

社群媒體分析 讀書會 1

Group3

B104020008 林冠儒 B104020044 黎懿綺
B114020046 蔡佳倚 M134020017 李恩榮
M134020024 盧菽原 M134020039 顏淳
如

目錄

1. 專案介紹 (含專案動機)
2. Week_3 文字雲分析
3. Week_4 情緒分析 1
4. Week_5 情緒分析 2 (含問題與解決辦法)

1. 專案介紹

專案動機

- 「進擊的巨人」電影版上映, 想要 查看PTT留言的感想及情緒反應

專案資料

- 從PTT電影&八卦版爬取「進擊的巨人」相關貼文

2. Week_3 文字雲分析

斷詞 (新增專有名詞、停用字)

```
# 你提供的專有名詞列表
```

```
proper_nouns = [
```

```
    # 角色名稱
```

```
    "艾倫・耶格爾", "艾倫", "米卡莎・阿克曼", "米卡莎", "阿爾敏・亞魯雷特", "阿爾敏",  
    "里維・阿克曼", "里維", "漢吉・佐耶", "漢吉", "萊納・布朗", "萊納",  
    "貝爾托特・胡佛", "貝爾托特", "尤彌爾・弗利茲", "尤彌爾", "艾爾文・史密斯", "艾爾文",
```

```
    # 地名與組織
```

```
    "帕拉迪島", "瑪雷", "艾爾迪亞", "調查兵團", "憲兵團", "訓練兵團",  
    "瑪利亞之牆", "羅塞之牆", "希娜之牆", "牆內",
```

```
    # 關鍵術語
```

```
    "巨人", "無垢巨人", "九大巨人", "始祖巨人", "進擊的巨人", "鎧之巨人",  
    "超大型巨人", "顎之巨人", "戰鎚巨人", "車力巨人", "座標", "自由之翼",  
    "立體機動裝置", "雷槍", "硬質化"
```

```
]
```

```
# 參數形式手動加入停用字
```

```
stopwords_manual = [
```

```
    '看到', '看完', '知道', '新聞', '記者', 'The', 'the', 'my', 'on', 'of', 'from', 'Sent', '一直', '一下', '根本', '完全', '有沒有', '現在', '一堆',
```

```
    # 通用中文停用詞
```

```
    "的", "是", "在", "了", "和", "就", "也", "很", "這", "那",  
    "我", "他", "你", "她", "它", "我們", "他們", "什麼", "怎麼",
```

```
    # 情境相關低資訊量詞彙
```

```
    "說", "看", "見", "走", "去", "來", "東西", "人", "時間",  
    "地方", "有", "沒有", "一個", "一些", "都", "還是", "啊", "吧"]
```

```
stopwords.extend(stopwords_manual)
```

