



運用數位人文工具探討檔案時事議論 主題及其情感分類之實作

張奕萱* 林巧敏** 

【摘要】

本文以中國大陸網路時事論壇中討論與檔案相關內容為例，分析 2019—2020 年有關檔案貼文呈現的主題概念及其內容情感傾向，運用 NLPiR 大數據語義智能分析平臺進行斷詞、詞性標註及詞頻統計等文本預處理，進而採用中文情感詞彙本體庫進行文本情感分析。研究結果顯示，民眾最關注的檔案議題歸為兩類，一是特定人物與事件議題；二是以「檔案工作實務」和「檔案公開」為代表的社會議題。在情感方面，檔案時事論壇在針對特定人物與事件的討論時，呈現負面情緒較為明顯；而在談及檔案工作實務時，則展現出對於檔案開放應用的正面期待。

關鍵詞

檔案貼文 文本情感分析 社群媒體 檔案主題 中國大陸

* 國立政治大學圖書資訊與檔案學研究所碩士生

E-mail: zyx19971015@163.com

** 國立政治大學圖書資訊與檔案學研究所教授

ORCID 0000-0002-9309-9884

通訊作者 E-mail: cmlin@nccu.edu.tw

壹、緒論

網路論壇成為民眾對於社會現象發表演論的舞台，這種由具有共同興趣愛好的網民自發形成的網路社區，成為獲取、交流和共享資訊的重要媒介。在網路論壇，網友可以自由地發表自己對事物的看法與見解，表達自己的情緒與思想，這些認知、態度、情感和行為傾向的集合被稱為網路輿情，當輿情大量聚集時便形成輿論，主導或控制社會的話語權。近年來，已有不少對於網路輿情的分析和研究，多數是對於政治、社會議題的討論（Dwianto, Nurmandi, & Salahudin, 2021; Sabatovych, 2019; Yaqub, Chun, Atluri, & Vaidya, 2021），藉此可即時瞭解民眾的看法和思想動態，還能夠從中提取出人們的訴求，有助於更好地制定公共決策。

在中國大陸，網路論壇是民眾議論時事的主要管道，是發掘社會大眾意見的重要來源。尤其在中國大陸的環境，新聞媒體多半帶有官方色彩，人們多半是藉由網路論壇發表演論，論壇是一種比較能呈現民眾意見的資訊管道。然而，網路論壇貼文的內容質量參差不齊，若進行文本分析必須先尋找和挑選具有深度評論時事的網站，而檢索與判斷內容相關程度亦需要人工參與。有鑑於社會大眾對於檔案及檔案工作存在著不同程度的刻板印象，時值中國大陸檔案工作的轉型成長階段，檔案的價值及開放應用雖受到社會關注，但也引發不同評論觀點，若能將網路論壇上議論的觀點加以分析，應可瞭解中國大陸檔案事業發展以及民眾期許的方向。因此，本文重點不在驗證數位工具的優越性，而是以人文角度觀察網路論壇呈現之檔案時事分析結果，希望藉此瞭解在中國大陸的網路時事論壇發表與檔案相關的言論內容主題為何？最有討論熱度的檔案議題為何？這些議論內容文字所傳達之情感意向為正面或負面？

貳、文獻探討

一、文本處理技術分析

進行文本情感分析（sentiment analysis）首先進行收集文本詞彙的預處理，目前中文環境已有相當成熟的中文斷詞（Chinese word segmentation）

系統，而進行情感分析的技術，多採用基於情感詞典的方式。相關技術發展以及文獻成果分述如下：

(一) 文本預處理

文本預處理是對於文本材料進行自然語言處理 (Natural Language Processing, NLP)，包括語句檢測 (sentence detection)、斷詞 (word segmentation)、詞性標註 (part-of-speech tagging) 等過程，過濾文本中無實際意義的詞彙，讓機器能更好識別中文的文本內容，提高分析準確率(丁蔚，2017)。中文不同於拉丁語系的語言如英語、法語等，在詞與詞之間有明顯的空格或以其它符號為間隔，可輕易區分詞彙，在中文文本因文字連續書寫的特性，必須先經過斷詞，區分出語彙的最小單元。在中文環境常見的斷詞語料處理平臺有 NLPIR 大數據語義智能分析平臺、Jieba(結巴)、CKIP 中文斷詞系統，略述差異如下：

1. NLPIR 大數據語義智能分析平臺

Natural Language Processing & Information Retrieval (NLPIR) 是由中國科學院計算技術研究所研發的漢語詞法分析系統，主要功能包括中文斷詞、詞性標註、命名實體識別、新詞識別等等，支持多種編碼、多種操作系統。採用最短路徑法進行斷詞，隱馬爾可夫模型用於詞性標註，被廣泛用於各個領域 (李颯，2006；韓雪丹，2017)，較適用於簡體中文。

2. Jieba (結巴)

Jieba (結巴) 是一款廣泛使用的 Python 第三方中文斷詞庫，是基於前綴詞典進行詞圖掃描，生成所有可能成詞狀況的有向無環圖 (directed acyclic graph)，然後採用動態規劃查找最大概率路徑，找出基於詞頻的最大切分組合。對於未登錄詞，採用基於漢字成詞能力的 HMM 模型和 Viterbi 算法，Jieba 雖為中國大陸發展，但可支援繁體中文斷詞和自定義詞典，亦可進行詞性標註和關鍵詞抽取 (石鳳貴，2020)。

3. CKIP 中文斷詞系統

Chinese Knowledge and Information Processing (CKIP) 中文斷詞系統由臺灣中央研究院詞庫小組於 2003 年開發，使用者除了對文本進行斷詞，亦可進行詞性標註 (Ma & Chen, 2003)，亦可自行定義推薦詞列表和必填詞列表，並在 GitHub 平臺上提供開源程式利用 (Li, Fu, & Ma, 2020)，但該系統目前僅限於處理繁體中文。

(二) 基於詞典的情感分析法

比對情感詞典是文本情感分析最常採用的方法，是將預處理後的文本與構建好的情感詞典進行匹配，找到對應的情感詞，從而計算文本的情感極性 (洪巍、李敏，2019)。

中文情感詞典的建構與英文相比起步較晚，資源也相對有限。姚天昉、程希文、徐飛玉、漢思·烏思克爾特、王睿 (2008) 通過對英文情感詞典的轉化和人工修整構建了包含 3,120 個褒義詞和 3,485 個貶義詞的漢語情感詞典。此外，尚有其他比較具有影響力的中文情感詞典，例如《知網情感分析詞語集》(HowNet) 是一個以揭示概念與概念之間及概念所包含的屬性關係為基本內容的知識庫，包含正負面評價詞、正負面情感詞、主張詞和程度級別詞等 6 種詞語共 12 個子文件，8,936 個中英文正度面情感／評價詞語以及 219 個中文程度副詞 (王李冬、張慧熙，2016；魏慧玲，2014)。這些由人工構建的情感詞典操控容易，也可以靈活地調整詞條內容，但人工成本相對較高。因此，對於自動構建情感詞典的討論也成為近年研究的重點。

