY = 'X4XksoWpBXXrKMG5Sk3bPmNvPplgckHy'  
client = AipNlp(APP_ID, API_KEY, SECRET_KEY)
```

想做情感分析的文本

```
text = "希望从情感的角度來看服務平臺誰可以成為未來的標準"  
client.sentimentClassify(  
client.emotion(text))
```

在網站上申請後
取得的資料

```
{'log_id': 4045458584401201752,  
'text': '希望从情感的角度來看服務平臺誰可以成為未來的標準',  
'items': [{  
'subitems': [], 'replies': [], 'prob': 0.997689, 'label': 'neutral'},  
{  
'subitems': [], 'replies': [], 'prob': 0.00117633, 'label': 'optimistic'},  
{  
'subitems': [], 'replies': [], 'prob': 0.00113456, 'label': 'pessimistic'}]}}
```

詞典法

前情提要

情感分析介紹

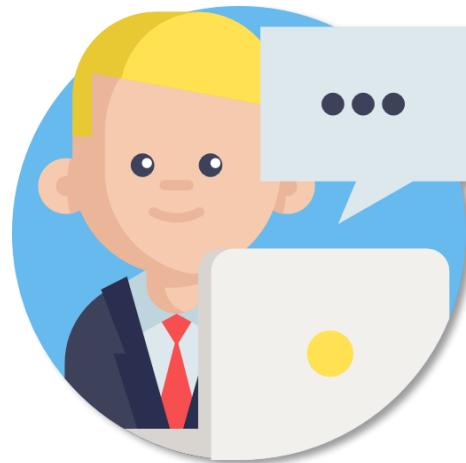
情感分析實務

發現

可以依產業/公司建立專屬詞庫
但需留意不同部門的想法不盡相同



業務部門



洗錢防制



風險審查

爬蟲法

前情提要

情感分析介紹

情感分析實務

發現

功能演示

请输入一段想分析的文本: [随机示例](#)

总而言之，是一家不会再去的店。

体验版最多可以输入 256 个字

分析结果:

情感偏负向



Ref: 1. [百度 / AI 開放平台 / 情感傾向分析](#)
2. wisers.ai



誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析(II)