斷詞結果



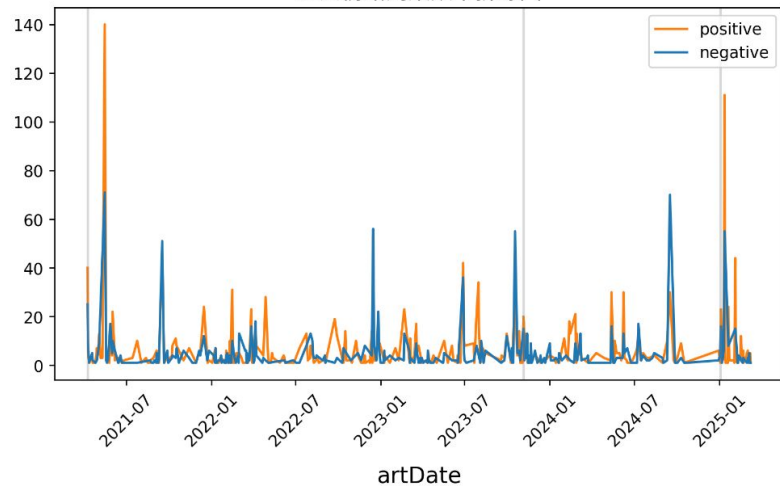
3. Week_4 情緒分析 1

3-1 lexicon-base 情緒分析

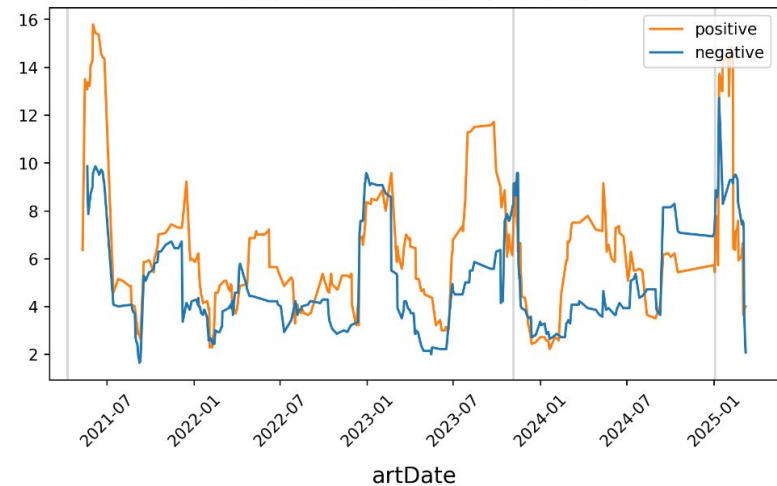
情緒分析指標

- 情感詞彙頻率
- 情感詞彙比例
- 情感詞彙分數
- 正負文章數趨勢
- 情緒代表字

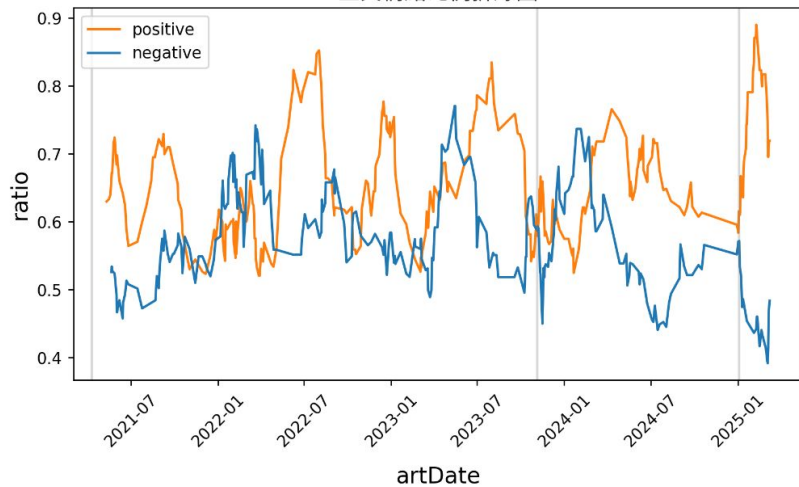
正負情緒詞彙頻率折線圖



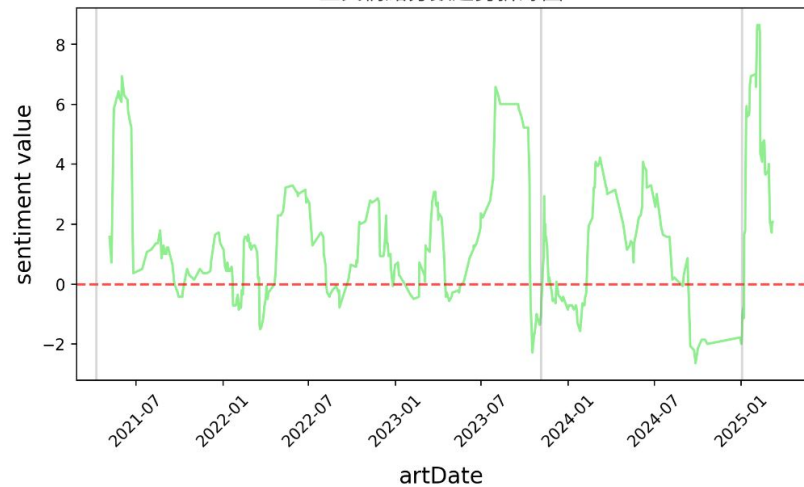
正負情緒詞彙頻率折線圖 (14天平均)