二、網路文本情感分析研究

網路論壇上由使用者發表的社評、影評、產品評價，或是由媒體刊載的新聞、文章等皆是分析社會大眾的意見資訊來源，對於這些文本材料進行挖掘與分析，不僅具有研究意義，更蘊含著巨大的社會價值。杜振雷 (2013) 曾以不同方法採集到 3,078 萬條微博訊息，在情感詞典的基礎上結合心理學對人類情感的劃分，進行細粒度情感分類的改造，構建出中文

微博情感詞典。隨後又提出基於六大輿情特徵的多特徵融合情感細分類方法，設計並實現微博情感挖掘系統，藉此搜索使用者關心的微博訊息，進行情感計算和分類。

紀雪梅(2014)通過計算情感詞在微博中的共現關係，並採位置算法將這些關係以圖形展示後，發現越靠近中心的高頻中心詞越能代表公眾的情感，將情緒詞分類統計後，便可看出使用者表達的各類情緒強度。利用社會網路分析法對於「奪命快遞」和「公務員加薪」兩起社會事件的傳播網路進行分析，發現那些表達「激動」、「詆毀」、「同意」或「反對」等情緒的使用者，在社會事件信息傳播中，往往起到關鍵作用，在事件發生時通過這種情感擴散規律，若能制定應對策略，就能避免公眾情感偏極化。

李孟潔(2009)提出一個擷取評論文章中有關情感意見的特徵集方法，先找出具有意見的詞彙、片段等特徵，利用機器學習方法訓練出一套評價分類系統，針對不同領域的評價文章予以分類分級，認為對於網路評價的分析應先區分出有情緒傾向的句子，再進行特徵詞選取訓練步驟，或加入句法結構與字詞間關係的考量分析。

Dwianto、Nurmandi 與 Salahudin (2021)觀察到美國與印尼在新冠肺炎(Covid-19)爆發期間，國家元首皆透過 Twitter 公開宣傳防疫相關政令及應對措施，在 2019 年 9 月 21 日到 10 月 21 日間，佐柯威(Jokowi)、川普(Trump)與新冠肺炎等關鍵字並列為熱門搜尋字詞；為瞭解民眾對兩位元首政策推文的情感傾向，利用 Brand 24 自動化軟體，蒐集 38,813 則推文，並進行情感分析，結果顯示，民眾對於兩位元首的推文，於該段期間的負面情感皆大於正面情感。

可知關注於網路文本情感分析的研究日漸增多，也發展出不同領域進行文本情感分析的成果。但對於文本斷詞工具和情感詞典的使用，仍有語言環境的差異，進行檔案時事議論內容的探討原本已不多見，更遑論對於群眾意見較為漠視的中國大陸。因此，本文基於關心兩岸檔案領域的社會輿論走向，嘗試運用數位工具分析中國大陸網路論壇議論之檔案內容重點，透過對於自身領域的實證應用研究，探討現有工具和技術在中文環境運用的經驗。

參、研究設計與實施

一、研究設計

本文選擇屬於議論時事及檔案議題相關之論壇，經檢索及人工判斷整理相關貼文後，進行文本內容探勘分析，瞭解這些貼文所呈現的重要詞彙，並探討貼文內容呈現之情感意向，從而知悉社會大眾對於檔案議題的認知與情感。研究實施分為兩個部分：第一部分中，根據設定的關鍵詞檢索中國大陸「風聞社區」網站與檔案相關議題的貼文，使用 NLPiR 大數據語義智能分析平臺對於文本進行預處理，以人工協助剔除無關主題文本，再進行斷詞和標註詞性，最後去除停用詞，通過統計詞頻得知檔案議題貼文所關注的核心主題。第二部分為情感分析，導入中文情感詞彙本體庫與處理後的文本詞彙比對，得到各詞彙的情感分數，結合詞頻計算文本總情感分數，將貼文分為正面、中性和負面 3 種情感類型。

二、研究對象

本文希望能夠獲取內容較有深度探討的貼文為研究對象，排除匿名性或是回覆句子過短之貼文，以接受讀者或專家投書，並有公開姓名或筆名，以示對文章內容負責之網路論壇為限。曾瀏覽強國論壇、風聞社區、天涯論壇、凱迪社區和四月網等屬於中國大陸受眾較廣且較為活躍的論壇，並逐一勘查其討論主題、管理模式、用戶類型以及文章質量。由於各論壇不同程度地存在缺乏用戶監管（如天涯網）、貼文時效性差或品質低（如四月網、凱迪社區）或檢索困難（如強國論壇）等問題，因此在綜合考慮與評判後，選擇以「風聞社區」為採集檔案議題貼文來源對象。「風聞社區」（<https://user.guancha.cn/>）是 2017 年由觀察者網推出的內容平臺，擁有超過 20 萬名註冊用戶，用戶可以發佈國際時政、社會熱點、娛樂八卦和日常生活等各方面有深度的內容。

檢索採用「檔案」、「檔案館」、「檔案工作」為檢索詞，但考慮到大陸民眾對於「檔案」的理解與表達，也可能採用「文件」、「記錄」等詞語，故加入「文件」、「記錄」兩個檢索詞，輔以人工判斷篩選出切合主題的貼文，以求最大範圍並精準囊括與「檔案」議題的相關貼文，以 2019—2020

兩年間所有貼文為範圍，經人工瀏覽判斷，共獲得 162 篇與檔案相關的貼文。

三、研究工具

本文欲探討中國大陸網路論壇對於檔案的議論主題與情感傾向，因文本內容有在地使用特性，故採適用於簡體中文環境的語庫及工具。

(一) 文本斷詞以大陸中科院 NLPIR 大數據語義智能分析平臺，此為北京理工大學張華平團隊設計研發的語義智能分析工具，其前身為 Institute of Computing Technology, Chinese Lexical Analysis System (ICTLAS) 中文斷詞平臺，2009 年開始更名為 Natural Language Processing and Information Retrieval (NLPIR)，並逐步擴展為包括採集、理解、挖掘和搜索等 13 項功能的操作平臺，最新版本發佈於 2019 年 (張華平、商建雲，2019；NLPIR1, 2021)。

