

使用情緒分析於圖書館使用者 滿意度評估之研究

A Study on Library Users' Satisfaction Evaluation Using
Sentimental Analysis

郭俊桔

June-Jei Kuo

國立中興大學圖書資訊學研究所 助理教授

Assistant Professor

Graduate Institute of Library and Information Science

National Chung Hsing University

張育蓉

Yu-Jung Zhang

亞洲大學圖書館 行政助理

Administrative Assistant

Library, Asia University

【摘要 Abstract】

本研究採用情緒分析探究社群網路上（如噗浪等）之使用者對於圖書館的意見屬於正向或負向，並以圖形方式顯示評估結果，可及時提供圖書館業務調整與改善使用者滿意度。蒐集網路上圖書館的使用者意見進行人工標記，建立相關情緒分析辭典後，使用 KNN、NB 和 SVM 等分群工具探討圖書相關語料分類的效果外，提出導入程度詞、否定詞與情緒詞權重、及情緒極性與情緒類別等兩種情緒分析方法。研究結果顯示 KNN 具有較佳的分類效果以及以情緒極性與情緒類別得到較佳的圖書館使用者

滿意度之評估結果。進一步，也證實所有建置之情緒分析辭典的可靠性。

This paper employs the sentimental analysis to study the library users' satisfactory evaluation, where users' opinions were classified as positive, neutral and negative, respectively. To provide the immediate library service improvement, the graphic representation is further used to show evaluation results. After collecting the library users' opinions from micro webs, i.e., Plurk, to be the corpus, the corpus was then pre-processed and annotated manually. Moreover, all the related dictionaries for sentimental analysis were constructed. accuracy each category, were employed to cluster the corpus into five categories. Furthermore, two new sentimental analyses for library satisfactory evaluation were proposed, which include introducing degree words, negatives and sentimental words weights, and introducing sentimental polarities and categories, respectively. Experimental results showed that both KNN and sentimental analysis introducing sentimental polarities and categories obtain better satisfactory evaluation results. Furthermore, all the constructed dictionaries for sentimental analysis were shown their feasibility as well.

[關鍵字 Keywords]

圖書館使用者；滿意度評估；情緒詞；情緒分析

Library users; Satisfaction evaluation; Sentimental words; Sentimental analysis

壹、緒論

每所大學都設有圖書館，支援學術研究、教學與推廣服務，可見圖書館在大學教育中占有重要地位，因此大學圖書館向來被喻為大學的心臟。由於大學數量激增，使得大學品質良莠不齊，追求卓越、提昇品質的訴求發聲，引發高等教育的改革並朝著提升大學教育品質的方向發展。近年來，各大學相互競爭激烈，急欲發展成為教學卓越，甚至是國際頂尖一流的大學，尤其需要圖書館作為後援，因此圖書館服務品質的良窳極為重要，而為此圖書館必須進行評鑑，具體呈現圖書館服務的成效與自身的核心價值。

圖書館屬於非營利機構，其經營目標乃是有效地應用有限的資源，

以達到最大的服務目的。由於圖書館所提供的產品除了有形的圖書及設備外，最主要的乃是無形的資訊、知識與服務（彭金堂、鄭姿均、卓欣姿、陳幸慈、程瑞琪，2010）。因此，要評估圖書館績效不能只以客觀的量化資料為準，如圖書借閱率、藏書量、或空間大小等，而更應重視使用者的整體感受與滿意程度。一般而言，許多研究在探討圖書館績效時，經常以一些量化資料去評估，而且一些具權威性的統計調查也都僅以量化資料為主，如圖書館每年進行的例行統計報告，甚少對使用者的真正需求與滿意度加以瞭解。圖書館評鑑從 1950 到 1960 年代，偏向以圖書館之輸入資源，如館藏、人員、設備等量的評估；1970 到 1980 年代，以輸入資源經由作業程式而轉化成的服務成效為評鑑重心；1990 年代開始，受管理學影響，圖書館開始強調服務品質與重視使用者需求，圖書館評鑑也轉以使用者需求為主，因此使用者滿意度開始受到重視（盧秀菊，2003）。瞭解使用者需求乃是目前圖書館經營管理的重要方向，因此使用者滿意度成為圖書館營運成效評估的主要指標。