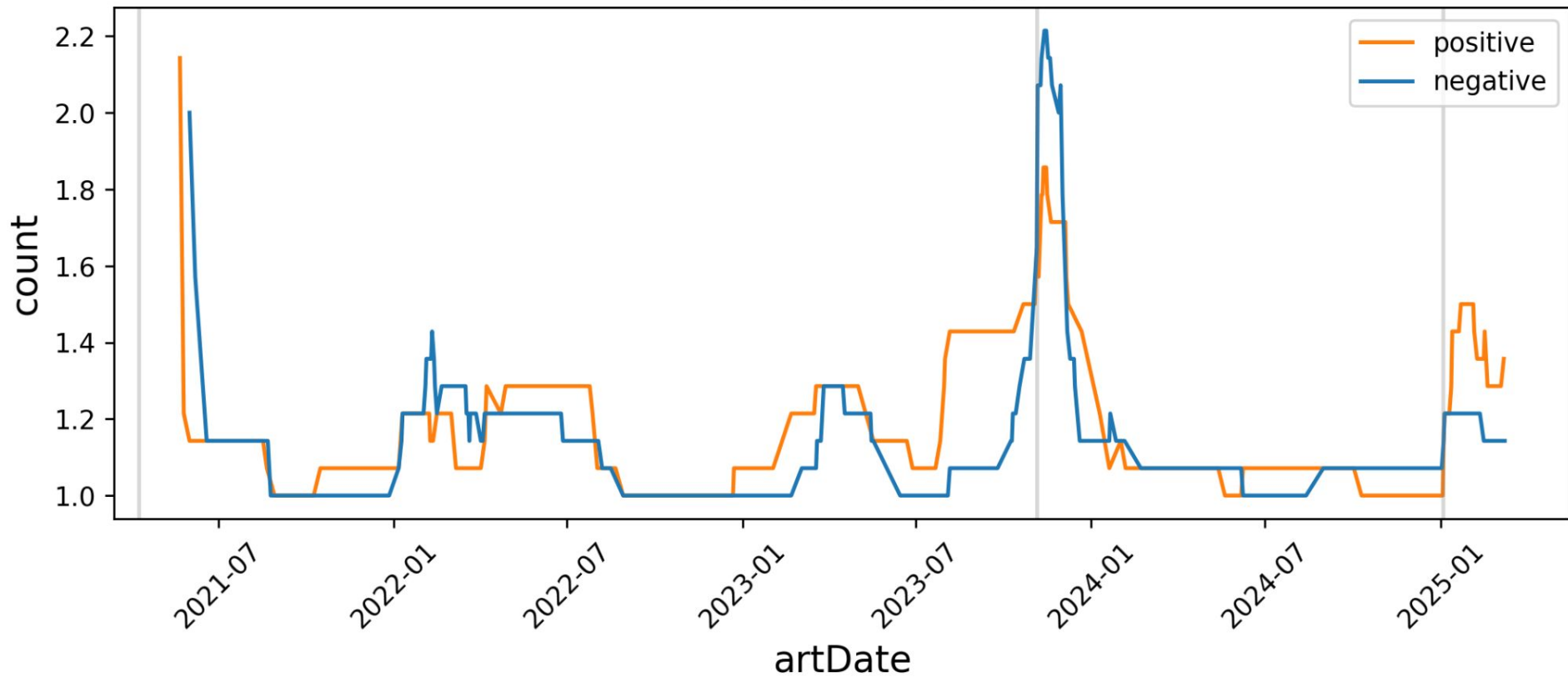
正負情緒比例折線圖



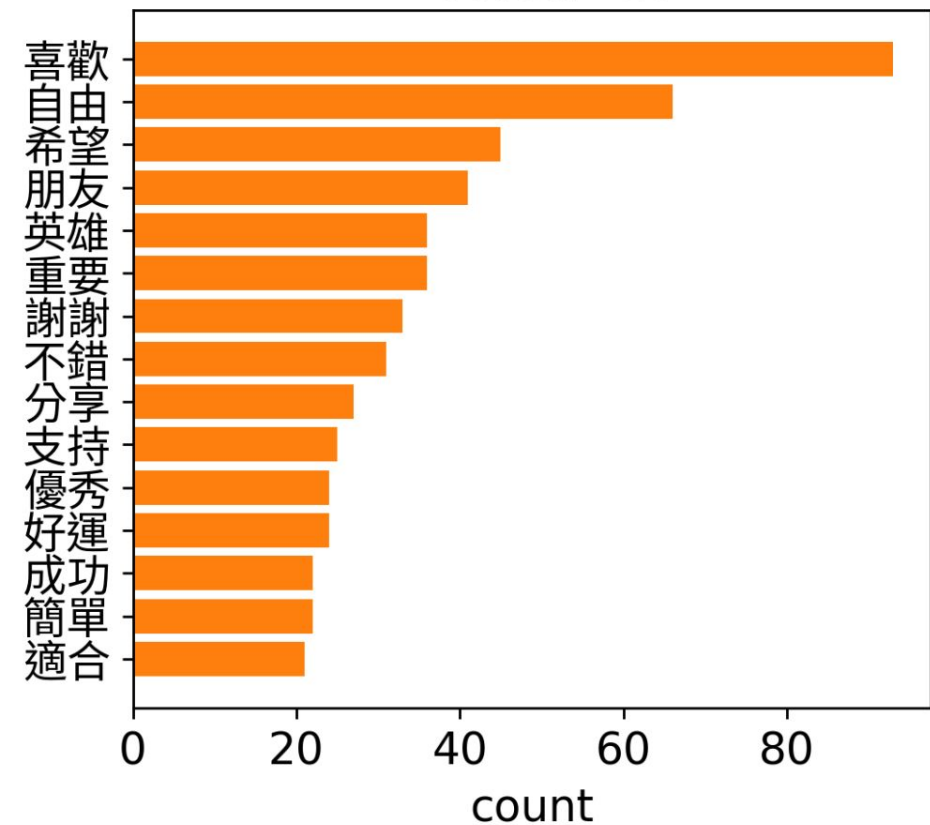
正負情緒分數趨勢折線圖



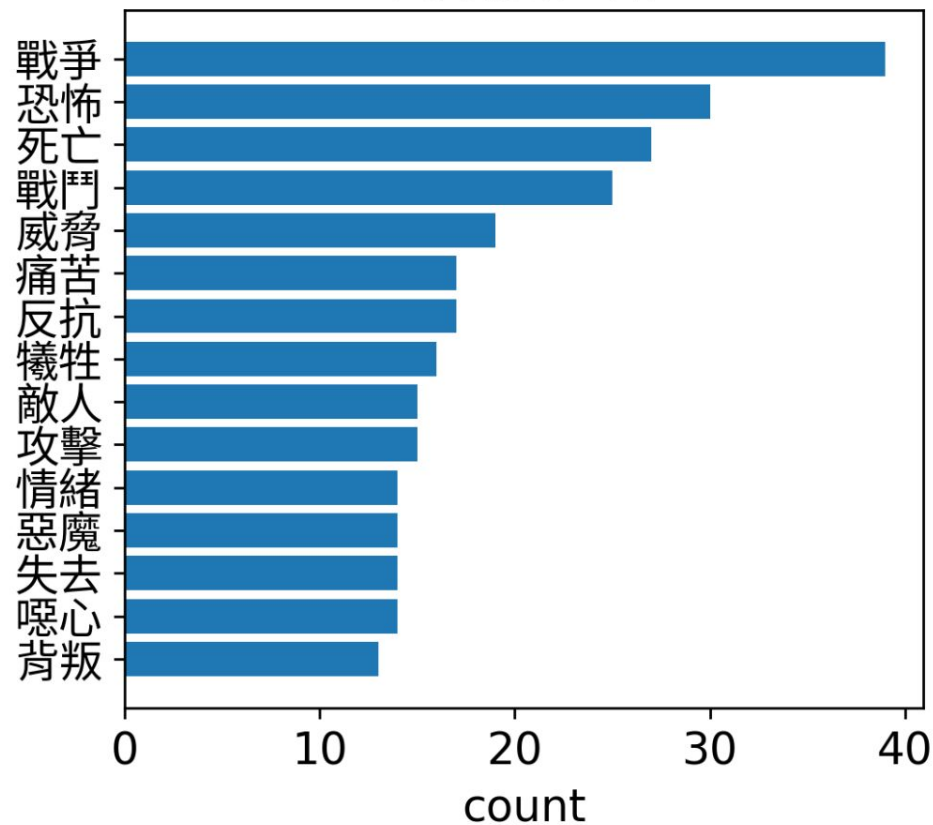
正負文章數趨勢折線圖



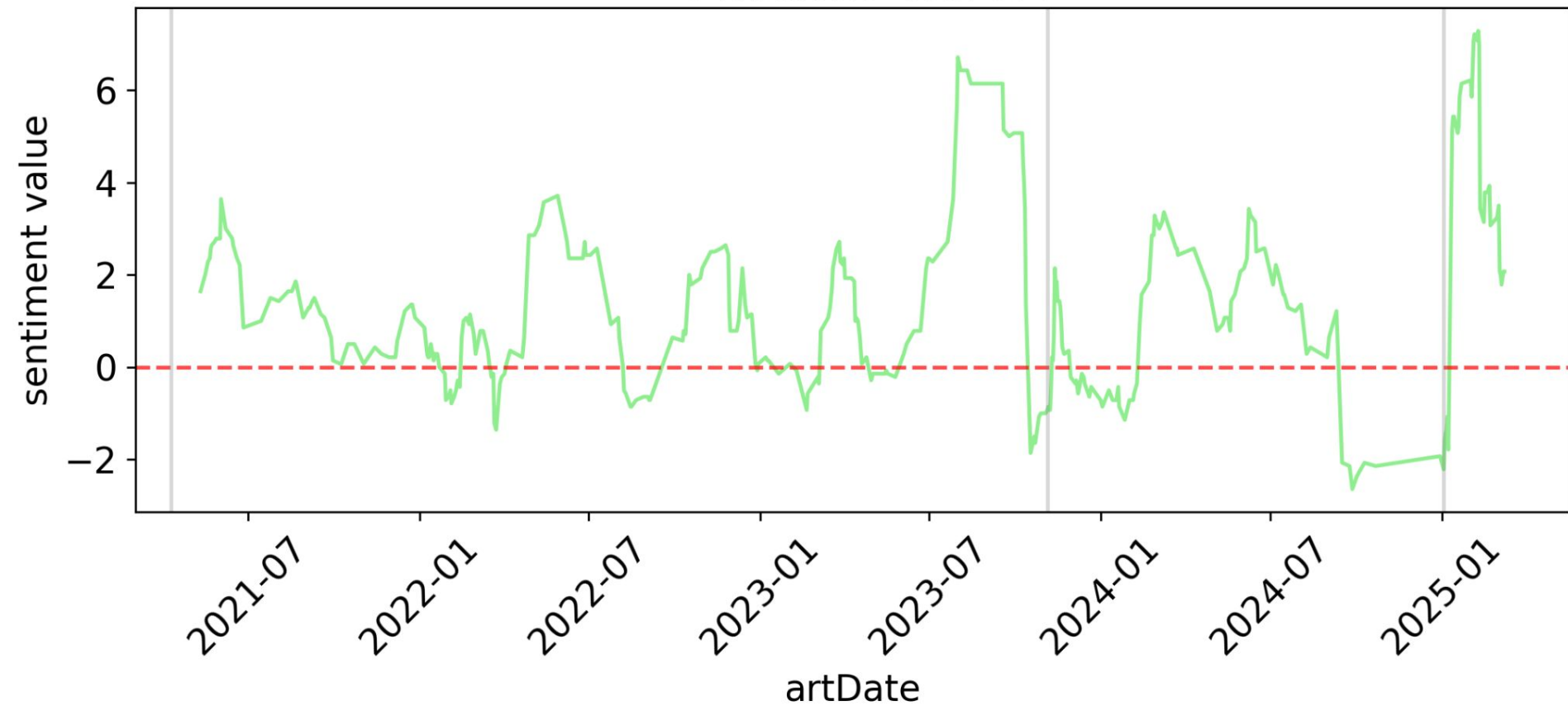
正向情緒總數



負向情緒總數



正負情緒分數趨勢折線圖



2023/11、12



2023/11、12

衝突 戰鬥 反抗 懷疑 恍惚 戰爭 對抗 擊敗 犧牲 打敗 擊敗 惡魔 騷擾 攻擊 訴苦 統治 諷刺 情緒 可憐 媽的 擔心 人敵 批評 傲慢 自私 震驚 迫害 瞧不起 嚴重 貪婪 該死 看不起 艱困 歧視 背叛 反叛 眼淚

	artTitle	sentence	sentiment_class
5674	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	也是極少數中的極少數吧	負向
5675	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	話說	負向
5676	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	最近結局	負向
5677	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	真要說動畫跟漫畫有啥不一樣的	負向
5678	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	就是漫畫中最後有幾幕	負向
5679	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	讓人有點懷疑米卡莎是不是在艾蓮死後	負向
5680	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	又跟約翰結婚了	負向
5681	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	但諫山創故意不說清楚	負向
5682	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	但動畫中	負向
5683	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	很明顯	負向
5684	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	把讓人懷疑的那幾幕刪掉了	負向
5685	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	且增加放上了米卡莎跟艾蓮	負向
5686	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	葬在一起的畫面如果說米卡莎後來又再結婚	負向
5687	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	不太可能又再跟艾蓮葬在一起	負向
5688	[問卦]進擊的巨人漫畫早完結，還在討論劇情？==	算是幫漫畫補完了最後諫山創留白的那個空間	負向
5689	[問卦]小便斗旁邊站著進擊的巨人，該怎麼辦？	如題沒有看動畫 真正的完結篇終於來了	負向

