

以部落格文本進行情緒分類之研究

楊昌樺 陳信希

國立台灣大學資訊工程學系

chyang@nlg.csie.ntu.edu.tw; hhchen@csie.ntu.edu.tw

摘要

本文從部落格文本中帶有情緒符號的文句出發，探討人們的溝通行為擴展至網路空間後，如何將情緒表達的需求反映在文字與情緒符號的使用上，並進一步以情緒符號的意涵作為文句表達情緒的分類依據。我們從雅虎奇摩部落格服務取得訓練與測試集，以向量支撐機(SVM)運用文句特徵設計情緒分類器，並藉由各項情緒分類器的實驗數據，研究是否可以利用情緒詞彙解釋人們在部落格中使用情緒符號的偏好與特徵，進而達成對網路空間人們情緒的解讀與分析。

1. 緒論

達爾文於 1872 年發表“The Expression of the Emotions in Man and Animals”一書，他以進化的觀點分析動物和人類在情緒質量上相似與相異之處。之後一百多年來，心理學、腦神經科學、認知科學等領域的學者也投入人類情緒分析的研究，並發展出透過腦部影像、心跳、血壓等生物訊號來判斷人類情緒變化的方法(Dolan, 2002)。情緒狀態的傳遞亦屬於人類溝通行為的一個重點成份，人類可以透過臉部表情、肢體動作、手勢、語言、聲調等訊息來表達自己的情緒狀態，溝通的對象也會根據自身的經驗、以及對表達者的理解來解讀對方的情緒。

情緒解讀的工作也吸引電腦科學領域學者的注意，隨著電腦辨識技術的演進，人類表達出的各項訊息轉換成數位資訊，研究人員採用機器學習等方式，計算這些資訊與各項情緒類別的關係。相關的研究包括：當判斷的類別定義成人類的情緒狀態後，如何挑選適當的資訊作為特徵，以訓練出相關模型來判斷人類的情緒。如 Chuang and Wu (2004)使用文

字及語音兩類特徵資訊對語音句子所隱含的情緒進行辨識，研究結果顯示同時使用兩類特徵，比單獨使用一類特徵對情緒辨識有更好的效果。Pao 等人 (2005)也使用語音特徵建立情緒分類器，開發的工具可協助聽障人士透過語音進行情緒的表達。

近代電腦科學界所興起的一項重大發明—網際網路，除了提供電子資訊分享的平台外，也提供人們不同於以往的溝通介面。追溯自網際網路開始用來傳遞資訊(如電子郵件、電子佈告欄、電子聊天)的年代，透過網際網路傳遞文字，可以在即時即地的基礎上，達成使用者互動溝通的便利性。然而當通訊的對象越來越習慣這項新的通訊方式後，也開始產生了需傳遞彼此情緒狀態的需求。通常在網際網路上傳遞訊息時，會缺乏夠多的前後文資訊來判斷對方的情緒狀態，為了避免因此造成溝通時的誤解，1981年CMU的Scott Fahlmany曾經設計「:-)」和「:-(」兩個可以放在文字中的符號，以象徵性的笑臉和哭臉來代表高興或不愉快的情緒。隨著類似象徵情緒或表情的符號不斷地被網際網路使用者創造與使用，甚至到了90年代unicode要擴充全球通用文字碼時，情緒符號也佔了一席之地，如unicode集合中三個字元「☺」、「☺」、「☹」的十六進位碼分別定為0x2639、0x263a、0x263b。

近年來，因為圖形使用介面的演進，讓更多表情文字得以圖示化的方式呈現，所產生的新興圖示集合通稱為Smiley或Emoticon，這些圖示被大量使用在以網際網路為主的通訊媒介中。而個人化訊息傳遞服務的流行更加速了這種圖示流行的趨勢，如雅虎即時通、微軟MSN Messenger等傳訊軟體，皆提供使用者可在傳訊的介面上使用許多不同情緒符號或是自訂圖示。此類圖形對應到一些表情文字或是自訂的字元順序，讓個人偏好的圖示與動畫得以即時內嵌在通訊的內文中間。Liu等人 (2003)曾以文字與生活常識建立模型來判斷人的情緒，並設計出一個電子郵件介面EmpathyBuddy，根據郵件中每一句話賦予一個Chernoff Faces，這些臉部表情也增進了使用者訊息互動的趣味性。

除了表情與情緒符號的使用外，網際網路使用者也會在既有語言範疇之外，創造新的文字表達規則。所創造的新詞彙或用語，形成網路方言的一部分。這些方言通常出現在以網際網路為主的電子媒介所使用的溝通信息中，夾雜於正常的語言文句之內。這樣的呈現通常會讓原句子不合於文法、或是讓網路方言本身形成了未知詞彙。高與倪(2003)透過網

際網路文章收集情緒用詞，並將這些詞彙歸納在憤怒、悲傷等情緒分類下。

發佈網路文章的介面、表達方式、語言，隨著更多使用者的參與而不斷地擴充、創新，藉由更多使用者的加入所引發的新創意，也產生了更多新興的溝通方式。這個由許多網際網路使用者彼此互動、溝通所形成的虛擬空間，有個特殊稱號為Cyberspace。最近Cyberspace下，又興起了一個子空間叫Blogosphere。所謂Blogosphere，是指由一種個人導向、可與使用者互動的媒體發佈介面Blog(或稱部落格)，所形成的集合。我們於前文曾敘述過使用者在適應了Cyberspace的溝通方式後，會開始有情緒狀態傳遞的需求。而Blog在開始吸引眾多使用者之後，由於其創立思維更貼近人們表達個人狀態的動機，也吸引更多使用者前來表達自己的觀念、想法、甚至個人情緒。

一些Blog服務提供使用者不同層次的情緒標記選項，例如：Xuite¹讓使用者選擇個人站台的心情分類，如「甜蜜中」、「憂鬱中」等共15種分類；LiveJournal²讓使用者以文章為單位，標記發表文章當時的心情，如sad(悲傷)、mischievous(淘氣)等共118種分類。Mishne(2005)使用由LiveJournal的文本所抽取的特徵，以SVM(Cortes and Vapnik, 2005)訓練以文章為單位的心情分類器。文中顯示實驗成果仍有很大的進步空間，也反映心情分類仍是個相當具有挑戰性的議題。本文從這個角度切入，進一步以文句為單位，觀察個人情緒在Blog文本上反映的現象，這是中文情緒分析的創新嘗試。