隨著電腦科技與網際網路的快速發展，改變了圖書館傳統服務的型態以及使用者利用圖書館的習慣，圖書館體認到唯有從使用者服務的角度切入圖書館滿意度研究，提供客製化服務，瞭解使用者的真正需求與意見，方能達到服務使用者之最終目的，並使圖書館達到提供盡善盡美的館藏與服務之境界（吳明德，2003；Drake,1993; Goodall, 1988）。傳統圖書館使用者滿意度評估方式多以問卷法、訪談法或焦點團體法等來瞭解使用者對圖書館滿意之處，以及須改進之處。然而，傳統方法有些實施上的限制，欲設計出具信效度的問卷須請專家檢視並進行前測，從問卷發放到分析結果出爐通常會耗費幾個月的時間；訪談法與焦點團體法要找到具有代表性研究對象並不容易，製作訪談大綱到錄音檔謄稿到分析結果也是一個費時又費力的過程。由上述可知，傳統使用者滿意度評估方法實施一次耗費的成本過高，無法立即反映出圖書館問題，也無法確實得知使用者之不滿意事項，並及時改善。本研究嘗試採用情緒分析解決前述缺失。

微網誌的蓬勃發展使得許多人們悅於在網路上以文字、圖片、影音等媒介抒發心情、分享對某些人事物的經驗或看法，這些意見有正面也有負面。因此，為解決傳統圖書館滿意度評估的課題，本研究提出一個自動使用者滿意度評估的架構，及時呈現使用者滿意度以供圖書館參考。本研究採用網路上針對圖書館所提出的意見，運用分類預測方法將意見依照不同面向加以分類，並以句為單位參照情緒分析辭

典與情緒分析方法，判斷情緒極性並彙總整句情緒之極性分數為各面向的滿意度。接著，與傳統滿意度評估所得結果相比對，探討本研究方法之可行性。本文第一節陳述研究動機與目的；第二節為文獻探討；第三節闡明研究方法；第四節為研究結果與分析；最後，第五節針對本研究的結論做綜合闡述後，提出未來可能的研究方向。

貳、文獻探討

一、圖書館使用者滿意度評估方法

近年，Web 2.0 概念廣泛傳播與應用，加上網際網路普及與使用者對網路使用習慣的改變，大學圖書館使用者對於圖書館提供的各項服務也產生了相當程度上的變化與期待，服務是提供給使用者，使用者直接接觸服務，欲評估圖書館服務效能，自然必須重視使用者的感受與意見。圖書館使用者滿意度常見的評估方法有問卷調查法（Questionnaire survey）、訪談法（Interview）、焦點團體法（Focus Group）、觀察法（Observational survey）以及 LibQUAL+®等。

（一）問卷調查法

問卷調查法是最普遍且最容易施行的研究方式，目前絕大多數圖書館使用者滿意度之相關研究皆以問卷調查法進行調查。問卷調查法量表編製容易，題目少，信度高與效度亦高，使用上相當易於上手，但研究者在設計問卷內容時卻易陷入主觀立場，使得問卷問題受限，使用者無法選擇適當的內容，所蒐集的資料僅能呈現使用者滿意與不滿意之項目與數量，無法表示使用者對不滿意項目的原因與建議（張郁蔚，2007；楊國樞、文崇一、吳聰賢、李亦園，1990）。要設計出信效度皆佳的問卷並不容易，正式問卷發放前會先測量問卷的信度與效度，為瞭解研究主題是否確實存在問題，且每個題項是否能確實反映出研究問題。此外還須進行前測，問卷內容應淺顯易懂，題項意涵必須讓使用者清楚瞭解，文字闡述避免模稜兩可。回收問卷經常會有回收率難以掌握的問題，尤其是以郵寄方式發放問卷，除非研究者親自發放問卷，否則無法強制使用者回覆問卷，而回收率過低則會產生研究結果是否對研究母體具有代表性的質疑。使用者滿意程度會隨時間不同而有所變化，傳統圖書館滿意度調查方法，無法觀察出滿意度之各個面向變化的趨勢。

