

國立臺灣師範大學理學院  
資訊工程學系  
碩士論文

指導教授：侯文娟 博士

以產品圖形為基的方法分析  
手機及相機功能之相關性研究

A Product Graph-based Method for Relevance Research  
on Mobile Phones and the Camera Functions

研究生：蔡鎮宇 撰

中華民國一零九年八月

## 摘要

隨著網路的普及，店家的銷售模式及消費者的購物模式逐漸改變，許多人購物前會事先在網路論壇等平台查詢產品相關評價後才決定是否購買。網路論壇常見的討論方式為發佈一篇與產品之間比較的文章，而許多網友則會在下方留下自己偏好的產品以及一些對產品的看法。

本論文使用的資料來自於批踢踢實業坊 Mobilecomm 看板文章，實驗語料選自文章中的網友所撰寫的文章推文。研究目的主要希望針對這些文章推文進行極性分析，透過分析結果進一步對產品進行排名。儘管產品的整體排名是一項重要的指標，但現在越來越多消費者會根據相機功能選購手機，因此研究中同時會對相機功能的部分進行排名。

在計算手機品牌分數上，本研究將建立有向加權圖確立每個手機品牌之間的關係，透過定義比較句和主觀判斷句的方式，以這些句子建立每個產品本身的分數以及有向邊的分數。

最後透過正向句和負向句不同的權重計算方式得出兩種不同結論。當使用正向句分數減去負向句分數時，得出當相機功能排名靠前時，手機品牌的分數不會靠後。當使用正向句分數除以負向句分數時，目前暫時無法判斷手機品牌與相機功能的相關性。

關鍵字：手機評論、產品推薦、加權有向圖、非監督式學習、自然語言處理

# Abstract

With the popularization of the Internet, the sales model of stores and the shopping model of consumers have gradually changed. Many people will check product-related reviews on online forums and other platforms before making decision. A common discussion method on online forums is to post an article to compare products, and many people will reply some opinions about the products below the article.

This thesis downloaded the review data from PTT Mobilecomm. The experimental corpus is selected from the article tweets written by people in the article. The main purpose of the research is to analyze the sentiment of these articles and tweets, and to further rank products according to the analysis results. Although the overall ranking of products is an important measure, more and more consumers choose mobile phones based on camera functions now. The research will also rank based on the camera functions.

In calculating the scores of mobile phone's brands, this research will establish a directed weighted graph to establish the relationship between each mobile phone's brand. By defining comparative sentences and subjective sentences, these sentences are used to establish the vertex score and directed edges score of each product.

Finally, two different conclusions are drawn through the different weight calculation methods of positive and negative sentences. When the weight is calculated by the positive sentence score subtracting the negative sentence score, it is concluded that when the camera function score ranks high, the score of the mobile phone brand will not fall behind. When the weight is calculated by the positive sentence score divided by the negative sentence score, it is currently impossible to determine the correlation between the score of mobile phone brand and the score of camera functions.

Keywords : Mobile reviews, Product recommendations, Weighted directed graphs, Unsupervised learning, NLP

## 誌謝

本論文得以完成，首先要感謝指導教授侯文娟老師，感謝老師這三年來的教導，引導我進入自然語言處理的領域。在研究所期間，為我的研究方向提供很多寶貴的建議；研究過程遇到瓶頸時，提供許多方向和想法使我得到啟發與收穫；在論文寫作上，老師也仔細提出用詞不完善或是明確點出不合理之處，使本篇論文更加完整。謝謝老師耐心與細心的指導和對學生的鼓勵與支持。

最後要感謝我的家人，無論是精神上的鼓勵或是實質上的支持，這些都是能讓我無後顧之憂地完成碩士學位的原因。

# 目錄

摘要 .....	I
Abstract .....	III
附表目錄.....	VII
附圖目錄.....	VIII
第一章 緒論 .....	1
第一節 研究背景與動機 .....	1
第二節 研究目的 .....	2
第三節 論文架構 .....	3
第二章 相關文獻探討 .....	4
第一節 中文斷詞系統 .....	4
第二節 情感語意字典 .....	9
第三節 PageRank 演算法 .....	11
第四節 產品推薦相關研究 .....	13
第三章 研究方法與步驟 .....	15
第一節 緒論 .....	15
第二節 實驗資料 .....	17
第三節 文章斷詞與詞性標記 .....	21
第四節 擴充情感字典語料庫 .....	22
第五節 建立加權有向圖 .....	24
第六節 排名分數計算 .....	30
第四章 實驗結果與分析 .....	32
第一節 情感字典數值化結果 .....	32
第二節 加權有向圖結果 .....	34
第三節 手機和相機功能排名 .....	37
第五章 結論和未來發展 .....	43
參考文獻 .....	44

## 附表目錄

表 1：中研院平衡語料庫詞類標記集 .....	5
表 2：Jieba 詞性類別 .....	8
表 3：NTUSD 正向詞彙節錄 .....	9
表 4：NTUSD 負向詞彙節錄 .....	10
表 5：ANTUSD 詞彙範例 .....	11
表 6：部分 Jieba 斷詞結果為形容詞和成語 .....	23
表 7：veRank 計算的分數結果 .....	31
表 8：節錄部分自建語料庫詞彙 .....	33
表 9：手機品牌比較句和主觀判斷句的正向句子個數 .....	35
表 10：手機品牌比較句和主觀判斷句的負向句子個數 .....	36
表 11：相機功能比較句和主觀判斷句的正向句子個數 .....	37
表 12：相機功能比較句和主觀判斷句的負向句子個數 .....	37
表 13：根據公式(5)正向和負向評論句為權重分數的結果 .....	38
表 14：根據公式(5)正向和負向評論句中的極性詞個數為權重分數的結果 .....	39
表 15：根據公式(5)正向和負向評論句中的極性詞分數為權重分數的結果 .....	39
表 16：根據公式(6)正向和負向評論句為權重分數的結果 .....	40
表 17：根據公式(6)正向和負向評論句中的極性詞個數為權重分數的結果 .....	41
表 18：根據公式(6)正向和負向評論句中的極性詞分數為權重分數的結果： .....	41



## 附圖目錄

圖 1：PageRank 示意圖(一).....	12
圖 2：PageRank 示意圖(二).....	12
圖 3：研究方法流程圖 .....	16
圖 4：節錄部分批踢踢實業坊 Mobilecomm 看板文章.....	18
圖 5：批踢踢實業坊文章範例.....	19
圖 6：文章標題含少於 2 個手機品牌且不含相機功能示例.....	20
圖 7：文章標題含少於 2 個手機品牌但含相機功能示例.....	20
圖 8：文章標題含大於等於 2 個手機品牌但不含相機功能示例.....	20
圖 9：文章標題同時含大於等於 2 個手機品牌和相機功能示例.....	20
圖 10：文章推文範例 .....	21
圖 11：文章推文範例斷詞後結果.....	21
圖 12：有向加權圖範例 .....	29
圖 13：部分加權有向示意圖.....	35

# 第一章 緒論

## 第一節 研究背景與動機

隨著網路的普及使現在取得資訊便利，過去消費者想獲取產品資訊需要透過報章雜誌、電視的購物頻道、或是直接進實體店面聽店員介紹。而現在只要接上網路，任何人都能接收到想要的資訊。一些網路商家透過消費者根據其自身需求(如價格、製造商和其他屬性)選擇產品，為消費者盡可能提供全面的購物體驗，並且允許消費者提供關於產品的相關評論。近年來在各大平台皆有手機相關的討論區，像是 PTT<sup>1</sup>、Mobile01<sup>2</sup>、SOGI 手機王<sup>3</sup>、ePrice<sup>4</sup>、巴哈姆特<sup>5</sup>等皆有提供討論區供消費者進行資訊交流與分享相關評論。一般而言，大多數的消費者認為網路上他人所提供的產品相關評論比賣方提供的描述更具準確性，消費者有可能與撰寫評論之人有類似需求，這些評論對潛在買家具有重大的影響。但是，隨著可用的消費者評論數量的增加，單一顧客幾乎不可能閱讀完所有的評論以做出明智的決定。因此，如何從這些評論中找出有用的資訊是一個重要且具挑戰性的問題。

---

<sup>1</sup> PTT <https://term.ptt.cc>

<sup>2</sup> Mobile01 <https://www.mobile01.com/>

<sup>3</sup> SOGI 手機王 <https://www.sogi.com.tw/>

<sup>4</sup> ePrice <https://www.eprice.com.tw/>

<sup>5</sup> 巴哈姆特 <https://www.gamer.com.tw/>

自然語言是人類在日常生活為了溝通自然發展出一套具有結構性的語言，而自然語言處理則是讓電腦理解這個語言的技術。不同國籍、不同文化背景，語言也會有所不同。以中文為例，必須先將句子透過斷詞處理，才能切割詞的單位，以便電腦後續處理，進而理解詞的意思。過去是透過人工的方式訂定規則，現在則可透過機器學習的方式讓電腦以自動化的方式歸納語言的特性。