本文內容安排如下：第二節介紹帶有情緒表達的Blog文本，第三節敘述以文本資訊進行情緒分類的實驗設計、與實驗結果，第四節總結成果、並指出未來的研究方向。

2. 部落格文本與情緒符號

由於本文探討的個人情緒分析所採用的語料來自 Blog 文本，因此先介紹建立 Blog 文本語料庫相關的議題。過去隨著 WWW 的發展，網站登錄與各家搜尋引擎相繼而起，提供使用者快速便利的資訊搜尋服務。Blogosphere 的發展有點類似，近年來陸續有許多網

¹ <http://blog.xuite.net/>

² <http://www.livejournal.com/>

站，甚至大廠開始提供 Blog 架設服務，吸引許多 Blog 登錄(Blog Ping)與搜尋服務的開設，如 Technorati³提供搜尋網頁介面及 API，Weblogs.com⁴提供登錄 Blog 的介面，並即時回報世界上最新更新的 Blog 文章。Kolari 等人(2006)的研究即利用 Technorati 及 Weblogs.com 提供的服務，取回英文文章加以分析。提到最受歡迎的 Blog 架設服務，分別是 Blogspot⁵與 MSN Spaces⁶。其中 Blogspot 文章發佈的介面僅提供基本的文字編輯，而 MSN Spaces 藉由著名的傳訊軟體 MSN Messenger 的附加功能，開始吸引使用者的投入。其發表文章的介面因而引入了 MSN Messenger 中慣用的情緒符號，這些符號可以讓發表 Blog 的使用者加在文章的任何部分。

台灣地區最大入口網站為雅虎奇摩，所提供的 Blog 服務⁷也可以讓使用者在發表文章的介面中，於文章任意部分加入由雅虎即時通軟體引入的情緒符號。雅虎 Blog 服務所提供的情緒符號共 40 種，其圖樣、代碼、意涵如表 1 所示，包括世界通用的笑臉「:)」和哭臉「:(」符號，也有自創的圖示如「@};-」代表一朵橫放的玫瑰花。由於使用者可自行決定要不要在發表的文章中加入情緒符號，因此並不是每篇文章都會含有情緒符號。

為了運用含情緒詞彙的文章來協助本研究的進行，我們首先建立一個前提：

(p1) 使用者情緒的傳遞會反映在這些情緒符號的使用上。

並提出一個假設：

(a1) 使用者伴隨著文本的情緒就是該情緒符號的意涵。

例如下列句子中：

(s1) 一次又一次的這樣~心中莫名的火大..每次他一回來~我都睡不飽 😡

(s2) 我們也...傻眼了 😞

(s3) 每次和妳講電話就是莫名的開心 😄

³ <http://www.technorati.com/>

⁴ <http://weblogs.com/>

⁵ <http://www.blogspot.com/>

⁶ <http://spaces.msn.com/>

⁷ <http://tw.blog.yahoo.com/>

參照表 1 我們定義(s1)句的使用者情緒為「大哭」、(s2)句的使用者情緒為「驚訝」、(s3)句的使用者情緒為「大笑」。經由這樣的設定，本研究即以雅虎 Blog 服務使用者所發表的文章為材料，分析與 40 種情緒符號(情緒)之間的關係。

2.1 實驗語料庫

為了提供本研究所需的實驗語料，我們透過雅虎 Blog 服務的搜尋引擎，取回 2006 年 6 月底至 7 月初共 47,570 篇 Blog 文章，相關統計資訊如表 2 所示。其中 2006 年 6 月份的文章共 9,668 篇為測試資料集、7 月份第一周文章共 29,528 篇為訓練資料集，兩項資料供實驗一至實驗四使用。7 月份第二周的文章共 8,374 篇為補充訓練資料集，供實驗三至實驗四使用。訓練資料集帶有情緒符號的文章數量為 4,146 篇、測試資料有 1,289 篇、補充訓練資料集有 1,241 篇，比例在 13%~15%之間。如果以去掉 HTML 標記後的 UTF-8 編碼

表 1、雅虎奇摩部落格服務所提供的情緒符號列表

符號	文字	意涵	符號	文字	意涵	符號	文字	意涵	符號	文字	意涵
	:)	微笑		:O	驚訝		0:)	天使		(:	呵欠
	:(難過		X-(生氣		:-B	戴眼鏡		=P~	流口水
	;))	眨眼		:>	得意		=;	再見		:-?	考慮
	:D	開懷		B-)	耍帥		I-)	睡著			偷笑
	::)	眨眨眼		:-S	擔心		8-)	環顧		=D>	鼓掌
	:-/	疑惑		>:)	邪惡		:-&	不舒服		[-o<	祈禱
	:x	愛意		:((大哭		:-\$	安靜		:-<	嘆氣
	:>)	害羞		:))	大笑		[-(不說了		>:P	吓
	:p	吐舌頭		:	呆住		:o)	小丑		@};-	花
	:*	親親		/:)	皺眉		@-)	神智不清		:@)	豬頭

表 2、訓練與測試資料集統計

	文章數	帶有情緒符號 文章數	比例	無情緒符號文章 平均長度	有情緒符號文章 平均長度
訓練資料集	29,528	4,146	14.04%	1,934	1,131
補充訓練集	8,374	1,241	14.82%	1,776	1,096
測試資料集	9,668	1,289	13.33%	1,614	1,058
合計	47,570	5,435	13.87%		



圖 1、Blog 文本⁸、發文介面範例、結構示意圖

長度衡量文章的相對長度，可發現通常沒有使用情緒符號達的文章平均長度較長，且與有使用情緒符號文章的長度有一定程度的差距。這個現象顯示情緒符號在文本中扮演重要角色，有些意涵隱含在符號中。

