楊亨利、林青峰（2017），『微網誌短句的情感指數分析－以新浪微博為例』，*中華民國資訊管理學報*，第二十四卷，第一期，頁 1-28。

微網誌短句的情感指數分析－以新浪微博為例

楊亨利*
國立政治大學資訊管理學系

林青峰
國立政治大學資訊管理學系

摘要

隨著個人網誌與社群網路的發展，從個人社群網誌去分析發言資料、互動記錄、交友狀況等最後找出可用的規則，已成為熱門的分析應用。本研究經由分析作者在微網誌發表的狀態文句，希望除了能找出作者的正 / 負面意見傾向外，更進一步能瞭解作者撰文時可能蘊含的情緒。我們提出一個新的方法，以大陸的新浪微博為例，首先利用演化策略的方法，我們可以建立對微網誌作者正向情緒分類器與負向情緒分類器。若有需要，正負向亦可區分為非常正 / 非常負向、正 / 負向兩類別。實驗結果顯示，我們分類的效果在精準率、召回率、F1 分數均達令人滿意水準。其次，我們開發了能找出作者的情感指數推估系統；該系統利用迴歸方法可經由分析作者在其微網誌上輸入的狀態文句，推估作者想表達的心情，給予一個幸福指數；其他的情感（如：喜樂、憤怒、悲傷、厭惡、恐懼）指數也能類似地建立。

關鍵詞：微網誌、情緒分析、意見分析、情感指數、演化策略

* 本文通訊作者。電子郵件信箱：yanh@nccu.edu.tw
2015/03/08 投稿；2015/07/28 修訂；2015/12/20 接受

Yang, H.L. and Lin, Q.F. (2017), 'Estimating emotion index of short sentences in a microblog website-taking weibo.com as an example', *Journal of Information Management*, Vol. 24, No. 1, pp. 1-28.

Estimating Emotion Index of Short Sentences in a Microblog Website-Taking Weibo.com as an Example

Heng-Li Yang*
Department of Management Information Systems, National Cheng-Chi University

Qing-Feng Lin
Department of Management Information Systems, National Cheng-Chi University

Abstract

Purpose－This study aims to propose an approach for mining positive/negative opinions and estimating an emotion index of sentences in microblog website.

Design/methodology/approach－After reviewing the related literatures, we proposed an ontology-based approach by using ConceptNet and evolution strategic for mining positive/negative opinions from short sentences posted in a microblog, Weibo.com. Applying regression analysis, we also built a prototype system to estimate its implied emotion.

Findings－Using the experiment data, we can build a positive classifier to provide positive sentiment cluster and negative classifier to provide negative sentiment cluster with five or three scales. The levels of precision and recall rates, and F1 scores for those classifiers are satisfactory. In addition, our system can give an index of happiness.

Research limitations/implications－The future study can collect more sentences for testing and try other micro-blog or regular blog sites. The efficiency can be also further enhanced.

Practical implications－Practically, businesses can apply our proposed approach

* Corresponding author. Email: yanh@nccu.edu.tw
2015/03/08 received; 2015/07/28 revised; 2015/12/20 accepted

to understand the emotion of the customers after purchasing their products/services. Social workers or police departments might identify persons with suicidal potentials at the early stage from the web.

Originality/value－The academic contribution is to propose a new approach to discover possible emotion.

Keywords: microblog, emotion mining, opinion mining, index of emotion

壹、緒論

隨著網路社群與個人部落格技術的不斷發展，社群網路現今已慢慢成為人們表達自我訊息的重要管道，其上的個人狀態文本資料對於了解該使用者當下情況，是個很好的可供機器自動辨識的資料來源。本篇研究所關注的問題是如何經由自動分析社群網路中使用者發表的動態來去了解其當下的意見傾向。網路資料量相當巨大，我們很難以人工方式去監控、辨識新增的意見。若透過機器人的自動爬文，從微網誌或其它部落格上收集到這樣的三則動態發文：甲：「好開心喔，我真的很久沒休息了。」、乙：「我忍了很久，真的夠了！」、丙：「為傷害我的人們，也為愛我的人們，想念我的話還有這些歌可以給你們聽喔！」，此時若能有一個系統經過自動推論與分析，便可如人類判斷出甲的心情動態應是正面喜樂的；相反的，乙可能是負面憤怒的，丙可能是負面悲傷的。這三則發言均是真實的網路案例，而實際狀況是乙在發言之後做了殺人的犯罪行為；而丙則是在發言之後隨即輕生了。

此種自動化取得作者意見傾向的研究可應用在找到重要人物心情側寫的分析上，如當代名人的心情分析與記錄，這對需要分析重要人物記錄的歷史工作者，亦或是追逐偶像動態的追星族來說，都是很有用的技術；另一個應用情境則是自殺或犯罪的預防，我們可以利用分析發言的動態即時找出潛藏在社群中需要幫助或是可能犯罪的使用者。

要能妥善解決這二個情境會面臨到的問題，我們除了要能自動、快速而又較準確的取得作者在網路上發表意見的正 / 負傾向之外。進一步的我們需要更明確的得到使用者揭露動態時想表現出的情緒類別與程度。在這樣的情況下，本研究試圖建立出一個利用機器學習方法來建立一個分類器，以處理微網誌的作者揭露訊息的意見正 / 負傾向，並且進一步可找出揭露訊息所可能蘊藏的作者情緒。

本論文接下來，將在第貳節探討相關文獻、第參節提出本研究之微網誌情緒分析架構、第肆節介紹情緒指數與實驗資料收集、第伍節報導實驗結果、第陸節提出結論與建議。

貳、文獻探討

一、微網誌的資料分析

微網誌（Micro-blog）是一種讓使用者用很簡短的文句去進行表達心情、分享資訊、提問、聊天等行為的社群網站；每一篇微網誌的文章包括回文都是以短句所組成的，系統也往往會限制使用者單篇的發言字數。在此種情況下，文句往

往是缺少足夠的語文結構線索，研究者也不應該將其當作一般的正規文章進行分析（Weng et al. 2011）。Kontopoulos 等人也提出這類微網誌如用傳統情感字典的方法去分析通常正確率偏低的原因是因為其較日常情境的內容充斥著非正規用語（Jargon）、錯字和表情符號（Kontopoulos et al. 2013）。賴正育與楊亨利（2012）研究指出，影響微網誌中自我揭露與即時分享資訊的行為之動機包含人氣需求、社交需求、娛樂需求以及追求流行。Java 等（2009）也指出大部份 Twitter 微網誌的使用者使用 Twitter 聊他們的日常生活，尋找及分享訊息。因為微網誌是如此直接的個人有關，所以若去分析微網誌上的文章，也應能得到貼文者之撰文時個人意見與情緒。Pak 與 Paroubek（2010）建立了一個能夠自動從 Twitter 微網誌上收集及分類包含正向字、負向字與只有中立事實的情感語料庫，進而他們也利用了這個語料庫及貝氏分類法建立了一個能分辨正、負意見傾向的分類器。Weng 等（2011）則提出了一個能摘要微網誌文章的系統，它能綜合、分類並分析微網誌文章及其回應的內容；能把微網誌上作者和回應者們對事物類似對話的文字/圖像記錄，轉換成較易閱讀的摘要。Kontopoulos 等（2013）利用知識本體技術與 OpenDover 情感服務將微網誌上關於特別商品的內容進行情感分類。

可惜，文獻上還沒有真能挖掘出微網誌作者單一特定情感（如幸福、悲傷）的作法。

二、意見挖掘

意見挖掘（Opinion Mining）、情感分析（Sentiment Analysis） 或稱為情感分類（Sentiment Classification）的研究，指的是經由處理網路上或其他來源搜尋到的對某個商品或服務的文本資料，產生對此商品或服務屬性的列表（品質、功能等），並且將每個屬性的意見彙總找出正負評價程度（Dave et al. 2003; Hu & Liu 2004; Pang & Lee 2008; Liu 2010）。意見挖掘有很廣泛的應用領域，例如，拿來分析較正規的新聞文本，可以有助於瞭解競爭者的運作模式（Ye et al. 2006）；而分析網路使用者評價的文本，對了解如電子產品（Turney 2003）、從商品評論裏面找出商品的特徵與評價（Hu & Liu 2004a; Hu & Liu 2004b; Zhang & Liu 2011）、電影（Ye et al. 2006; Chaovalit & Zhou 2005）、餐廳評價（Yan et al. 2013）等商品的網路評價也有很好的效果。近年隨著社群網路的發達，更有以此類微網誌為分析對象，用來更深入挖掘是否有客戶間的品牌情感等的商業資訊（Kontopoulos et al. 2013; Mostafa 2013）。

