

# 使用者意見探勘-以 Facebook 粉絲專頁為例

林冠呈  
朝陽科技大學  
資訊工程系

吳世弘  
朝陽科技大學  
資訊工程系

陳良圃  
財團法人資  
訊工業策進  
會

谷圳  
中央大學

陳國棟  
中央大學

lkcheng0829@gmail.com shwu@cyut.edu.tw eit@iii.org.tw cujing@gmail.com chen@csie.ncu.edu.tw

## 摘要

社群網站如 Facebook、Plurk...等已成為熱門的行銷工具之一，許多企業也漸漸發現網路行銷所帶來的商機。本研究以 Facebook 粉絲專頁的留言為語料，分析使用者的情緒。過去分析使用者情緒主要是透過情緒字典例如 NTUSD 來辨識文件中的情緒。本文增加作者分析的情緒規則，並結合情緒詞辭典以及機器學習的方法來探究使用者的意見屬於正向、負向或中立。研究結果顯示增加規則可有效提升分類效果。所得到的資訊將可提供企業業務參考，用以改善使用者滿意度。

**關鍵詞：**Facebook 粉絲專頁、意見探勘、情緒分析、社群網站分析。

## Abstract

Social network websites such as Facebook, Plurk, etc. have become a useful marketing toolkit. Many companies find that it can provide new opportunities. In this paper, we focus on the sentiment analysis on the Facebook Fans pages. Such analysis in previous work were often based on a sentiment dictionary, for example NTUSD. We add more domain rules for sentiment analysis in our system. We find that these rules can work with the traditional approach and get better recognition results. The results can be used for advance marketing and to satisfy more users.

**Keywords:** Facebook Fans page, opinion mining, sentiment analysis, social network analysis.

## 1. 前言

在這資訊發達的年代，我們隨時可以透過網際網路取得各式各樣的訊息，也讓以網際網路為基礎的商業應用蓬勃發展。人多的地方就有商機，網際網路也不例外，其中如 Facebook、

Twitter 等「社群網站(social networking sites)」的崛起，更使得現實生活中的人際互動與社交活動正逐漸移轉至虛擬的社群網路平台進行，更成為熱門的網路行銷工具之一[6]。

近年來在網路行銷方面，依據台北市數位行銷經營協會(DMA)發佈的觀測數據；2012 全年台灣整體數位媒體廣告市場達 116.01 億，成長 13.57%。其中「社群口碑行銷」之規模便達 9.46 億，成長 13.16%，如圖 1 所示，若進一步預測今(2013)年可能市場規模，則高達 18% 的成長率並使規模突破 11 億，同時超越一般常見的「網站廣告(10%)」與「關鍵字廣告(15%)」。此調查數據顯示社群網站行銷逐漸成為企業關注的焦點。[4]

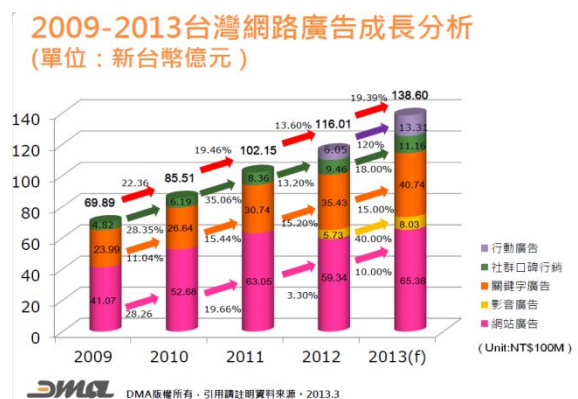


圖 1 2009-2012 年台灣網路廣告成長分析

根據市調機構創市際 ARO(2013)「台灣社群媒體網站使用概況」的研究報告[5]指出，2013 年 6 月台灣整體社群媒體類別共有 1,154 萬位不重複使用者，為 97% 到達率，顯示每 100 位網友只有 3 位不曾造訪社群媒體。其中 Facebook 在到達率、平均造訪次數、平均使用時間、平均瀏覽網頁數的四項數據皆高於其它社群網站如：Google、Plurk...等；顯示台灣 Facebook 社群網站的流量與使用者黏度極高。