2.2 部落格使用者心情之反應與文本風格

前文提及以歐美地區使用者為主的 Blog 服務 LiveJournal，提供使用者針對發表文章當時的心情給予一個標記，並搭配一個情緒符號加以表達。本研究的語料來源雅虎 Blog 服務，則提供使用者以平時在網路表達的習慣在文句中加入自己的情緒。圖 1 為本研究所取回的 Blog 文本範例，文本一個頁面中包含有標題、分類、發佈時間、及內文等資訊。使用者在發佈內文時，可自由加入相片、文字、情緒符號，發文的介面也提供變更文字格式、加入連結、變更排版樣式等功能。在這個例子中，發文作者曾使用過「微笑」、「驚訝」、「大笑」、「花」等四種情緒符號，相當程度地表達了其開心的情緒。

如同一般作文時，有諸如記事、抒情、論說、公文等體裁。使用者利用 Blog 發佈文

⁸ 文本來自 <http://tw.myblog.yahoo.com/jw!IlyqF8KeFRLqmezKFIeTaK5./article?mid=9470>

章時，對於內文的呈現，也有許多不同的表達方式，例如：

- 日記、抒發心情、詩賦、創作、評論，可能會引用一些相片。
- 純遊記、遊記加上許多相片、或是純粹只有相片分享。
- 資訊、教學、轉貼，通常有說明圖片和連結，也可能只給連結沒內文。
- 獨白、對話、留言、溝通，可能會擺一幅近照。

不同使用者都會有自己偏好的表達方式。另外，Blog 網站通常會提供預設、或是可自行設定的分類，讓使用者標記所發表的文章。

以上一節介紹的訓練資料集為例，蒐集到的 29,528 篇文章中，使用者自行標記的分類達 10,306 種之多，其使用前 20 名的分類名稱如表 3 所示。其中「未分類資料夾」六字為雅虎 Blog 系統預設的分類，此類文章共有 32.21%，表示約三成使用者沒刻意將文章賦予分類資訊。其他分類除了「心情日記」這四字獲得最多人使用，使用率比例大於 1% 外，其餘分類的的使用比例皆小於 1%。這些分類字面使用最多的詞彙是「心情」，反映出使用者願意透過 Blog 達到抒發心情的目的。表 4 將所有使用到情緒符號的文章，依照其所屬分類出現次數順序列出前 30 名，可發現使用者會透過部落格，在抒情、日記等這樣的體裁下抒發自己的情感，並使用適當的情緒符號描述自己的心情。

表 3、雅虎 Blog 分類使用比例前 20 名列表

分類	篇數	比例	分類	篇數	比例	分類	篇數	比例
未分類資料夾	9190	32.21%	心情點滴	105	0.37%	心情記事	61	0.21%
心情日記	348	1.22%	文章分享	104	0.36%	心情故事	55	0.19%
歌詞	178	0.62%	笑話	95	0.33%	星座透視	55	0.19%
心情	162	0.57%	生活點滴	84	0.29%	小說	49	0.17%
日記	152	0.53%	文章	74	0.26%	生活	48	0.17%
心理測驗	138	0.48%	蔡依林	73	0.26%	心情寫真	48	0.17%
星座	124	0.43%	心情札記	70	0.25%			

表 4、雅虎 Blog 使用情緒符號文章所屬分類前 30 名列表

分類	篇數	分類	篇數	分類	篇數	分類	篇數
未分類資料夾	1387	心情雜記	16	文章分享	11	自言自語	7
心情日記	84	生活	16	我的日記	11	心情手札	7
日記	35	心情札記	14	心情寫真	9	結婚篇	6
心情	34	心情記事	12	心情筆記	8	笑話	6
生活點滴	26	雜記	12	心理測驗	8	生活 543	6
心情點滴	26	生活雜記	12	我的心情	8	分享	6
生活日記	22	生活記事	12	日記	8	心情日誌	6
心情故事	21	心情分享	11	生活札記	8		

3. 情緒分類實驗

針對從雅虎 Blog 服務所蒐集的文本特性，我們將情緒分析的問題轉換成針對文句特徵來判斷情緒類別的問題。本文採用 SVM 做為建立情緒分類器的核心，SVM 的工具套件為 Fan 等人(2005)所提出的 Libsvm。本節實驗中所提到的**基準值**，即是 Libsvm 在不下任何參數的情況下所做出的分類器，對測試資料集所能作出的分類效能，這種情況下分類器通常會把類別歸給實例最多的一類。參照 Libsvm，本節實驗統一設定的訓練參數為 $c=10$ 、 $g=1.6$ 訓練出各分類器，之後各小節所回報**內部測試分類正確率**、**外部測試分類正確率**是該分類器分別應用在訓練、測試資料集後所得到的分類效能，而**內測提昇率**和**外測提昇率**則是內部測試分類正確率和外部測試分類正確率分別減去基準值。

SVM 透過各個實例所帶有的特徵向量進行訓練集分類的動作，我們參考第一節所提高與倪(2002)關於網路情緒字眼的研究，使用其蒐集的分類，包括憤怒類、害怕類、悲傷類、同情類、愛與喜悅類、謾罵類，共 2,659 個詞彙作為特徵向量，因此每個文句的特徵有 2,659 個維度，特徵值是有(1)或沒有(0)出現過該特徵詞彙。

我們進一步把訓練、測試集中帶有表情的文句取出。由於有些文句中會連續使用多個相同或不同情緒符號，為了簡化問題排除歧義性，我們只挑出帶有一個情緒符號的文句，也就是沒有分類的模糊性，這些文句以 HTML 的標記 `<p>` 為界分隔。如果文句中至少包含一個非零的特徵值，則形成訓練或測試實例。經由這樣的設定，我們分別從訓練資料集和測試資料集中獲得了 4,049 筆訓練實例，以及 1,234 筆測試實例。

3.1 實驗 1—綜合分類器

在這一組實驗中，我們將 4,049 筆訓練資料，直接訓練出一個能標記出 40 種情緒類別的分類器。根據訓練標記數量的分佈，「大笑」類的 342 筆最多，約佔 8.45%；「豬頭」類的 8 筆最少，僅佔 0.2%。訓練出來的分類器直接套用在原始的訓練集，得到的內部測試分類正確率是 55.32%；套用在測試集所得到的外部測試分類正確率，僅有 14.02%。雖然比基準值 8.45%高，卻沒有實際上應用的空間。