Mail: TLYu0419@gmail.com

34

爬蟲法

前情提要

情感分析介紹

情感分析實務

發現

```
import requests
```

```
url = 'https://www.wisers.ai/?api=ailab-demo-apilb.wisers.com:8000/senti/api/processtext'  
data = { 'text' : '幾十年來從沒像此時這麼的風雨飄搖',  
        'model' : 'dl',  
        'output_level' : 'subject',  
        'show_scores' : 'true'}  
resp = requests.post(url, json=data)  
resp.json()
```

```
{'docid': None,  
 'is_successful': True,  
 'message': '',  
 'overall_res': {'sentiment': -1,  
                 'sentiment_score': 0.29},  
 'detail_res': []}
```

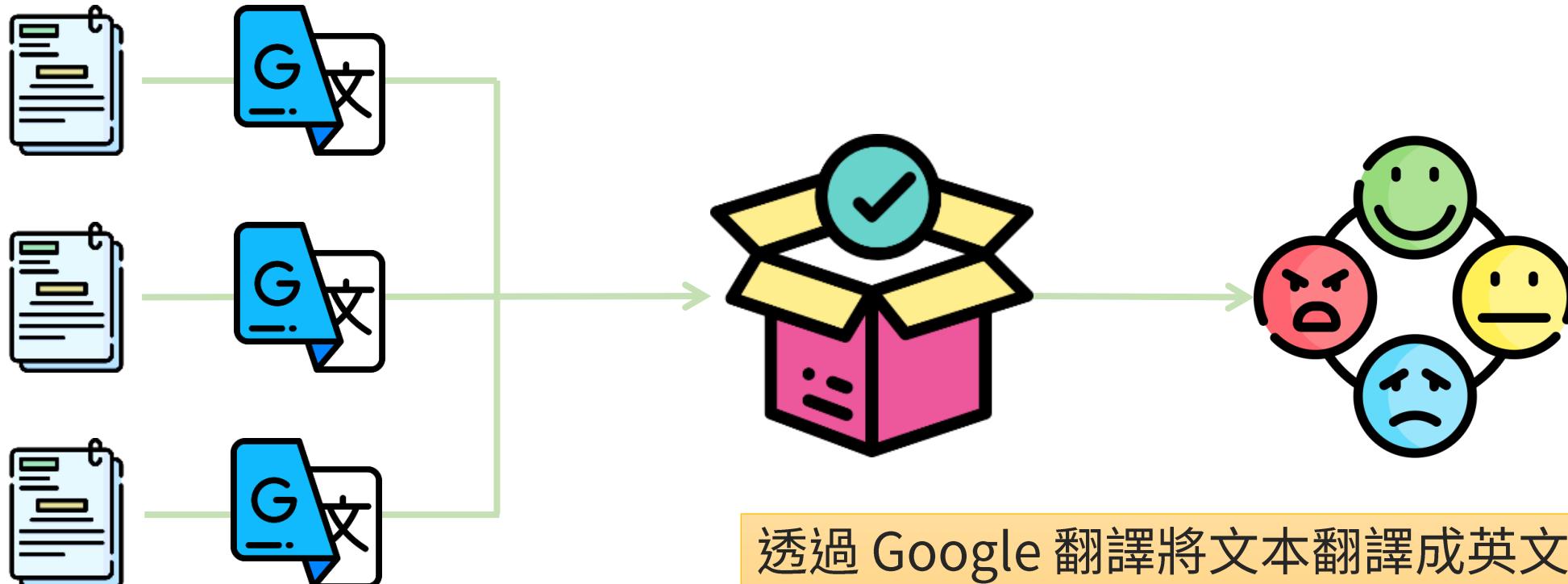
翻譯 + 套件法

前情提要

情感分析介紹

情感分析實務

發現



翻譯 + 套件法

前情提要

情感分析介紹

情感分析實務

發現

```
from googletrans import Translator  
import nltk
```

```
translator = Translator()  
sid = nltk.sentiment.vader.SentimentIntensityAnalyzer()  
#zh-tw to en
```

```