## 第二節 研究目的

近年來對評論中的句子做情感分析的研究與日俱增，但根據客戶的評論自動對產品進行排名的研究相對較少。本研究針對網路上手機評論進行分析，進一步的為手機品牌做排名。

儘管整體排名是一項重要的指標，但近年來的消費者可能會依據相機功能選購手機，而非僅僅依據整體排名做選擇。因此除了手機的整體排名之外，本研究同時也依據各手機的相機功能進行排名。由於各家廠牌手機的設計風格有所不同，消費者在考量品牌時，除了相機功能以外，或許不會考量其他功能的好壞。例如手機整體排名靠前，但相機功能表現不佳，如此可能會影響消費者的選擇。

本實驗進行方式如下，實驗中的每則評論是由文章推文中所組成，針對每篇評論先利用自然語言處理技術，對評論中的句子進行極性分析並給予分數。再將這些評論依手機和相機功能進行分類，分別建立有極性分數和沒有分數的加權有向圖。權重計算的方式又分為兩種：分為(1)正向分數與負向分數的相減，及(2)正向分數與負向分數的相除。如此一來，消費者在選擇手機廠牌時，可以根據手機排名以及相機功能排名的圖表，做為選購的參考依據。

### 第三節 論文架構

本論文的組織，第一章為緒論，說明研究背景與動機、研究目的；第二章為相關文獻探討，闡述與本研究相關的文獻或方法；第三章為研究方法與步驟，介紹本論文所提出的系統架構及流程；第四章為實驗結果與分析；最後的章節則將說明本論文的結論與未來發展。

## 第二章 相關文獻探討

### 第一節 中文斷詞系統

無論是哪種語言的處理系統，都必須先將詞抽取出來才能進行分析，因為詞是有意義且最小的語言單位。但不同於英文語料，中文語料的詞與詞之間並無空白做分隔，因此將中文文本『斷詞』，擷取出句子中正確的詞組並進行詞性的標注，便是現在語言處理的基本目標。

#### (一) 中研院中文斷詞系統(CKIP)

由於中文詞即是一個開放集合，沒有任何的辭典和方法可以列出所有的中文詞彙，所以在中文斷詞中最大的困難就是碰到未知詞。一般而言，一篇文章中大約含有 3%~5% 的未知詞，而一些類型的詞結構相當複雜，且不一定有強烈的統計特性。因此在中文語料的處理上，未知詞的擷取相當的困難。

針對以上問題，最有效的解決方式是加強詞彙的搜集，CKIP<sup>6</sup>提供一個具有新詞辨識能力並附加詞類標記選擇性功能的線上即時分詞系統。除了基本的詞彙

---

<sup>6</sup> 中央研究院中文斷詞系統 <http://ckipsvr.iis.sinica.edu.tw/>

庫以外，使用者可依自身需求添加專屬領域的詞庫。詞類所代表意義由如表 1 所

示。

表 1：中研院平衡語料庫詞類標記集

精簡詞類	簡化標記	對應的CKIP詞類標記 <sup>1</sup>	
A	A	A	/*非謂形容詞*/
C	Caa	Caa	/*對等連接詞，如：和、跟*/
POST	Cab	Cab	/*連接詞，如：等等*/
POST	Cba	Cbab	/*連接詞，如：的話*/
C	Cbb	Cbaa, Cbba, Cbbb, Cbca, Cbcb	/*關聯連接詞*/
ADV	Da	Daa	/*數量副詞*/
ADV	Dfa	Dfa	/*動詞前程度副詞*/
ADV	Dfb	Dfb	/*動詞後程度副詞*/
ASP	Di	Di	/*時態標記*/
ADV	Dk	Dk	/*句副詞*/
ADV	D	Dab, Dbaa, Dbab, Dbb, Dbc, Dc, Dd, Dg, Dh, Dj	/*副詞*/
N	Na	Naa, Nab, Nac, Nad, Naea, Naeb	/*普通名詞*/
N	Nb	Nba, Nbc	/*專有名稱*/
N	Nc	Nca, Ncb, Ncc, Nce	/*地方詞*/
N	Ncd	Ncda, Ncdb	/*位置詞*/
N	Nd	Ndaa, Ndab, Ndc, Ndd	/*時間詞*/
DET	Neu	Neu	/*數詞定詞*/
DET	Nes	Nes	/*特指定詞*/
DET	Nep	Nep	/*指代定詞*/
DET	Neqa	Neqa	/*數量定詞*/
POST	Neqb	Neqb	/*後置數量定詞*/
M	Nf	Nfa, Nfb, Nfc, Nfd, Nfe, Nfg, Nfh, Nfi	/*量詞*/
POST	Ng	Ng	/*後置詞*/
N	Nh	Nhaa, Nhab, Nhac, Nhbc, Nhc	/*代名詞*/
Nv	Nv	Nv1, Nv2, Nv3, Nv4	/*名物化動詞*/
T	I	I	/*感嘆詞*/

P	P	P*	/*介詞*/
T	T	Ta, Tb, Tc, Td	/*語助詞*/
Vi	VA	VA11,12,13,VA3,VA4	/*動作不及物動詞*/
Vt	VAC	VA2	/*動作使動動詞*/
Vi	VB	VB11,12,VB2	/*動作類及物動詞*/
Vt	VC	VC2, VC31,32,33	/*動作及物動詞*/
Vt	VCL	VC1	/*動作接地方賓語動詞*/
Vt	VD	VD1, VD2	/*雙賓動詞*/
Vt	VE	VE11, VE12, VE2	/*動作句賓動詞*/
Vt	VF	VF1, VF2	/*動作謂賓動詞*/
Vt	VG	VG1, VG2	/*分類動詞*/
Vi	VH	VH11,12,13,14,15,17,VH21	/*狀態不及物動詞*/
Vt	VHC	VH16, VH22	/*狀態使動動詞*/
Vi	VI	VI1,2,3	/*狀態類及物動詞*/
Vt	VJ	VJ1,2,3	/*狀態及物動詞*/
Vt	VK	VK1,2	/*狀態句賓動詞*/
Vt	VL	VL1,2,3,4	/*狀態謂賓動詞*/
Vt	V_2	V_2	/*有*/
T	DE	/*的, 之, 得, 地*/	
Vt	SHI	/*是*/	
FW	FW	/*外文標記*/	
COLONCATEGORY			/* 冒號 */

<sup>1</sup> 斜體詞類，表示在技術報告#93-05 中沒有定義，即後來增列的。

COMMACATEGORY	/* 逗號 */
DASHCATEGORY	/* 破折號 */
ETCCATEGORY	/* 刪節號 */
EXCLAMATIONCATEGORY	/* 驚嘆號 */
PARENTHESISCATEGORY	/* 括弧 */
PAUSECATEGORY	/* 頓號 */
PERIODCATEGORY	/* 句號 */
QUESTIONCATEGORY	/* 問號 */
SEMICOLONCATEGORY	/* 分號 */
SPCHANGECATEGORY	/* 雙直線 */

## (二) 結巴斷詞系統(Jieba)

Jieba 斷詞系統<sup>7</sup>最初是由中國百度的一位開發者所寫，是以 python 進行開發的開源程式碼，由於是中國人所撰寫的程式，所以核心是簡體中文。但因為它是一個開放原始碼的 project，任何人都可以進行修改，目前也可以針對繁體中文進行斷詞，並且支援多種程式語言。

Jieba 的演算法是利用語料庫建立 Trie 字典樹，Trie 字典樹是由詞組和機率所構成，可以增加字詞比對的速度。Jieba 斷詞系統先將句子所有可能構成詞的情況形成一個有向無環圖，利用動態規劃的方式尋找機率最大的路徑，並以此作為斷詞的依據。若是出現不在語料庫中的詞組，則將這些詞組載入 HMM 模型，使用 Viterbi 演算法進行斷詞處理。Jieba 詞性類別如表 2 所示。

---

<sup>7</sup> Jieba 中文斷詞系統 <https://github.com/fxsjy/jieba>



表 2：Jieba 詞性類別

詞性代碼	詞性名稱	詞性代碼	詞性名稱
a	形容詞	o	擬聲詞
ag	形語素	p	介詞
ad	副形詞	q	量詞
an	名形詞	r	代詞
b	區別詞	rg *	茲
c	連接詞	rr *	多數代詞
d	副詞	rz *	這位
dg	副語素	s	處所詞 (方位名詞)
df *	能願動詞: 不要	t	時間詞
e	嘆詞	tg	時語素
eng *	外語	u	助詞
f	方位詞(趨向動詞)	ud *	結構助詞: 得
g	語素	uj *	結構助詞: 的
h	前接成分	uv *	結構助詞: 地
i	成語	ug *	動態助詞: 過
j	簡稱略語	ul *	動態助詞: 了
k	後接成分	uz *	動態助詞: 著
l	慣用語	v	動詞
m	數詞	vg	動語素
mg *	干支	vd	副動詞
mq *	指示代詞	vi *	不及物動詞(內動詞)
n	名詞	vn	名動詞
ng	名語素	vq *	完成動詞
nr	人名	w	標點符號
nrfg *	完整人名	x	非語素字
nrt *	外國名詞	y	語氣詞(語氣助詞)
ns	地名	z	狀態詞
nt	機構團體	zg *	副狀態詞
nz	其他專名		