4. Week_5 情緒分析 2

4-1 利用CKIP進行自然語言處理

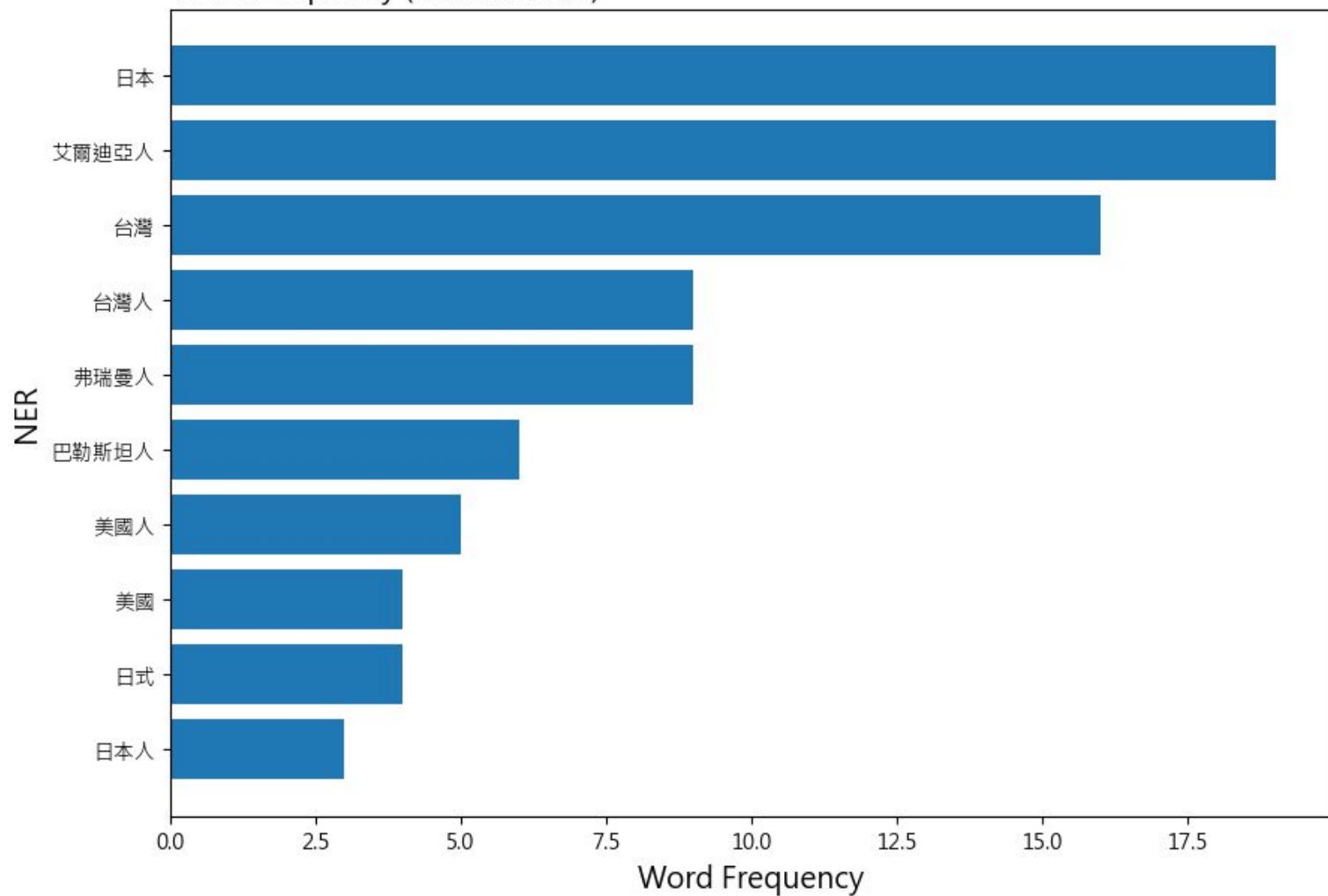
1. NER 找出文章裡各類實體

- 涉及到的民族、宗教、政治團體 (NORP)
- 涉及到的組織 (ORG)
- 涉及到的人名 (PERSON)

2. POS 實體型態識別

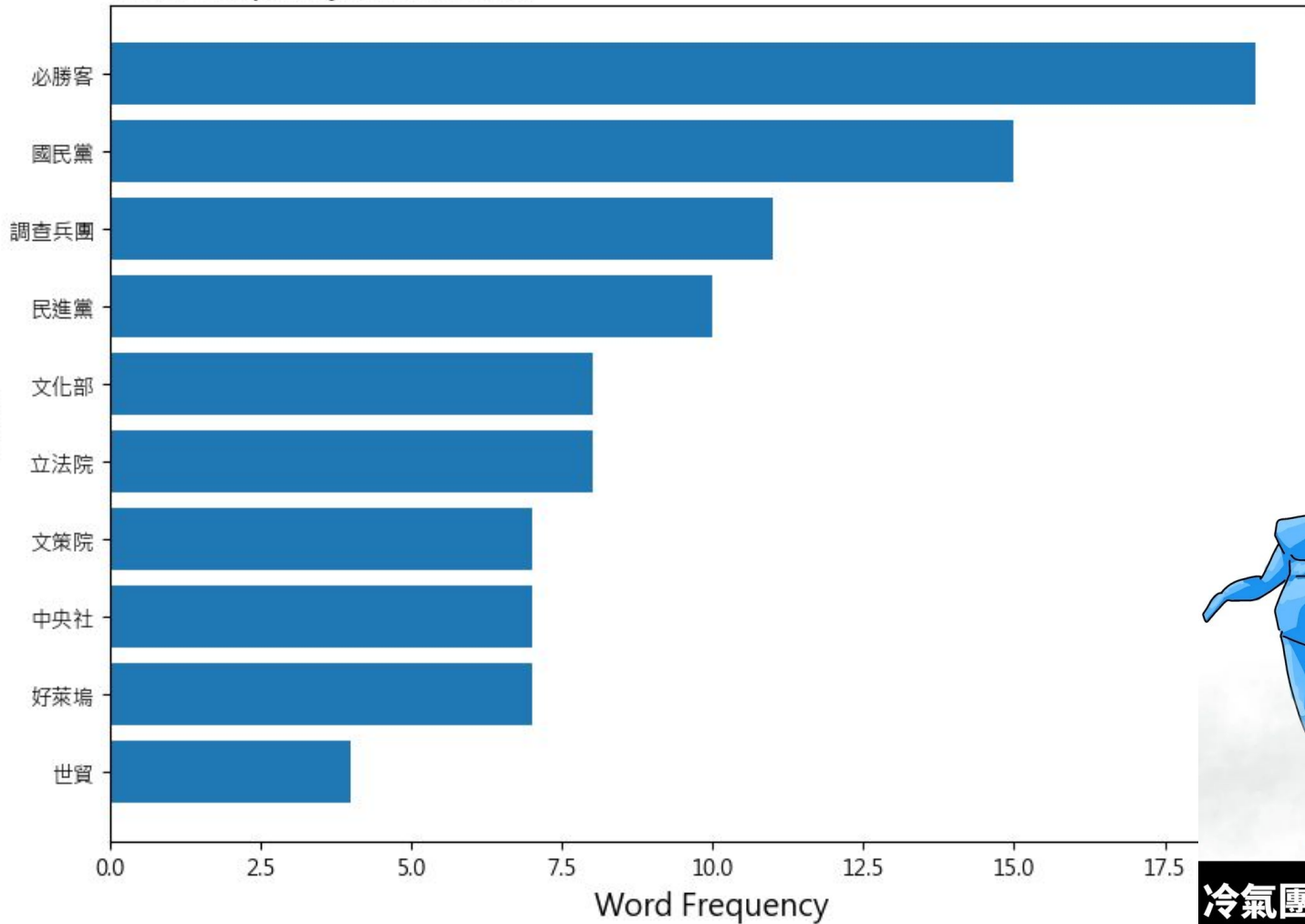
- 涉及到的名詞 (Noun)
- 涉及到的動詞 (VERB)