以下，我們進一步從意見挖掘的三個層面：分析資料的層級、分析的演算、中文分析，去作相關的研究回顧。

（一）分析資料的層級

依照不同的分析目標及不同的使用目的，可將過去在意見挖掘上的文獻依其資料分析顆粒的大小，分成片語（Phase-level）、單句（Sentence-level）及通篇（Document-level）三種。不過，欲解決越大層級的意見分析，還是得先從較小的資料進行分析。有些較早的文章用一個屬性列表的方式，略過句子層級只去計算通篇文章出現主觀字詞的次數來決定通篇的意見（Pang et al. 2002）。但大部份想要解決通篇文意的學者還是會依片語、單句及通篇意見這樣的方法來進行研究。在分析句子等級的資料時，研究者通常都已能解決片語等級的意見傾向問題，研究者經由分析句子中片語間的詞性、位置、密度、樹狀關係結構等資料來加強單句意見傾向分析的正確性（Zhang et al. 2009; Missen et al. 2013；蕭瑞祥等 2015）。

本研究鎖定的研究目標主要為使用者發表在微網誌社群網路中的個人狀態，此類的文字通常是單句或數句的中文短文。所以在本研究的資料層級是處於單句的顆粒層次，將不需要討論通篇文意的部份。

（二）分析的演算方法

目前用分析的方法來區分意見挖掘的文獻，大概可以分為規則式（Rule-based）分析法以及學習式（Learning-based）分析法二種的演算方法。規則式的分析法通常需要一個專家定義好的情感字典，經由分析句子或文章與這些情感字的關係，來預測出作者的意見傾向（Wiebe & Riloff 2005）；也就是說它須掃描句子來確定是否符合特定情感特徵以找出意見傾向。

相對地，學習式的分析法則不需要預先定義的字典；它乃經由每次輸入已經被標記好結果的訓練資料去自我調整內部的學習參數，經過多次、全面的學習及正確率評估之後，便得到一個有預測能力的模型（Pang et al. 2002; Li & Wu 2010）。學習式演算法雖然通常需要較大量的時間和被打好分類的訓練資料來進行訓練，但因為其不需要專家先定義好的字典就可以進行分析，也有越來越多的研究者使用這類方法（Kontopoulos et al. 2013），本研究也是採用此種方法。

（三）中文的分析

雖然早期意見挖掘研究者所研究的語言是以英文為主的，但經過許多研究者的努力，目前也已有不少以中文為分析目標的文獻（Zhang et al. 2009）。中文與西方語言最大不同的地方在於中文並不像西方語言用空白將每一個單詞隔開；所以中文文章在分析之前，需要先經過「分詞斷句」的過程。另外中文有各式各樣的副詞，這些副詞很容易可以使句子有時變得舉足輕重，有時又模稜兩可。比方說：「這部電影好刺激，我看的好不快樂」，這樣的中文文字在分類器中會把「不快樂」找出來並歸成負面意見，但其實「好不快樂」的這種特殊用法，在中文中

的意思是「非常快樂」的正面意見。

為了解決中文所面臨到的問題，Yuen 等（2004）提出了一個使用中文單字語素及它與強烈情感字詞的統計關係，以期找出能區分中文字詞意見傾向的作法。Tsou 等（2005）延伸了這種作法，多考慮了中文文章中情感字詞的散布狀況、密度和強烈程度以加強正確率。

然而，就我們所知，針對網路文句特定情緒之發掘，不管是英文或中文，文獻上尚無相關的研究。

參、本研究提出的微網誌之情感分析

一、資料的收集與處理

基於上述文獻回顧，本研究試圖建立起一個能夠在微網誌的環境中取得作者意見傾向，以及其深入蘊含的情感之機制。首先，我們從知名的微網誌－新浪微博中利用網路爬蟲工具收集到真實的中國、香港、台灣明星名人分享的狀態，如「今天早上出發工作後看到的美麗日出，到現在依舊高掛天際，散發著光芒和溫暖著大地。」。這類從微網誌收集到的狀態句通常都只有單句，用以表達發表者當下的心情，而非實際討論什麼特別的內容。

我們利用網路爬蟲工具於三天內共收集到 572 則狀態句後，即進行資料前處理的過程；首先我們會過濾掉過長（組合句超過三句或是超過 150 字）及過短（5 個字以內）的文句。過濾掉過長的狀態句是因為除了會有上下句文意關係等較難分析的問題外，通常過長的狀態句並非太專注在表達心情，而是在說明事件。另外，過短的狀態句，如「睡了」，雖然它不一定無法充份表達作者很多的情緒，但它因為過於簡單，做為分析目標與訓練資料的成效不大；而且就算訓練期不考慮這種短句，將來正式分析時如果遇到這種簡單的句子也可以分析。所以在本研究中，還是選擇將其捨棄；在前處理的過程中，共過濾掉了 43 句的狀態句。

海峽兩岸的中文慣用語問題大概可以分為四類：(1)同實異名：如台灣用語的「軟體」在大陸被稱為「軟件」；(2)同名異實：如「土豆」這個詞，在大陸指的是「馬鈴薯」，而在台灣指的是「花生」；又如「窩心」這個詞，在大陸指的是「鬱悶」，而在台灣指的卻是「感動而開心」；(3)同中有異：如「菜單」這個詞，大陸與台灣都可以當作「點菜用的清單」，而大陸又可當作台灣的「系統選單」使用；(4)一方特有詞：如大陸用語中的「個體戶」，或是台灣用語中的「博愛座」都是只有一方特有的用詞。本研究資料收集的平台因為是由大陸開發經營，用戶大多數亦為大陸用戶，就算我們偏向於收取台灣名人的微網誌，但諸如「這是『神馬』(台灣：『什麼』) 爛名稱！」、「有很多『服務員』(台灣：『服務生』)

他們態度都不好…」這些兩岸用詞不同的狀況，在資料中還是頻繁出現。本研究的處理方式是為標記出來的大陸常用語，建立對應台灣用語的單向替換字典進行常用語的直接替換；雖然這是比較簡單的處理方法，但已能大部份解決本研究的問題。

最後我們人工從 529 句中挑選出用語與長度最適當的狀態語句共 100 句，利用中央研究院的分詞斷句系統[1] 進行分詞斷句。經由完成分詞斷字後，將其放至本研究所建置的語句情感意向測試網站（圖 1），進行語意傾向的專家評定，此網站有以下幾個功能：(1)依不同的策略平衡的挑選出應被測試的語句；(2)確保同樣的語句在同一次的受測任務中不會被重複測試；(3)具有諸如記錄答題間隔時間的功能，可做判斷胡亂答題受測者的依據；(4)能進一步直接收集觸發情緒的資料。

圖 1：本研究的中文情感意向測試網站的使用者畫面

利用這個網站，本研究一方面利用 Facebook 社群網站散播連接給社群成員邀請其填寫成為受測者，另方面並 email 邀請台灣北部某大學之學生自願當受測者進行意向判斷測驗。測驗的一開始是由受測者決定是不是要註冊（具名），如果受測者不希望註冊，系統會利用 Cookie 技術來減少同一受測者接受到同一訓練題目的狀況；如果受測者願意簡單註冊的話，則可以完全避免重複題目的問題。完

1 中央研究院的分詞斷句系統 SINICA CKIP 在 at http://ckipsvr.iis.sinica.edu.tw/

成選擇正式開始時，系統會依目前資料庫及受測者的目前的答題狀況，指派一個受測者還沒有答過，又可以最平均化整個資料評定作業的題目給受測者。受測者會在畫面上看到題目與提示，並依其知識在線上判斷此句話原作者想要表達的意見傾向。受測者可依感受到的意見傾向正負及大小程度從{非常負面、負面、中性、正面、非常正面、無法分辨}這些答案中擇一回答。當受測者作完傾向程度的判定之後，系統會對同一題，進一步的詢問這樣的文句會讓人們覺得作者現在具備有那些情緒。