## 第二節 情感語意字典

### (一) NTUSD

在中文自然語言處理中，NTUSD(National Taiwan University Semantic Dictionary) (Ku & Chen, 2007) 是由台灣大學建立的情緒詞字典，此字典包括詞彙及對應的極性，並整理成兩個檔案。檔案分別為包含正向極性的 NTUSD positive、負向極性的則為 NTUSD negative，前者記錄所有的正向極性詞彙；後者則是記錄所有負向極性詞彙。此字典紀錄了 2810 個正向詞彙和 8276 個負向詞彙。表 3 節錄部分正向詞彙；表 4 節錄部分負向詞彙。

表 3：NTUSD 正向詞彙節錄

大膽	勻稱的	可信任的	全神貫注的
小心	天真的	可敬重的	合乎邏輯的
才能	主動的	未被污染	好追根究底
中肯	出色的	生氣勃勃	有才幹的人
天才	可忍受	光芒四射	有益於健康

表 4：NTUSD 負向詞彙節錄

上鉤	大膽的	極愚蠢的	詭計多端的
乞討	不公平	煙霧瀰漫	道德不好的
口渴	不友好	煞風景的	過分單純化
大雨	使憂傷	禁止使用	漠不關心的
大錯	極苦悶	罪孽深重	漫不經心的

## (二) ANTUSD

ANTUSD (Augmented NTU Sentiment Dictionary) (S.-M. Wang and L.-W. Ku, 2016) 為上述 NTUSD 字典的進階版，總共收集了 26021 個詞彙，目前詞彙仍隨著研究持續增加中。對於每個詞皆標注了六個欄位，分別是：CopeOpi 數值情感分數、正向標記數、中立標記數、負向標記數、非意見詞標記數、非詞標記數。本實驗將使用具有正向標記數和負向標記數的詞，並以這些詞的 CopeOpi 數值情感分數進行研究分析。表 5 節錄部分 ANTUSD 詞彙。

表 5：ANTUSD 詞彙範例

詞	CopeOpi數值情感分數	正向標記數	中立標記數	負向標記數	非意見詞標記數	非詞標記數
一下子爆發	-0.38874592	0	0	1	0	0
一下子爆發的一連串	-0.208389089	0	0	1	0	0
一夕成名	0	1	0	0	0	0
一大打擊	-0.4059861	0	0	1	0	0
一大步	0.048718	1	0	0	0	0
一大指標	0.02775835	1	0	0	0	0
一大突破	0.2054757	1	0	0	0	0
一大問題	-0.06945255	0	0	1	0	0
一大勝利	0.32151185	1	0	0	0	0
一大障礙	-0.04172545	0	0	1	0	0
一己之力	0.15985245	1	0	0	0	0
一分打點	0.287740425	1	0	0	0	0
一反	-0.4540008	0	0	1	0	0
一反前態	-0.2520004	0	0	1	0	0
一夫當關	0.224588275	1	0	0	0	0
一巴掌	-0.221000867	0	0	1	0	0
一心想	0.293363867	1	0	0	0	0
一手造成	-0.174794575	0	0	1	0	0
一文不值	-0.1126813	0	0	1	0	0
一日千里	0.04946985	1	0	0	0	0
一片譁然	-0.00426785	0	0	1	0	0
一再	0	0	0	1	0	0
一再	0	1	0	0	0	0
一再出現	0.02988815	1	0	0	0	0
一再叮囑	-0.0085089	0	0	1	0	0

### 第三節 PageRank 演算法

PageRank (L. Page, S. Brin, R. Motwani, and T. Winograd) (1999) 是一種透過客觀的方式對網頁進行排名的方法，可以有效的衡量使用者對網頁的興趣和關注度。每個頁面的 PageRank 是以所有連向它的頁面的重要性來做決定，一個有較多連入的頁面會有較高的 PageRank，相反的若是一個頁面完全沒有任何連入的頁面，則將不會產生 PageRank。

假設有四個網頁，B、C、D 均只有一個連結連入 A，示意圖如圖 1，則 A 的 PR (PageRank) 則會是 B、C、D 的 PR 總和。如公式(1)：

$$PR(A) = PR(B) + PR(C) + PR(D) \quad (1)$$

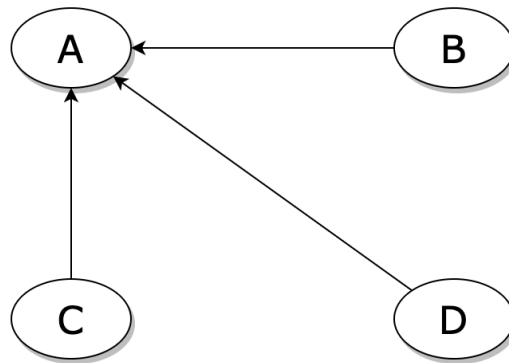


圖 1：PageRank 示意圖(一)

繼續假設 B 也有連結到 C，並且 D 也有連結到包括 A 的 3 個頁面，此時 B 的對外連結數為 2，C 的對外連結數為 1，D 的對外連結數為 3。示意圖如圖 2，則 A 的 PR 如公式(2)：

$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3} \quad (2)$$

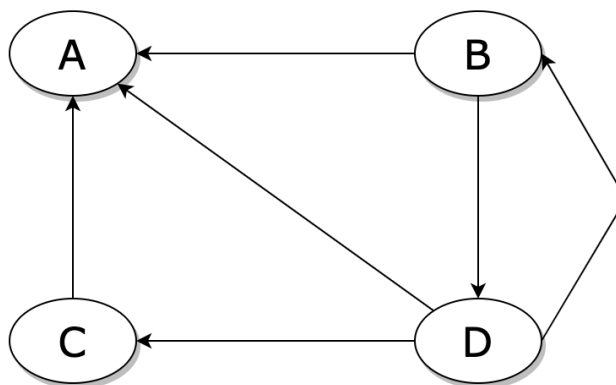


圖 2：PageRank 示意圖(二)

避免出現 PR 為 0 的頁面，PageRank 演算法引入阻尼係數  $d$ ，目的是為了解決有些網頁沒有其他頁面連入的問題。通常將  $d$  設為 0.85，表示使用者在點選網頁時，有 85% 的機率會透過連結進到下一個頁面，有 15% 的機率會隨機進入沒有外部頁面連入的網頁。改良後的演算法如公式(3)，其中  $L(P_j)$  為  $P_j$  所有的外部連結個數， $M_i$  為所有指向網頁  $P_i$  的網頁所形成的集合， $N$  代表所有網頁的個數：

$$PR(P_i) = d * \sum_{P_j \in M_i} \frac{PR(P_j)}{L(P_j)} + \frac{(1-d)}{N} \quad (3)$$

#### 第四節 產品推薦相關研究

Decker 和 Trusov (2010) 提出一個方法，統整網路上的消費者評論，這個方法可以用來推斷消費者對產品的滿意程度是否會因為品牌或是產品功能而有所不同。

Miao (2013) 等人發現，消費者評論可以使一些潛在的消費者購買未使用過的產品品牌，Miao 等人觀察到，消費者在發佈評論時可能會涉及不同的面向，因此他們在不同面向中設定關鍵詞，使關鍵詞代表各種不同面向。透過獲取各面向之間的評論，藉此分類消費者可能會對哪些產品面向感興趣。

Netzer 使用 CRF(Conditional Random Fields)( John D. Lafferty, Andrew McCallum, Fernando C. N. Pereira, 2001) 方法針對網路評論獲取產品名稱及產品功能，透過這個方法比較兩個不同產品之間的優劣。

Li 等人(2011)提出圖論的方法比較產品之間的好壞，方法是透過網路評論以及論壇上問與答的方式取得比較句子，接著分析這些比較句定義圖中每個產品的權重。

Zhang 等人(2013)使用網路中的比較句子建立產品之間的網路比較圖，這些圖包含單邊連通圖(single-link graph)、雙向連通圖(dichotomic-link graph)、以及多邊連通圖(multi-link graph)，此外也使用不同迴歸模型分析影響產品排名的因素。

### 第三章 研究方法與步驟

#### 第一節 緒論

本論文研究目的如下：從手機相關討論區評論中，找出情緒詞彙，並對文章推文進行情緒分析，最後根據情緒傾向計算分數，對手機品牌進行排名，並分析手機及相機功能排名相關性。

第一步驟先對句子進行前處理，將不必要的推文者名稱去除，並將相同的連續推文者的文章推文視為一篇文章推文。

第二步驟將句子做斷詞處理，由於具有情緒意義的詞彙大多由形容詞組成，中研院中文斷詞系統會將大多數的形容詞斷為動詞，經分析後發現中研院中文斷詞系統中的動詞中有許多詞彙並不具有情緒意義，故本研究以 Jieba 斷詞系統進行斷詞任務。本研究使用 Jieba 斷詞系統擷取具有情緒意義的詞彙，並與 ANTUSD 的情緒詞字典比對，若不在 ANTUSD 中則根據規則計算詞的極性與分數後自建詞彙庫進行擴充。