translations = translator.translate('對的事情做，不對的事情不要做 -> Do the right things, don't do the wrong things')  
{'neg': 0.0, 'neu': 0.758, 'pos': 0.242, 'compound': 0.3724}  
=====
```

不要為了選舉而選舉 -> Don't vote for election
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
=====

```
for translation in translations:  
    print(translation)  
    print(sid.polarity_scores(translation))  
    print('=====')
```

心存善念，盡力而為 -> Do your best
{'neg': 0.0, 'neu': 0.323, 'pos': 0.677, 'compound': 0.6369}
=====

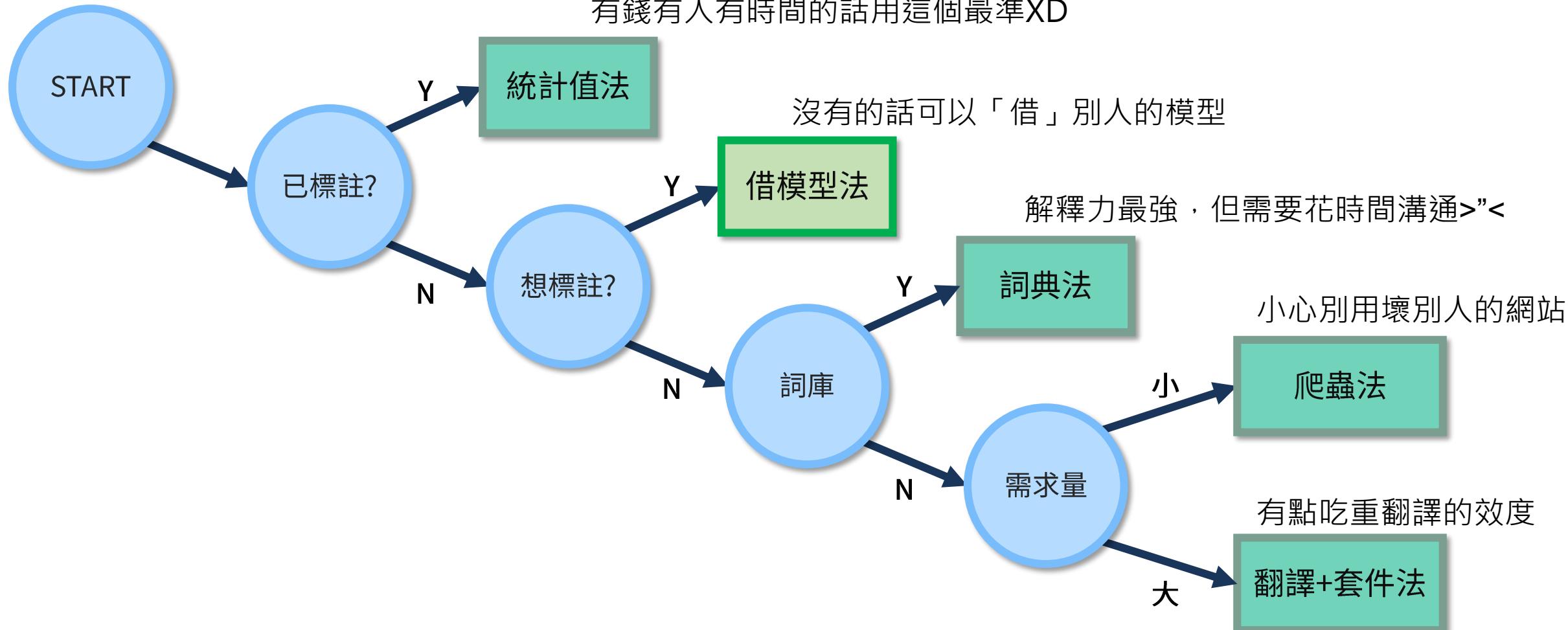
情感分析策略

前情提要

情感分析介紹

情感分析實務

發現





分析發現

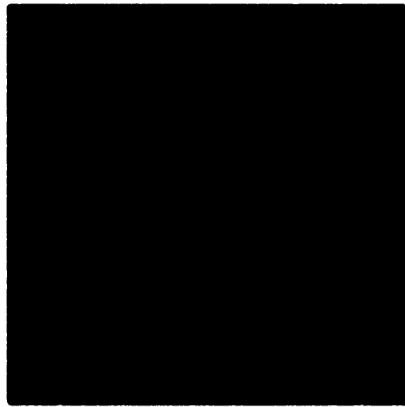
SNA + 情感分析

前情提要

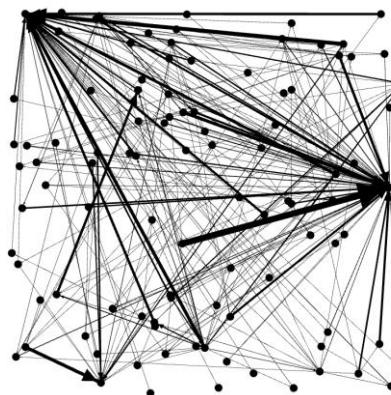
情感分析介紹

情感分析實務

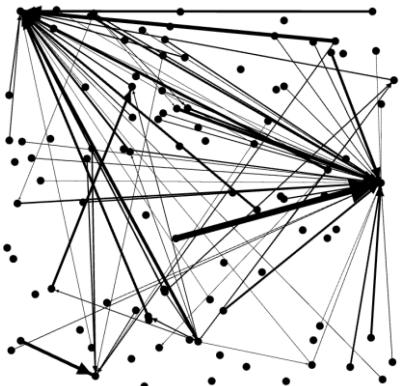
發現



1. 原始圖



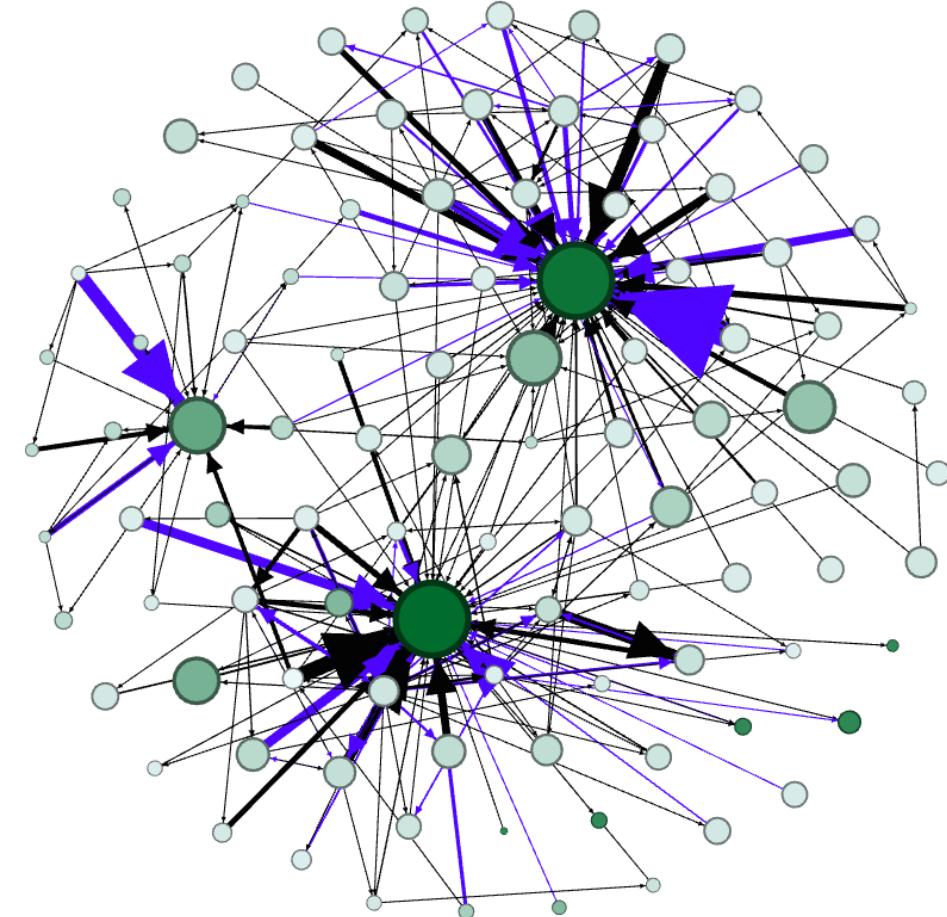
2. 保留重要節點



3. 濾除弱連結



4. Layout



5. 視覺化調整

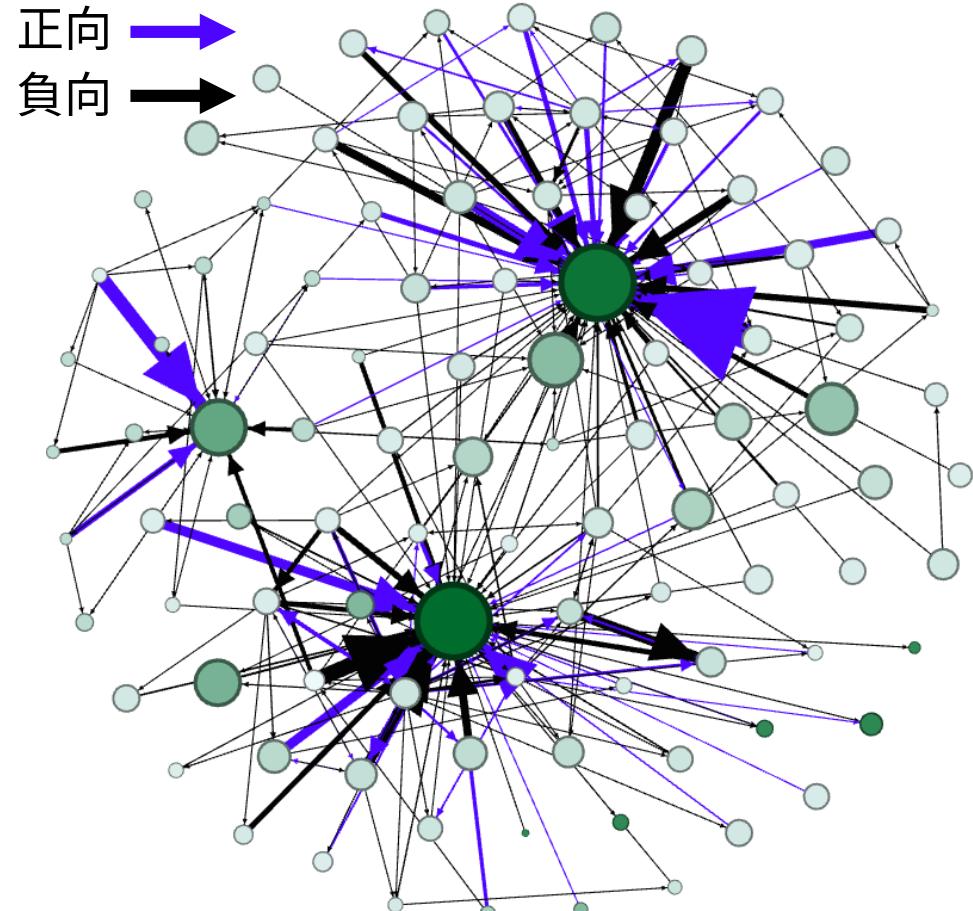
從互動行為找出 KOL

前情提要

情感分析介紹

情感分析實務

發現