在傳統心理學對人類基本情緒的研究，許多學者都提出過他們自己對人類情緒的分類（Ortony & Turner 1990），例如，生氣（Anger）、厭惡（Disgust）、興奮（Elation）、恐懼（Fear）、服從（Subjection）、溫柔感覺（Tender-emotion）、懷疑（Wonder）、焦慮（Anxiety）、幸福（Happiness）、難過（Sadness）、愛（Love）、喜悅（Joy）、驚訝（Surprise）、信任（Trust）或預期（Anticipation）等都有學者曾經提出認為是基本情緒。這些情緒中有些是我們這次的實驗情境上比較不在意的，像是服從、懷疑等。在本研究中我們採用人類五種會出現在面部表情的情緒當做基本情緒，這五種表情情緒不論什麼種族或文化程度，人類看到了對應的表情圖，不用語言的溝通，也都能明白這個人現在的感受。這五種情緒分別是：{喜（開心、想笑、Joy）、怒（憤怒、生氣、Anger）、悲（難過、哀傷、Sadness）、噁（厭惡、噁心、Disgust）、懼（擔心、恐懼、Fear）}；最後我們加上一個我們有興趣且較複雜的情緒{幸（感動、幸福、Happiness）}，一共六個情緒作為本研究的基本情緒。受測者對每一種情緒都可以選擇三種強度{無（預設）、中、強}。在受測者做完進一步的問題後，受測者可以選擇再做一題，或是中斷測驗。如果資料庫中已沒有受測者未曾回答過的問題，系統會顯示目前已沒有問題可供作答。

經過利用此網站，配合從作者的社群網站進行傳播，收集了二個星期的資料後。系統共收集了 7431 筆的意向資料，其中社群網站傳播上共有 1120 位被系統視為的不同受測者，共完成了 3021 筆完整的意向資料；大學生受測者共有具名者 128 人，共完成了 4410 筆完整的意向資料。

但這些資料當中，並不是所有的資料都是正確可用的。我們使用下列二個策略來做資料的清洗：第一是參考平均值與極端值的標準，首先我們先計算每一個答題的平均值與變異數，再將每一題的平均值加減變異數 3 倍的區間設定為正常區間。當答題者的回答如果在正常區內，那就是代表這個答題是正常的，否則就是極端值；當同一個答題者回答的問題有超過 1/5 是極端值的，那這個答題者的所有答題，我們都將其捨棄不用。第二個方式是參考回答的停留時間；當某題回答停留時間少於 3 秒時，我們認為該題的答案是沒有經過正常理性判斷，故捨棄之。在經過輔助資料的計算與清洗之後，我們得到了 4532 筆的有效回答資料

（社群 1122 筆 / 大學生 3410 筆），這些資料我們視為是從社群網站中爬蟲過濾得來使用者狀態句的專家意見基準值，將做為訓練資料及驗證資料的資料集。

二、利用 ConceptNet 組合文句特徵值

接下來我們要利用知識本體去決定每句語句個別的特徵。由於沒有類似的中文本體論可用，我們採用了 ConceptNet。ConceptNet 是一個以英語為主的日常生活常識知識本體；它是一個具有推論能力的一般常識知識庫，可支援現實世界的實際文字處理與推論工作、情感分析、類比決策（Analogy-Making）、文本摘要、情境內容擴張、因果投影、冷文件分類和其它語意導向的推論（Liu & Singh 2004）。利用 ConceptNet 的推論引擎，我們可以計算出目標語句中每一個經過英文翻譯後的標定字詞，與本研究選定的六個情感字詞的各別推論距離：喜（Joy）、怒（Anger）、悲（Sadness）、惡（Disgust）、懼（Fear）及幸（Happiness）的各別推論距離。我們只鎖定有可能會影響一個句子評價的字詞進行標定；通常動詞、形容詞、副詞、否定詞與程度量詞這幾類的字詞是主要可能會影響意見傾向的。

ConceptNet 的推論引擎可以計算出二個字詞的在日常知識的所有推論，像是如「Discount」和「Joy」，我們經由使用 ConceptNet 的推論引擎可以找出「Discount」到「Joy」共有 10 個路徑可以連接。例如其中有一個距離為 3 的推論路徑為：discount(Desires By) ➔ person (Have) ➔ human experience(Part Of) ➔ Joy。其意為：discount 是 person 所想要的東西，person 會有 human experience，而 Joy 是 human experience 的一部份。如果這個字詞和情緒字詞的關係越接近，推論的路徑就會越短。把所有的情緒字都計算一次後，就可以得到 Discount 這個單詞與各情緒間各有多少推論路徑，及每一條路徑的長度。本研究採用如下平均推論距離的方式去計算每一句話的情緒強度：

假設某句話裏含有 n 個被標記的字詞 $word_i$,$i = 1..n$ ，利用 ConceptNet，可各別計算出這一句話中每一個被標記的字詞 $word_i$ 與每一個基本情緒 $Emotion_j$ 的如下平均推論距離：

$$AVG.Dis(word_i, Emotion_j) = \frac{\sum Length.Of.Inference.Rule}{Count.Of.Inference.Rule}$$

假設某句話 X 裏有三個字詞{A,B,C}被標記，在本研究中為了要計算這句話的綜合特性，我們會計算出 18 個平均推論距離，分別是 *AVG.Dis* (A,喜)、*AVG.Dis* (A,怒)… *AVG.Dis* (B,喜)、*AVG.Dis* (B,怒)… *AVG.Dis* (C,喜)、*AVG.Dis* (C,怒)…。

當完成單句中每一字與每一情緒的平均推論距離計算之後，我們定義此句話的單一情緒平均推論距離為將同一個情緒所包含的字詞平均推論距離的平均：

$$AVG.Dis.Emotion_j = \frac{\sum_{i=1}^{n} AVG.DIS(word_i, Emotion_j)}{n}$$

在上述例句 X 裏，喜的情緒平均推論距離為 *AVG.Dis* (A,喜)+ *AVG.Dis* (B,喜)+ *AVG.Dis* (C,喜)再除於標記字詞總數 3；另外五種情緒也可以一併計算出。

最後我們利用以下公式定義此句話對此單一情緒的情緒強度：

$$EmoStr_j = \frac{U - AVG.Dis.Emotion_j}{U}$$

其中 U 是一個自定的、計算情緒強度是否小到可被忽略的推論距離門檻值；當一組字詞對某情緒字計算出的推論距離越長，也代表此字詞該情緒的強度也就越弱；利用 ConceptNet 計算出來的推論距離如果超過這個距離 U 的時候，我們視為「這組字詞對這個情緒是完全沒有強度的」。U 這個最大推論距離門檻，也能當作計算情緒強度的界限使用。在實驗中，經過不同數值的測試，我們發現當某字詞蘊含的情緒所計算出的推論距離大於 20 時，該字詞蘊含的該情緒已可被忽略，是以 U 值被設定為 30。

最後，對於每句話我們可以得到如表 1 的單句情緒特徵值，我們假設表 1 的例句為「[終於][拍完了]讓我[超緊張]的[主打 MV]，[太開心]啦！」。

表 1：某一語句對六大情緒的特徵分數

句子	喜樂	幸福	悲傷	厭惡	恐懼	憤怒
S001043	0.4857	0.5428	0.2457	0.2428	0.2171	0.1628

三、權重表格的最佳化

上一節的工作結束之後，我們可以得到每一單句的情緒特徵。在本研究的設定裏，我們認為這些與文字相關的情緒特徵是從字面上分析而來的，這是呈現的文字事實。但人們在解讀文字時，會依任務的不同，對文字的意見會有不同傾向的解讀；如在微博上看到的一句話，和在報紙上看到的同一句話，人們的解讀會有不同。也就是說，本研究認為除了文句的情緒特徵表（Characteristic matrix）$C_{1,6}$（表 1）外，另存在一個可以轉換不同意見傾向的權重表（Weight Table of Emotions）$WTE_{1,6}$（如表 2）存在，使得：

某句話套用某候選權重表的意見傾向數值 = $C \times WTE^T$ (Function 1)

C 為一句話的文句的情緒特徵分數矩陣；而 WTE 則是轉換不同意見傾向的權重表，這二個矩陣的維度皆為1×6，$C \times WTE^T$ 會變成一個純數值。

而這個意見傾向數值，可被使用當作某語句的意見值。

表 2：某個候選的情緒權重表

ID	喜樂	幸福	悲傷	厭惡	恐懼	憤怒
Can001	0.6	0.7	-0.7	-0.4	-0.5	-0.9

情緒權重表在本研究中被當做主要的訓練目標，其中的每一個權重元素的值域都在[-1~1]之間，在最佳化的過程中，會產生許多情緒權重表做為候選者，我們再從其中選出表現較好的進行下階段的最佳化。

如何順利得到一個有效的情緒權重？我們將利用演化策略的演算法（Beyer & Schwefel 2002）來進行最佳化訓練，圖 2 是本研究此演算法最佳化過程的示意圖。