(二) 訪談法

訪談法是一種相當常見的質性研究方法，除一對一個人訪談外，尚包括團體訪談之焦點團體法。採用訪談法獲取的資訊主要為使用者對於某些事物的個人觀點（張郁蔚，2007）。訪談法的執行需要投入較多人力、時間與金錢成本，如訪談後須將錄音轉譯成文字形式並謄寫訪談內容報告、邀約訪談花費的時間、聯絡訪談事宜所需的電話費、前往訪談地點的交通費等等成本項目。

(三) 觀察法

觀察法屬於一種科學的研究方法，由觀察者從正面觀察被觀察者的語言、臉部表情、肢體語言、情緒表現、外在行為等，或從側面打聽被觀察者的言行表現等訊息，以便得知受試者的滿意度（劉峻霖，2000）。研究者的判斷尤其重要，必須進行事前的教育訓練，而且觀察員不能過於投入，以免產生干擾效果。

(四) LibQUAL+®¹

LibQUAL+®是1999年美國研究圖書館學會（Association of Research Libraries，簡稱ARL）進行一項研究計畫所發展出評量圖書館工作績效與服務品質的新方法，以SERVQUAL為基礎，經過多次修訂，最終確定分為三項構面以及22項核心指標，其構面為服務影響（Affect of Service）、資訊控制（Information Control）以及圖書館實體環境（Library as Place）。由於服務不但是一種無形的產品及非實體的東西，更是沒有固定的標準模式，不僅是產品本身，更強調無形的服務品質，服務品質的優劣則無法以具體的量化數字來判斷，而是視服務對象感覺滿意與否而定。

整體而言，傳統評估圖書館使用者滿意度的方法皆必須耗費人力與時間蒐集資料，通常都是幾個月之後獲得研究結果，無法立即得知近期使用者對圖書館服務滿意與否，而且圖書館例行的使用者滿意度調查通常都是1年僅1次，使用者的聲音才會被聽見，也才會瞭解圖書館服務的成效。由於圖書館無法即時反應與處理問題，造成使用者滿意度下降。

1 LibQUAL+®為美國研究圖書館學會的一項計畫，詳細資料參見計畫網站 <http://www.libqual.org>

二、情緒分析與滿意度

隨著 Web 2.0 概念與應用普及化，愈來愈多使用者悅於在網路上以文字、圖片、影音等媒介抒發自己的心情或發表對於某些人事物的看法，這些數位內容包含多元化主題的內容和蘊含豐富的主觀意見，以及個人化的情緒和想法，與傳統媒體生產相似度高的內容或具有某種程度的客觀性的事實陳述有所不同。近年相當熱門的微網誌（MicroBlog），如噗浪²（Plurk）和推特³（Twitter），其文章沒有完整結構，加上規定須在有限字數下發表文章，使用者須在簡短範圍內完整表達其情緒與意見，因此，更加突顯使用者的情緒與意見表達。每日有成千上萬的使用者透過微網誌分享他們的意見，可以說微網誌是進行情緒分析所需的豐富資源。

因此，從網路大量資料中分析出對商品業者或消費者有用的資訊，或者瞭解某個人或組織對某個東西或某件事的看法是相當重要的，此為情緒分析（Sentimental Analysis）。圖書館的商品就是所提供的服務，因此要如何利用網路資源瞭解使用者對圖書館服務的意見與滿意程度，就是本研究之主軸。