Word Frequency (NER is NORP)



Word Frequency (NER is ORG)

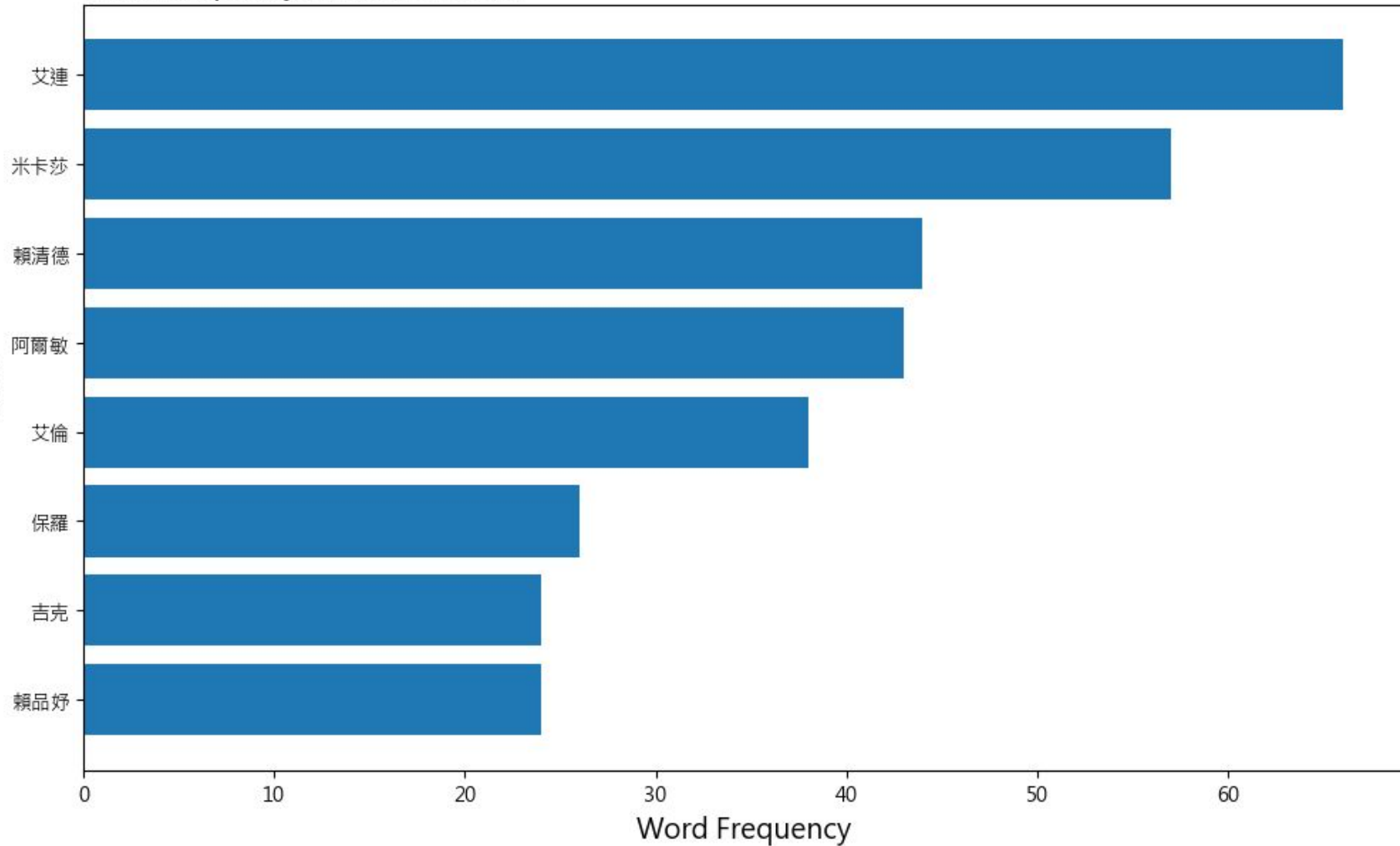
NER



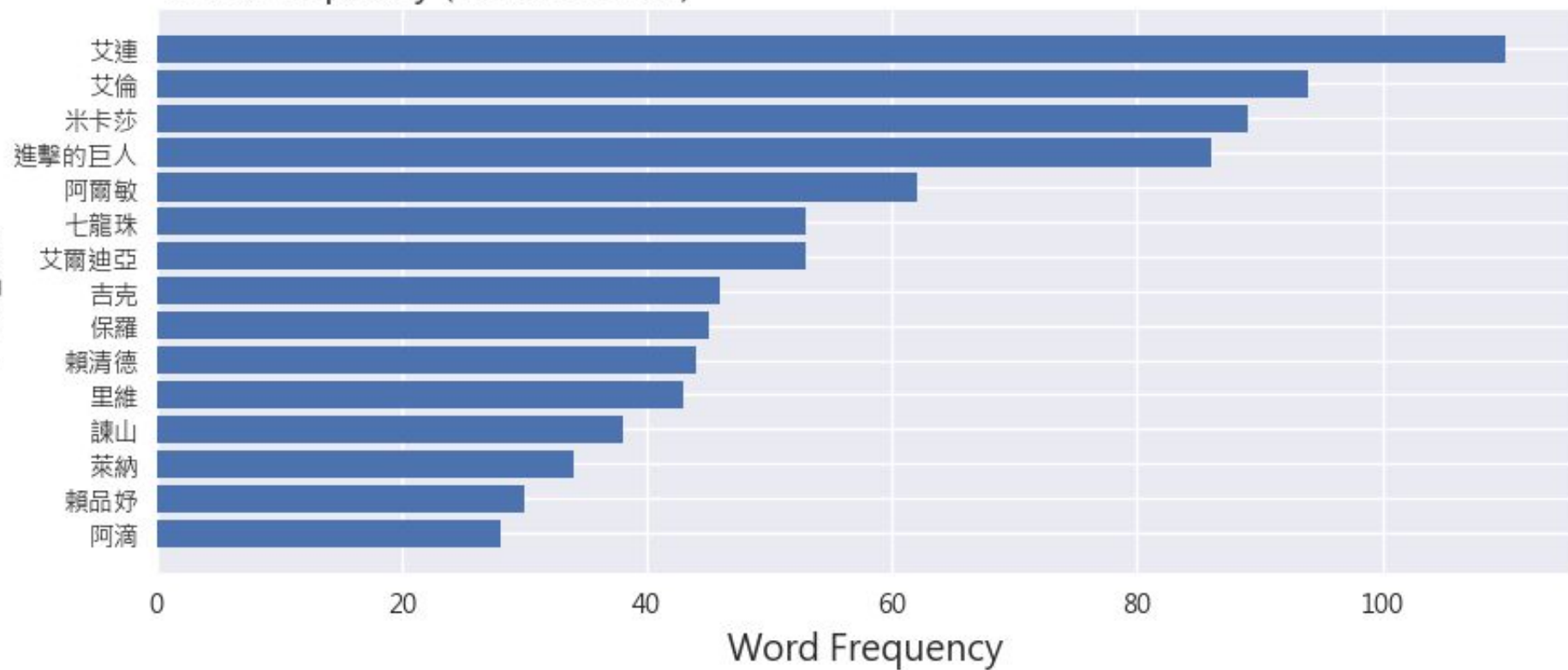
冷氣團襲來 拿出你的厚外套吧

Word Frequency (NER is PERSON)