第三步驟利用所選取的產品功能和句子的極性建立有向加權圖。



第四步驟則是利用有向加權圖計算出基於產品功能的排名和品牌整體性的

排名。流程圖如圖 3 所示。

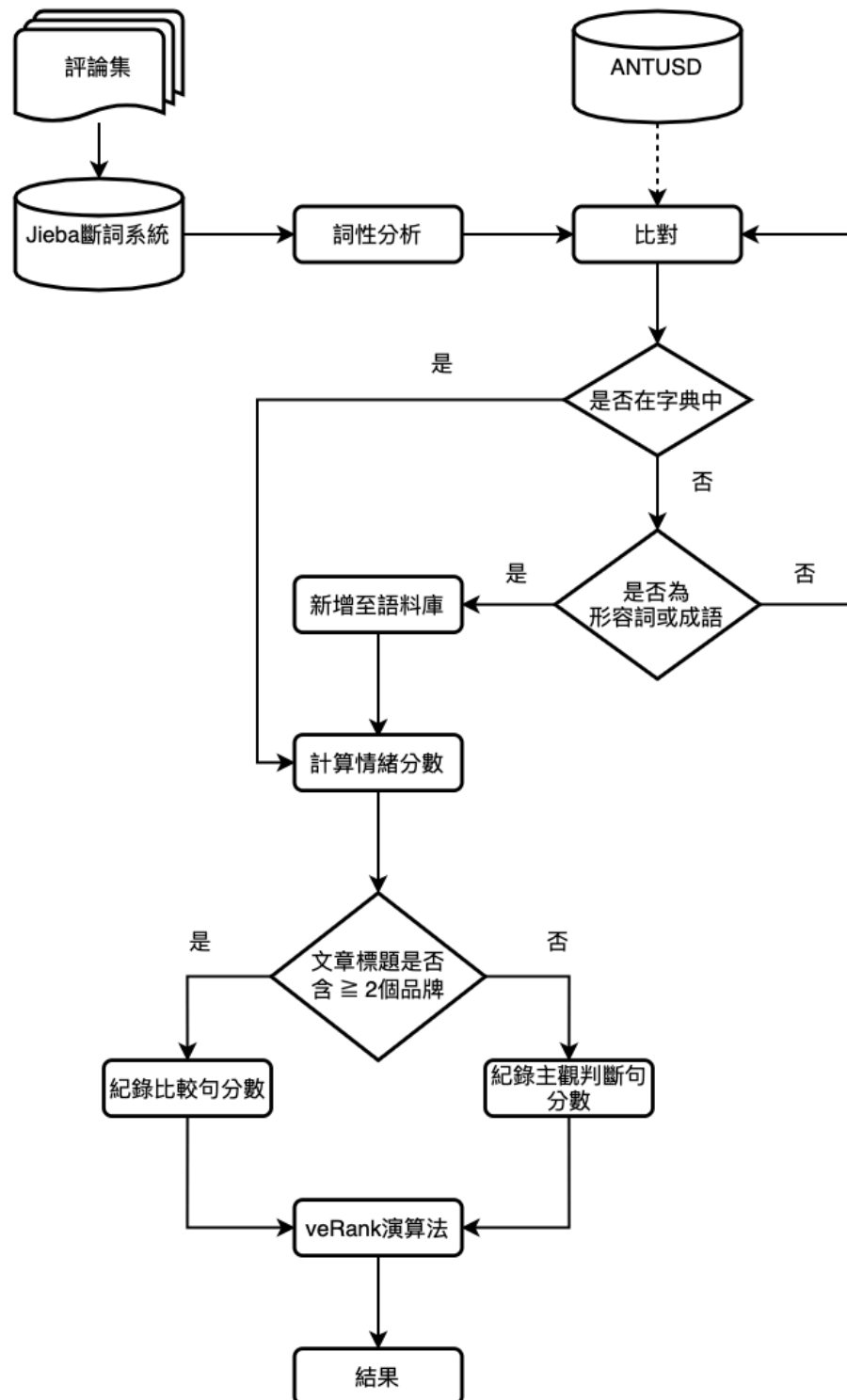


圖 3：研究方法流程圖

## 第二節 實驗資料

本論文實驗資料是選用各大知名品牌手機作為實驗對象，資料取自於批踢踢實業坊<sup>8</sup>。批踢踢實業坊是一個台灣電子佈告欄，大部分的系統原始碼由台灣大學資訊工程學系之學生和校友進行維護，並邀請專業法律人士擔任顧問。批踢踢實業坊擁有超過 2 萬個不同主題的看板，每日超過 2 萬篇新文章及 50 萬則推文被發表，是台灣使用人次最多的網路論壇之一。

實驗語料選自批踢踢實業坊 Mobilecomm 看板文章中，選擇分類為[問機]和[問題]且文章標題與手機相關之文章推文，根據本研究發現，分類為[問機]的文章標題皆與手機相關，但分類為[問題]的文章標題並不是全部皆與手機相關，因此需另外篩選與手機相關之文章。依據手機品牌在文章推文出現次數由高至低，選取排行前 11 名的知名廠牌手機作為主要的實驗對象，11 個選用廠牌手機分別是 ASUS、APPLE、GOOGLE、HTC、HUIWEI、LG、NOKIA、OPPO、SAMSUNG、SONY、小米。圖 4 節錄部分批踢踢實業坊 Mobilecomm 看板文章，其中紅色框的部分為選用的實驗文章，這些文章的分類皆為[問機]或[問題]且文章標題與手機相關，以圖 4 第二個紅色框為例：「[問題] 適合熱點分享的手機?」，由於分類為

---

<sup>8</sup> 批踢踢實業坊 <https://term.ptt.cc>

[問題]且是詢問手機，因此將選用此篇文章的推文作為實驗語料。若分類為[問題]，但文章標題與手機並無關聯，例如圖 4 的第三篇文章標題：「[問題] 收不到 LINE 通知」，這篇文章標題與通訊軟體相關，但卻與手機無關，因此這篇文章的推文將不會被選用作為實驗語料。其他的分類，例如[討論]、[新聞]、[心得]、[閒聊]、[公告]等分類的文章皆不會被選用。

【板主:Honorguard06/RG5678】【通訊】方案分類新上路,看板規					看板《MobileComm》
[←]離開 [→]閱讀 [Ctrl-P]發表文章 [d]刪除 [z]精華區 [i]看板資訊/設定 [h]說明					
編號	日期	作者	文章標題	人氣:835	
129924	+ 6 8/09	imhorny	<input type="checkbox"/> [問題] 純遊戲大螢幕手機請益		
129925	+ 7 8/09	Howard1420	<input type="checkbox"/> [閒聊] Nearby share		
129926	+14 8/09	g0935375046	<input type="checkbox"/> [問題] 收不到Line通知		
129927	+10 8/09	olmtw	<input type="checkbox"/> [新聞] 小米的 MIX Alpha 概念機最終還是放棄量		
129928	+ 2 8/09	olmtw	<input type="checkbox"/> [新聞] 傳iPhone 12搭128GB登場? 爆料客曝蘋果		
129929	+ 6 8/09	nocturne	R: [問題] 要綁5G還是4G?		
129930	+ 7 8/09	longkiss5566	<input type="checkbox"/> [問題] 5G速度比4G慢是正常的麻		
129931	+ 1 8/09	kukutoto5007	R: [心得] 無框行動 3個月試用 開箱		
129932	+ 2 8/09	three88720	<input type="checkbox"/> [問題] realme的系統有辦法在程式集建立資料夾嗎		
129933	+ 5 8/09	pupuliao	R: [閒聊] 5G上路後 原4G用戶有感覺網路變慢嗎?		
129934	+ 4 8/09	solumate	<input type="checkbox"/> [問題] 適合熱點分享的手機?		
129935	+ 6 8/09	olmtw	<input type="checkbox"/> [新聞] Google 似乎準備在 Android 上「好心提		
129936	+ 8/09	winwinwin777	<input type="checkbox"/> [問題] 中華續約吃到飽問題		
129937	8/09	honoYang	<input type="checkbox"/> [問題] 6~8k 長輩機 螢幕取向為主		
129938	+ 8/09	yougenmi	<input type="checkbox"/> [問題] HTC D10i 觸控失靈問題		
129939	+ 5 8/09	elynn889	<input type="checkbox"/> [心得] Realme X3 原廠、GCAM白天測試		
129940	+ 8/09	jjjbig119	R: [問題] 適合熱點分享的手機?		
129941	+ 8/09	weber18767	<input type="checkbox"/> [討論] 對於基地台的疑慮 不同國家人民有哪些迷		
★ !	6/05	RG5678	<input type="checkbox"/> [公告] MobileComm板規 20200605版		
★ M10	8/03	RG5678	<input type="checkbox"/> [公告] 置底閒聊區 2.0		
文章選讀 (y)回應(X)推文(^X)轉錄 (=[]<>)相關主題(/?a)找標題/作者 (b)進板畫面					

圖 4：節錄部分批踢踢實業坊 Mobilecomm 看板文章

由於各篇文章的討論度不同，為了增加可信度，僅針對推文數超過 50 則以上的文章進行搜集，並且僅搜尋 2016 年至 2019 年的文章推文。這些推文將為本實驗的語料庫。在批踢踢實業坊中，推文有字數上的限制，一行之中最多僅能有 80 個半型字元，因此根據每個帳號長度不同，推文中不同帳號可以顯示的字數也不盡相同。本實驗中，一個帳號的連續推文代表一筆語料，當上下行的推文帳號不同時，新帳號的推文則為一筆新的語料。圖 5 為批踢踢實業坊文章範例，語料庫