（一）情緒分析之相關研究

情緒分析是指透過一些主題或文件的整體脈絡極性去判斷說話者或作者的態度。此種態度可能是作者的評價或意見、情緒狀態，或是預期與使用者間的情感交流。廣義而言，情緒分析又可稱為意見分析（Opinion Analysis），是近年自然語言處理與資訊檢索領域新興且相當熱門的研究主題之一。情緒分析之相關研究包含情緒分析辭典的建置、情緒之辨識與分類（Sentimental Identification and Classification）以及情緒辨識方法（Sentimental Identification Method）等。

1. 情緒分析辭典的建置

目前情緒詞辭典的建置多半是先由大型資料庫中蒐集大量詞彙，再以人工標記情緒詞與情緒極性。然而，由於人工標記方式較耗費時間、人力與金錢，且標記的詞彙涵蓋率有限，因此為了要蒐集到更

2 噗浪為新興的微網誌，於 2008 年由雲惟彬先生與其朋友共同創立，詳細資料參見網站 <http://www.plurk.com/>

3 推特是基於 web2.0 技術的一個社交網路及微型部落格服務，詳細資料參見網站 <http://twitter.com/>

多情緒詞彙，發展出兩種情緒辭典建置的方法：字典法（Ku & Chen, 2007）與機器學習法（吳蕙欣，2011；李孟潔，2009）。字典法是將人工標記完成的情緒詞作為種子（Seed），再利用有語意架構的詞典（如 WordNet⁴、HowNet⁵ 等）進行擴展比對，找出種子詞彙的同義詞或下位關係詞，全部標記為與種子詞彙相同的極性，便會產生收錄詞彙量較多的情緒辭典。字典法能擴展的範圍有限，且相同詞彙出現於不同主題，可能會產生不同的極性，字典法無法準確標記出極性，因此有研究開始嘗試利用機器學習法自動建置情緒辭典，以人工標記完成的情緒詞作為種子，蒐集網路上大量語料（如新聞），再藉此計算情緒候選詞的 PMI（Point-wise Mutual Information）值，若 PMI 值為正數，歸類為正向情緒詞，反之歸類為負向情緒詞，而趨近於零值則歸於中性情緒詞（Manning & Schütze, 1999）。然而，訓練資料不均衡容易造成機器學習結果偏頗，大多數的研究仍使用字典法。

2. 情緒之辨識與分類

情緒之辨識與分類的研究又包含三個子研究：主觀性分析（Subjectivity Analysis）、情緒極性分類（Polarity Classification），以及情緒極性之強度辨識（Strength），屬於情緒分析中最基礎之研究。情緒辨識的第一步主觀性分析就是要先找出文本中個人情緒表達的主觀意見或是針對事實陳述的客觀描述，適用的範圍是以句子層次、段落層次和文件層次作為情緒判斷單位（Pang & Lee, 2008）；即是要辨識文本中哪些字詞或句子具有表達意見、立場、感受、情緒等意思（李孟潔，2009）。本研究使用微網誌上的發文為實驗語料，語料中的各語句皆可視為情緒分析的基本單位。再者，情緒極性分類的目的是辨識個人對特殊事件的感受，將詞語、句子區分為正向（Positive）、負向（Negative）或中立（Neutral），以瞭解文本中的觀點與情緒傾向（楊昌樺 & 陳信希，2006；Hu & Liu, 2004）。

目前針對網路使用者評論文章之情緒分析研究很少考慮到使用者的情緒強度，尤其強度是衡量產品或服務的整體品質之相當重要的因素（Wilson, Wiebe, & Hwa, 2004）。一般而言，不同評論的相同情緒詞應該有完全不同的情緒強度，儘管兩個情緒詞可能有相同的情緒極性

4 Wordnet 是美國普林斯頓大學認知實驗室所創建的英語語意辭典，詳細資料參見計畫網站 <http://wordnet.princeton.edu/>

5 董振東教授 1988 年所創建的漢語語意辭典，詳細資料參見計畫網站 <http://www.keenage.com/>

(Lu, Kong, Quan, Liu, & Xu, 2010)。有鑑於情緒詞大多是由形容詞組成，許多研究便根據意見詞中形容詞的強度去得知情緒強度。語言學中，副詞經常用於修飾形容詞，在使用者提出意見與評論時，採用的副詞可以增加形容詞的強度。