演化策略演算法是一種模擬生物演化的演算法，這種演算法的每一個迭代（Iteration）我們也可叫它世代（Generation）。從一開始產生的一組初始個體開始，每一個世代都會依造物競天擇的設定，進行個體競爭、繁殖子代、淘汰個體及調整演化壓力等動作。在每一世代裏，都會進行競爭能力的計算，這個計算，競爭能力上較優秀的個體，會有較高的機會能繁殖下一代，也會有較高的機率可以留存到下一個世代；這並不是絕對的，只是說比較有機會而已。在每一個世代中，父代都會繁殖出一些下一代個體。這些下一代個體的基因組成，是由父母基因的交換以及基因的突變這二種變化所決定的。「虎父無犬子」這句話，放在這個可用來說明經由交換，父代和子代的樣子應該是較相似的；另外「歹竹出好筍」這句話，可用來說明經由突變，子代還是有可能突然變化的。當世上存在的個體數增加了，在每一個世代的最後，為了平衡資源，會進行個體的淘汰，競爭力較好的個體，有比較大的機率會被留下來，反之亦反。完成淘汰之後，可以視留下來個體的狀況，調整演化環境的壓力。如果希望個體的變異大一點，增加多樣性，可以把演化的壓力調大；如果希望個體的表現能夠較收斂一致，則可以將演化的壓力調小。完成了這個世代的演化，就可以進行下一個世代的演化。一直到滿足我們的演化需求為止。

我們的演算法分成幾個主要步驟來進行：

1. 亂數產生初始的權重表。

2. 主要的迭代訓練程序，包含：

(1) 生成子代權重表。

(2) 計算所有候選權重表的表現，並刪掉表現不佳的權重表。

(3) 調整演化壓力。

3. 終止條件檢查與正確率評估。

以下，針對這些步驟再逐一進行詳細的說明。

（一）初始的權重表

採用表 2 權重表的概念，利用上述得到的情感特徵向量，我們希望能訓練出在微網誌的情境下較適當的權重表。因為是權重的概念，所以這個表當中的每一個情緒權重元素的值域範圍是在[-1,1]之間。經由正負數及比例的調整，讓這個總合評價值可以去近似真正的評價值。

所以在本研究中，訓練權重表的第一步，是產生初始的權重表。我們採用的是利用亂數產生每個元素皆為[-1,1]間的 10 組（父代數量參數 μ=10）的權重值表做為訓練的一開始的初始權重表，計算各初始表的表現(*)，然後將這 10 組權重放進生殖池中。計算各初始表表現的方法容後說明。

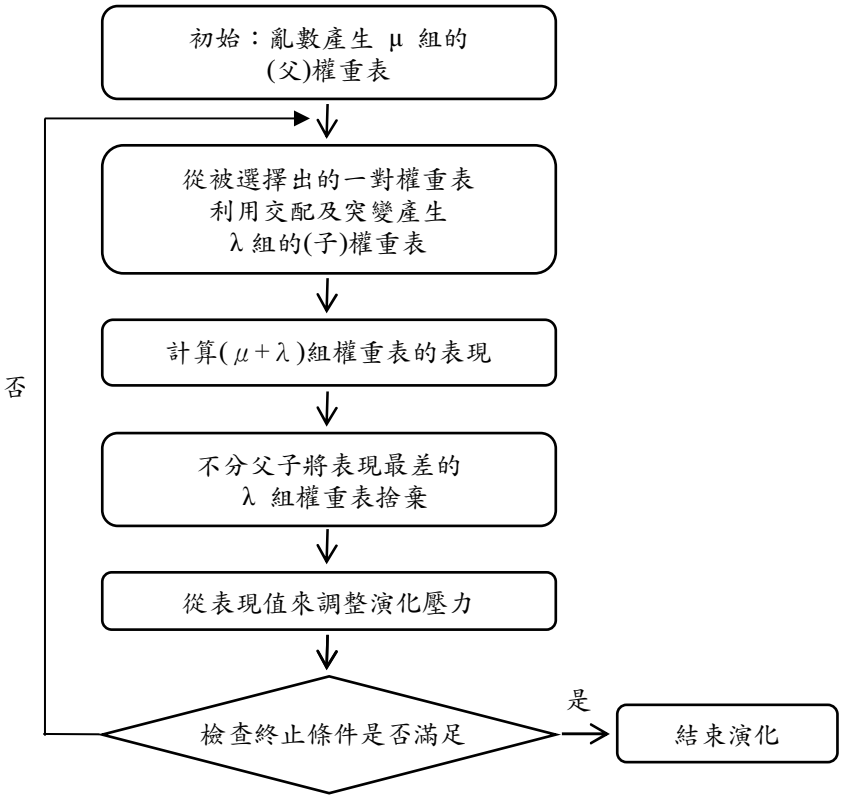

圖 2：本研究的最佳化示意圖

（二）主要的迭代訓練流程

迭代訓練流程為本研究的主要核心，這個迭代流程可能的進入點有二個，一是剛完成初始權重表之後，另一個則是在一代的訓練結束後，若還不能滿足終止條件時，均會再次進入這個迭代訓練流程。不管那一種狀況，在剛進入這個流程的時候，生殖池中都會有 10 個已經計算過表現的父代權重表。接下來的動作就是去計算生成子代權表。

1. 生成子代權重表的方法：首先，我們介紹一個參數 λ，這個參數代表的是在每一代演化時，所要新產生的子代數量。在本研究中我們選擇 λ=10，也就是說我們在這階段的目標是生成另外十組有遺傳到父代、但與父代又不盡相同的子權重表。

欲產生子代，得先決定要由生殖池中那一對父代來進行生殖。我們希望能有的模擬效果有二：(1)有比較好表現的父代，會有比較大的機率被選到；(2)表現較差的父代，雖然他的表現不好，但因為他還是有可能會含有一部份能改進整體表現的優良基因，我們也不希望他完全沒有機會被選到。

能夠同時滿足這二種效果的有效方法，是一種被稱為輪盤法的方法。本研究依目前生殖池中每一權重表的表現函數（*Performace*）的大小來決定輪盤分割的大小；因為我們表現函數的算法是去計算如果利用這個權重表配合所有的訓練資料，會產生的所有誤差加總；所以當權重表的表現函數值越小的時候，它的表現越好，它也應該有更大的機率被選擇出來。在選擇第一位父代，每個父代權重表會被選中的機率可以被如下表達：

$$P(WTE_i) = 1 - \frac{Performace(WTE_i)}{\sum_{i=1}^{\mu} Performace(WTE_i)}$$

類似地，選擇第二位父代的機率即為剩下權重表的表現函數所構成的輪盤機率。順利完成兩位父代的選擇後，接下來才能進行生成子權重表。

在本研究中，子權重的產生會經過二個步驟，每一對被選擇的父權重表，會同時產生一對的候選子權重表。第一個步驟是父母權重值對應數值的調整交換（Crossover）行為；交換行為完成後，二個候選子代中的每一個元素會再個別經過較小機率（0.5%）發生的突變（Mutation）行為進行調整後，即可完成生成。二個行為的詳細說明如下：

(1) 交換（Crossover）：二個父代情緒權重表的交換行為，指的是在交配池中的二個父代情緒權重表在產生候選的子權重表時，同一位置的二個情緒權重元素，會有一定的機率會進行互換（如圖 3），本研究這個「交換機率參數 cp」初始值設定為 50%。

ID	喜	幸	…	怒
Can001	0.6	0.3	…	0.8

父權重表

ID	喜	幸	…	怒
Can002	0.4	0.5	…	0.9

→交換

ID	喜	幸	…	怒
Can102	**0.4**	0.3	…	0.8

候選權重表

ID	喜	幸	…	怒
Can102	**0.6**	0.5	…	0.9

圖 3：情緒權重表「交換」的示意圖

(2) 突變：除了交換之外，父母權重表數值在遺傳給候選權重表時，每一個單一情緒元素還是有一個較小的可能會發行數值上的調整。本研究的「突變機率參數 mp」的初始值設定為 0.5%；「突變值幅度 mv」的初始值設定為 0.1。在演化壓力較大的狀況之下（也就是，最佳的權重表表現離目標值還很遠的時候），調整的幅度 mv 和機率 mp 都會比較大；反之，則較小。