- 透過 Pagerank 計算節點的重要性
 - 不只考量 In-Link 的節點數量
 - 還會考量 In-Link 的節點重不重要
- SNA 的視角較能真實反映影響力
 - 可以減少受到網軍/假帳號的影響

SNA 可以幫助候選人 / 公司找到
真正有影響力的KOL 來幫忙行銷、傳達理念

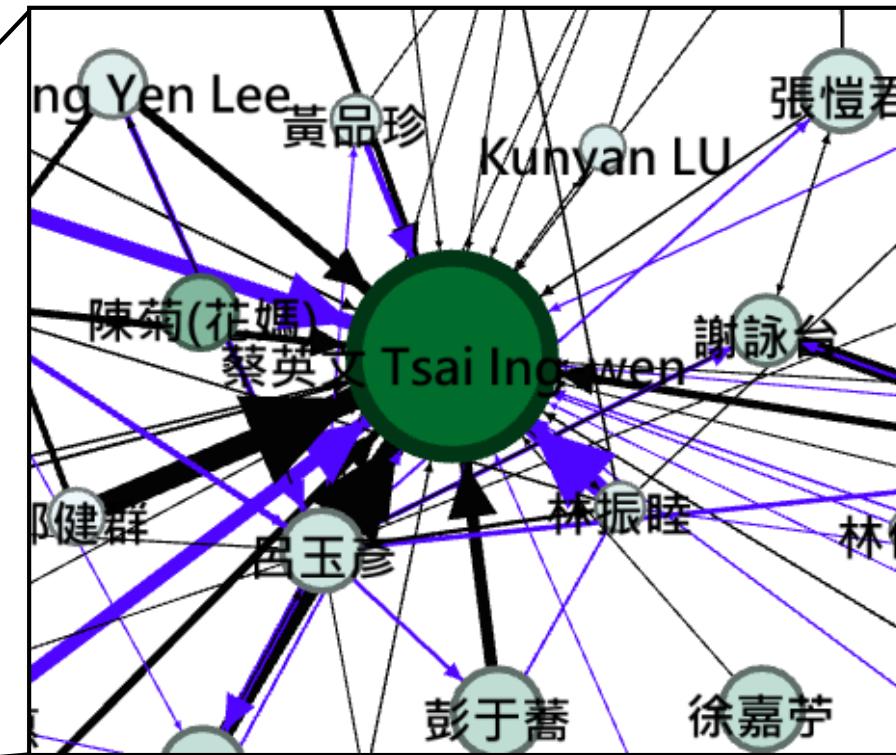
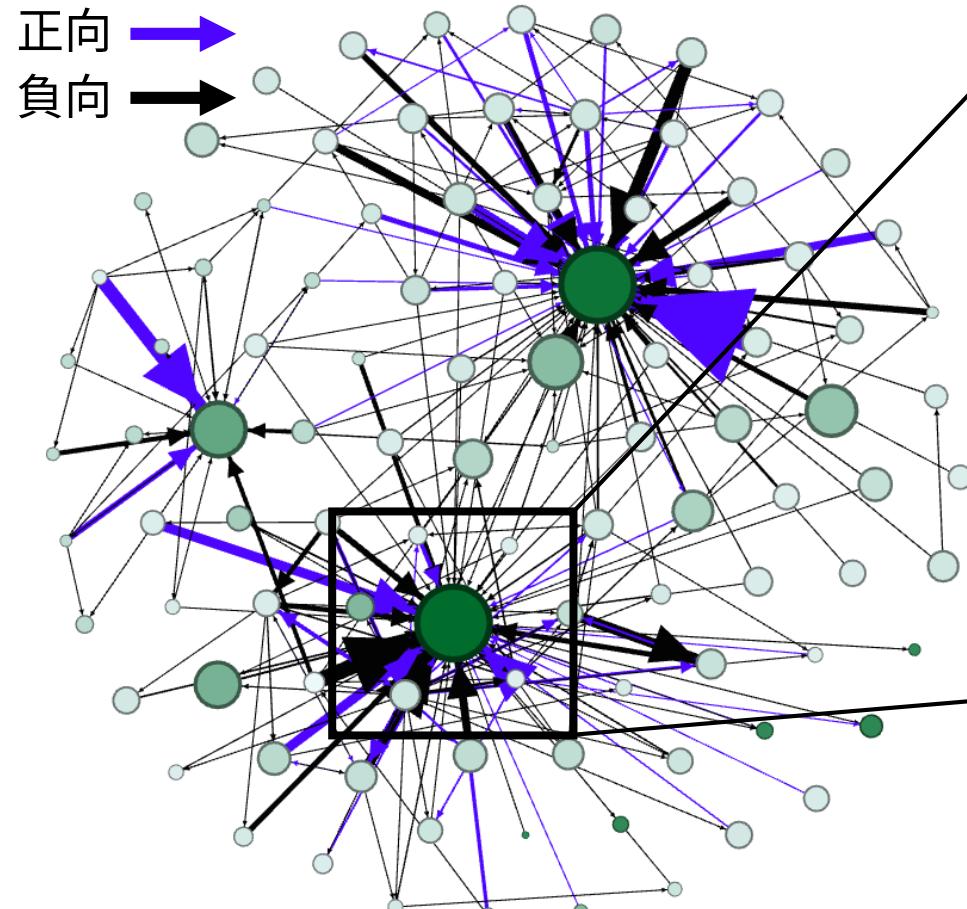
KOL 級候選人的聲音

前情提要

情感分析介紹

情感分析實務

發現



每位候選人身邊都不乏有正/負向的聲音
可以再進一步找出應該關注/經營的對象

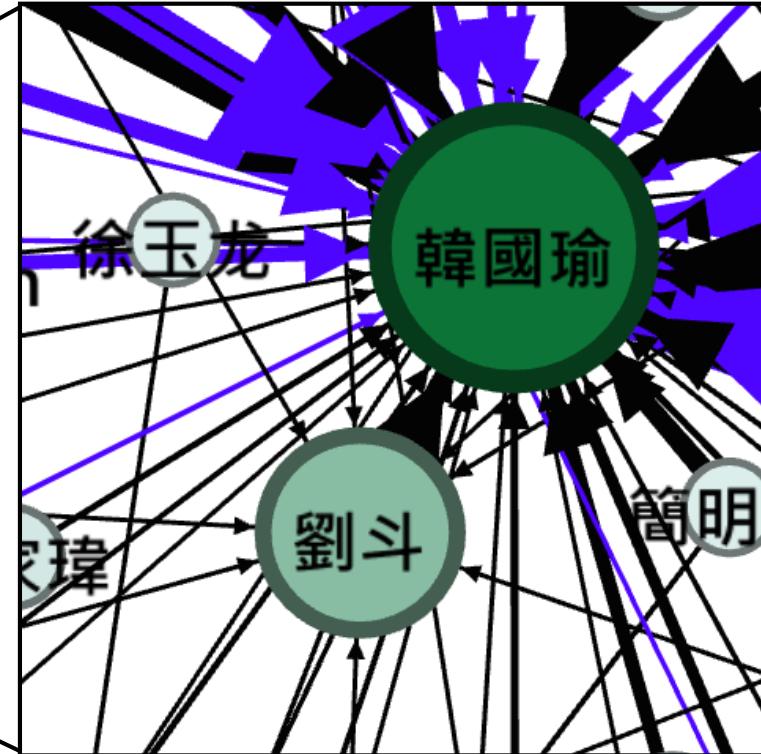
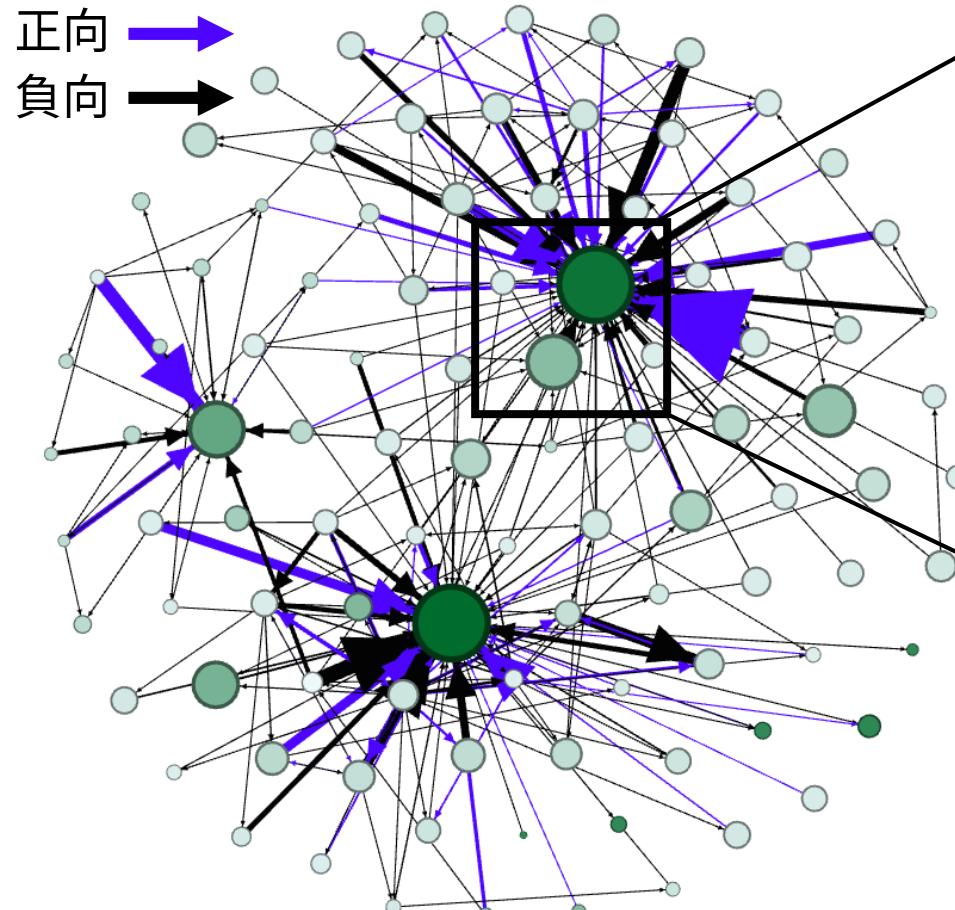
KOL 與周遭人的互動

前情提要

情感分析介紹

情感分析實務

發現



也可以觀察候選人身邊的人附近的情感態度