表 5、40 類綜合分類器錯誤分析

正確答案	數量	標記答案															
		大笑	開懷	愛意	大哭	微笑	吐舌	驚訝	花	生氣	祈禱	害羞	難過	鼓掌	擔心	疑惑	眨眼
大笑	109	55	11	9	15	5	3		1	2	2		1		1	1	
開懷	81	28	9	15	13	4	3		1	2	2	1	1			1	
愛意	90	17	9	47	9	1	2			2	2						
大哭	75	17	2	3	31	2	1	2	1	5			2		3	2	
微笑	75	22	6	22	8	3	1			1	6	2				1	
吐舌	71	23	9	17	9	2				2	1	2	1	1			
驚訝	43	13	3	3	16					1			1		3	1	
花	51	13	2	20	5	1			2		5				1	2	
生氣	41	17	1		7	1	1		1	2	2		2	1	1		
祈禱	52	5	4	5	17					2	13	1	2		1		
害羞	36	9	5	13	4	2			1			1					
難過	33	6		3	7			1		1	2		8		3	2	
鼓掌	29	9	6	5	4	2					2						
擔心	37	10	1	2	18			2			1		1				
疑惑	28	7	7	2	9		1										
眨眼	30	11	5	6	2					1	3						

表 6、不同類別個數的綜合分類器的效能

分類器	訓練量	測試量	基準值	內測 正確率	外測 正確率	內測 提昇率	外測 提昇率
40 類	4,049	1,234	8.45%	55.32%	14.02%	46.87%	5.57%
32 類	3,877	1,177	8.82%	56.10%	14.70%	47.28%	5.88%
16 類	2,937	881	11.77%	61.94%	19.98%	50.17%	8.21%
8 類	1,860	595	18.38%	68.98%	28.91%	50.60%	10.53%
4 類	1,225	355	27.92%	79.51%	46.76%	51.59%	18.84%
2 類	628	190	54.46%	85.03%	56.84%	30.57%	2.38%

表 7、16 類綜合分類器錯誤分析

正確答案	數量	標記答案															
		大笑	開懷	愛意	大哭	微笑	吐舌	驚訝	花	生氣	祈禱	害羞	難過	鼓掌	擔心	疑惑	眨眼
大笑	109	56	12	9	15	5	3		1	2	2		1		1	1	
開懷	81	28	9	15	13	4	3		1	2	2	1	1	1		1	
愛意	90	17	9	47	9	1	2			2	3						
大哭	75	18	2	3	33	2	1	2	1	5	1		2		3	2	
微笑	75	23	6	22	10	3	1			1	6	2				1	
吐舌	71	24	9	17	10	2				2	1	2	1	1	1	1	
驚訝	43	13	3	3	16					1			3		3	1	
花	51	13	2	20	5	1			2		5				1	2	
生氣	41	17	1		11	1	1	1	1	2	2		2	1	1		
祈禱	52	6	4	5	17					2	14	1	2		1		
害羞	36	9	5	13	5	2			1			1					
難過	33	6		3	7			1		1	2		8		3	2	
鼓掌	29	9	6	5	4	2					2			1			
擔心	37	12	1	2	18			2			1		1				
疑惑	28	8	7	2	9		1						1				
眨眼	30	12	5	6	3					1	3						

為了調整分類器設計的策略，我們先針對綜合分類器表現不佳的原因進行探討，進行錯誤分析。表 5 統計訓練集實例數量分佈前 16 名情緒分類，在測試集所得到的表現，以進行錯誤分析。留意訓練集分佈數量排名，跟測試集排名可能有些許差異，例如「愛意」分類的測試實例有 90 筆，比「開懷」類 81 筆多。表格中對角線上顯示粗體字的格子，表示該分類分到正確類別的數量。標示灰底的格子，表示該分類在綜合分類器處理過後分到最多的地方。表現最好的分類是「大笑」和「愛意」，但正確率僅有 5 成上下。根據對角線往右下角觀察，甚至可以發許多分不到正確類的情況。以斜線標記的格子代表測試後，沒有任何一筆分到該類別，我們可以發現大部分的文句都有機會分到前四類，卻很難歸類到如「驚訝」、「眨眼」等類別。為了進一步了解類別個數是否會影響到分類器表現，我們分別保留排名分類前 32、16、8、4、2 名的分類，來篩選出不同的訓練測試集實例，以這些實例所做出的分類結果如表 6 所示。

表 6 顯示隨著分類篩選數目的減少，訓練量和測試量也隨之下降，但是在基準值、內部測試分類正確率、外部測試分類正確率部份，皆隨著分類難度的減少而提昇，這個結果指出分類類型的減少有助於分類器效能的提昇。另外內、外部測試分類正確提昇率為內、外部分類測試正確率減去基準值，提昇的幅度隨著分類的減少也呈現爬升的狀態，唯獨到了二元分類器時提昇程度僅剩 2.38%。分析其原因是由於前 2 類「大笑」和「開懷」是相當接近的情緒，造成分類器難以區分。另外表 7 類似表 5，列出 16 類綜合分類器分類正確和錯誤的數量統計，各類正確分類數(對角線上) 提昇約 0 到 2 筆不等。雖然觀察到的正確分類所提昇的數量有限，但是綜合表 5、表 7 我們可以發現—儘管文句很容易分類到前四類，但是屬於「難過」類的就不會被標成「開懷」類，屬於「生氣」類的就不會被標成「愛意」類。

3.2 實驗 2—各種二元分類器效能之比較

根據上一節對表 6 的觀察顯示二元分類器提昇效能的潛力，而針對表 5、表 7 的交叉分析，也讓我們推測特定類別特徵之間可能有互斥的表現。因此我們接著將 40 個類別的任兩類的資料，來訓練出共 $C_2^{40}=780$ 個二元分類器。為了符合頁面顯示，我們先將前 16

類配對形成的 $C_2^{16}=120$ 個二元分類器的實驗數據歸納在表 8，其中右上角為外部測試分類正確率、左下角為基準值，標示灰底的格子表示外部測試分類正確率比基準值還要差。對角線上的格子沒有特殊意義，例如不需要訓練「大笑 vs. 大笑」分類器。