3. 情緒辨識方法

情緒辨識目前有兩個主要的方法，一是字典法 (Lexicon-Based Method)，二是機器學習法 (Machine Learning Approaches)。字典法是利用事先已具有標示情感訊息之情緒辭典，對一個未知情緒極性的新詞，從現有的辭典或知識庫比對與該詞語意相近，且出現在情緒辭典中的若干詞彙，根據這些詞彙去決定未知新詞的極性，甚至可推斷出句子、段落、文本或網路語料的情緒極性。另外，機器學習法是利用已分類極性的語料進行訓練，歸納出語料特徵，再根據特徵規則去自動判斷學習字詞、句子或文本的情緒極性 (孫瑛澤、陳建良、劉峻杰、劉昭麟、蘇豐文，2010)。李啟菁 (2010) 採用字典法比對分析文章的程度詞與情緒詞，並進一步計算出文章整體評價，實驗結果顯示以字典法進行情緒辨識的準確度與人工閱讀給予的評價相似。

(二) 情緒分析之相關應用

意見中的情緒資訊不論對於提供商品的業者、使用服務的客戶、人際關係或社群網路分析都具有高度的應用價值，尤其資訊科技發達，網路資訊數以百計不斷出現，人們的想法、經驗或意見四處散佈，造就一個情緒分析的有利場所。例如，對企業而言，瞭解企業本身形象及其產品與服務之優劣在市場行銷與顧客關係管理上都是很重要的一環。為了在競爭激烈的商場環境中擁有立足之地，形象是企業很重視的一項無形資產，代表著企業的信譽、產品質量、人員素質、股票的漲跌，是公眾對企業的整體認知與評價，企業無不積極瞭解大眾對企業在整體社會觀感上的定位與價值 (孔誠志，1998)。情緒分析除了用以判斷使用者的意見和情緒之外，還可以進行情緒趨勢分析。所謂情緒趨勢是指某段時間內，使用者對某個話題的情緒變化走向。楊昌樺、高虹安與陳信希 (2007) 基於部落格文章通常具有時間標記，由新至舊的排列方式，透過部落格蒐集使用者觀點，不僅分析意見極性，也可分析出不同時間域使用者的特徵，並且以視覺化的方式呈現，更能清楚顯示使用者對某個議題或人物的情緒反應。

總而言之，情緒分析已應用於許多領域，不論在情緒辨識或呈現

上皆有很好的成效，然而，圖書資訊學界尚未有相關研究，因此本研究採用之繁體中文情緒分析辭典由實驗語料中自動擷取而建立。建立此一情緒分析辭典需先將語料之情緒詞進行人工辨識與極性標記，標記的來源為結合 NTUSD⁶ 且具情緒標記之 E-HowNet（李政儒，2011）。針對實驗語料中各句的情緒詞，參照情緒分析之字典法以及導入副詞強度，提出兩種新的句情緒極性分類方法，自動衡量各句的情緒傾向並以圖形化方式針對五大面向（館員、館藏、服務、設備和空間與環境）呈現使用者的情緒傾向，以供圖書館業務及時調整與改善。

參、研究方法

本研究提出如圖 1 所示之系統架構圖，主要分成 2 大部分，特徵詞擷取與情緒分析辭典建立過程以及情緒分析過程。特徵詞擷取與情緒分析辭典建立過程中，首先蒐集相關語料，經中文斷詞和人工標記後，可由人工或程式擷取預測分類用的特徵詞辭典和情緒分析用的情緒分析辭典。另一方面，情緒分析過程中，測試語料經中文斷詞和預測分類模組，分成館員、館藏、服務、設備以及空間與環境等 5 大類，並分別進行情緒分析後，進行各類別的情緒綜合評分。在以下各節中將分別說明語料蒐集、預測分類和情緒分析的執行方法。雖然分別以人工處理方式或軟體工具說明，若將各模組程式化即可成為自動化系統。