在生殖池中的一對父代權重表，要開始產生子權重表時，在本研究中會先依序對每一種單一情緒位置用系統亂數去判定是否要進行交換，這樣的判定每一對會做十次。確定要交換的情緒位置，就會進行二種情緒元素的交換（如圖 3）。當完成交換後，就初步完成了一對候選子權重表；接下來對這一對權重表當中的每一個情緒元素都會進行突變判斷，也就是說，在每一對的生殖過程中，突變判斷總共會進行 12 次（一候選表 6 次）。每一個情緒權重元素的突變判斷都要用亂數決定「是否要突變」與「突變幅度值與正負」。在這裏會有需要處理每一個元素的邊界值：為了要讓每一個權重元素在突變之後，數值還能落在值域的[-1~1]之間，當突變後的數值會超過邊界時，系統會自動修正為其較靠近的邊界值。例如說已經是 0.95 的權重元素，當確定要增加 0.1 的幅度時，系統應能夠修正其只突變到 1.0 的值域邊界，在-1.0 的邊界也是類似的處理方法。當完成所有元素的突變判斷後，即能完成一對候選權重表的產生。

所以利用上述的方法，重複操作五次，我們就能順利的得到 λ=10 組的子權重表，並再一起加入生殖池中。這一階段完成後，生殖池中會有 10 組的第 n 代的父權重表，還有 10 組的子權重表。這 20 組權重表將會一起被衡量表現，在此時，我們稱這 20 組權重表，為 n+1 代的候選權重表。

2. 表現函數的計算：在策略演化的演算法中，各個候選權重表的表現值的計算方法是非常重要的項目。從單一候選權重表來看，一個候選權重表的表

現值，除了會直接影響到它是否還有機會持續存在於生殖池中，另外也會影響到它被選出來產生後代的機率；從整個演化策略演算法來看，表現值也有衡量整個訓練過程好壞的效果，比方說目前最好的表現值可做為衡量整個訓練過程是否可以停止的依據。甚至可以這樣思考，整個訓練過程的最終目的就是為了能找出最好表現值的權重表。

在進行訓練前，我們需要收集專家對於句子的評價傾向資料做為訓練的基準。如同我們在資料收集章節中所報導的，本研究從受測者所收集到的五級評價編碼資料，依清洗規則清洗之後，我們將其平均，可以得到這 100 句的專家評價傾向基準，我們標記其中第 j 句的狀態句其專家評價的傾向值為 $Target_j$。

另一方面，從前述 Function 1 中，我們利用某句話的情感特徵矩陣及一個情感權重表，即可以算出該句話在這個情感權重表下系統計算的傾向數值，我們認為，這句話在目前這個候選權重下的傾向分數，我們標記第 j 句子在權重表 WTE_i 的訓練傾向數值為 $Train_{ij}$。

因為我們希望讓這個訓練傾向數值能夠越貼近專家的判斷越好，所以一個情緒權重表 WTE_i 的表現值，我們可以定義為：

$$Performace(WTE_i) = \sum_{j \in \text{ all training data}} \left| Train_{ij} - Target_j \right|$$

其意義為：套用此權重表所得到的訓練傾向單句分數與該句專家評價基準值可計算出單句絕對訓練誤差，而這個權重表的表現就是所有訓練資料的單句絕對訓練誤差加總，這個表現值越大，指的是誤差越大，表現就越不好；相反的這個表現值越小，誤差也越小，表現就越好。

有了計算表現值的方法之後，在每個迭代進行到這個步驟時，我們都會同時考慮生殖池裏所有 20 組候選權重表的表現值。並且將其中表現最差的 10 組候選權重表移除；留下的 10 組表現較好的權重表就成為下一世代的父權重表。

3. 調整演化壓力參數：演化壓力在演化策略演算法當中，具有調節演化強度的作用。本研究的作法是，當目前的最佳表現（總誤差）低於 50%以下時，將突變機率參數 mp 調成 3%, 突變頻度參數 mv 調成 0.08，當總誤差低於 30%以下時，再將突變頻度參數 mv 向下調成 0.05。這樣的設定是為了當演化目標還很遠時，我們希望子代能有比較多的突變去試各種不同的基因組合；而當演化目標已離目標較近時，為了可以穩定收斂，我們則希望能有相對穩定的子代。這部份的設定會影響到最佳化過程是不是會太快收斂，造成陷入區域最佳解的現象。

（三）終止條件的檢查與正確率評估

當演算進行到這個步驟的時候，一次的迭代訓練運算已經完成。接下來就是要檢查目前的訓練成果，是不是已經滿足研究者所設定的停止條件。在本研究中，我們所設定的停止條件是整個系統要進行 100 次迭代的訓練才能終止。如果訓練次數未滿足此條件的話，系統流程將再回到主要訓練流程進行下一代的訓練；若條件滿足了，接著會進行權重表正確率的評估。

四、利用權重表進行正確率評估

訓練完畢後，最後我們會得到訓練完成的權重表（*WTE*），我們將利用剩下的 40 句測試語句資料進行正確率和誤差的計算。在前一節中我們已計算取得測試資料每一句的情緒強度矩陣（*C*）；將*WTE* 及 *C* 套用 Function 1 公式：

$$\text{測試資料套用最佳化權重表的意見傾向數值} = C \times WTE^{T}$$

我們可以計算出 40 句測試資料各自在此最佳權重表下意見傾向數值。舉例說 TestSent_1 計算出的意見傾向值如果是-0.125，這代表這是我們系統的預測值。但還不確定-0.125 的意見傾向值應該被歸屬在那一類較合適，在此我們採用的是利用訓練資料意見傾向值值域的方法，來確定系統意見傾向的歸類方法。

因為評價的分數在本研究中是以{非常正面(1)、正面(0.5)、中立(0)、負面(-0.5)、非常負面(-1)}來計算，所以在同一類的訓練資料狀態句中，我們可以得到一個平均評價的範圍，以這次的實驗的資料來舉例，被分類成{非常正面, 正面}這二類的句子平均值範圍是[0.12,0.91]，而被分類成{負面、非常負面}這二類的平均值範圍是[-0.94,0.04]。因為二區的範圍沒有重疊，所以我們可以定義如果測試資料計算出的評價值範圍是小於等於 0.04 為負面；大於 0.12 為正面，在[0.04,0.12]間的話則是中立區域。進一步要區分{非常負面}及{負面}也是用類似的方法，{非常負面}的句子意見平均值範圍是[-0.94,-0.79]，{負面}的句子意見平均值範圍是[-0.71,0.04]，可用-0.79 與-0.71 二數的中間值為負面程度的歸類分界。

利用這個歸類準則，我們可以知道前例的 TestSent_1 可被歸類成{負面}意見，我們也可以得到所有測試狀態句的系統分類結果。

如果這組權重表所產生的分類器在測試資料的表現不是可以接受的，就必須修改參數再從頭訓練，直到有可接受正確率的權重為止。

肆、情緒指數

一、情緒評價值的資料收集

在本研究的情境裏，我們不只希望能得到作者的正負意見傾向；更希望能進一步的指出，作者所想要表達的情緒類別及其強度。因此，在收集資料時，在受測者對於一個問題回答了正負傾向的問題之後，我們的資料收集系統會進一步的詢問受測者「作者想要表達的情緒，你覺得是什麼？」，受測者可以從五大基本情緒「喜、怒、悲、噁、懼」及我們有興趣的「幸福」中選擇他認為作者想要表達的一或多種情緒，每種情緒可細分成{無、輕微、強烈}三種；換句話說可視為受測者對 100 題中的每一個問題都又面臨了有三個選項的六個進一步的關於情緒的子問題。

在未清洗的資料中，我們收集到 7,431 人次（44,586 項次）的情緒評價資料。我們採用比清洗傾向資料更嚴格的條件來清洗這類的資料；(1)在清洗傾向資料被歸類成極端值的受測者，其所有的情緒評價資料也視為極端值捨棄不用、(2)因為情緒觸發的問題比傾向難的多，在頁中的停留時間我們較嚴格的將停留 5 秒以內就回答完畢的資料，視為太快就下決定，缺乏思考的答題，該題的資料捨棄不用。採用這樣的清洗原則來清洗資料後，我們得到 1,819 人次（10,914 項次）的有效情緒分項評價資料。

二、找到感興趣的情緒指數

有了情緒分項評價資料後，接下來就是做數值化的計算。每一位受測者評價每一句話的某個情緒評價，有三個可能{無(0)、輕微(1)、強烈(2)}。利用取得的資料，我們可以計算出每一句話蘊含的特定情緒的平均評價數值。在此我們以「幸福」的情緒指標為例，如果狀態句 $Sent_1$ 共有 30 個有效的受測者幸福評價，這 30 個數值的總合為 40，狀態句 $Sent_1$ 的幸福平均評價數值即為 1.333。我們可以計算出所有 100 句狀態文句專家評定的幸福平均評價值。