(二) 中文情感詞彙本體庫採用大連理工大學信息檢索研究室林鴻飛教授及其團隊，以 Ekman 的 6 大類情感分類體系為基礎，將情感分為 7 大類 21 小類，此詞彙庫將詞彙劃分為名詞、動詞、形容詞、副詞、網路詞語、成語、介詞短語等 7 類，每個詞對應一種極性，1 為褒義 (正面)，2 為貶義 (負面)，0 則表示中性，情感強度由弱到強分為 1、3、5、7、9，共收錄情感詞 27,466 個 (徐琳宏、林鴻飛、潘宇、任惠、陳建美，2008)。本文依照原本詞庫情感詞標示之情感強度加以統計，情感分數亦採 1、3、5、7、9 之 5 種等級。

肆、研究結果分析

一、檔案貼文斷詞結果分析

本文完成 162 篇網路論壇貼文的收集、整理與匯總後，使用 NLPIR 大數據語義智能分析平臺對於貼文進行斷詞處理，162 篇貼文共涵蓋 20,999 種詞彙，詞性分布涵蓋 15 大類 43 個子類，總詞頻達 157,043 次 (表 1)。觀察在檔案貼文中使用的詞彙，以名詞 (47.207%)、動詞 (34.849%) 出現數量最多。名詞當中又以人名 (8.424%) 和地名 (3.410%) 最為常見，動

詞則是動名詞（6.300%）和不及物動詞（5.638%）使用最多。可知在網路論壇中，檔案貼文習慣以人事物為論述主題，搭配使用動詞進行文字表述，整體文本詞語風格偏向於敘述性而非論證性內容。

表 1

所有貼文經斷詞分析之詞性及詞頻統計表

詞性	詞彙數	百分比% (N=20,999)	詞頻數	百分比% (N=157,043)
名詞	9,913	47.207	70,933	45.167
動詞	7,318	34.849	51,736	32.944
形容詞	1,428	6.800	9,532	6.071
數詞	387	1.843	6,542	4.166
量詞	224	1.067	5,695	3.626
副詞	624	2.972	5,219	3.324
區別詞	431	2.052	3,580	2.280
時間詞	432	2.057	2,686	1.711
處所詞	114	0.543	786	0.500
狀態詞	97	0.462	202	0.129
代詞	7	0.033	90	0.057
歎詞	4	0.019	4	0.002
擬聲詞	6	0.029	12	0.008
前綴	1	0.005	3	0.001
字符串	13	0.062	23	0.014
總計	20,999	100.000	157,043	100.000

名詞之詞彙數及詞頻皆為各詞性之首，在總數 9,913 個名詞中，以普通名詞數量最多，其次是人名、地名、團體機構名以及其他專有名詞。將此類屬於可明確呈現主題概念之名詞，統計呈現前 10 個高頻詞，可知在普通名詞詞彙中以「人」、「檔案」、「歷史」、「國家」為多，人名中「毛澤東」、「蔣介石」、「斯大林」（史達林）（註 1）等政治人物最為突出，與之相對應的地名中以「中國」、「美國」、「蘇聯」為提及較多的概念，團體機構名最常出現「中共中央」、「中國共產黨」、「聯合國」等機構。至於其他

專有名詞，則包括：「聊齋」、「永樂」、「抗日戰爭」、「法西斯」、「土地革命」等，根據高頻名詞統計可呈現文本內容概念多是與討論中國歷史或政治議題相關者。

表 2

各類名詞前 10 個高頻詞排序表

序 列	普通名詞		人名		地名		團體機構名		其他專名	
	詞彙	詞頻	詞彙	詞頻	詞彙	詞頻	詞彙	詞頻	詞彙	詞頻
1	人	1,188	毛澤東	262	中國	1,050	中共中央	45	聊齋	123
2	檔案	831	蔣介石	131	美國	757	中國 共產黨	37	永樂	47
3	歷史	810	斯大林	129	蘇聯	337	聯合國	25	抗日 戰爭	33
4	問題	414	趙君邁	63	日本	332	西路軍	21	微信	30
5	文件	414	毛岸英	62	朝鮮	177	黨中央	18	法西斯	24
6	紙	349	蒲文珊	56	俄羅斯	131	北京大學	17	中華	24
7	國家	330	特朗普	55	埃及	130	太平洋	13	土地 革命	22
8	戰爭	328	胡適	54	印度	129	中央軍委	11	嘉靖	21
9	書	322	蒲松齡	50	上海	124	中國人民 解放軍	10	乾隆	21
10	中央	301	尚俊俊	47	英國	121	大西洋	10	英文	20

二、檔案貼文議論主題分析

本文將人名、地名和其他專有名詞等此類具有專指性、特殊性的詞彙劃為一類，將普通名詞所包含與社會議題相關的詞彙歸為另一類，從中歸納檔案貼文最常討論的社會話題。

(一) 特定人物與事物

對於人名、地名、團體機構名、其他專有名詞和新名詞進行綜合排序，前 20 項高頻詞彙如表 3 所示。

表 3

名詞高頻詞綜合排序表

序列	詞彙	序列	詞彙
1	中國	11	印度
2	美國	12	上海
3	蘇聯	13	聊齋
4	日本	14	英國
5	毛澤東	15	北京
6	朝鮮	16	德國
7	俄羅斯	17	趙君邁
8	蔣介石	18	毛岸英
9	埃及	19	衢州
10	斯大林	20	蒲文珊

根據詞彙統計，可知國家和人物是檔案主題貼文最為核心的議題，並且名詞之間存在著較強的關聯性，如蘇聯—斯大林、中國—毛澤東—蔣介石、朝鮮—毛岸英等等，本文從「毛澤東」、「蔣介石」、「斯大林」、「趙君邁」、「毛岸英」和「蒲文珊」等人名，對照貼文內容瞭解論述主題。發現圍繞著「毛澤東」、「斯大林」、「蔣介石」幾位人物，而與之相關之事物為「中蘇（中俄）」、「國民黨（國共）」、「抗美援朝」等主題，並出現與中國互動密切之「美國」、「日本」等國家主題詞。此類主題貼文內容例如：

毛澤東致電史達林，指出中國政府成立第二天即獲得蘇聯的無條件承認，並很快即獲得各新民主國家同樣的承認。（風聞社區，2019c）

或是

出自《俄羅斯解密檔案》所載……中國出兵抗美援朝完全是自主決策，根本不存在有史達林的“逼”或是毛澤東是出自維護中蘇友好同盟互助條約而不得不出兵等因素。（風聞社區，2020b）

可知特定人物與事件的連結，往往是回顧歷史事件，並根據過往檔案史料所載，據以評論時事，呈現檔案史料憑證的功能。

（二）社會話題

在普通名詞中往往出現社會關注的話題，前述表 2 針對普通名詞進行

高頻詞排序，將包含這些高頻詞的貼文分別篩選出來，排除前述所總結的特定歷史人物和事物話題後，可以歸納出檔案貼文經常提及的社會話題，為「檔案公開」和「檔案工作實務」2 個主題。