分析表 8 的數據，以正面情緒「大笑」為例，跟「開懷」(56.8%)、「吐舌」(58.9%、小於基準值 66.3%)表示最不容易區分，而跟負面情緒「難過」(78.2%)、「疑惑」(78.8%、小於基準值 79.9%)最容易區分開來。另外，正面情緒更強烈的「愛意」，跟其他負面情緒「大哭」(79.4%)、「擔心」(79.5%)、「難過」(80.5%)、「生氣」(81.7%)都容易區分開來。但是「愛意」跟「微笑」(56.4%)和「花」(64.5%)，除了基準值低外，外部測試分類正確率也無法超越，這顯示跟這兩組正面情緒難以區分開來。

我們另外製作一張類似表 8 的分析表格，首先把相同的灰色區塊複製到表 9，接著在右上三角填入對應二元分類器內部測試分類正確率、左下三角填入對應內部測試分類提昇率(內部測試分類確率減掉基準值)，其提昇率範圍分佈在 13.0%到 43.9%之間，提昇率小於 25%者以粗體字表示。表 9 的數據在訓練後的內部測試階段即可獲得，藉由這樣的觀察我們獲得在內部測試階段可以應用上的經驗法則為：

- (h1) 正面情緒和負面情緒比較容易區分開來，其二元分類器所表現的內部測試分類提昇率上升的幅度較高。
- (h2) 如果某二元分類器內部測試分類提昇率過低，如「大笑 vs. 疑惑」為 14.0%、「愛意 vs. 花」為 18.8%，可以預期該分類器在外部測試時也很難有好的表現。
- (h3) 如果內部測試後，該類所有分類器平均提昇率過低，如「疑惑」類，則該類在外部測試時，也很難有好的表現。
- (h4) 內部測試分類提昇率皆為正數，表示以 SVM 來解決本研究議題有一定程度的強健度，或許有機會再透過訓練資料的增加，進一步提昇正確率。

表 8、二元分類器外部測試正確率與基準值實驗結果

	大笑	開懷	愛意	大哭	微笑	吐舌	驚訝	花	生氣	祈禱	害羞	難過	鼓掌	擔心	疑惑	眨眼
大笑		56.8%	72.9%	69.6%	62.5%	58.9%	69.7%	73.1%	74.7%	72.7%	75.2%	78.2%	76.1%	74.7%	78.8%	73.4%
開懷	54.5%		64.9%	70.5%	56.4%	49.3%	66.1%	63.6%	73.8%	69.9%	68.4%	74.6%	71.8%	76.3%	72.5%	68.5%
愛意	52.1%	52.4%		79.4%	56.4%	62.1%	74.4%	64.5%	81.7%	77.5%	71.4%	80.5%	73.1%	79.5%	78.8%	74.2%
大哭	54.8%	50.4%	52.8%		74.7%	64.4%	63.6%	69.8%	62.9%	70.1%	80.2%	70.4%	75.0%	64.3%	72.8%	79.0%
微笑	64.7%	60.5%	62.7%	60.1%		62.3%	75.4%	62.7%	81.0%	66.1%	66.7%	74.1%	69.2%	77.7%	73.8%	65.7%
吐舌	66.3%	62.2%	64.4%	61.8%	51.8%		66.7%	62.3%	69.6%	70.7%	58.9%	71.2%	56.0%	71.3%	73.7%	65.3%
驚訝	72.2%	68.4%	70.5%	68.1%	58.6%	56.9%		68.1%	61.9%	66.3%	70.9%	53.9%	63.9%	50.0%	63.4%	64.4%
花	70.7%	66.8%	68.9%	66.5%	56.8%	55.1%	51.8%		66.3%	59.2%	60.9%	59.5%	57.5%	70.5%	58.2%	53.1%
生氣	69.7%	65.7%	67.9%	65.4%	55.7%	53.9%	53.0%	51.2%		64.5%	81.8%	63.5%	65.7%	64.1%	62.3%	73.2%
祈禱	67.2%	63.1%	65.4%	62.8%	52.8%	51.0%	55.9%	54.0%	52.8%		68.2%	58.8%	63.0%	61.8%	63.8%	61.0%
害羞	69.7%	65.7%	67.9%	65.4%	55.7%	53.9%	53.0%	51.2%	63.8%	52.8%		79.7%	58.5%	80.8%	62.5%	63.6%
難過	69.9%	66.1%	68.2%	65.7%	56.0%	54.2%	52.7%	50.9%	50.3%	53.2%	50.3%		67.7%	42.9%	55.7%	69.8%
鼓掌	75.5%	72.0%	73.9%	71.8%	62.8%	61.1%	54.3%	56.1%	57.3%	60.1%	57.3%	57.0%		71.2%	50.9%	52.5%
擔心	74.2%	70.6%	72.6%	70.3%	61.1%	59.4%	52.6%	54.4%	55.6%	58.4%	55.6%	55.3%	51.7%		55.4%	67.2%
疑惑	81.2%	78.4%	79.9%	78.1%	70.3%	68.8%	62.6%	64.3%	65.4%	67.9%	65.4%	65.0%	58.4%	60.1%		62.1%
眨眼	73.2%	69.6%	71.6%	69.3%	59.9%	58.2%	51.4%	53.2%	54.4%	57.2%	54.4%	54.0%	53.0%	51.2%	61.3%	