一、語料蒐集

根據創市際⁷（InsightXplorer）（2011）市場研究顧問公司的 7 月份月刊報告書中針對台灣地區微網誌服務使用狀況進行調查，Plurk 為台灣網友到達率最高的微網誌平台，且花費時間最多，而 Twitter 目前

6 NTUSD 為台灣大學資訊工程學研究所自然語言處理實驗室所建置的中文情緒詞辭典，詳細資料參見計畫網站 <http://nlg18.csie.ntu.edu.tw:8080/opinion/index.html>

7 創市際採用網路測量研究互動資料庫、線上市場研究等多項專業技術，提供行銷市場整合性的解決方，案詳細資料參見網站 <http://www.insightexplorer.com/index.html>

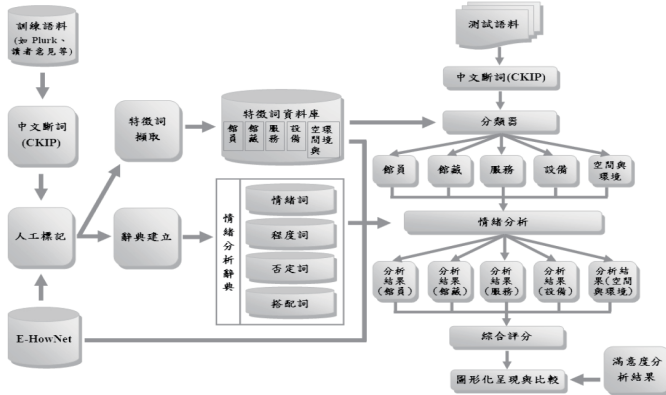


圖 1 系統架構圖

尚無繁體中文介面，微博⁸中少有提及台灣地區大學圖書館之文章。再者，Facebook 目前最多只能搜尋過去 30 天發佈的內容⁹，無法完整蒐集 Facebook 上針對圖書館服務之使用者意見，因此本研究選用 Plurk 作為語料來源。實驗語料和蒐集語料之對象為參照 2010 年獲得教育部邁向頂尖大學計畫¹⁰補助名單，蒐集 12 所頂尖大學圖書館於 Plurk 上的語料。此外，針對設有圖書資訊學相關系所之 6 所大學圖書館蒐集 Plurk 語料，共有國立台灣大學圖書館、國立政治大學圖書館、國立台灣師範大學圖書館、淡江大學圖書館、輔仁大學圖書館、國立中興大學圖書館。合併重複者，總計蒐集 14 所大學圖書館 Plurk 語料。語料蒐集時間為截至 2011 年 12 月 31 日的所有語料。

Plurk 語料的蒐集是利用其搜尋功能，例如，蒐集中興大學圖書館之 Plurk 語料時，輸入關鍵字「中興大學 and 圖書館」、「興大 and 圖書館」搜尋所有的公開噗文，去除中興大學圖書館本身所發的噗文，以及有關圖書館活動宣傳的噗文。然而，針對中興大學圖書館進行之個案實驗研究，在實驗語料上可能出現 Plurk 語料不足的問題，則會加

8 微博是新浪公司於 2009 年 8 月推出，中國大陸使用率最高的微網誌，詳細資料參見網站 <http://www.weibo.com>。

9 臉書的創辦人是 Mark Zuckerberg，一個社交網絡服務網站，於 2004 年 2 月 4 日上線，詳細資料參見網站 <http://www.facebook.com/help/21255448209369>

10 教育部撥款 5 年 500 億新臺幣用以補助重點大學成為世界一流學府的計畫，詳細資料參見網站 http://www.edu.tw/news.aspx?news_sn=4387