會放入的語料為圖 5 下方中黃色句子的文章推文部分，文章推文中含有 9 行，但由於第 7 行至第 8 行為同一個帳號連續推文，因此視為一筆語料，圖 5 中共含 8 筆語料。



圖 5：批踢踢實業坊文章範例

實驗的語料庫分為四大類：1.文章標題含少於 2 個手機品牌且不含相機功能，2.文章標題含少於 2 個手機品牌但含相機功能，3.文章標題含大於等於 2 個手機品牌但不含相機功能，4.文章標題同時含大於等於 2 個手機品牌和相機功能。本文所使用與相機的相關詞包括：「相機」、「相片」、「拍照」、「拍攝」、「照

片」、「照相」、「實拍」、「夜拍」、「攝影」、「錄影」、「鏡頭」。當文

章標題含有相機相關詞時將被分類到含有相機功能。圖6至圖9為分類類型示例。

[問機] 12000-15000元的孝親機
[問機] 25K上下 希望三年內可以順用的手機
[問機] U11瀕死，換機
[問題] 麻煩推薦手機，沒有品牌及預算限制

圖6：文章標題含少於2個手機品牌且不含相機功能示例

[問機] 求推薦相機比U11好的手機，不要中國品牌
[問題] 夜拍強的ANDROID手機
[問題] 請推薦照相為主，全新空機預算10K
[問題] 自拍跟 HTC A9 一樣的手機推薦

圖7：文章標題含少於2個手機品牌但含相機功能示例

[問機] 6S換機考慮ZF6/NOTE10/S10/IPHONE 11PRO
[問機] IPHONE 11 / XPERIA 5 / ZENFONE 6
[問機] IPHONE 8+和XR選擇
[問題] IPHONE XR或NOTE9孝親機選擇

圖8：文章標題含大於等於2個手機品牌但不含相機功能示例

[問題] 預算2萬 拍照好U12+還是MATE20
[問機] 訴求拍照 要換IPHONE11還是GALAXY系列？
[問題] 拍照手機選擇 U11/ZF5/R11S/A8+
[問題] 請問自拍的話U11與XA1 ULTRA哪個好？

圖9：文章標題同時含大於等於2個手機品牌和相機功能示例

### 第三節 文章斷詞與詞性標記

本研究對象為中文評論的文章，為了分析文章推文的極性，須將推文進行斷詞。中文斷詞和英文斷詞的區別在於，由於中文的句子間沒有空格，無法以英文斷詞的方式進行，因此在技術上較為困難。文章推文範例如圖 10 所示。

本研究使用 Jieba 斷詞系統將收集的語料庫做斷詞與詞性標記，Jieba 內建的字典為簡體詞庫，由於本研究的實驗語料為繁體中文，因此需使用繁體詞庫進行斷詞。為了使 Jieba 斷詞系統能更精確判斷出情緒詞，本研究也將自定義的詞庫載入到 Jieba 斷詞系統中，自定義詞庫中包含 ANTUSD 裡所有的詞、批踢踢實業坊中的慣用的負向詞「不推」。經過設定的詞庫斷詞之後，每個詞彙後面會加上所屬的詞性標記。圖 11 為圖 10 經過斷詞後加上詞性標記的結果。

萬元以下好像還是中國機的天下啊 我還是建議撐到 i6s 出來再買新手機

圖 10：文章推文範例

萬元(m) 以下(f) 好像(v) 還是(x) 中國(x) 機(ng) 的(uj) 天下(s) 啊(zg) (x)  
我(r) 還是(x) 建議(x) i6s(eng) 撐到(x) 出來(x) 再(d) 買(v) 新手機(x)

圖 11：文章推文範例斷詞後結果

## 第四節 擴充情感字典語料庫

本研究以 ANTUSD 情緒字典為基礎建立擴充情感字典語料庫，首先將 ANTUSD 正向和負向極性的詞取出，加入語料庫中。另一部分，文章推文經由 Jieba 斷詞系統進行斷詞之後，可以瞭解每個字彙所屬的詞性，常被標記的詞性有動詞、名詞、形容詞、副詞、成語、助詞等。其中以形容詞和成語在中文句子中較常帶有情緒意義，因此在斷詞後將形容詞和成語取出，並根據規則計算情緒分數後加入語料庫。最後取出詞性為成語和形容詞共有 1226 個詞彙，符合規則加入語料庫的詞彙共有 918 個，其中有 263 個是 ANTUSD 字典中有存在的詞彙，655 個詞需要另外根據規則計算情緒分數，需要計算情緒分數的形容詞佔 541 個，成語佔 114 個，表 6 節錄部分形容詞和成語。

表 6：部分 Jieba 斷詞結果為形容詞和成語

形容詞	成語
厚重	寸土寸金
潮濕	不共戴天
最爛	美中不足
很滑	忠言逆耳
超慘	反唇相譏
高大	惺惺相惜
迷糊	光明正大

為了使每句推文能表示客觀的極性強度，因此需要將語料庫中沒有分數的詞加上分數。如果語料庫的詞彙本身存在 ANTUSD 裡，則直接沿用 ANTUSD 的分數。針對沒有出現在 ANTUSD 的分數則根據以下規則自動化給予分數。

情緒詞算分規則：

- (1) 並非所有的情緒詞都會被加入語料庫之中。當 Jieba 斷詞系統斷詞結果為一個字的詞將不會被加入。另外，由於 Jieba 斷詞系統有時會將少於四個字的詞誤判為成語，這類的詞也不會被加入。



(2) 根據本研究分析，當詞彙字數為兩個字時，詞彙的第二個字通常較具有情

緒意義，因此將第二個字與 ANTUSD 裡所有詞彙進行比對，將所有匹配

到的字做算術平均。

(3) 當詞彙字數大於等於三個字時，因無法判斷哪些字較具有情緒意義，因此

將所有字與 ANTUSD 裡所有詞彙進行比對，將所有匹配到的字做算術平

均。

## 第五節 建立加權有向圖

本實驗根據手機和相機功能分別建立出相對應的加權有向圖。建立圖表時，

文章標題含 2 個以上手機品牌和文章標題含 2 個以下手機品牌的文章推文需分開

進行。當文章標題含有 2 個以上品牌時，這些文章推文代表文中推文中手機品牌

和文章標題中手機品牌之間的比較句。文章標題含 2 個以下手機品牌時，文章推

文中僅代表該文章推文中的手機品牌對於該品牌的主觀偏好句。在本實驗中，僅

針對推文含一個品牌的句子進行研究，當一則推文包含超過一個品牌或沒有品牌

時將不處理。

本實驗中，以上述的比較句和主觀偏好句共會建立 2 個有向加權圖，分別為手機的有向加權圖和具有相機功能建立有向加權圖，以文章標題是否含相機功能為區分，分別做兩個有向加權圖的分數計算。這個圖可以表示出每個品牌的分數。而圖的定義如公式(4)：

$$G = (V, E) \quad (4)$$

$V$  是每個點的集合， $E$  則是每個有向邊的集合。在推文中，每一個點代表一個品牌，有向邊則是兩個品牌產品之間的比較。若產品 A 和產品 B 進行比較，則增加一條 B 到 A 的有向邊，接著要計算 A、B 兩點和邊的權重。

為了便於計算，定義有向邊和點如下： $PE(B,A)$ 代表 B 到 A 這條有向邊的正向分，同時表示 A 產品品質優於 B 產品； $NE(B,A)$ 代表這條邊的負向分，同時表示 A 產品品質劣於 B 產品； $PV(A)$ 代表 A 點的正向分， $NV(A)$ 代表 A 點的負向分，接著再為每個點和有向邊分配權重。

(1) 每個產品所代表的點之正向和負向分計算方式如下：

假設文章推文中的品牌為「A」，當情緒詞前含有否定詞時，極性將作轉換，原本的正向極性詞將轉為負向極性詞，反之亦然，原本的情緒詞分數直接加上負號。其中否定詞包含「沒有」、「不是」、「不會」、「沒」、「不」、「非」、「無」。計算該文章推文分數的方式有以下三種：

(1) 當文章推文正向極性詞多於負向極性詞時，則  $PV(A)$  分數加 1 分，反之則

$NV(A)$  加 1 分。若文章推文中沒有任何極性詞，則  $PV(A)$  加 1 分。

(2) 若文章推文正向極性詞的數量為  $n$ ，負向極性詞數量為  $m$ ，則  $PV(A)$  分數

加  $n$  分， $NV(A)$  加  $m$  分。若文章推文中沒有任何極性詞，則  $PV(A)$  分數加

1 分。

(3) 將文章推文中所有的極性詞載入語料庫的分數，將所有極性詞的分數相

加後結果為  $s$ ，若  $s$  大於 0 則  $PV(A)$  為  $s$  分，若  $s$  小於 0，則  $NV(A)$  為  $|s|$  分。