表 9、二元分類器內部測試正確率與內部提昇率實驗結果

	大笑	開懷	愛意	大哭	微笑	吐舌	驚訝	花	生氣	祈禱	害羞	難過	鼓掌	擔心	疑惑	眨眼
大笑		85.0%	89.6%	94.1%	90.0%	87.4%	93.0%	94.0%	94.1%	95.9%	92.3%	94.1%	93.6%	94.6%	95.2%	89.5%
開懷	30.6%		86.5%	93.8%	85.8%	87.0%	92.6%	92.8%	94.0%	93.6%	90.3%	92.8%	91.7%	95.8%	94.8%	88.6%
愛意	37.6%	34.1%		95.1%	84.5%	89.4%	94.6%	87.7%	95.3%	91.9%	86.4%	93.9%	92.5%	97.2%	96.2%	88.9%
大哭	39.3%	43.5%	42.4%		93.0%	91.2%	90.6%	94.8%	92.8%	93.5%	95.1%	89.5%	94.4%	93.5%	93.9%	93.9%
微笑	25.3%	25.4%	21.7%	32.8%		85.0%	92.8%	89.1%	94.0%	91.0%	87.8%	91.3%	91.3%	96.4%	94.4%	86.9%
吐舌	21.1%	24.8%	24.9%	29.4%	33.2%		90.8%	92.7%	93.5%	92.4%	88.5%	90.7%	91.2%	94.5%	92.9%	87.3%
驚訝	20.9%	24.2%	24.2%	22.5%	34.2%	34.0%		94.2%	92.2%	93.6%	93.6%	88.9%	93.0%	92.0%	90.0%	91.1%
花	23.3%	25.9%	18.8%	28.3%	32.2%	37.7%	42.3%		95.5%	92.6%	90.0%	93.4%	92.9%	98.1%	95.9%	90.3%
生氣	24.4%	28.3%	27.4%	27.4%	38.4%	39.6%	39.1%	44.3%		94.6%	93.0%	93.2%	95.0%	94.4%	89.5%	94.2%
祈禱	28.7%	30.5%	26.6%	30.7%	38.1%	41.3%	37.8%	38.5%	41.8%		93.0%	93.3%	92.8%	96.5%	92.7%	91.8%
害羞	22.6%	24.6%	18.5%	29.7%	32.1%	34.7%	40.6%	38.8%	29.2%	40.2%		92.6%	90.8%	94.4%	94.3%	90.1%
難過	24.1%	26.8%	25.8%	23.8%	35.3%	36.4%	36.2%	42.6%	42.9%	40.1%	42.2%		91.9%	93.2%	91.6%	92.6%
鼓掌	18.1%	19.6%	18.5%	22.6%	28.5%	30.2%	38.7%	36.8%	37.7%	32.7%	33.5%	34.9%		95.7%	94.7%	90.7%
擔心	20.4%	25.2%	24.7%	23.2%	35.3%	35.2%	39.4%	43.7%	38.8%	38.1%	38.8%	38.0%	43.9%		93.9%	93.9%
疑惑	14.0%	16.4%	16.2%	15.8%	24.1%	24.1%	27.5%	31.7%	24.1%	24.8%	28.9%	26.5%	36.3%	33.8%		92.2%
眨眼	16.3%	19.0%	17.3%	24.6%	26.9%	29.1%	39.7%	37.1%	39.8%	34.6%	35.8%	38.6%	37.7%	42.6%	30.9%	

3.3 實驗 3—增加訓練資料量後效能之比較

根據經驗法則(h4)，本實驗引入 2.1 節所介紹的補充訓練資料集，藉此訓練實例可增加至 5,279 筆，各類的訓練實例數量也會有所更動。參照 3.1 節以排名篩選出不同實例集合，所做出的分類結果如表 10 所示。首先從表 6 引入增加訓練量前的分類器效能數據，接著也以不同分類個數套用在增量後的訓練資料，列出了一組外部測試分類正確率參考數據。最後列出增加訓練資料量後，以不同分類個數篩選實例後所作的各項測試，相對提昇率是外部測試分類正確率減去增加訓練量前外部測試分類正確率的數值。與表 6 類似的狀況，是各項評估都隨著分類數量的縮小而呈現上升的趨勢。但是與增加訓練量前的數據比較起來，在 16 類分類器與 4 類分類器的外部測試分類正確率卻降低了。另外二元分類器外部測試分類正確率大幅提昇至 70.85%，是因為「大笑 vs. 愛意」比原先「大笑 vs. 開懷」還容易區分的關係，因此我們仍需針對二元分類器提昇效能的表現作進一步的分析。

表 10、綜合分類器增加訓練量後之效能

分類器	增加訓練量前		增加訓練量後			
	基準值	外測正確率	基準值	外測正確率	外測提昇率	相對提昇率
40 類	8.45%	14.02%	8.94%	14.26%	5.32%	0.24%
32 類	8.82%	14.70%	9.33%	14.93%	5.60%	0.23%
16 類	11.77%	19.98%	12.35%	19.55%	7.20%	-0.43%
8 類	18.38%	28.91%	18.73%	30.20%	11.47%	1.29%
4 類	27.92%	46.76%	28.59%	45.92%	17.33%	-0.84%
2 類	54.46%	56.84%	53.58%	70.85%	17.27%	14.01%

表 11、二元分類器增加訓練量前後「單類 vs. 其他類」平均外部測試正分類確率

分類	增量前	增量後	分類	增量前	增量後	分類	增量前	增量後
小丑	89.06%	88.99%	親親	72.29%	72.56%	神智	69.03%	68.58%
大笑	81.34%	81.38%	得意	72.08%	71.60%	不清		
愛意	81.13%	81.25%	嘆氣	72.04%	72.82%	戴眼鏡	69.01%	70.12%
大哭	79.41%	79.28%	驚訝	71.88%	71.85%	難過	68.94%	69.30%
睡著	79.04%	76.96%	花	71.78%	72.03%	邪惡	68.73%	68.40%
微笑	78.20%	77.66%	擔心	70.79%	71.67%	不說了	67.58%	67.22%
開懷	77.88%	78.22%	流口水	70.60%	71.69%	鼓掌	67.26%	67.39%
天使	75.21%	75.44%	安靜	70.54%	70.76%	疑惑	66.59%	68.11%
豬頭	75.16%	74.39%	呆住	70.30%	71.00%	環顧	66.48%	66.20%
吐舌頭	75.15%	75.36%	呸	69.90%	71.24%	皺眉	66.42%	68.16%
生氣	74.60%	74.95%	再見	69.65%	68.25%	偷笑	66.04%	66.44%
呵欠	74.10%	72.49%	眨眼	69.06%	69.64%	考慮	65.04%	66.02%
祈禱	72.96%	71.93%	眨眨眼	69.04%	69.29%			