「檔案公開」主要針對國內外檔案館或相關機構公開珍貴檔案，討論檔案展出、出版等檔案公開使用活動。此類討論例如：

通過對館藏有關九一八事變後日本侵華檔案的挖掘整理，瀋陽市檔案館首次公開 106 件珍貴館藏歷史檔案。這批檔案真實記錄了九一八事變以來，日軍在瀋陽地區實施殘暴的侵略和殖民統治的史實。（風聞社區，2019b）

或是

紀念中國人民志願軍抗美援朝出國作戰 70 周年，中央廣播電視總台《國家記憶》欄目攝製 20 集大型紀錄片《抗美援朝保家衛國》，將於 10 月 12 日起在 CCTV-4 晚 8 點檔黃金時段連續播出。（風聞社區，2020a）

「檔案工作實務」則是探討檔案工作中涉及的諸如檔案修復、檔案用紙、檔案管理等內容。爰引相關摘文為例：

你們現在看到的檔案都很平整漂亮，常引起一般人誤會說，檔案這麼好，也要裱褙嗎？其實不是，剛送來裱褙前的檔案並不是這樣的。老舊檔案很髒，原先存放在各機關陰暗庫房中，很多剛開始裱褙機關檔案的師傅要揭頁、表面除塵與字跡加固等。（風聞社區，2019a）

或是

現存的紙莎草紙卷文獻都是隨出土文物一道發現的，因此從其物質形態來說，是不可多得的文物，而從其所載內容來看，則又是彌足珍貴的歷史文獻。所以，修繕與保管工作尤其重要。過去由於沒能正確掌握修繕、保護紙莎草紙文獻的方法，致使大量紙莎草紙文獻遭到毀滅。（風聞社區，2019d）

歸納檔案貼文呈現的核心主題，可分為 2 類，一類是特定人物與事物，主要包括「毛澤東」、「蔣介石」、「斯大林」幾位人物，以及與之相關的「中蘇（中俄）」、「國民黨（國共）」話題和「抗美援朝」事件。因檔案往往與歷史密不可分，檔案內容揭露歷史事件與人物關係，自是成為輿論關注討論的核心議題。另一類則是社會話題，是社會大眾關注「檔案公開」和「檔案工作實務」的討論。人們關心哪些檔案資訊可公開，或是舉辦檔案展覽

揭示的內容，或是對於檔案工作感興趣的環節產生討論共鳴。社會大眾在網路論壇針對檔案展開的討論，不僅是重大歷史人物事件這樣的特殊題材，也包含著與日常生活和檔案工作相關的社會話題，兩者組成了網路議論檔案的核心主題與概念。

三、檔案貼文情感分析

本文將經過 NLPIR 的斷詞處理產生 20,999 個詞彙與中文情感詞彙本體庫進行比對，共找到 2,763 個具有情感的詞彙，總詞頻數達到 27,649 次，扣除 11 個兩義詞及其 98 次詞頻後，其中具有正向情感的詞彙 1,049 個，詞頻數 12,210 次；具有負向情感的詞彙 908 個，詞頻數 6,397 次；中性情感詞 795 個，詞頻數達到 8,944 次。統計正向、負向及中性等 3 種極性出現詞彙之前 10 個高頻詞及其強度，可知正向情感詞彙出現頻率高於負向詞彙，在正向高頻詞中有諸如：「重要」、「必須」、「起來」、「解放」等高強度情感詞彙，負向高頻詞有：「帝國主義」、「爆發」等高強度負向情感詞彙。

表 4

各極性高頻詞強度及詞頻次數統計表

序列	正向	強度	頻次	負向	強度	頻次	中性	強度	頻次
1	重要	7	536	侵略	3	158	作為	3	320
2	通過	3	348	犯罪	5	156	發展	3	248
3	革命	3	276	嚴重	3	128	回憶	3	236
4	必須	7	206	前科	3	88	實際	1	220
5	負責	5	194	洩露	5	82	教育	1	162
6	起來	7	182	帝國主義	7	80	一定	5	152
7	真實	3	182	敵人	5	72	其實	5	146
8	安全	3	170	爆發	7	70	不過	5	130
9	解放	7	144	人事	5	64	解決	1	120
10	犧牲	5	138	手段	1	60	根本	1	114

在 162 篇貼文中，正向情感詞彙出現的總詞數和詞頻皆比其他 2 個屬

性（負向、中性）高，但負向情感詞彙的詞頻不高，只有正向詞的一半。雖然中性詞數量少，但頻率卻不低，甚至超過負向情感詞的詞頻。在情感詞強度方面，貼文中使用的正向或負向情感詞往往都具有較高的強度，強度值達到 5 和 7，意味著貼文用語包含強烈情感，代表著極端的「好」或「不好」，對於貼文情感傾向影響更為顯著。而中性情感詞雖然詞頻不低，但強度較低，僅僅為 1 或 3，情感色彩相對寡淡，不具有正向或負向情感的影響力，對於文章情感傾向的整體影響就更低。

找出情感詞彙後，將情感詞彙回推各篇文本中出現的頻次並乘以情感分數計算各文本情感傾向。據此計算 162 篇貼文所對應的情感傾向，得出正面情感貼文共 83 筆，百分比 51.23%，而負面情感文本共有 73 筆，百分比 45.06%，中性情感貼文 6 篇，僅占 3.71%。

表 5

162 篇網路貼文情感傾向統計表

情感傾向	文本數	百分比(%)
正面	83	51.23
負面	73	45.06
中性	6	3.71
總計	162	100.00

通過對於情感詞彙使用情況的分析可發現，網民在網路論壇上編寫和發表文章時，對於各種情感詞能夠加以利用，不乏以極端強度的情感詞表達自己的觀點，以致於讓貼文內容產生具有明顯的正向或負向情感引導傾向。

四、檔案核心議題的情感分析

探討各核心議題下社會大眾表現出的情感態度，可發現：

（一）圍繞特定人物與話題而展開討論時，貼文常呈現負向情感

圍繞特定人物貼文的正向情感較弱，無論是哪位人物或哪個主題皆屬

如此，社會大眾對於這些人物與話題的討論，往往以澄清謠言、還原真實歷史故事或是批判某種行為而展開，例如：以澄清為目的的例子，有解釋毛澤東派兵發起抗美援朝並非迫於斯大林壓迫、國民黨軍血戰諾曼第的檔案是無從考證的空穴來風等內容；或是以還原歷史為目的的例子，例如：揭秘新中國與蘇聯簽訂《中蘇友好條約》背後的複雜過程、講述國共鬥爭過程、中央文庫保管檔案面臨的艱辛困難等；若以批評為目的之類型，通常是指責社會上出現對於毛澤東、蔣介石等相關檔案內容史事進行批評等。因此，社會大眾的討論不免有屬於負面情感的詞語，例如「侵略」、「敵人」、「謠言」、「不可」、「篡改」等等，貼文最終也在整體上呈現出負向情感。