若文章推文中沒有任何極性詞，則  $PV(A)$  分數加 0.9 分，0.9 為語料庫中

正向極性詞的最高分數。

(2) 每個產品與產品之間所代表的邊之正向和負向分計算方式如下：

假設文章標題包含 2 個品牌「A」、「B」，文章推文中的品牌為「A」。當

情緒詞前含有否定詞時，極性將作轉換，原本的正向極性詞將轉為負向極性詞，

反之亦然，原本的情緒詞分數直接加上負號。其中否定詞包含「沒有」、「不是」、

「不會」、「沒」、「不」、「非」、「無」。計算該文章推文分數的方法有以下三種：

- (1) 當文章推文正向極性詞多於負向極性詞時，則  $PE(B,A)$  分數加 1 分，反之則  $NE(B,A)$  分數加 1 分。若文章推文中沒有任何極性詞，則  $PE(B,A)$  分數加 1 分。
- (2) 若文章推文正向極性詞的數量為  $n$ ，負向極性詞數量為  $m$ ，則  $PE(B,A)$  分數加  $n$  分； $NE(B,A)$  分數加  $m$  分。若文章推文中沒有任何極性詞，則  $PE(B,A)$  分數加 1 分。
- (3) 將文章推文中所有的極性詞載入語料庫的分數，將所有極性詞的分數相加後結果為  $s$ ，若  $s$  大於 0 則  $PE(B,A)$  分數給  $s$  分；若  $s$  小於 0，則  $NE(B,A)$  分數給  $|s|$  分。若文章推文中沒有任何極性詞，則  $PE(B,A)$  分數加 0.9 分，0.9 為語料庫中正向極性詞的最高分數。

假設文章標題包含 2 個品牌「B」、「C」，文章推文中的品牌為「A」。當情緒詞前含有否定詞時，極性將作轉換，原本的正向極性詞將轉為負向極性詞，反之亦然，原本的情緒詞分數直接加上負號。其中否定詞包含「沒有」、「不是」、「不會」、「沒」、「不」、「非」、「無」。計算該文章推文分數的方法有以下三種：

(1) 當文章推文正向極性詞多於負向極性詞時，則  $PE(B,A)$  分數加 1 分、 $PE(C,A)$

分數加 1 分；反之則  $NE(B,A)$  分數加 1 分、 $NE(C,A)$  分數加 1 分。若文章

推文中沒有任何極性詞，則  $PE(B,A)$  分數加 1 分、 $PE(C,A)$  分數加 1 分。

(2) 若文章推文正向極性詞的數量為  $n$ ，負向極性詞數量為  $m$ ，則  $PE(B,A)$  分

數加  $n$  分、 $PE(C,A)$  分數加  $n$  分； $NE(B,A)$  分數加  $m$  分、 $NE(C,A)$  分數加  $m$

分。若文章推文中沒有任何極性詞，則  $PE(B,A)$  分數加 1 分、 $PE(C,A)$  分數

加 1 分。

(3) 將文章推文中所有的極性詞載入語料庫的分數，將所有極性詞的分數相

加後結果為  $s$ ，若  $s$  大於 0 則  $PE(B,A)$  分數為  $s$  分、 $PE(C,A)$  分數為  $s$  分；

若  $s$  小於 0，則  $NE(B,A)$  分數為  $|s|$  分、 $NE(C,A)$  分數為  $|s|$  分。若文章推文中

沒有任何極性詞，則  $PE(B,A)$  和  $PE(C,A)$  分數各加 0.9 分，0.9 為語料庫中

正向極性詞的最高分數。

上述分數在計算文章標題不含相機功能時，若文章推文中含有相機功能，則

該分數計算在相機功能的有向加權圖上。當有向邊的計算為兩個相同手機品牌時，

則不列入計算。

假設經由上述計算結果，正向分為  $x$  分，負向分  $y$  分，則點和有向邊的權重

計算方式分為以下兩種：

(i) 根據正向和負向的分數做加減，假設最後的權重為  $W_1$ ，則結果如公式

(5)：

$$W_1 = \begin{cases} x - y & \text{if } x \geq y \\ 1 & \text{if } x < y \end{cases} \quad (5)$$

(ii) 根據正向和負向的分數做相除，假設最後的權重為  $W_2$ ，則結果如公式

(6)：

$$W_2 = \begin{cases} x / y & \text{if } y \neq 0 \\ x & \text{if } y = 0 \end{cases} \quad (6)$$

本研究製作了一個範例圖如圖 12。其中每個邊代表著比較句的分數，點為主觀偏好句的分數。以圖 12 為例： $PV(A)=3$ 、 $PV(B)=2$ 、 $PV(C)=5$ 、 $PV(D)=2$ 、 $NV(B)=7$ 、 $PE(A,B)=2$ 、 $PE(A,D)=9$ 、 $PE(B,D)=3$ 、 $PE(C,D)=4$ 、 $NE(A,D)=4$ 。當使用公式(5)時，若正向邊減負向邊小於 0，則權重等於 1。當使用公式(6)時，若負向邊分數等於 0 時，在相除前令負向邊分數等於 1，避免產生數學無意義的情形。

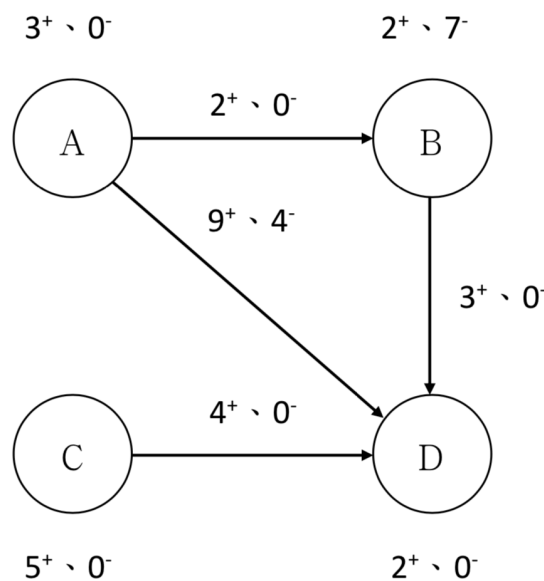


圖 12：有向加權圖範例

## 第六節 排名分數計算

在本實驗中，將使用下列 veRank 公式評估手機以及相機功能的排名。傳統的 PageRank 演算法僅平等的對待每個有向邊，並且不考慮每個點的權重。但若從某個相對重要的點指向該點，那該點應該要具有更高的重要性。在 veRank 中，不僅考慮產品之間的相對重要性，也考慮了每個產品本身的重要性。其中 veRank 如公式(7)：

$$\text{veRank}(P_i) = \left[ d * \sum_{P_j \in M_i} \text{veRank}(P_j) * E(P_j) + \frac{(1-d)}{N} \right] * V(P_i) \quad (7)$$

其中  $\text{veRank}(P_i)$  代表產品  $P_i$  的最終排名的分數； $\text{veRank}(P_j)$  代表產品  $P_j$  的最終排名的分數，而  $M_i$  為所有指向產品  $P_i$  的有向邊所形成的集合； $E(P_j)$  代表  $P_j$  指向  $P_i$  的有向邊在所有  $P_j$  指向外部節點的有向邊中所佔的比例； $V(P_i)$  代表在產品  $P_i$  該點的分數權重； $N$  代表產品的個數； $d$  為阻尼係數，與一般的 PageRank 演算法一樣設為 0.85。在計算每個品牌的 veRank 時，會先將所有品牌的 veRank 設為 1，每次計算後的結果將進行標準化，標準化的方式如公式(8)， $\text{last\_veRank}(P_i)$  代表該點計算前的 veRank 分數， $\text{new\_veRank}(P_i)$  代表標準化後

該點的 veRank 分數，V 代表品牌。將公式(7)重複執行後，每個產品的分數會趨於穩定，穩定後的分數即為每個產品的分數。表 7 為圖 11 加權有向圖權重為公式(8)重複計算至各點分數穩定後之結果。

$$new\_veRank(P_i) = \frac{last\_veRank(P_i)}{\sum_{j \in V} last\_veRank(P_j)} \quad (8)$$

表 7：veRank 計算的分數結果

點	分數
A	0.112132
B	0.064520
C	0.186886
D	0.636462



## 第四章 實驗結果與分析

### 第一節 情感字典數值化結果

經由 Jieba 斷詞系統將所有文章推文進行斷詞，經由和 ANTUSD 比對之後，發現存在 655 個形容詞或成語的詞彙不在 ANTUSD 字典中。將這些詞用第三章所提出的方法計算情緒分數，並根據情緒分數的正負判斷這些詞所代表的極性。

以「紅」為例：

雖然「紅」被斷詞 Jieba 斷詞系統判斷為形容詞，但由於這是屬於一個字的詞，根據規則，「紅」這個詞並不會被加入語料庫中。

以「忍耐力」為例：

經過 Jieba 斷詞系統斷詞後，這是一個三個字的詞，但被系統判定為成語，由於在成語大多都是四個字以上所組成，因此這個詞將被略過，不會被加入語料庫之中。

以「硬朗」為例：

這是屬於兩個字的詞，因此取第二個字與 ANTUSD 裡的詞做比對，其中 ANTUSD 裡含有「朗」的詞有：「明朗」、「明朗化」、「健朗」、「清朗」、

「爽朗」、「晴朗」、「開朗」、「樂觀開朗」、「豁然開朗」。這些詞分別對應的分數為：0.47、0.35、0、0、0.18、0.1、0.17、0.33、0.14，將這些分數相加後做算術平均 $(0.47+0.35+0+0+0.18+0.1+0.17+0.33+0.14) / 9=0.19$ 。因此硬朗的情緒分數為 0.19 分。表 8 節錄部分自建語料庫詞彙。