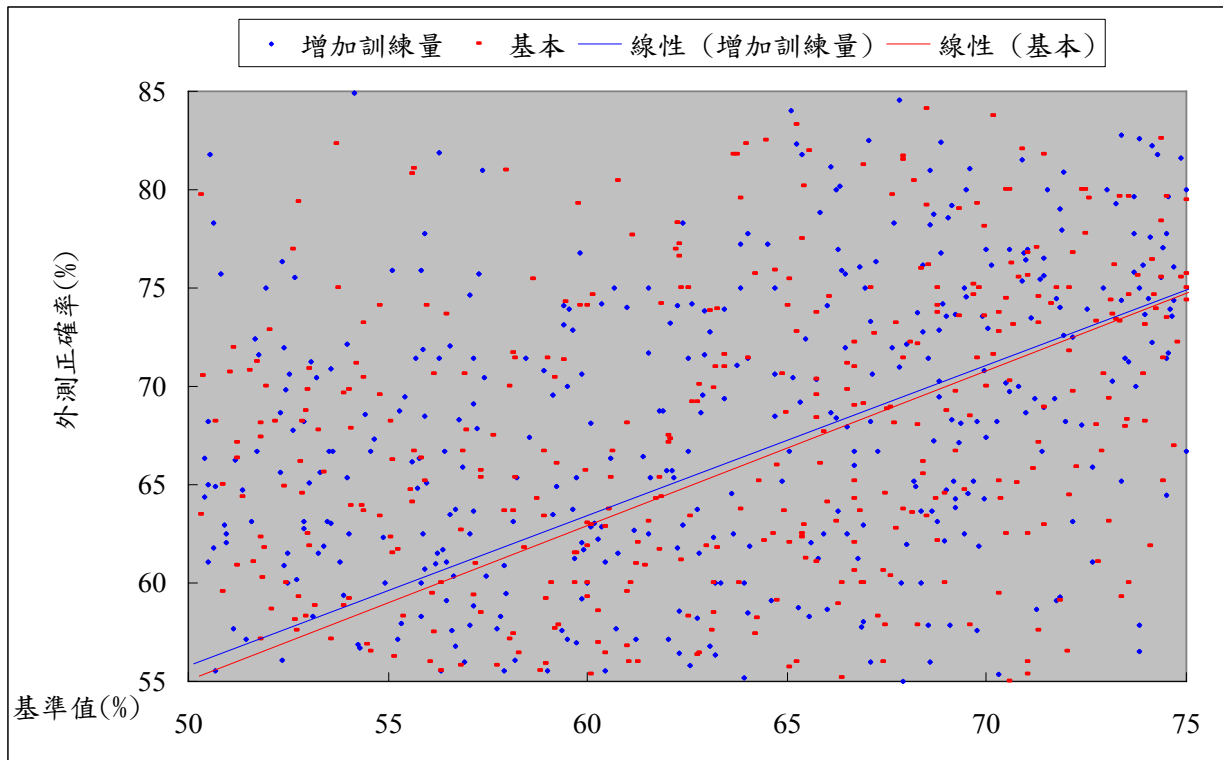


圖 2、增加訓練量前後分類器效能分佈與線性趨勢

我們首先將 40 類別中各個單一類別，與其他 39 個類別所對應的二元分類器之外部測試分類正確率，在增加訓練資料量前後的兩種狀態下取一個平均數值，並依照增加訓練量前的外部測試分類正確率由高到低的順序列出如表 11 所示，其中增加訓練量後外部測試分類正確率下降的分類以灰色網底標示之。由表 11 我們發現 40 個分類中，有 27 個分類的外部分類測試正確率在訓練量增加後上升、13 類下降。接著，我們於圖 2 將 720 個分類器的基準值與外部測試分類正確率的分佈，根據增加訓練量前後的狀況以紅、藍兩種顏色標示，並各列出一個線性趨勢線，趨勢線驗證了隨著分類基準值上昇，外部測試分類正確率也會上升的現象，並且說明增加訓練量對分類器效能的影響是正面的。

3.4 實驗 4—貪婪歸類演算法

透過上文對二元分類器效能的掌握，我們藉此設計一套貪婪演算法，試著將所有情緒歸納到正面或負面情緒類。我們曾在第 1 節提到網路界首先發明的情緒符號是微笑「:)」和難過「:(」，分別用來溝通人們最基本的正面和負面情緒。經由實驗 3 我們得到「微笑

vs.難過」分類器的基準值是 55.09%，內部測試分類正確率是 90.74%，外部測試分類正確率是 75.93%，對情緒具有一定程度的區分能力。因此，我們首先設定正面的種子情緒是「微笑」，負面的種子情緒是「難過」，並依據以下的演算法把其他 38 類情緒歸納進來：

1. 正面情緒集合 P 設成種子情緒如 {微笑}，負面情緒集合 N 如 {難過}。
2. 從剩下未歸類的情緒中挑出一個情緒 e：
 - 2.1 訓練「{P∪e} vs. N」分類器，得到內部測試分類提昇率 increase_P。
 - 2.2 訓練「P vs. {N∪e}」分類器，得到內部測試分類提昇率 increase_N。
3. 如果 increase_P > increase_N，則 e 加入集合 P，反之加入集合 N。
4. 重複步驟 2，直到所有的情緒都被歸類完畢。

歸類後的結果如表 12 所示，另外也挑選兩組種子，包括「愛意 vs.大哭」，以及都是正面的情緒「大笑 vs.開懷」，以觀察不同種子的對歸類結果的影響。前兩組結果可以發現偏正面情緒 {微笑、得意、花、愛意}，兩回都可以歸在正面情緒類。相對地，{難過、大哭、嘆氣、不說了、驚訝、生氣}，兩回都歸在負面情緒類。但是，{開懷、大笑}也各有一次歸到相反類別。我們接著應用歸類結果作成正負情緒分類器，分別將(增量後)訓練與測試集原始情緒標記對應成正負類情緒標記，以分類器效能衡量貪婪歸類演算法的實用性，三組種子最後所形成的分類器效能如表 13 所示。