在本研究中，我們認為一句話的情緒平均推論距離是有能力去猜測專定評定的情緒平均評價值的。為了建立這個推估方程式，我們利用自然對數迴歸分析的方法進行處理，本研究將 100 句狀態文句的幸福平均推論距離當做自變數（x），去估計其各自專家評定的幸福平均評價值（y）。實際上我們對每一個狀態句都計算(x,y)的資料值對；x 為單句中的所有標註詞的平均幸福推論距離，y 即為上段所得的專家評定幸福平均評價值。完成這 100 對的資料對之後，我們利用自然對數迴歸分析，可以得到以下迴歸函式的參數(a, b)，使得整體的誤差是最小的。

$$y = a \times ln(x) + b$$

在實際要使用這個幸福指數時，使用者只要將某句話輸入到本研究的比對系統，因為幸福的平均推論距離是可以被計算出來的，也就是說系統可以得到目前輸入文句的 x_t 值，帶入得到的迴歸函式之後，就可以得到此推估方程式預估出該句可能蘊含的幸福平均評價值。

雖然幸福評價值很有意義，但是對使用者來說並不便於解讀。為了提供適合使用者解讀，我們以實驗的 100 句狀態文句的幸福情緒評價值排序，建立出如表 3，如此可將「幸福平均評價值」轉換為「幸福情緒指數」。表中的數值為實驗句子的幸福平均評價值；有兩句指數為 0，被判定是沒有幸福；若指數為 60 代表其推估的幸福評價值大於本實驗的 60 個句子的專家所認定的幸福值；若指數為 95 代表其幸福評價值大於本實驗的 95 個句子的幸福值。所以指數越高，該幸福情緒程度越大。當然，這只是一個參考，而且隨著不同的網站、乃至不同的時間應會有不同的參考表。

表 3：本研究的幸福情緒平均評價值轉換幸福指數表

評價值範圍	大於實驗的句數	顯示的幸福指數
0	0	0
$0 < x \le 0.01819$	2	2
$0.01819 < x \le 0.01853$	3	3
$0.01853 < x \le 0.01867$	4	4
⋮	⋮	⋮
$0.9815 < x \le 1.0183$	97	97
$1.0183 < x \le 1.2501$	98	98
$1.2501 < x \le 1.4260$	99	99
$1.4260 < x \le 2$	100	100

有了這個平均評價值轉換指數表，系統就能先將線上計算出的平均推論距離利用函數估計算出可能的平均評價值，再利用查表的方法顯示出適當的指數值，提供使用者能容易解讀的指數值。利用相同方法，我們也可得到其他的幾種情緒的指數，例如開心指數、憤怒指數、難過指數、厭惡指數、恐懼指數等。

伍、實驗結果

一、微網誌意見分析的實驗結果

首先我們亂數選擇 60 句狀態句（2,816 筆資料）做為實驗的訓練資料，剩下的 40 句狀態句（1,716 筆資料）做為測試資料。這些句子歸屬於那一個分類，是由受測者投票決定；也就是說當有較多的受測者認為這句話是{非常正面}，這句話的專家看法就屬於{非常正面}；而這句話的評價數值，則是由所有受測者評價的平均值決定。例如說：Sent001 共有 50 個人評價其意見傾向，其中有 26 個人認為它是屬於{正面}的，Sent001 的歸屬即為{正面}；而 Sent001 的傾向評價數值，則由這 50 個人給的評價利用{非常正面(1)、正面(0.5)、中立(0)、負面(-0.5)、非常負面(-1)}去計算平均。

從實際的專家資料收集結果顯示，在本實驗爬蟲過濾取得的 100 句狀態句中，所有的句子依多數投票原則都只被歸類到{非常正面、正面、負面、非常負面}這四個類別，完全沒有句子被多數的受測者歸類到{中立}區。這可能原因有二個：其一是微博狀態的短句多是在進行情緒的表達，這也是我們想要研究的目標；如果作者的心情「沒什麼起伏」，通常也不會發表狀態，是故我們從網站收集到的文章多是有情緒的文章。第二可能的原因是受測者的填答行為所造成。雖然我們在受測網站上有說明受測者可以選擇中立的選項，也希望受測者對傾向的強度能有回答；但我們從評價數值較中立的句子上的停留數據可以觀察到，受測者在這些句子的平均停留時間為較長的 7.3 秒，而有明顯正負評價數值句子的平均停留時間只有 4.6 秒。這意味受測者在這些句子上會停下來思考一下，當分辨出結果之後，受測者往往會比較偏向選擇{正向、負向}這二個結果，而不去會去選擇讓他暫時停下來考慮的中立因子；可能是因為還有{非常正向、非常負向}這二個選項會讓受測者下意識地稀釋了他會選擇完全中立的機會。

利用此 60 句訓練資料，利用第參節所述的最佳化方法，我們在第 91 代得到此次實驗最佳的權重矩陣，如表 5。這個最佳化 100 代的訓練時間共花了為 4 分 48 秒。原則上，每個微網誌網站只需要訓練一次這個權重表，除非網站內對語言字面的內涵有很大的衝突和前後改變，否則不需再次訓練。

表 5：實驗訓練出的權重表

權重表 ID	喜樂	幸福	悲傷	厭惡	恐懼	憤怒
Can918*	0.71	0.68	-0.56	-0.66	-0.37	-0.63

*Can918 為第 91 代的第 8 個候選權重表

因目的不同，若想關心偶像或研究目標是否有正向情緒，我們可以建立正向分類器（可分出「非常正向」、「一些正向」、「非正向」三類）；反之，若在意負向情緒，我們可以建立負向分類器（可分出「非常負向」、「一些負向」、「非正向」三類）。

將系統分類結果和專家判斷結果進行比對後，我們可以計算系統的正確率。由於這屬多分類問題，我們參考 Sokolova 與 Lapalme（2009）建議，分別提供巨觀平均（Macro-Averaging）與微觀平均（Micro-Averaging 的精準率）Precision）、召回率（Recall）、F1）兩種數值。表 6 呈現當考慮五程度分類時，我們的分類器的實驗分類狀況：也呈現若只像傳統一樣只分為三等級（正面、中立、負面）的程度時的實驗分類狀況[2、3]。根據該表數值，相較文獻分為三等級，我們正向分類器、負向分類器的 F1 數據，分別為 86.49%、88.24%。從文獻來看，目前意見分析研究情感分類的正確率往往在 60%～80%之間已是可以接受，如 Liu 與 Zheng 整理不同電影類別的分類器的正確率是落在 66%～84%之間（Liu & Zheng 2012）；Zhang 在 2009 年提出不同計算方式與不同類別商品的意見分析正確率亦是落在 64%～83.8%之間（Zhang 2009），是以本研究的正確率應已達令人滿意水準。

表 6：本實驗各種類分類器的正確率整理表

分類數	種類	平均種類	精準率	召回率	F1
五等級	正向分類	巨觀	84.59%	77.78%	81.04%
		微觀	82.5%	82.5%	82.5%
	負向分類	巨觀	78.0%	72.2%	75.0%
		微觀	80.0%	80.0%	80.0%
三等級	正向分類	-	100%	76.19%	86.49%
	負向分類	-	100%	78.95%	88.24%

2 若採 10 次交叉驗證〔即平分十等份，選一份（為 10 句）測試，剩下的（為 90 句）訓練〕三等級的正向分類器與負向分類器的精準率、召回率與 F1 值分別為 90.83%、71.67%、80.12%及 95.50%、83.14%、88.89%。若採 5 次交叉驗證〔即平分五等份，選一份（為 20 句）測試，剩下的（為 80 句）訓練〕三等級的正向分類器與負向分類器的精準率、召回率與 F1 值分別為 97.14%、81.35%、88.55%及 97.50%、84.17%、90.34%。

3 為了檢視資料是否有過度擬合（Overfitting）問題，我們再從同一社群網站，不同時間利用爬蟲再次取得了一組 30 句狀態句的資料，作為評估。我們得到額外取得的狀態句的正確率並沒有明顯下降的結論，也就是此分類器並沒有過度擬合的現象。(在五等級的正向分類器上，我們得到巨觀的精準率、召回率、F1 分數分別為 83.33%、77.98%、80.57%，而微觀的三項分數皆為 83.33%；同樣的分數在五等級的負向分類器上巨觀為 74.07%、74.71%、74.39%，微觀分數皆為 80.0%；三等級正向分類的精準率、召回率與 F1 分數為 100.0%、80.00%、88.89%；三等級負向分類的三項分數分別為 91.67%、84.62%、88.0%）