目前國內、外的企業也發現網路社群所帶來的商機，並著手經營；以 Facebook 來說「粉絲專頁」是各企業普遍選擇的行銷工具，對內

提供社群成員間溝通與知識分享平台，對外則將網路社群視為與消費者互動溝通管道，以滿足顧客需求與增加企業品牌知名度[1]。

使用者們透過 Facebook 粉絲專頁互相交流、分享心情以及對企業、商品的經驗或看法，這些留言有正面情緒也有負面情緒；而正、負向情緒的產生對企業顧客管理而言是重要的。因此本研究以 Facebook 粉絲專頁使用者留言為單位，透過分類器預測的方式將留言依照正向(Positive)、負向(Negative)以及中立(Neutral)進行分類，讓企業可以快速理解使用者的情緒，並觀察意見背後的原因。

本篇除第一節為前言外，第二節為情緒分析相關文獻探討，於第三節陳述研究之方法，並於第四節就實驗結果進行討論，最後則為結論及未來研究方向。

## 2. 意見探勘相關文獻探討

隨著社群網站的熱門，許多使用者喜歡在社群網站上透過簡短的文字、圖片...等分享意見或是記錄生活，因此，更能顯示使用者的情緒，每日在 Facebook、Plurk...等社群網站上有成千上萬的訊息產生，可以說社群網站擁有進行情緒分析所需要的豐富資源。[2]

### 2.1 意見探勘

意見探勘又可稱作情緒分析(sentiment analysis)，是國內、外自然語言處理與資訊檢索領域近年來相當熱門的研究議題之一。

在國際公開評測 NTCIR-7 與 NTCIR-8[12]皆包含意見探勘的項目，項目目標要受測者定義語料的正向(Positive)、負向(negative)。Huang et al.[10]在 NTCIR-8 MOAT(Meaning Of A Tag)項目使用情緒句提取(opinion sentence extraction)和詞性標註(polarity identification)模型，並使用支援向量機(support vector machine.SVM)進行情緒分類。Kuo et al.[2]等人收集 Plurk 上的留言來分析圖書館使用者的滿意度，使用 K-最鄰近分類法(KNN)、簡單貝氏分類法(Naïve Bayes)和 SVM 等機器學習法分類。

情緒分析發展至今已有很好的成效，但針對 Facebook 粉絲專頁尚無相關研究，因此本研究使用中文情緒字典 NTUSD 以及自行制定特定粉絲專頁的情緒規則進行實驗。

## 3. 研究方法

本研究選擇熱門社群媒體 Facebook 做為研究對象。我們所提出之系統架構如圖 2 所示，首先收集來自 Facebook 粉絲專頁的資料，包含貼文(posts)、留言(comment)、按讚(like)、分享(share)，而因為這些收集到的語料含有許多雜訊，如停用字、標點符號...等，必需先由程式過濾；接著將留言(comment)人工標註情緒，並將語料分為訓練資料(traning set)及測試資料(test set)，實驗方法分為單以情緒規則和情緒規則加 NTUSD 情緒字典二種，再經 SVM 分類預測模組得至結果。