表 8：節錄部分自建語料庫詞彙

怕怕	-0.19341331
超貴	0.22522917
滑滑	-0.15382127
退而求其次	0.04535573
蠻慘	-0.39865991
偏弱	-0.33351662
平常	0.05179044
很貴	0.22522917
自爆	-0.16957823
浪大	0.0131666
如履薄冰	-0.08851722
如履平地	0.0265857

使用上述規則為情緒詞計算分數時分別在不同的計算情緒分數時可能會產生誤判。

(1) 成語或是三個字以上所組成的形容詞詞彙採用的計分方式是將每個字與

ANTUSD 做比對，但有些詞彙每個字的極性是正向，但組合成詞彙後便

有了和原本的字不相關的意思。例如「五十步笑百步」這個詞的每個字匹

配的的情緒分數皆是偏正向，但通常我們會將這個詞彙視為負向詞彙。

(2) 兩個字所組成的詞彙皆是採第二個字和 ANTUSD 的詞彙比對，通常兩個

字的詞彙第一個字通常帶有強度的意味，第二個字較具有情緒意義。但若

發生兩個字皆具有情緒意義，或是第一個字為否定詞時，將有可能會造成

誤判。

由於在本研究中誤判的詞彙較少，因此並無特別處理誤判的情形。

## 第二節 加權有向圖結果

在本實驗中文章推文各個手機品牌與其他產品比較次數資訊如下：APPLE 品

牌和其他產品比較 2843 次，GOOGLE 品牌和其他產品比較 499 次，HTC 品牌和

其他產品比較 5124 次，LG 品牌和其他產品比較 750 次，小米品牌和其他產品比

較 357 次，NOKIA 品牌和其他產品比較 196 次，OPPO 品牌和其他產品比較 184

次，SAMSUNG 品牌和其他產品比較 3600 次，HUIWEI 品牌和其他產品比較 647

次，SONY 品牌和其他產品比較 1646 次，ASUS 品牌和其他產品比較 2241 次。

在非比較的部分，APPLE 被提到 2139 次，GOOGLE 被提到 304 次，HTC 被提到 2394 次，LG 被提到 1087 次，小米被提到 429 次，NOKIA 被提到 143 次，OPPO 被提到 316 次，SAMSUNG 被提到 1777 次，HUIWEI 被提到 218 次，SONY 被提到 1159 次，ASUS 被提到 1445 次。圖 13 為部分加權有向圖示意圖，表 9 和表 10 為手機品牌比較句和主觀判斷句的正向句子和負向句子的個數。

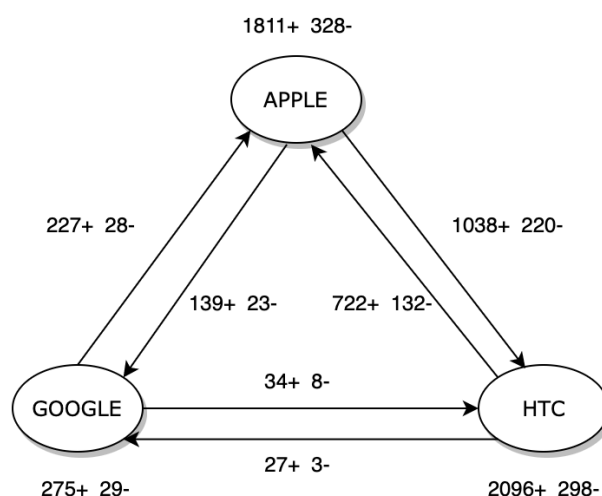


圖 13：部分加權有向示意圖

表 9：手機品牌比較句和主觀判斷句的正向句子個數

手機品牌正向分														
		APPLE	GOOGLE	HTC	LG	小米	NOKIA	OPPO	其他品牌	SAMSUNG	HUAWEI	SONY	ASUS	
比較句	APPLE	0	227	722	28	38	5	35	10	737	23	314	308	
	GOOGLE	139	0	27	6	9	2	3	1	143	21	39	46	
	HTC	1038	34	0	375	46	43	137	29	1407	93	553	533	
	LG	24	5	320	0	2	0	6	6	140	13	101	31	
	小米	14	3	51	3	0	9	32	4	19	12	2	174	
	NOKIA	3	2	74	3	2	0	25	1	7	0	45	14	
	OPPO	70	13	143	14	9	11	0	11	121	30	20	191	
	其他品牌	25	1	53	1	0	0	13	0	7	0	14	37	
	SAMSUNG	753	112	1288	144	20	5	68	33	0	47	294	274	
	HUAWEI	58	28	231	33	4	2	43	1	80	0	52	30	
	SONY	283	78	385	92	9	3	17	5	342	17	0	176	
ASUS	367	14	596	84	56	19	115	18	271	20	301	0		
主觀句		APPLE	GOOGLE	HTC	LG	小米	NOKIA	OPPO	其他品牌	SAMSUNG	HUAWEI	SONY	ASUS	
		1811	275	2096	944	379	130	265	89	1547	201	976	1273	

表 10：手機品牌比較句和主觀判斷句的負向句子個數

手機品牌負向分													
	APPLE	GOOGLE	HTC	LG	小米	NOKIA	OPPO	其他品牌	SAMSUNG	HUAWEI	SONY	ASUS	
比較句	APPLE	0	28	132	4	5	1	3	0	100	4	68	51
	GOOGLE	23	0	3	2	1	1	2	0	14	5	9	3
	HTC	220	8	0	99	10	7	32	6	257	18	104	75
	LG	2	1	46	0	0	0	3	1	30	4	8	7
	小米	4	0	10	2	0	0	3	0	3	4	1	7
	NOKIA	1	0	9	0	0	0	3	1	2	0	2	2
	OPPO	12	3	19	1	0	1	0	4	30	4	1	21
	其他品牌	7	1	14	1	0	0	0	0	2	0	6	2
	SAMSUNG	168	12	211	42	2	0	14	4	0	12	73	24
	HUAWEI	8	4	15	8	2	1	23	0	18	0	4	2
主觀句	SONY	55	16	61	24	2	1	5	1	56	2	0	16
	ASUS	78	8	132	26	8	3	23	6	40	4	52	0
	APPLE	328	29	298	143	50	13	51	17	230	17	183	172

而本實驗中文章推文各個相機功能與其他產品比較次數資訊如下：APPLE 品牌和其他產品比較 2536 次，GOOGLE 品牌和其他產品比較 472 次，HTC 品牌和其他產品比較 4432 次，LG 品牌和其他產品比較 668 次，小米品牌和其他產品比較 339 次，NOKIA 品牌和其他產品比較 183 次，OPPO 品牌和其他產品比較 655 次，SAMSUNG 品牌和其他產品比較 3152 次，HUIWEI 品牌和其他產品比較 571 次，SONY 品牌和其他產品比較 1453 次，ASUS 品牌和其他產品比較 1904 次。

在非比較的部分，APPLE 被提到 296 次，GOOGLE 被提到 159 次，HTC 被提到 639 次，LG 被提到 292 次，小米被提到 71 次，NOKIA 被提到 37 次，OPPO 被提到 162 次，SAMSUNG 被提到 546 次，HUIWEI 被提到 103 次，SONY 被提到 342 次，ASUS 被提到 339 次。表 11 和表 12 為相機功能比較句和主觀判斷句的正向句子和負向句子的個數。

表 11：相機功能比較句和主觀判斷句的正向句子個數

相機功能正向分													
	APPLE	GOOGLE	HTC	LG	小米	NOKIA	OPPO	其他品牌	SAMSUNG	HUAWEI	SONY	ASUS	
比較句	APPLE	0	227	722	28	38	5	35	10	737	23	314	308
	GOOGLE	139	0	27	6	9	2	3	1	143	21	39	46
	HTC	1038	34	0	375	46	43	137	29	1407	93	553	533
	LG	24	5	320	0	2	0	6	6	140	13	101	31
	小米	14	3	51	3	0	9	32	4	19	12	2	174
	NOKIA	3	2	74	3	2	0	25	1	7	0	45	14
	OPPO	70	13	143	14	9	11	0	11	121	30	20	191
	其他品牌	25	1	53	1	0	0	13	0	7	0	14	37
	SAMSUNG	753	112	1288	144	20	5	68	33	0	47	294	274
	HUAWEI	58	28	231	33	4	2	43	1	80	0	52	30
主觀句	SONY	283	78	385	92	9	3	17	5	342	17	0	176
	ASUS	367	14	596	84	56	19	115	18	271	20	301	0
	APPLE	GOOGLE	HTC	LG	小米	NOKIA	OPPO	其他品牌	SAMSUNG	HUAWEI	SONY	ASUS	
	240	148	544	242	58	32	143	21	470	89	293	305	