上述的分類器是綜合情緒的正負向分類器，其實，依我們的演算法也可以得到特定情緒的分類器，如幸福分類器、喜樂分類器、悲傷分類器、厭惡分類器、恐懼分類器、憤怒分類器等。

最後在演化策略演算的效果評估上，圖 4 展示了本研究訓練代數與訓練結果的收斂狀況。從關係圖上看，隨著進行的演化代數增多，平均誤差穩定的下降，在 40 代之後誤差就較穩定的降低到 30%左右，並不再有太大的起伏，雖然訓練到達 80、90 代後誤差還有下降的現象，但還在 25%的範圍中。

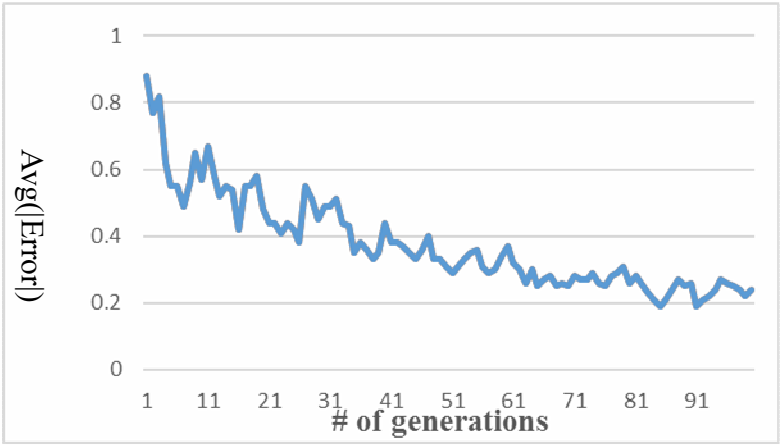

圖 4：訓練的代數與平均絕對誤差的關係圖

二、情感指數

我們以幸福指數為例，本研究可以對某狀態句利用 ConceptNet 先計算出對「幸福」的平均推論距離數值（x），而後利用迴歸分析，去找出依變數的「幸福估算值」（y）。我們將受測者回答的 10,914 項情感觸發資料跑迴歸分析，得到如下的迴歸式[4]：

$$y = -0.322\ln(x) + 1.0021\ , \quad R^2 = 0.8184$$

圖 5 為此 100 句狀態句推論距離、幸福評價值分佈狀況及迴歸線。這個估算的幸福評價值 y 如果是在 0 附近為低幸福感，在 1 附近為有幸福的感覺，在 2 附

4 對於此一迴歸式的測試，我們利用前述註腳 3 所提到為了檢驗 Overfitting 而額外下載的 30 句狀態句。利用專家評定其幸福程度{無、幸福、非常幸福}，並用以測試其幸福指數的正確率。我們發現若以評定等級來看，對額外下載的 30 句狀態句，有 24 句狀態句的幸福程度等級是完全正確的、其餘 6 句均只差一等級；若轉化為幸福指數的分數，此 30 句實際與預測的平均絕對誤差為 7.37 分；相對若沒有系統，亂猜的平均絕對誤差會是 33.3 分，導入我們的系統可以減少這個誤差 4.5 倍；若以平均實際值是 60.5，計算預測準確率為 87.82%。

近則為非常幸福。但由於這樣的估算值對一般網友可能不易理解，我們進一步以原訓練資料庫為參考，依據表 3，去找出這個估算值在原訓練資料 100 句中所的排名，給予一個百分比，當成一般人較易理解的「幸福指數」。

我們利用上述觀念製作了幸福指數線上分析器，此分析器可依即時根據計算出的平均推論距離，來去估算這句狀態所蘊含的幸福估算值，進一步給予幸福指數。如圖 6 所示，該句話的「幸福評價值」為 0.827，這個數值依原 100 句排名第 94，所以我們顯示「幸福指數」為 94。

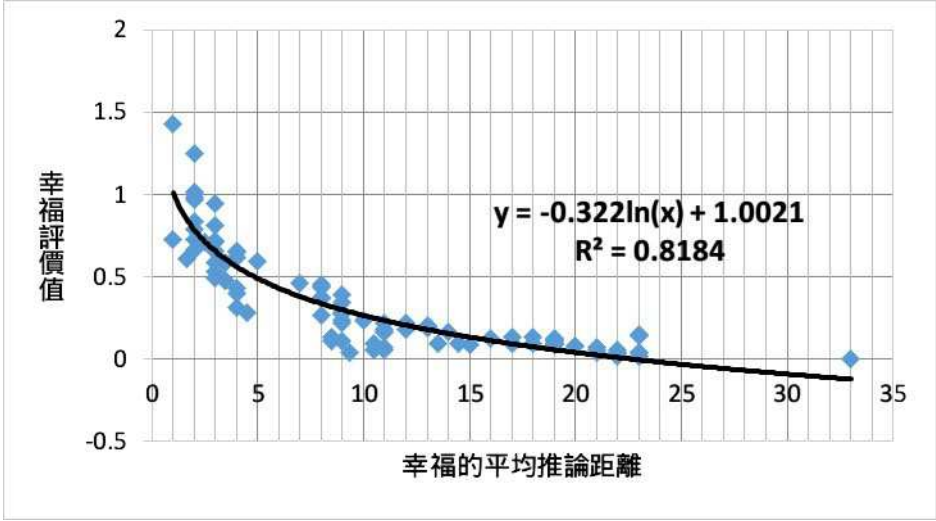

圖 5：實驗分布狀況及迴歸線

圖 6：幸福指數的雛形系統

此情感指數分析器的分析效能瓶頸主要在於對單句分詞斷句及推論距離的計算，表 5 的權重表只需針對網站作一次。訓練出的在第一版雛型系統中，我們的

系統平均回應時間為 13 秒（等待分詞斷句 7.2 秒、推論距離計算 5.8 秒）。為了改善這個問題，我們將已計算過的推論距離便依表 7 的格式儲存下來，當每次要計算推論距離時，先去檢查資料庫是否有可用的數值；若無，才會利用 ConceptNet 進行計算。當完成應用此技術後，系統平均回應時間減少到了 10 秒。如果後續研究想要進一步的改善效能，可考慮採用非線上的分詞斷句系統，如 ICTCLAS，並將常用字詞的推論距離都完成建檔後，此雛型系統效能應會有明顯的改善。

表 7：推論距離歷史資料的記錄

SN	情感字詞	比對字詞	規則數	總推論距離	平均推論距離
22	幸福	餐廳	4	15	3.75

雖然我們在這裡只有提出「幸福指標」當作例子，利用類似的方法與數據，我們也可以建立其它很有用的指標，如「悲傷指標」便有可能做為自殺預防偵測的情緒指標。

陸、結論

微網誌等社群網路已漸漸成為人們不可或缺的一種溝通管道。在其中具時間性、可自動化且有大量的意見資料是了解社群使用者的重要資料。本研究利用演化策略的最佳化方法提出了一個能有效分析微網誌狀態文句的意見分析架構。從實驗數據的結果看來，我們可以確定使用這個方法能夠有效的找到作者的意見傾向。另外本研究也提出了一個建立情感指標的方法，並建立了一個線上分析情感指標的雛型系統，這個系統可即時分析使用者所輸入的文句，提供即時情感指標的數值做為使用參考。

在學術貢獻上，在本研究中，我們除了提出了一個機器學習的演算法去訓練分類器，與一般機器學習演算法直接從分類的結果來訓練不同，我們提出的演算法保留了文句字面上的意涵而去訓練一個情感權重的對應表；我們認為在不同的微網誌社群、不同的主題，雖然說的是同一句話，都有可能會有很不同的意見傾向。這是本研究提出的這個演算法與一般演算法最大的不同之處。就我們所知，針對網路文句特定情緒之發掘，不管是英文或中文，文獻上尚無相關的研究，本研究對中文的情感分析提出語句蘊含某特定情緒的分析作法，可建立各種特定情緒的分類器，這應是首創。而且，不只是建立正向情緒分類器與負向情緒分類器；而若有需要，正負向亦可區分為非常正 / 非常負向、正 / 負向兩類別，這是以往未見的作法。另外，本研究也提出了一個較有效可以計算出特定情感指數