（二）圍繞社會議題而開展的討論，整體情緒較為積極，情感傾向因討論主題不同而有差別

在針對各類檔案工作實務而展開的討論中，正向情感較佔上風，相關貼文的比率近 60%。社會探討有關如何建立完善檔案制度，或是檔案領域出現的新技術新變革等，在這些話題中所流露出的都是正面情緒。只有在提及檔案工作出現疏忽大意、保密失誤或肆意捏造等不當行為時，才會出現批判態度。而對於檔案公開所進行的討論中，當介紹某一檔案公開資訊時，通常屬於正面情感，但伴隨檔案公開所揭露的歷史背景介紹，不可避免地出現具有負向情感的詞彙。換言之，在一篇計算結果為負向情感的貼文，很有可能是因為作者在文章中大量介紹與之相關的灰色歷史背景，才導致貼文內容整體偏向負面情感，卻不意味著作者在反對或批評檔案公開行為。因此，對於文本情感分析之計算，無論是採用情感詞典或是機器學習分析，只能對於文本採用之詞彙加以歸類，對於推知文本作者陳述內容之傾向仍然會有差距。

伍、結論與建議

基於前述對於網路論壇檔案文本主題及情感分析結果，本文歸納結論如下：

一、運用數位工具進行文本內容分析是人與機器協作的過程，需要輔以人文研究者參與解讀和判斷

在人文研究中引入和使用數位工具，其目的是為了簡化人工操作流程，同時實現一些傳統人文研究無法完成的計算和分析過程。但這並不意味著研究者只需要將材料交給工具就可以袖手旁觀，目前對於文件格式的調整和轉換、文字符號的清洗和替代等都需要由人工來完成。因此，開展數位人文研究的人文學者需要具備數位工具之實踐操作能力，基於自身對於資料內容之判斷，修正調整資料分析過程，以回應研究目的之資料分析解釋。

二、分析網路論壇檔案主題貼文，可呈現出人物與事件之明顯核心議題，也能反映出一段時間內社會大眾比較關心討論的檔案工作問題

本文以檔案貼文進行主題觀察，發現貼文內容呈現兩個明顯的核心議題，一是以「毛澤東」、「斯大林」、「蔣介石」以及「中蘇」、「中共」和「抗美援朝」為代表的政治人物與事件議題；另一類則是以「檔案工作實務」和「檔案公開」為代表的社會議題。核心議題的出現具有一定主題性和階段性，伴隨著時間和社會環境的變化，民眾會不斷產生新的關注熱點，討論的內容也相應發生轉變。本文雖以風聞社區為研究標的，提出一種探勘論壇議題的思路和方法，但此法可運用於不同年代或地區之分析。

三、分析檔案貼文主題之情感，顯示對於時事議題偏向負面情感，是以批判角度檢視問題；對於檔案工作有較多正面情感，呈現對於檔案工作的興趣

分析檔案主題貼文，發現在針對特定人物和話題的討論，往往充斥著負向情感。但不必將這種負面情緒視為民眾對於檔案的反感或厭惡，因為有很大的程度是陳述內容的灰色屬性，更多的是一種對於歷史的惋惜和感慨，這種情緒不可避免和負面情感用語連結，雖然議論人物或是時事偏向負面，但也正是這種情感，才激發出民眾更想瞭解檔案內容的興趣，讓檔

案受到更多關注和討論。

對於檔案工作實務，社會大眾展現比較多的正面情感，對於檔案機構而言是值得欣慰的現象。此類主題貼文的內容，反映出民眾不僅僅關心檔案管理政策，對於檔案資料和檔案工作的開展，保持好奇和期許的態度。而在檔案公開主題的討論中，儘管沒有呈現明顯的情感偏向，但仍然能夠從中窺見社會整體對於各種檔案公開資訊的關注和訴求，檔案開放使用應是檔案工作未來推動的重點。

根據本次實作過程經驗，對於研究發現及不足之處，提出進一步研究建議：

- (一) 本研究以風聞社區收集研究材料，樣本量顯有不足。數位工具可以進行大量文本材料的處理和分析，未來可以考慮擴展蒐集資料來源，加入微博、Facebook 等積攢海量內容的網站，以擴大樣本範圍。
- (二) 嘗試基於深度學習的文本情感分析，本文採情感詞典比對計算文本情感，易受詞彙缺陷而影響分析效果，未來可採用基於深度學習的文本情感分析，分層次進行特徵學習方法，將文本表示為稠密、連續、低維度的向量，構建深度神經網絡模型，自動提取文本特徵並完成情感分類，降低文本特徵構建的複雜性（楊開漠、吳明芬、陳濤，2019；季立堃，2019）。近年來，已有研究將深度學習方法引入自然語言研究，並取得不錯的研究成果，未來可以考慮引入這種新的文本情感分析技術，為數位人文研究注入新的活力。

（接受日期：2022 年 12 月 13 日）

註釋

註 1：「斯大林」在臺灣常被中譯為史達林，但為保留大陸文本原來內容，仍以「斯大林」呈現統計。

參考文獻

- 丁蔚 (2017)。基於詞典和機器學習組合的情感分析 (未出版之碩士論文)。西安郵電大學, 西安市。
- 王李冬、張慧熙 (2016)。基於 HowNet 的微博文本語義檢索研究。《情報科學》, 34(9), 134-137。doi:10.13833/j.cnki.is.2016.09.027
- 石鳳貴 (2020)。基於 Jieba 中文分詞的中文文本語料預處理模塊實現。《電腦知識與技術》, 16(14), 248-251、257。doi:10.14004/j.cnki.ckt.2020.1579
- 杜振雷 (2013)。面向微博短文本的情感分析研究 (未出版之碩士論文)。北京信息科技大學, 北京市。
- 李孟潔 (2009)。利用機器學習作法之中文意見分析 (未出版之碩士論文)。清華大學, 新竹市。
- 李颯 (2006)。基於 GATE 的中文信息抽取系統的開發和實現 (未出版之碩士論文)。中國科學院研究生院文獻情報中心, 北京市。
- 季立堃 (2019)。基於深度學習的文本情感分析技術研究 (未出版之碩士論文)。北京郵電大學, 北京市。
- 洪巍、李敏 (2019)。文本情感分析方法研究綜述。《計算機工程與科學》, 41(4), 750-757。doi:10.3969/j.issn.1007-130X.2019.04.024
- 姚天昉、程希文、徐飛玉、漢思·烏思克爾特、王睿 (2008)。文本意見挖掘綜述。《中文信息學報》, 22(3), 71-80。
- 紀雪梅 (2014)。特定事件情境下中文微博用戶情感挖掘與傳播研究 (未出版之博士論文)。南開大學, 天津市。
- 風聞社區 (2019a)。孫中山手繪民生主義圖手稿曝光, 原來長這樣 ↓ ↓ [線上論壇]。檢自 <https://user.guancha.cn/main/content?id=108951&s=fwzxfbbt>
- 風聞社區 (2019b)。瀋陽首次公布記載“平頂山慘案”經過的歷史檔案 [線上論壇]。檢自 <https://user.guancha.cn/main/content?id=175233>
- 風聞社區 (2019c)。中俄檔案: 蘇聯第一個與新中國建交背後“鮮為人知”的故事…… [線上論壇]。檢自 <https://user.guancha.cn/main/content?id=181061>
- 風聞社區 (2019d)。論紙莎草紙的興衰及其歷史影響 [線上論壇]。檢自 <https://user.guancha.cn/main/content?id=189798>
- 風聞社區 (2020a)。國家記憶推出 20 集大型紀錄片《抗美援朝保家衛國》 [線上論壇]。檢自 <https://user.guancha.cn/main/content?id=394130>
- 風聞社區 (2020b)。張興德: 抗美援朝並非迫於史達林的壓力而出兵 [線上論壇]