我們可以知道他的聲音應該被重視或者忽略

留意跨陣營的 KOL

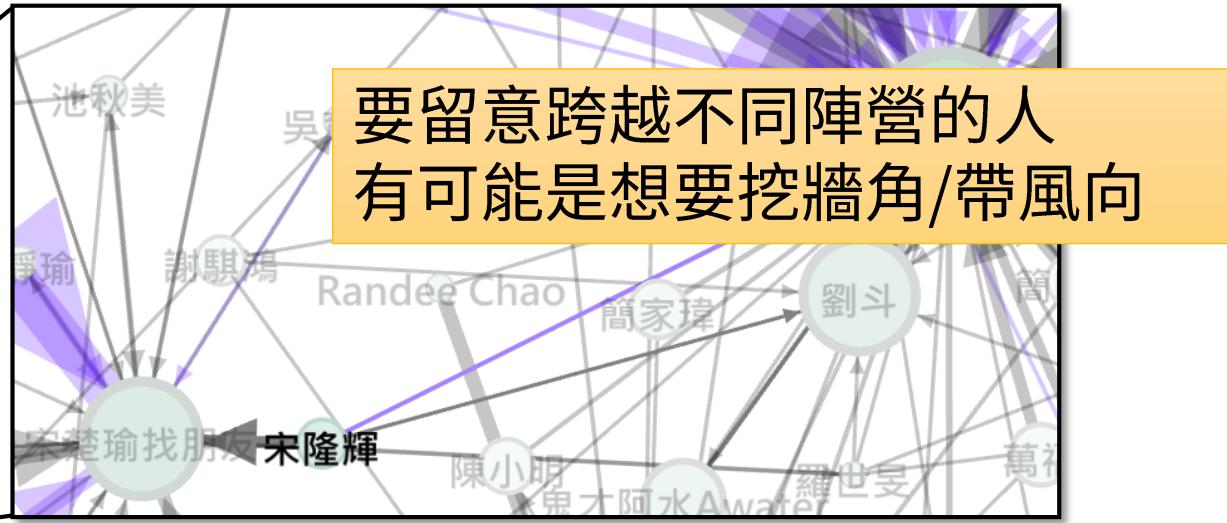
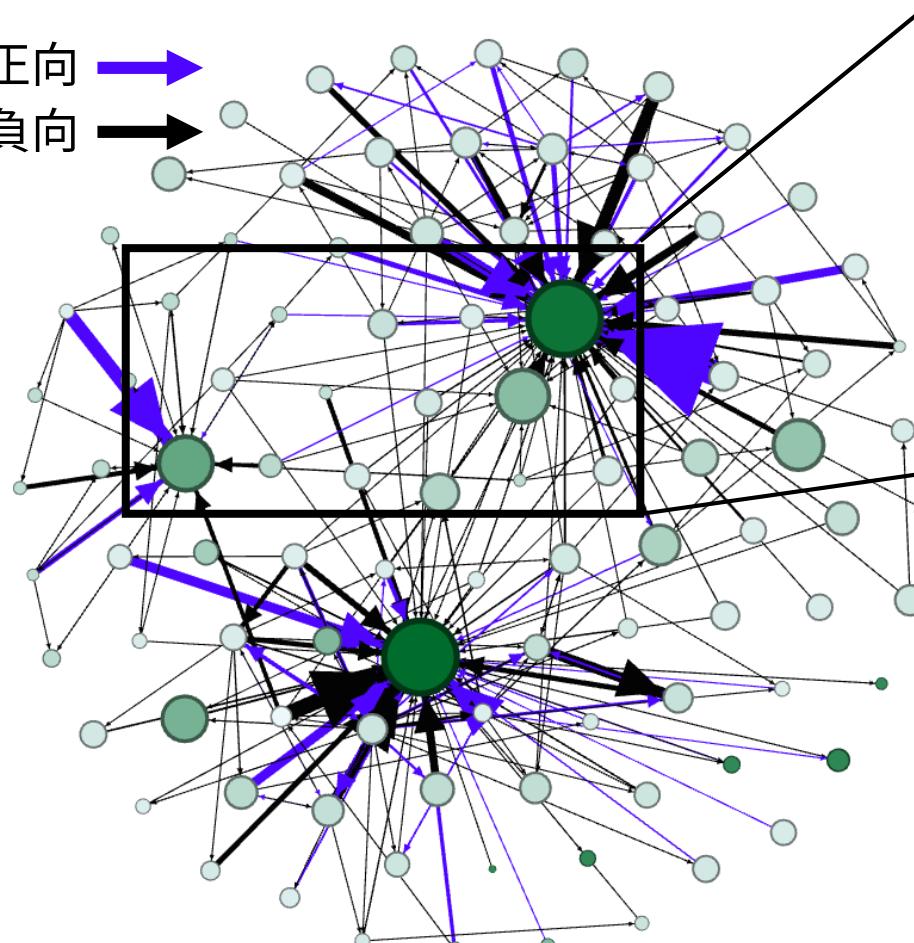
前情提要

情感分析介紹

情感分析實務

發現

正向 →
負向 ←



- To 韓國瑜:
中華民國保衛戰加油 百萬人重返凱道
- To 宋楚瑜:
我要投韓國瑜 團結才能下架民進黨蔡英文

總結

前情提要

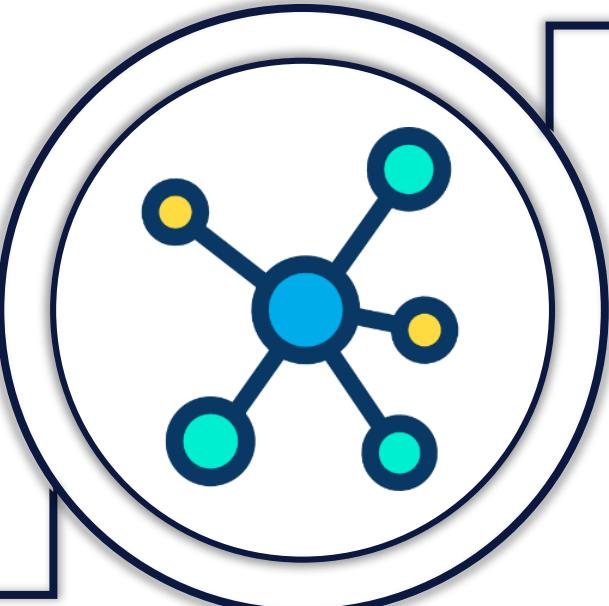
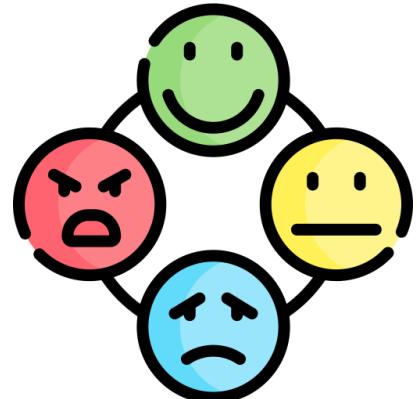
情感分析介紹

情感分析實務

發現

情感分析

- 統計值法
- 爬蟲法
- 借模型法
- 翻譯+套件法
- 辭典法



社群網絡分析

- 節點重要性的評估方式
- 檢視 KOL 對候選人/產品的態度
- 觀察 KOL 與周遭人的互動
- 留意跨陣營的 KOL