的簡單方法。

在實務貢獻方面，綜合本研究的發現與方法，對於製造商掌握現有商品的網路評論輿情的能力會有很大的幫助，對於警政體系掌握危險份子，或是社會工作單位由網路主動搜尋可能亟需協助的人也會有很大的幫助。不過，若要應用到別的微網誌，會需要重新搜集該微網誌的代表文句、重新訓練、與專家評估。

應用本研究開發的這二個工具配合簡單的關鍵字過濾，對於社群經營者可辨認社群中特定狀況的使用者，進而提供新的服務。經營者利用我們的工具，可建立出一系列的「情緒名單」，例如最近五次發言的狀態句的悲傷指標都超過 60、且悲傷指數的平均趨勢是增加的、狀態句中有出現與「死亡」或「厭倦」的相關字；這種人可能屬「自殺高危險群」的名單。社群管理者在系統自動辨視出這群人後，可進一步提供新的服務，例如，對其有互動的親友提供及時的系統提醒：「您的朋友○○○，最近看起來好像有點低落。您要不要試著多關心他一下？」或是系統可調整客製內容將較正面陽光的內容優先提供給高危險使用者。

對於當代人物的研究者而言，利用本研究所開發的二個工具，研究者也可以更容易的分析出研究標的當下可能有的意見傾向及情感狀況，如果配合時間區段，研究者也能更明確的了解研究標的心情的轉變。

後續研究者可思考對本研究之核心演算法的演化參數調整，也可在進一步調校情緒指數線上分析器之效率，以求其更為實用。更可考量將本研究所提的方法嘗試其他微網誌，或用到更長的網路貼文、部落格文章的分析，做到並非只是單句而是全文的情緒分析。

誌謝

作者感謝匿名評審及主編寶貴意見，本文為科技部補助研究計畫（NSC 101-2410-H-004-015-MY3）成果之一部份，特此致謝。

參考文獻

蕭瑞祥、姜青山、曹金豐、陳柏翰（2015），『基於中文語法規則的情感評價單元抽取方法之研究』，*中華民國資訊管理學報*，第二十二卷，第三期，頁 243-272。

賴正育、楊亨利（2012），『微網誌使用的需求動機及其影響』，*中華民國資訊管理學報*，第十九卷，第一期，頁 81-103。

Beyer, H.G. and Schwefel, H.P. (2002), ‘Evolution strategies-a comprehensive introduction’, *Natural Computing*, Vol. 1, No. 1, pp. 3-52.

Chaovalit, P. and Zhou, L. (2005), ‘Movie review mining: a comparison between

supervised and unsupervised classification approaches', *Proceedings of The 38th Hawaii International Conference on System Science*s(HICSS'05), Hilton Waikoloa Village, Island of Hawaii, Hawaii, U.S.A., January 03-06, pp.1-9.

Dave, K., Lawrence, S. and Pennock, D.M. (2003), 'Mining the peanut gallery: opinion extraction and semantic classification of product reviews', *Proceedings of International Conference of World Wide Web*, Budapest, Hungary, May 20-24, pp. 519-528.

Hu, M. and Liu, B. (2004a), 'Mining and summarizing customer reviews', *Proceedings of The 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)*, Seattle, WA, USA, August 22-25, pp. 168-177.

Hu, M. and Liu, B. (2004b), 'Mining opinion features in customer reviews', *Proceedings of the 19th International Conference on Artifical Intelligence* (AAAI'04), San Jose, California, July 25-29, pp. 755-760.

Kontopoulos, E., Berberidis, C., Dergiades, T. and Bassiliades, N. (2013). 'Ontology-based sentiment analysis of twitter posts', *Expert Systems with Applications*, Vol. 40, No. 10, pp. 4065-4074.

Li, N. and Wu, D.D. (2010), 'Using text mining and sentiment analysis for online forums hotspot detection and forecast', *Decision Support Systems*, Vol. 48, No. 2, pp. 354-368.

Liu, B. (2010), 'Sentiment analysis and subjectivity', in Nitin, I. and Fred, J.D. (Eds.), *Handbook of Natural Language Processing*, CRC Press, Taylor and Francis Group, Boca Raton, FL, pp. 627-666.

Liu, H. and Singh, P. (2004), 'Focusing on ConceptNet's natural language knowledge representation', *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems* (KES'2004). Wellington, New Zealand, September 22-24.

Liu, B. and Zhang, L. (2012), 'A survey of opinion mining and sentiment analysis', in Aggarwal, C.C. and Zhai, C-X (Eds.), *Mining Text Data*, Springer, New York City, pp. 415-463.

Java, A., Song X., Finin, T. and Tseng, B. (2009), 'Why we twitter: An analysis of a microblogging community', in Zhang, H. and Spiliopoulou, M. (Eds.), *Advances in Web Mining and Web Usage Analysis*, Springer, Berlin Heidelberg, pp. 118-138.

Missen, M.M. S., Boughanem, M. and Cabanac, G. (2013), 'Opinion mining: reviewed from word to document level', *Social Network Analysis and Mining*, Vol. 3, No. 1, pp. 107-125.

Mostafa, M.M. (2013), 'More Than Words: Social Networks' Text mining for consumer brand sentiments', *Expert Systems with Applications*, Vol. 40, No. 10, pp. 4241-4251.

Ortony, A. and Turner, T.J. (1990), 'What's Basic About Basic Emotions?' , *Psychological Review*, Vol. 97, pp. 315-331.

Pak, A. and Paroubek, P. (2010), 'Twitter as a corpus for sentiment analysis and opinion mining', *Proceedings of the 7th International Conference on Language Resources and Evaluation* , Malta , European Union, May 19-21, Vol. 10, pp. 1320-1326.

Pang, B. and Lee, L. (2008), 'Opinion mining and sentiment analysis', *Foundations and Trends in Information Retrieval*, Vol. 12, No. 1-2, pp. 1-135.

Pang, B., Lee, L. and Vaithyanathan, S. (2002), 'Thumbs up? Sentiment classification using machine learning techniques', *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP). East Stroudsburg, PA: Association for Computa-tional Linguistics, July 6-7, pp. 79-86.

Sokolova, M. and Lapalme, G. (2009), 'A systematic analysis of performance measures for classification tasks', *Information Processing and Management*, Vol. 45, No. 4, pp. 427-437.

Tsou, B.K.Y., Yuen, R.W.M., Kwong, O.Y., Lai, T.B.Y. and Wong, W.L. (2005), 'Polarity classification of celebrity coverage in the Chinese press', *Proceedings of International Conference on Intelligence Analysis*, McLean, VA, USA., May 2-6.

Turney, P.D. and Littman, M.L. (2003), 'Measuring praise and criticism: Inference of semantic orientation from association', *ACM Transactions on Information Systems*, Vol. 21, pp. 315-346.

Weng, J.Y., Yang, C.L., Chen, B.N., Wang, Y.K. and Lin, S.D. (2011), 'IMASS: an intelligent microblog analysis and summarization system', *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations*, PA: Association for Computational Linguistics, Oregon, Portland, June 19-24, pp. 133-138.

Wiebe, J. and Riloff, E. (2005), 'Creating subjective and objective sentence classifiers from unannotated texts', *Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics* (CICLoing 2005), Lecture Notes in Computer Science, 3406, Mexico City, Mexico, February 13-19, pp. 486-497.

Yan X., Wang J. and Chan M. (2013), 'Customer revisit intention to restaurants: Evidence from online reviews', *Information Systems Frontiers*, Vol. 17, No. 3, pp.

645-657.

Ye, Q., Shi, W., and Li, Y. (2006), 'Sentiment classification for movie reviews in Chinese in improved semantic oriented approach', *Proceedings of the 39th Hawaii International Conference on System Sciences*, Hyatt Regency, Kauai, Hawaii, U.S.A., January 4-7, Vol. 3, pp. 53b-53b.

Yuen, R.W.M., Chan, T.Y.W., Lai, T.B.Y., Kwong, O.Y. and T'sou, B.K.Y.(2004), 'Morpheme-based derivation of bipolar semantic orientation of Chinese words', *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*, East Stroudsburg, PA: Association for Computational Linguistics, Geneva, Switzerland, August 23-27, pp.1008-1014.

Zhang, C., Zeng D., Li, J., Wang, F.Y. and Zuo, W. (2009), 'Sentiment analysis of Chinese documents: From sentence to document level', *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 12, pp. 2474-2487.

Zhang, L. and Liu, B. (2011) 'Indentifying noun product features that imply opinions', *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (ACLHLT '11), Oregon, Portland, June 19-24, pp. 575-580.