- 壇]。檢自 <https://user.guancha.cn/main/content?id=398862>
- 徐琳宏、林鴻飛、潘宇、任惠、陳建美（2008）。情感詞彙本體的構造。《傳報學報》，27(2)，180-185。doi:10.3969/j.issn.1000-0135.2008.02.004
- 張華平、商建雲（2019）。NLPIR-Parser：大數據語義智能分析平臺。《語料庫語言學》，6(1)，87-104。
- 楊開漢、吳明芬、陳濤（2019）。廣義文本情感分析綜述。《計算機應用》，39(S2)，6-14。
- 韓雪丹（2017）。《結合 ICTCLAS 方法的 APP 用戶體驗評價方法研究》（未出版之碩士論文）。天津大學機械工程學院，天津市。
- 魏慧玲（2014）。《文本情感分析在產品評論中的應用研究》（未出版之碩士論文）。北京交通大學，北京市。
- Dwianto, R. A., Nurmandi, A., & Salahudin, S. (2021). The Sentiments analysis of Donald Trump and Jokowi's Twitters on Covid-19 policy dissemination. *Webology*, 18(1), 389-405. doi:10.14704/WEB/V18I1/WEB18096
- Li, P.-H., Fu, T.-J., & Ma, W.-Y. (2020). Why attention? Analyze BiLSTM deficiency and its remedies in the case of NER. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(5), 8236-8244. doi:10.1609/aaai.v34i05.6338
- Ma, W. Y., & Chen, K. J. (2003). Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff. In Association for Computational Linguistics (Ed.), *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*, 17 (pp. 168-171). Sapporo, Japan. doi:10.3115/1119250.1119276
- NLPIR1 (2021). NLPIR 大數據語義智能分析平臺。檢自 <http://www.nlpir.org/wordpress/2021/04/12/%e5%a4%a7%e6%95%b0%e6%8d%ae%e8%af%ad%e4%b9%89%e6%99%ba%e8%83%bd%e5%88%86%e6%9e%90%e5%b9%b3%e5%8f%b0/>
- Sabatovych, I. (2019). Do social media create revolutions? Using Twitter sentiment analysis for predicting the Maidan Revolution in Ukraine. *Global Media and Communication*, 15(3), 275-283. doi:10.1177/1742766519872780
- Yaqub, U., Chun, S. A., Atluri, V., & Vaidya, J. (2021). Analyzing social media messages of public sector organizations utilizing sentiment analysis and topic modeling. *Information Polity*, 26(4), 375-390. doi:10.3233/IP-210321



Using Digital Humanities Tools to Explore the Implementation of Topic and Sentiment Analysis in Archival Affairs Commentary

Yi-Xuan Zhang * Chiao-Min Lin **

【 Abstract 】

This paper collected forum discussions related to archives in mainland China and analyzed the topics and sentiment trends presented in archives posts in 2019-2020. This paper used the NLPiR platform to perform text preprocessing such as word segmentation, part-of-speech tagging, and word frequency statistics. Chinese sentiment vocabulary ontology is then introduced to precede the sentiment analysis. The results showed that the topics that people are most concerned about are generally divided into two types. One is the specific figures and events, the other is the social issues represented by “archival work practice” and “archives access”. In terms of sentiment analysis, the archival affairs forum showed obvious negative emotions when discussing specific people and events; and when it came to the practice of archival work, it showed positive expectations for the archival access.

* MA Student, Graduate Institute of Library, Information and Archival Studies,
National Chengchi University
E-mail: zyx19971015@163.com

** Professor, Graduate Institute of Library, Information and Archival Studies,
National Chengchi University
ORCID 0000-0002-9309-9884
Principal author for all correspondence E-mail: cmlin@nccu.edu.tw

Keywords

Archive posts, Text sentiment analysis, Social media, Archive topics, Mainland China

[Summary]

Introduction

Internet forum, spontaneously formed by netizens, is an important medium to discuss social issues and exchange information, where public cognition, attitude, and sentiment trend are collected to lead or control social discourse power as well as influence public policy making. Archives work in Mainland China is in the transformation and growth period, when the development conditions of archives and the direction of people' expectations in China could be understood through forum content analyses.

Taking 162 posts about archival issues discussed on “Rumor Community”, China, in 2019-2020 as the objects, NLPIR is applied to preprocess word segmentation, part-of-speech tagging, and word frequency statistics in the texts. Chinese sentiment vocabulary ontology is then introduced to precede the sentiment analysis to explore archival issues mostly concerned by people and the sentimental attitude.

Methodology

Internet forums are restricted to post contents with deep discussions and open names for being responsible for the study. After browsing and selection, “Rumor Community” is used as the post source. After searching and manually organizing the content, text content mining analysis is preceded. The research is divided into two parts. First, posts related archives in “Rumor Community”, China, are searched according to the set keywords. NLPIR (Natural Language Processing and Information Retrieval) is then used for text preprocessing.

Irrelevant texts are eliminated manually, and word segmentation and part-of-speech tagging are preceded. Post topics related to archival issues are organized through word frequency statistics. Second, Chinese sentiment vocabulary ontology is used for comparing processed text vocabulary to acquire the sentiment score of vocabulary, for sentiment analysis. Word frequency is combined to calculate the sentiment scores of texts to classify posts into positive, neutral, and negative types.