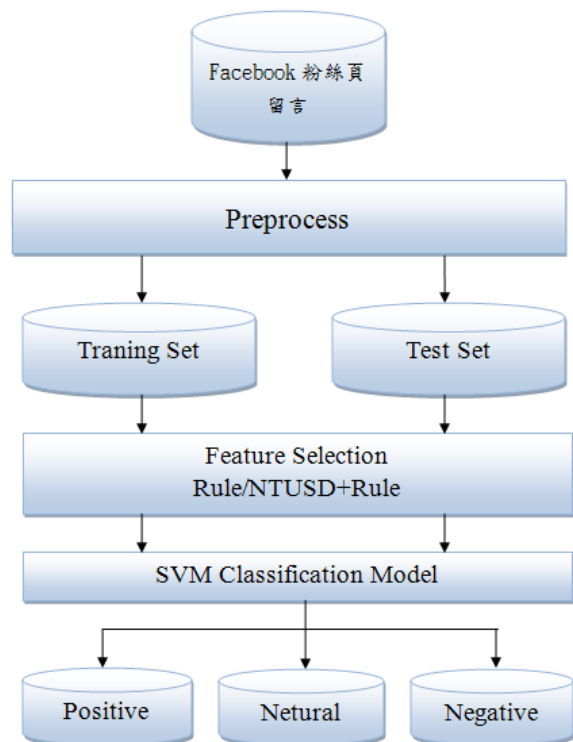


圖 2 系統流程圖

### 3.1 語料蒐集

以「Huggies 好奇 台灣」粉絲專頁為對象從 2013 年 1 月 1 日到 2013 年 12 月 13 日該粉絲專頁的貼文(posts)以及使用者(粉絲)按讚(likes)、留言(comments)、分享(share)四項數據，如表 1。

圖 3、4、5、6 為「Huggies 好奇 台灣」粉絲專頁 2013 年各資料的的走勢圖，我們選擇粉絲專頁的留言(comments)做為本研究的語料庫。

表 1 「Huggies 好奇 台灣」(2013)

	posts	likes	Comments	share
1月	23	2557	765	857
2月	23	1982	1481	304
3月	46	1981	109	303
4月	22	3649	2669	777
5月	44	7206	320	995
6月	39	6879	355	451
7月	30	2569	127	395
8月	23	1741	152	279
9月	25	2765	199	323
10月	24	1839	243	200
11月	23	1797	177	147
12月	18	852	69	22

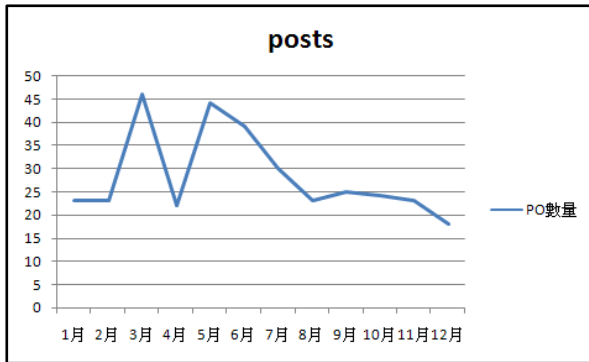


圖 3 各月粉絲專頁貼文數(2013)

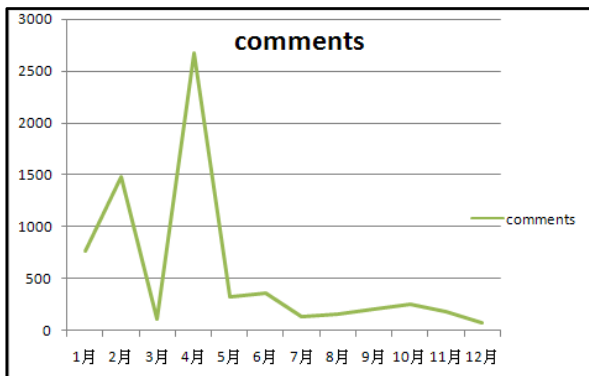


圖 4 各月粉絲專頁留言數(2013)

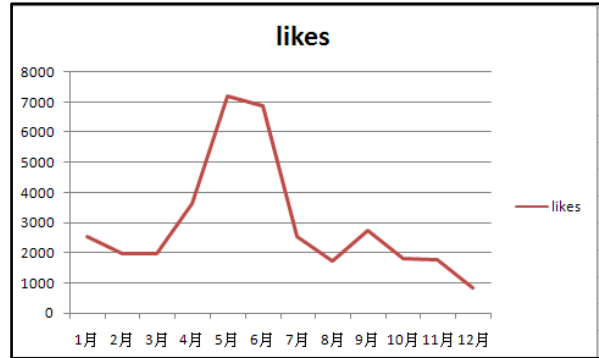


圖 5 各月粉絲專頁按讚數(2013)