表 12：相機功能比較句和主觀判斷句的負向句子個數

相機功能負向分													
	APPLE	GOOGLE	HTC	LG	小米	NOKIA	OPPO	其他品牌	SAMSUNG	HUAWEI	SONY	ASUS	
比較句	APPLE	0	30	17	5	1	0	3	3	16	0	10	4
	GOOGLE	6	0	5	2	0	0	0	2	9	7	4	1
	HTC	19	5	0	10	4	2	5	0	44	3	29	23
	LG	0	7	3	0	0	0	0	2	1	1	6	0
	小米	2	1	3	2	0	2	2	0	2	0	0	2
	NOKIA	0	1	4	0	0	0	2	0	0	0	0	0
	OPPO	5	0	7	0	1	1	0	0	6	1	0	1
	其他品牌	0	5	3	2	0	0	0	0	1	0	1	0
	SAMSUNG	20	13	24	8	1	0	2	0	0	6	23	17
	HUAWEI	0	0	2	0	0	0	0	0	4	0	2	1
主觀句	SONY	11	6	9	1	1	0	1	0	11	1	0	5
	ASUS	1	0	17	1	1	1	5	0	10	0	7	0
	APPLE	GOOGLE	HTC	LG	小米	NOKIA	OPPO	其他品牌	SAMSUNG	HUAWEI	SONY	ASUS	
	56	11	95	50	13	5	19	2	76	14	49	34	

### 第三節 手機和相機功能排名

最後顯示出的結果為經由 veRank 公式計算之後進行標準化，使每個手機品

牌和相機功能的分數介於 0%至 100%之間。

直觀上，手機品牌的排名與相機功能的排名不應該有顯著的差異，在使用公

式(4)時，可以觀察到手機品牌與相機功能的排名並不一致。但可以發現到，手機

品牌排名的前五名和相機功能排名的前五名具有高度相關性，也就是說在相機功

能排名靠前半的產品，在手機品牌的排名並不會落到後半。公式(4)的實驗結果如

表 13 至表 15：

表 13：根據公式(5)正向和負向評論句為權重分數的結果

	手機		相機功能
HTC	34.72%	HTC	35.37%
SAMSUNG	21.89%	SAMSUNG	29.17%
APPLE	18.69%	ASUS	9.96%
ASUS	10.75%	APPLE	8.61%
SONY	6.57%	SONY	7.12%
LG	3.72%	LG	3.03%
OPPO	0.90%	GOOGLE	2.62%
MI	0.84%	OPPO	2.36%
GOOGLE	0.80%	HUAWEI	1.02%
HUAWEI	0.75%	MI	0.33%
NOKIA	0.23%	NOKIA	0.21%

表 14：根據公式(5)正向和負向評論句中的極性詞個數為權重分數的結果

	手機		相機功能
HTC	35.06%	HTC	35.76%
SAMSUNG	21.75%	SAMSUNG	29.39%
APPLE	18.57%	APPLE	8.57%
ASUS	10.63%	SONY	8.25%
SONY	6.66%	ASUS	8.16%
LG	3.69%	LG	3.24%
OPPO	0.98%	GOOGLE	2.64%
MI	0.87%	OPPO	2.03%
GOOGLE	0.73%	HUAWEI	1.10%
HUAWEI	0.71%	MI	0.36%
NOKIA	0.22%	NOKIA	0.21%

表 15：根據公式(5)正向和負向評論句中的極性詞分數為權重分數的結果

	手機		相機功能
HTC	33.73%	HTC	35.48%
SAMSUNG	22.33%	SAMSUNG	30.11%
APPLE	18.79%	ASUS	9.84%
ASUS	11.09%	APPLE	8.20%
SONY	6.64%	SONY	6.90%
LG	3.69%	GOOGLE	2.79%
OPPO	0.96%	LG	2.55%
MI	0.85%	OPPO	2.31%
HUAWEI	0.79%	HUAWEI	1.03%
GOOGLE	0.76%	MI	0.33%
NOKIA	0.23%	NOKIA	0.21%



由於公式(5)的計算結果會因為手機被討論到的次數多寡而直接影響結果，因此較常被提到的手機品牌會佔有較高的分數。經過研究可以發現，APPLE、HTC、SAMSUNG、SONY、ASUS 的比較句遠遠大於其他品牌的討論度。

由於以公式(5)的方式計算手機排名，討論度較低的手機無法得到較好的排名，但這不一定代表產品本身品質不佳，因此使用公式(6)的方式計算排名並進行分析，結果如表 16 至表 18。

表 16：根據公式(6)正向和負向評論句為權重分數的結果

	手機		相機功能
HUAWEI	13.19%	HUAWEI	13.60%
GOOGLE	10.64%	HTC	12.09%
MI	8.98%	ASUS	11.46%
NOKIA	8.81%	GOOGLE	9.93%
SAMSUNG	8.72%	SAMSUNG	9.63%
OPPO	8.65%	OPPO	8.76%
APPLE	8.12%	LG	8.09%
ASUS	8.08%	APPLE	7.62%
HTC	7.75%	SONY	5.98%
LG	6.70%	NOKIA	4.30%
SONY	6.55%	MI	4.12%

表 17：根據公式(6)正向和負向評論句中的極性詞個數為權重分數的結果

	手機		相機功能
GOOGLE	12.00%	ASUS	12.32%
MI	10.57%	OPPO	12.02%
HUAWEI	10.50%	GOOGLE	11.40%
NOKIA	9.82%	HTC	8.91%
OPPO	8.88%	SAMSUNG	8.64%
SAMSUNG	8.45%	HUAWEI	8.48%
APPLE	7.84%	LG	6.90%
ASUS	7.71%	SONY	5.97%
HTC	7.32%	APPLE	5.50%
SONY	6.77%	NOKIA	3.41%
LG	5.85%	MI	1.86%

表 18：根據公式(6)正向和負向評論句中的極性詞分數為權重分數的結果：

	手機		相機功能
HUAWEI	13.08%	GOOGLE	16.78%
OPPO	9.45%	SAMSUNG	11.67%
GOOGLE	9.30%	ASUS	10.47%
APPLE	8.63%	HUAWEI	10.04%
NOKIA	8.59%	HTC	8.55%
SAMSUNG	8.53%	OPPO	6.82%
MI	8.31%	LG	5.35%
ASUS	8.27%	APPLE	4.74%
HTC	6.89%	SONY	4.43%
LG	6.48%	MI	1.83%
SONY	5.69%	NOKIA	1.80%

使用公式(6)進行計算之後，發現某些討論度較低的手機品牌獲得了較好的排名，但手機品牌和相機功能排名極為不一致，並且幾乎沒有相關性。本研究認為是因為語料庫的選擇所導致的。由於語料庫的選擇大多為詢問推薦之手機，因此正向極性的句子會遠多於負向極性的句子，造成語料庫中負向極性的句子過少，甚至有些手機與手機之間的比較沒有出現負向句的情形。若以公式(6)的方式進行計算，將無法真正有效判斷每個產品比較句和主觀判斷句的正向句和負向句的比例，因為在論壇上幾乎不會有人詢問不推薦的手機品牌。因此在本研究中，公式(6)的做法目前無法提供有效的排名給消費者。

## 第五章 結論和未來發展

本研究探討與現代生活息息相關的手機產品，因為網路的發達，現代人幾乎離不開智慧型手機，在購物前，在網路上尋找相關評論是消費者常見的行為，但消費者很難閱讀所有評論之後做出公正的比較。因此本研究提出產品與產品的比較方式，並在每則評論具給予情緒分數，研究中除了使用 ANTUSD 字典的分數以外，也從評論資料中擷取未出現在字典裡的情緒詞並給予情緒分數。根據這些句子建立類似 PageRank 的演算法進行排名，並得出當手機品牌的討論度越高時，容易獲得較靠前的排名。但若是以正向分除以負向分的方式進行研究時，目前由於語料庫受限，仍較難以觀察出結果。

未來的發展如下：

- (1) 收集更多的語料。
- (2) 提出蒐集產品特徵的方法
- (3) 將本論文提出的方法延伸應用到不同的產品特徵。
- (4) 將本論文提出的方法延伸應用到其他的領域。
- (5) 使用監督式的方式實作評估效能。

## 參考文獻

Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293-307.

Ku, L. W., & Chen, H. H. (2007). Mining opinions from the Web: Beyond relevance retrieval. *Journal of the Association for Information Science and Technology*, 58(12), 1838-1850

Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Li, S., Zha, Z. J., Ming, Z., Wang, M., Chua, T. S., Guo, J., & Xu, W. (2011, July). Product comparison using comparative relations. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 1151-1152).

Miao, Q., Zhang, S., Meng, Y., & Yu, H. (2013, May). Domain-sensitive opinion leader mining from online review communities. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 187-188).

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521-543.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab.

Wang, S. M., & Ku, L. W. (2016, May). ANTUSD: A large Chinese sentiment dictionary. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 2697-2702).

Zhang, Z., Guo, C., & Goes, P. (2013). Product comparison networks for competitive analysis of online word-of-mouth. *ACM Transactions on Management Information Systems (TMIS)*, 3(4), 1-22.

中研院中文斷詞系統：<http://ckipsvr.iis.sinica.edu.tw/>