Results

Word Segmentation Result of Archive Posts

The 162 archive posts cover 15 categories and 43 sub-categories, with total 20,999 types of vocabulary; total word frequency achieves 157,043 times, mostly including nouns (47.207%) and verbs (34.849%). Common nouns or politicians' names appear the most in nouns, revealing that archive posts are used to use persons, affairs, and objects as the topics, and verbs are matched for word expression. The entire text word style tends to narrative, rather than argumentative contents.

Archive Post Topics

Core topics appeared in archive posts are classified into two categories. Specific figures and affairs mainly contain figures of “Mao Zedong”, “Chiang Kaishek”, and “Stalin” and the related topics of “sino-soviet (China-Russia)” and “Kuomintang (Kuomintang and Chinese communist party)”, as well as events like “resist U.S. Aggression and Aid Korea”. Since archives are closely related to history, archival contents become the core issues concerned and discussed in public opinions, due to the exposure of historical events and figures. Social topics are discussions about “archival access” and “archival work practice” concerned by the mass society. People care about archive information

which could be publicized, contents which could be shown in archives exhibition, or interesting parts of archives work.

Archive Post Sentiment

The posts include 2,763 sentimental words, where positive words appear the most (1,049 words for total 27,649 times), followed by negative and neutral vocabulary. With the combination of frequency calculation, 83 posts show positive sentiment, 73 posts are negative, and 6 posts are neutral. Since negative sentimental words in posts present higher strength, the influence on post sentiment trend is comparatively significant. It reveals that netizens could apply various sentimental words to express personal points of view and particularly prefer using high sentimental words to highlight strong attitude; the posts therefore show obvious sentiment trend.

Archive Core Issue Sentiment

People' discussions about specific figures or events often start from the perspectives of rumor clarification, truth restoration, or historical issue criticism to present remarkably negative sentiment. For archives work, they would present positive emotions of curiosity and study, and merely appear critical attitude towards negative events with major mistakes. It is worth discussing that the post authors of archival access issues do not show opposed or critical behavior on archival access. However, when the post contents mention about archival contents involving in past grey historical situations, negative vocabulary would be inevitably used to result in different sentiment calculation results from the authors' actual attitudes. Apparently, there are still differences in deriving an author's statement content tendency simply from sentiment scores.

Conclusion

The application of digital tools to text content analysis is the cooperation process between people and machine that it requires humanities researchers' participation in interpretation and judgment for constant correction and perfection.

The analysis of archive topic posts in Internet forums would present the core issues of figures and events as well as reflect the archive work focus concerned by the mass society within a period of time.

Analyses of sentiment corresponding to archive post score issues reveal that people, when facing archival affair issues, tend to negative attitude and often view social issues from critical perspectives, but appear more positive emotions on archives work to reveal the interests and expectations on archives work.

Romanized & Translated Reference for Original Text

- 丁蔚 (2017)。基於詞典和機器學習組合的情感分析 (未出版之碩士論文)。西安郵電大學, 西安市。【Ding, Wei (2017). *Sentiment analysis based on the combination of dictionary and machine learning* (Unpublished master's thesis). Xi'an University of Posts and Telecommunications, Xi'an. (in Chinese)】
- 王李冬、張慧熙 (2016)。基於 HowNet 的微博文本語義檢索研究。《情報科學》, 34(9), 134-137。【Wang, Li-Dong, & Zhang, Hui-Xi (2016). Microblog text semantic retrieval based on HowNet. *Information Science*, 34(9), 134-137. (in Chinese)】 doi:10.13833/j.cnki.is.2016.09.027
- 石鳳貴 (2020)。基於 Jieba 中文分詞的中文文本語料預處理模塊實現。《電腦知識與技術》, 16(14), 248-251、257。【Shi, Feng-Gui (2020). Ji yu Jieba zhong wen fen ci de zhong wen wen ben yu liao yu chu li mo kuai shi xian. *Computer Knowledge and Technology*, 16(14), 248-251, 257. (in Chinese)】 doi:10.14004/j.cnki.ckt.2020.1579

Using Digital Humanities Tools to Explore the Implementation of Topic and Sentiment Analysis in Archival Affairs Commentary

- 杜振雷 (2013)。面向微博短文本的情感分析研究 (未出版之碩士論文)。北京信息科技大學, 北京市。【Du, Zhen-Lei (2013). *Sentiment analysis towards microblog short text* (Unpublished master's thesis). Beijing Information Science and Technology University, Beijing. (in Chinese)】
- 李孟潔 (2009)。利用機器學習作法之中文意見分析 (未出版之碩士論文)。清華大學, 新竹市。【Lee, Meng-Chieh (2009). *Opinion analysis of Chinese text using machine learning* (Unpublished master's thesis). National Tsing Hua University, Hsinchu. (in Chinese)】
- 李颯 (2006)。基於 GATE 的中文信息抽取系統的開發和實現 (未出版之碩士論文)。中國科學院研究生院文獻情報中心, 北京市。【Li, Sa (2006). *The implementation of the Chinese Information Extraction system based on GATE* (Unpublished master's thesis). National Science Library, Chinese Academy of Sciences, Beijing. (in Chinese)】
- 季立堃 (2019)。基於深度學習的文本情感分析技術研究 (未出版之碩士論文)。北京郵電大學, 北京市。【Ji, Li-Kun (2019). *Research on text sentiment analysis technology based on deep learning* (Unpublished master's thesis). Beijing University of Posts and Telecommunication, Beijing. (in Chinese)】
- 洪巍、李敏 (2019)。文本情感分析方法研究綜述。《計算機工程與科學》, 41(4), 750-757。【Hong, Wei, & Li, Min (2019). A review: Text sentiment analysis methods. *Computer Engineering & Science*, 41(4), 750-757. (in Chinese)】
doi:10.3969/j.issn.1007-130X.2019.04.024
- 姚天昉、程希文、徐飛玉、漢思·烏思克爾特、王睿 (2008)。文本意見挖掘綜述。《中文信息學報》, 22(3), 71-80。【Yao, Tian-Fang, Cheng, Xi-Wen, Xu, Fei-Yu, Uszkoreit, Hans, & Wang, Rui (2008). A survey of opinion mining for texts. *Journal of Chinese Information Processing*, 22(3), 71-80. (in Chinese)】
- 紀雪梅 (2014)。特定事件情境下中文微博用戶情感挖掘與傳播研究 (未出版之博士論文)。南開大學, 天津市。【Ji, Xue-Mei (2014). *Users' sentiment mining and spreading among Chinese microblogs in the context of specific events* (Unpublished doctoral dissertation). Nankai University, Tianjin. (in Chinese)】
- 風聞社區 (2019a)。孫中山手繪民生主義圖手稿曝光, 原來長這樣 ↓ ↓ [線上論壇]。檢自 <https://user.guancha.cn/main/content?id=108951&s=fwzxfbb> 【Feng Wen She Qu (2019a). *Sun zhong shan shou hui min sheng zhu yi tu shou gao pu guang, yuan lai chang zhe yang* ↓ ↓ [Online forum comment]. Retrieved from