表 12、根據不同情緒種子套用貪婪歸類演算法對正負面情緒歸類後結果

種子	正面情緒	負面情緒
微笑 vs.難過	微笑、天使、睡著、呆住、不舒服、呵欠、邪惡、得意、耍帥、親親、考慮、眨眨眼、疑惑、鼓掌、害羞、花、吐舌頭、愛意、開懷	難過、小丑、安靜、再見、戴眼鏡、豬頭、流口水、皺眉、環顧、嘆氣、呸、偷笑、不說了、神智不清、眨眼、擔心、祈禱、生氣、驚訝、大哭、大笑
愛意 vs.大哭	愛意、睡著、豬頭、戴眼鏡、皺眉、環顧、得意、邪惡、耍帥、呸、神智不清、親親、眨眼、擔心、花、害羞、祈禱、微笑、大笑	大哭、小丑、安靜、再見、天使、流口水、呆住、呵欠、不舒服、嘆氣、考慮、偷笑、疑惑、不說了、眨眨眼、鼓掌、驚訝、生氣、難過、吐舌頭、開懷
(反例) 大笑 vs.開懷	大笑、戴眼鏡、呆住、不舒服、環顧、嘆氣、得意、耍帥、親親、考慮、眨眨眼、疑惑、擔心、害羞、生氣、花、微笑、愛意	開懷、小丑、安靜、再見、天使、睡著、豬頭、流口水、皺眉、呵欠、邪惡、呸、偷笑、不說了、神智不清、眨眼、鼓掌、難過、祈禱、驚訝、吐舌頭、大哭

表 13、種子初始二元分類器與演算法歸類後之正負情緒分類器效能比較

		基準值	內測 正確率	外測 正確率	內測 提昇率	外測 提昇率
微笑 vs. 難過	初始分類器	55.09%	90.74%	75.93%	35.65%	20.84%
	正負分類器	50.31%	80.28%	63.45%	29.97%	13.14%
愛意 vs. 大哭	初始分類器	53.58%	88.42%	70.85%	34.84%	17.27%
	正負分類器	51.81%	78.71%	57.78%	26.90%	5.97%
大笑 vs. 開懷	初始分類器	54.25%	83.10%	56.84%	28.85%	2.59%
	正負分類器	50.41%	78.20%	56.16%	27.79%	5.75%

表 13 顯示原種子所對應的二元分類器分類效能，以及藉由該種子分類經由本節所述演算法學出最後正負情緒分類器的效能，第一組以典型的正反面情緒為種子，最後得到外部測試分類正確率最高，達 63.45%，其各項效能衡量也比其他兩組為高。第三組反例選的都是正面情緒種子，最後得到的效能最低，第二組的效能則介於其他兩組之間。

3.5 小結

實驗 1 我們看到應用在全體文章的 40 類綜合分類器的效能僅有 14.02%，並不理想。篩選成二元分類器後，雖然有 56.84%的效能，但是只能應用在兩種分類所涵蓋的實例。實驗 2 我們進一步探討 780 個二元分類的效能，發現其中帶有正反面情緒分辨意義的分類器效能較高。透過實驗 3 驗證了增加訓練資料量可以提昇效能後，實驗 4 使用貪婪歸類演算法利用世界通用「:)」和「:(」情緒符號為種子，最後得到一個能套用在全部測試實例，且效能達 63.45%的正負面情緒分類器。

4. 結論與未來研究方向

本研究探討人們的溝通媒介擴展到網際網路後，溝通行為中的情緒傳遞需求，也反應在媒介的使用上。部落格作為一個新興的媒介代表，也協助我們觀察到使用者在站台、文章、文句各個層次表達情緒的行為。我們進一步探討中文情緒處理、且由文句層次出發，試著對使用者所表達的情緒加以分類。利用雅虎奇摩提供的部落格服務，取得包含情緒符號的文本來源。並進一步透過情緒分類器的設計，以實驗數據分析問題的難度、效能提昇的方式、以及應用在情緒歸類的方法。

透過實驗章節的討論，我們最後以二元分類作為現階段歸納正面及負面情緒的方法，未來待更多訓練資料的取得，以及更多文本特徵選取的設計後，我們將進一步討論以機器協助分析情緒的效能上限、設計更高效能的多類情緒分類器。本研究以 Blog 文本進行中文情緒分析的率先嘗試，初期以分析人們使用情緒符號的現象使用出發，往後將可以此為基礎，建立對應於語言處理各階段可協助分析的情緒知識庫，幫助了解人類以使用語言來表達複雜情緒的方式。

感謝

本文部分成果由國科會計畫 NSC95-2752-E001-001-PAE 支持，在此致謝。

參考文獻

- Ze-Jing Chuang and Chung-Hsien Wu, "Multi-Modal Emotion Recognition from Speech and Text," *International Journal of Computational Linguistics and Chinese Language Processing*, Vol. 9, No. 2, pp. 45-62, 2004.
- C. Cortes and V. Vapnik, "Support-Vector Network," *Machine Learning*, Vol. 20, pp. 273-297, 1995.
- R. J. Dolan, "Emotion, Cognition, and Behavior," *Science*, Vol. 298, No. 8, pp. 1191-1194, 2002.
- Rong-En Fan, Pai-Hsuen Chen and Chih-Jen Lin, "Working Set Selection Using Second Order Information for Training Support Vector Machines," *Journal of Machine Learning Research*, Vol. 6, pp. 1889-1918, 2005.
- Pranam Kolari, Tim Finin and Anupam Joshi, "SVMs for the Blogosphere: Blog Identification and Splog Detection," *Proceedings of AAI 2006 Symposia on Computational Approaches to Analyzing Weblogs*, 2006.
- Hugo Liu, Henry Lieberman and Ted Selker, "A Model of Textual Affect Sensing using Real-World Knowledge," *Proceedings of 2003 International Conference on Intelligent User Interfaces*, pp. 125-132, 2003.
- Gilad Mishne, "Experiments with Mood Classification in Blog Posts," *Proceedings of Workshop on Stylistic Analysis of Text for Information Access*, 2005.
- Tsang-Long Pao, Yu-Te Chen, Jun-Heng Yeh and Yuan-Hao Chang, "Emotion Recognition and Evaluation of Mandarin Speech Using Weighted D-KNN Classification," *Proceedings of Rocling 2005*, pp. 203-212, 2005.
- 高台茜、倪珮晶。"華語文網路言論負向情緒用詞檢核軟體研發，"第三屆全球華文網路教育研討會 (ICICE2003) 論文集，393-402 頁，2003。