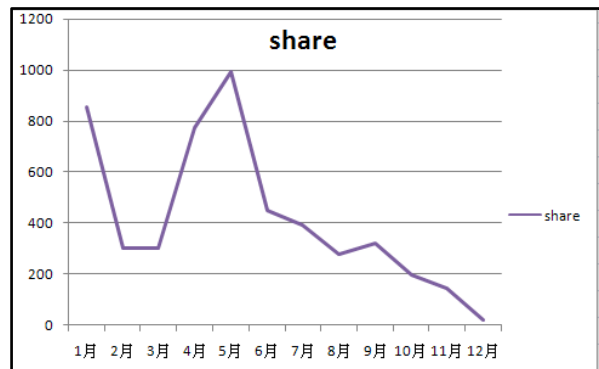


圖 6 各月粉絲專頁分享數(2013)

「Huggies 好奇 台灣」粉絲專頁中的貼文可區分四大類，分別是圖片、影片、育兒心得和活動消息。而粉絲專頁管理者的貼文數的多寡不一定能引起討論的熱度，能引起使用者熱烈討論的貼文，大部份是活動消息，從3月、4月、5月的數據觀察，3月的貼文數為全年最多，但是討論的熱度(like、comments)並不高。粉絲專頁在4月有舉辦贈獎活動，貼文數雖然不多，卻能引起使用者的熱烈討論。5月不僅貼文數多也有舉辦活動，所引發討論的熱度顯而易見。

由於粉絲專頁的活動消息，如贈獎活動，粉絲專頁管理者會要求使用者對在活動消息的貼文回覆設定好的留言，以當做參與活動的資格，所以我們在語料的選擇會先濾除活動消息的留言。

### 3.2 人工標註

取得語料後，以人工標記的方式針對語料庫中的 500 筆留言如表 2 進行情緒標註(P: Positive, 0: Netural, N: Negative)。人工標記統計於表 3。

表 2 人工標註範例

留言	情緒
我喜歡好奇 Huggies，因為可以讓寶寶的屁股好乾爽不會有尿布疹	P
真的很好用，寶寶包一個晚上起床時換尿布屁屁是乾乾的	P
請問用電腦看了之後怎麼領試用包和抽尿布?	0
剛生完的媽咪可以參加此活動嗎?	0
超級會漏尿	N
每次點選都一直無法進入也退不出來!造成當機!	N

表 3 人工標記統計

情緒極性	數量
正向	389
中立	101
負向	10

### 3.3 前處理

為方便程式處理，我們首先對原始語料進行刪去標點符號、去除停用字(如：的、嗎、喔...等)，接著執行斷詞；我們採用中央研究院詞庫小組 (Chinese Knowledge Information Processing Group ,CKIP)所開發的中文斷詞 API 進行斷詞。

- (1)原句：我喜歡好奇 Huggies 因為可以讓寶寶屁股好乾爽不會有尿布疹
- (2)斷詞結果：我(N) 喜歡(Vt) 好奇(Vi) Huggies(FW) 因為(C) 可以(ADV) 讓(Vt) 寶寶(N) 屁股(N) 好(ADV) 乾爽(Vi) 不會(ADV) 有(Vt) 尿布(N) 疹(N)

### 3.4 情緒分析方法

#### 3.4.1 情緒字典

本研究提出的系統在判斷語料情緒時，使用由台大資工所自然語言處理實驗室所建置的中文情緒詞辭典 NTUSD(<http://nlg18.csie.ntu.edu.tw:8080/lwku/pub1.html>)，NTUSD 共包含正向詞 21,056 詞、負向詞 22,750 詞，用此辭典與測試語料比對結果做為機器學習所需要的特徵值。