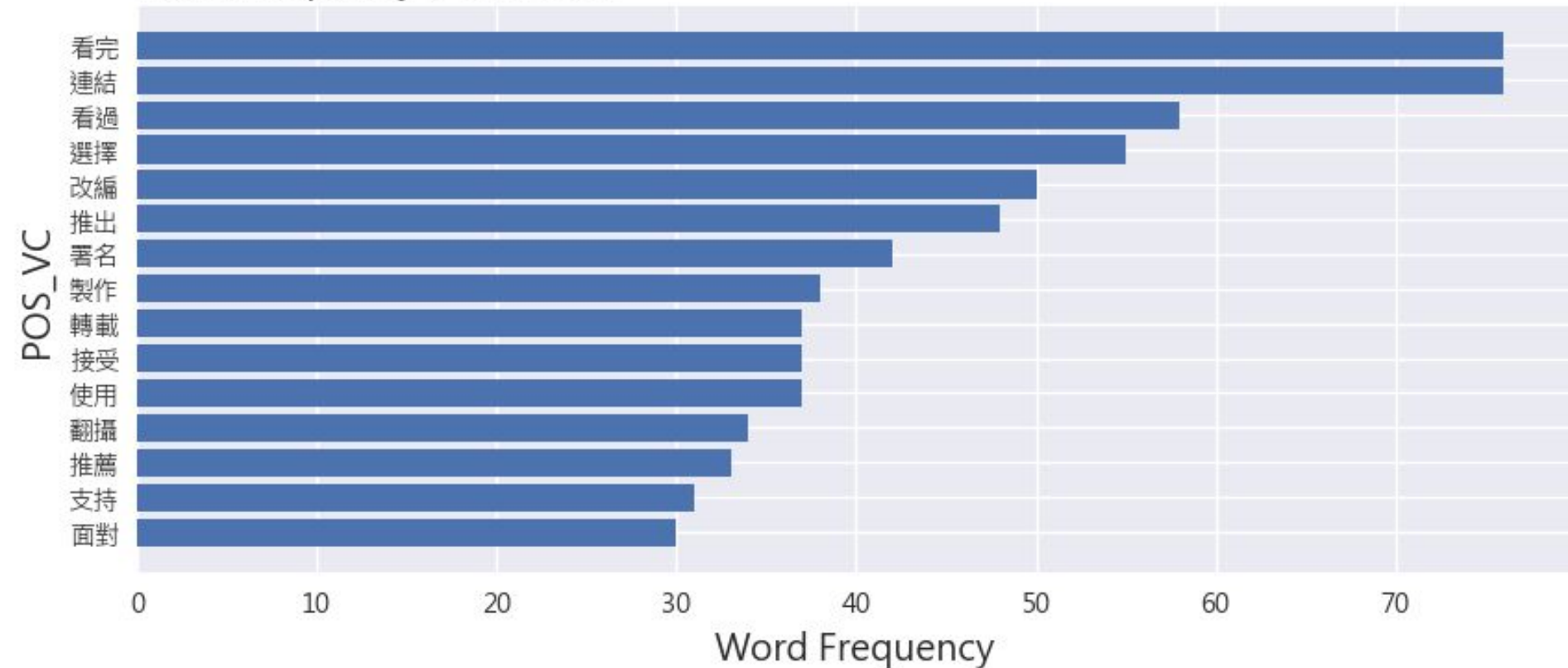
NER



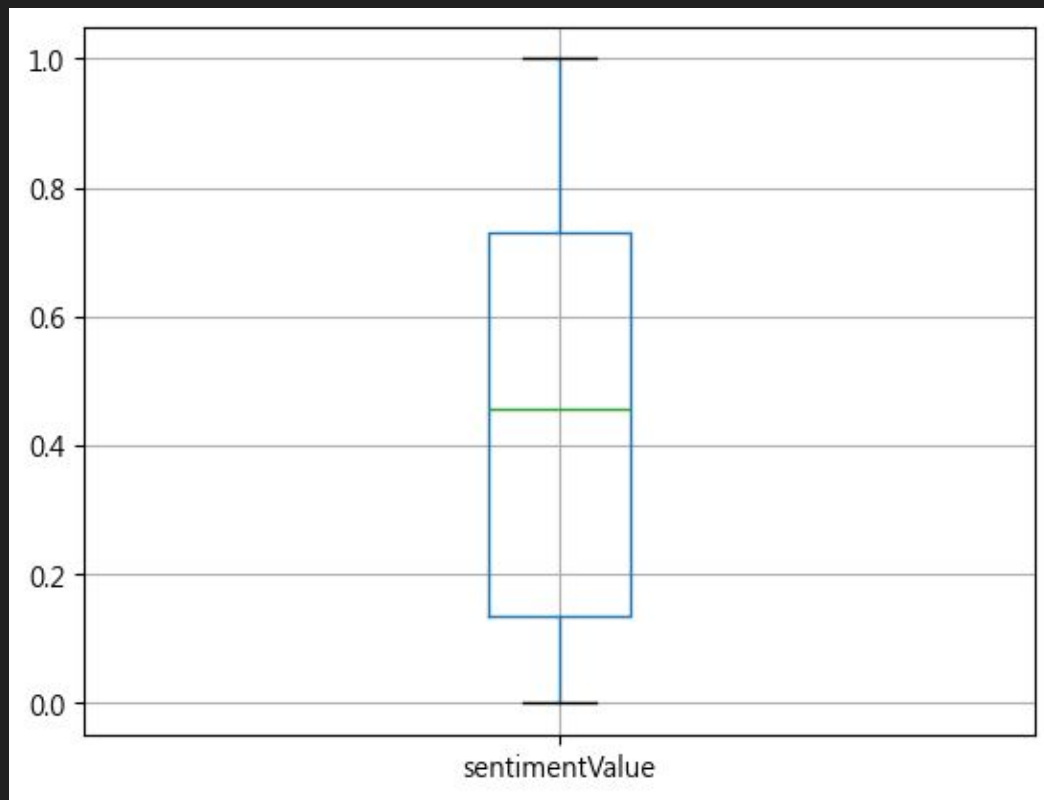
Word Frequency (POS is NOUN)



Word Frequency (POS is VC)



4-2 利用SnowNLP進行情緒分析





問題

1. 取平均的計算導致 大部分月份的情緒分數皆為中間 值
2. 情緒分數 未反映實際情緒
 - 抓出2024/07的資料，並未看出情緒高分的原因
 - 2025/01電影上映，我們預測情緒分數應該會較高，但是圖表呈現上並不是高峰 值
3. 情感分析誤判：劇情討論本身並不等於情感評價
 - 討論戰爭、死亡等 內容可能會包含許多負面詞彙，但是這不代表觀眾不喜歡這部作品，他們可能是在讚賞故事的深度

解決辦法

1. 針對**原本負面詞彙改為正面**（戰爭，痛苦，戰鬥，犧牲，死亡，威脅，復仇）
 - 手動修改LIWC字典：修改後發現結果差異不大，故認為此分析方法較無上述推測的錯誤分類
 - 調整SnowNLP詞彙權重：當句子包含上述詞彙時，強制把情緒分數提高



2023/08



2023/10

現實 不可能 國家 馬萊人 馬萊樣 鴿派 艾連 蝦 艾爾迪亞 想成 和平 蛋糕 內部 老實說 欺負 希望 鷹派 巴 情況 對錯 存在 無辜 進擊的巨人 利用 生日 出 來 原來 復仇派 以色列 過去 說說 世界 滅世 一樣 軍事 圍牆 哈瑪斯 對了 宗教 嚴重 痛苦 平民 過往 進擊 操弄 殺人 賈碧

2025/01



感謝聆聽