- <https://user.guancha.cn/main/content?id=108951&s=fwzxfbbt> (in Chinese)】
風聞社區 (2019b)。瀋陽首次公布記載“平頂山慘案”經過的歷史檔案〔線上論壇〕。檢自 <https://user.guancha.cn/main/content?id=175233> 【Feng Wen She Qu (2019b). *Shen yang shou ci gong bu ji zai “ping ding shan can an” jing guo de li shi dang an* [Online forum comment]. Retrieved from <https://user.guancha.cn/main/content?id=175233> (in Chinese)】
- 風聞社區 (2019c)。中俄檔案：蘇聯第一個與新中國建交背後“鮮為人知”的故事……〔線上論壇〕。檢自 <https://user.guancha.cn/main/content?id=181061> 【Feng Wen She Qu (2019c). *Zhong e dang an: Su lian di yi ge yu xin zhong guo jian jiao bei hou “xian wei ren zhi” de gu shi...* [Online forum comment]. Retrieved from <https://user.guancha.cn/main/content?id=181061> (in Chinese)】
- 風聞社區 (2019d)。論紙莎草紙的興衰及其歷史影響〔線上論壇〕。檢自 <https://user.guancha.cn/main/content?id=189798> 【Feng Wen She Qu (2019d). *Lun zhi suo cao zhi de xing shuai ji qi li shi ying xiang* [Online forum comment]. Retrieved from <https://user.guancha.cn/main/content?id=189798> (in Chinese)】
- 風聞社區 (2020a)。國家記憶推出 20 集大型紀錄片《抗美援朝保家衛國》〔線上論壇〕。檢自 <https://user.guancha.cn/main/content?id=394130> 【Feng Wen She Qu (2020a). *Guo jia ji yi tui chu 20 ji da xing ji lu pian “kang mei yuan zhao bao jia wei guo”* [Online forum comment]. Retrieved from <https://user.guancha.cn/main/content?id=394130> (in Chinese)】
- 風聞社區 (2020b)。張興德：抗美援朝並非迫於史達林的壓力而出兵〔線上論壇〕。檢自 <https://user.guancha.cn/main/content?id=398862> 【Feng Wen She Qu (2020b). *Zhang xing de: kang mei yuan zhao bing fei po yu shi da lin de ya li er chu bing* [Online forum comment]. Retrieved from <https://user.guancha.cn/main/content?id=398862> (in Chinese)】
- 徐琳宏、林鴻飛、潘宇、任惠、陳建美 (2008)。情感詞彙本體的構造。《情報學報》，27(2)，180-185。【Xu, Lin-Hong, Lin, Hong-Fei, Pan, Yu, Ren, Hui, & Chen, Jian-Mei (2008). Constructing the affective lexicon ontology. *Journal of the China Society for Scientific and Technical Information*, 27(2), 180-185. (in Chinese)】 doi:10.3969/j.issn.1000-0135.2008.02.004
- 張華平、商建雲 (2019)。NLPIR-Parser：大數據語義智能分析平臺。《語料庫語言學》，6(1)，87-104。【Zhang, Hua-Ping, & Shang, Jian-Yun (2019). NLPIR-Parser: An intelligent semantic analysis toolkit for big data. *Corpus Linguistics*, 6(1), 87-104. (in Chinese)】

Using Digital Humanities Tools to Explore the Implementation of Topic and Sentiment Analysis in Archival Affairs Commentary

- 楊開漢、吳明芬、陳濤 (2019)。廣義文本情感分析綜述。《計算機應用》，39(S2)，6-14。【Yang, Kai-Mo, Wu, Ming-Fen, & Chen, Tao (2019). Generalized text sentiment analysis review. *Journal of Computer Applications*, 39(S2), 6-14. (in Chinese)】
- 韓雪丹 (2017)。《結合 ICTCLAS 方法的 APP 用戶體驗評價方法研究》(未出版之碩士論文)。天津大學，天津市。【Han, Xue-Dan (2017). *Research on user experience evaluation of applications based on ICTCLAS method* (Unpublished master's thesis). Tianjin University, Tianjin. (in Chinese)】
- 魏慧玲 (2014)。《文本情感分析在產品評論中的應用研究》(未出版之碩士論文)。北京交通大學，北京市。【Wei, Hui-Ling (2014). *Application research of text sentiment analysis in product reviews* (Unpublished master's thesis). Beijing Jiaotong University, Beijing. (in Chinese)】 doi:10.7666/d.Y2602847
- Dwianto, R. A., Nurmandi, A., & Salahudin, S. (2021). The Sentiments analysis of Donald Trump and Jokowi's Twitters on Covid-19 policy dissemination. *Webology*, 18(1), 389-405. doi:10.14704/WEB/V18I1/WEB18096
- Li, P. H., Fu, T. J., & Ma, W. Y. (2020). Why attention? Analyze BiLSTM deficiency and its remedies in the case of NER. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(5), 8236-8244. doi:10.1609/aaai.v34i05.6338
- Ma, W. Y., & Chen, K. J. (2003). Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff. In Association for Computational Linguistics (Ed.), *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing - Volume 17* (pp. 168-171). Sapporo, Japan. doi:10.3115/1119250.1119276
- NLPIR1 (2021). NLPIR 大數據語義智能分析平臺。檢自 <http://www.nlpir.org/wordpress/2021/04/12/%e5%a4%a7%e6%95%b0%e6%8d%ae%e8%af%ad%e4%b9%89%e6%99%ba%e8%83%bd%e5%88%86%e6%9e%90%e5%b9%b3%e5%8f%b0/> 【NLPIR1 (2021). NLPIR intelligent semantic analysis toolkit for big data. Retrieved from <http://www.nlpir.org/wordpress/2021/04/12/%e5%a4%a7%e6%95%b0%e6%8d%ae%e8%af%ad%e4%b9%89%e6%99%ba%e8%83%bd%e5%88%86%e6%9e%90%e5%b9%b3%e5%8f%b0/> (in Chinese).】
- Sabatovych, I. (2019). Do social media create revolutions? Using Twitter sentiment analysis for predicting the Maidan Revolution in Ukraine. *Global Media and*

Communication, 15(3), 275-283. doi:10.1177/1742766519872780

Yaqub, U., Chun, S. A., Atluri, V., & Vaidya, J. (2021). Analyzing social media messages of public sector organizations utilizing sentiment analysis and topic modeling. *Information Polity*, 26(4), 375-390. doi:10.3233/IP-210321