### 3.4.2 特定企業的情緒規則

以 Facebook 粉絲專頁而言，會有只用 NTUSD 情緒詞辭典無法辨識特定的情緒詞，如表 1 所標註的負向情緒句。因此，除了 NTUSD 之外，我們加入特定企業的情緒規則(表 4)，以提高系統正確率。

規則邏輯為以第一列為例，"不 回滲"~2 代表空白處可以插入 0~2 個字元，第三列乾爽-不乾爽，"- "為 NOT，出現乾爽則不能同時出現不乾爽。情緒標註 1 為正向情緒、0 為負向情緒。

表 4 特定粉絲專頁情緒規則範例

規則	情緒標註
"不 回滲"~2	1
"不 過敏"~2	1
乾爽-不乾爽	1
"摩擦傷 屁屁"~4	0
"不 透氣"~2	0
外漏-不外漏	0

### 3.4.3 支援向量機(SVM)

SVM 是一種效率、成效很高的監督式機器學習方法，特別是處理非線性資料分類的問題。SVM 可輸入一群以未知方式已分類好的資料做為訓練集，經由訓練後獲得一個模型，再輸入有尚未分類的資料時，SVM 便可依據先前資料訓練出的模型去判斷該資料所屬的類別[3]。

我們實驗使用的 SVM 是採用 Chang&Lin[9]所發表的 LiBSVM 支援向量機，核心函數 RBF 的 C 與  $\gamma$  採用預設值。

## 4. 實驗

我們將 500 句語料取 350 句為訓練集，150 句為測試集。並設計二組實驗比較加入特定情緒規則後是否對辨識 Facebook 粉絲頁有所幫助。

### 4.1 實驗(一)

單以特定企業的情緒規則來辨識情緒，正確率為 68%。



表 5 混淆矩陣(單以情緒規則判斷)

Predicted	Actual			Total
	Class1	Class2	Class3	
Class1	62	2	0	64
Class2	40	38	2	80
Class3	1	2	3	6
Total	103	42	5	150

## 4.2 實驗(二)

實驗(二)結合 NTUSD 以及特定企業情緒規則，正確率為 77%。

表 6 混淆矩陣(NTUSD + 情緒規則)

Predicted	Actual			Total
	Class1	Class2	Class3	
Class1	73	2	0	75
Class2	29	39	2	70
Class3	1	1	3	6
Total	103	42	5	150

## 4.3 討論

單以特定企業的情緒規則進行情緒分析即可有一定程度的效果，原因是會在粉絲專頁留言的使用者大都有使用過該企業所提供之服務，或是使用相關的商品經驗。以本研究的對象「Huggies 好奇」是生產嬰兒用品的企業，如表 2 情緒規則所示，特定企業的情緒規則能精準的判斷使用者對商品的評價；而為了更全面的辨識情緒加入 NTUSD 情緒辭典，實驗結果正確率也有所改善。

## 5. 結論及未來研究方向

本研究使用 Facebook 粉絲頁的留言為語料，建立情緒規則，並加上 NTUSD 情緒詞辭典，用監督式機器學習的方法來預測留言情緒極性，得到不錯的效果。這代表我們可以用意見探勘的方法自動在 Facebook 粉絲專頁中分析留言情緒，找到對企業有幫助的意見，也能從長時間的留言、按讚的資料中統計使用者關注的話題，並根據其話題將使用者進行分群。

未來研究方向，觀查該粉絲專頁的使用者消費個性，了解其背後的原因，提供企業做為參考。除了對留言(comment)意見探勘之外，我們會使用其它所蒐集的資料(likes、posts、share)進行研究。從數據走勢圖分析貼文與留言、按讚的關連，將貼文品質對粉絲專頁的影響程式化執行。而在 LibSVM 的部份，參數我們是採用預設值，從這部份著手調整找到最佳參數，

對於結果也能有所提升。

目前資訊視覺化(Infographic)技術也廣泛使用在社群媒體之中，在 Twitter 上有許多相關的應用，如圖 7 是對 Nokia 的 Twitter 應用的結果，其圖表顯示在一段時間之中，各種主題被使用者所討論的熱度，可以觀察不同主題討論的幅度變化。我們也會加入資訊視覺化技術在之後的研究中。

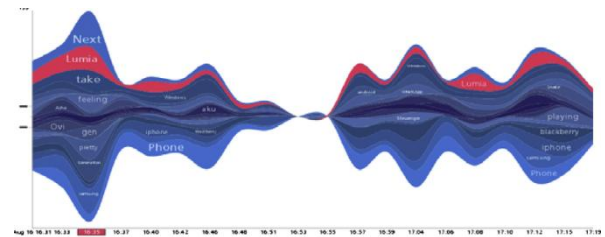


圖 7 Twitter stream graph (1000 tweets each) for Nokia.[11]

## 誌謝

本研究依經濟部補助財團法人資訊工業策進會「103 年度虛實整合智慧商務關鍵技術與平台研發計畫(1/4)」辦理。

## 參考文獻

- [1] 白榮吉、蕭穎謙、陳楠熹、陳意婷、劉忠輔、黃宛婷、杜育帆，”線上購物網站網路社群經營關鍵成功因素之多重個案研究-以 Facebook 粉絲頁為例”，*創新研發學刊*，第九卷，第一期，pp. 1814-1927，2013。
- [2] 郭俊桔、張育蓉，”使用情緒分析於圖書館使用者滿意度評估之研究”，*圖書資訊學研究*，pp. 153-197，2013。
- [3] 李政儒、游基鑫、陳信希，”廣義知網詞彙意見極性的預測”，*中文計算語言學期刊*，第 17 卷，第二期，pp. 21-36，2012。
- [4] 台北市數位行銷經營協會，**2012 全年台灣整體數位媒體廣告量**，2013 年 3 月 25 日，取自 <http://www.iama.org.tw/upload/ResourceTrend/20130325035256917.pdf>
- [5] 台北市數位行銷經營協會，**台灣社群媒體網站使用概況**，2013 年 7 月 31 日，取自 <http://www.iama.org.tw/upload/ResourceReport/20130913101003132.pdf>
- [6] 楊舜慧、楊千慧、林玉凡、潘金谷、劉瀚文，”從口碑傳播軌跡初探社群行銷效益-以消費打卡為例”，*第十八屆人工智慧與應用研討*

- 會，2013。
- [7] 吳文貴、林孟陞，”Facebook 粉絲專頁訊息廣告效果之研究：以 7-Eleven 與麥當勞粉絲專頁為例”，**行銷評論**，Vol. 10, No 1, pp. 43-60，2013。
- [8] 曾議慶、吳世弘，”定義有用之意見在線上消費者評論”，朝陽科技大學，2012。
- [9] C. C. Chang, C. J. Lin, “LIBSVM: a library for support vector machines”，2001
- [10] Wan-Chi Huang, Meng-Chun Lin, Shih-Hung Wu, ”Opinion Sentences Extraction and Polarity Classification Using Automatically Generated Templates”，**Proceedings of NTCIR-8 Workshop Meeting**, June 15-18,2010
- [11] Mohamed M. Mostafa, “More than words: Social networks’ text mining for consumer brand sentiments” **Expert Systems with Applications** **40**, pp.4241-4251, 2013
- [12] NTCIR.(Online). Available: <http://research.nii.ac.jp/ntcir/index-en.html>
- [13] NTUSD(National Taiwan University Semantic Dictionary). Available: <http://nlg18.csie.ntu.edu.tw:8080/lwku/pub1.html>
- [14] Pang B., Lee L. “Opinion mining and sentiment analysis.” *Foundations and Trends in Information Retrieval*, pp. 1-135.2008.
- [15] Wu He, Shenghua Zha, Ling Li, “Social media competitive analysis and text mining: A case study in the pizza industry”，**International Journal of information Management** **33** , pp. 464-472, 2013.