入中興大學圖書館自動化系統以及中興大學問題建議與回答論壇之使用者意見與回覆。

本研究所蒐集到的原始實驗語料總共 3392 筆，經過前處理斷句後總句數為 3649 句。其中，自微網誌 Plurk 蒐集 14 所大學圖書館 2008 年至 2011 年之使用者語料，共 1557 筆。國立台灣大學圖書館有 106 筆、國立成功大學圖書館 161 筆、國立清華大學圖書館 127 筆、國立交通大學圖書館 152 筆、國立中央大學圖書館 49 筆、國立陽明大學圖書館 64 筆、國立中山大學圖書館 56 筆、國立中興大學圖書館 425 筆、長庚大學圖書館 56 筆、國立台灣科技大學圖書館 24 筆、國立政治大學圖書館 146 筆、國立台灣師範大學圖書館 20 筆、淡江大學圖書館 80 筆、輔仁大學圖書館 91 筆。中興大學自動化系統和問題建議與回答論壇的語料數量分別為 1994 句和 98 句，共 2092 句。然而，有些語料情緒無法由語境判斷出來，或者是本研究不處理之英文語料，不列入實驗語料中，共 16 筆（Plurk：5，自動化系統：5，論壇：6），實際語料共 3633 句，如表 1 所示。依照類別劃分，館員類 26 句、館藏類 363 句、服務類 501 句、設備類 394 句、空間與環境類 301 句、整體性 247 句、其他 93 句，無分類 1521 句、無關係 196 句。另外，因 Outside Test 實驗所需將中興大學 Plurk 語料分成訓練語料與測試語料。中興大學 Plurk 為測試語料共 226 句；其餘的所有圖書館 Plurk 為訓練語料共 1326 句。

表 1 斷句後實驗語料總數量

語料來源	語料數量		
	1552		
	測試語料		訓練語料
Plurk 語料	中興大學圖書館 2010/08-2010/12	中興大學圖書館 2011/01-2011/07	所有圖書館 2008-2010/07、 2011/08-2011/12
	133	93	1326
中興大學圖書館 自動化系統	1989		
中興大學問題建 議與回答論壇	92		
總筆數	3633		

二、前處理與人工標記

語料蒐集完成後進行斷詞、切成長句，以及去除停用字。斷詞利用中央研究院詞庫小組（Chinese Knowledge Information Processing Group，簡稱 CKIP）開發的中文斷詞系統¹¹，預先將語料進行斷詞與斷句，然為避免句子過短或句子語意不清楚則會彈性合併句子。此外，若單筆語料中提及 2 個以上與主題相關之情緒目標，則會依主題斷成多句，以便於語料分類。接著，去除停用詞（Stop Word），如文章中的介系詞、連接詞、代名詞或語助詞，較常出現在字串的首尾。本研究以中央研究院平衡語料庫詞集及詞頻統計¹² 頻率最高的前 100 個詞作為去除停用詞之依據，若詞彙為情緒詞則不列入停用字，並刪去程度副詞與否定詞，以備後續實驗之用，可以進一步執行未知詞的處理。

（一）前處理

1. 斷句

以句號（。）、驚嘆號（！）、問號（？）、分號（；）、連接號（…）為基礎進行語料斷句，然為避免句子過短或句子語意不清楚則會彈性合併句子。此外，若單筆語料中提及 2 個以上與主題相關之情緒目標，則會依主題斷成多句，以便於語料分類。基於圖書館自動化系統之讀者意見系統中的語料，主要為使用者向圖書館提出意見或問題之用，使用者經常以多句詳細陳述一個事件，而嘖浪多是發表心情與情緒之用，通常一筆資料中會包含一個或多個主題。因此本研究將圖書館自動化系統之讀者意見系統的語料以單筆為單位，倘若單筆語料內含兩個以上的主題，方進行斷句。

▶ 斷句範例 1：

「中興圖書館的「公文」書籍，也太舊了吧?? 1987 年 ~? 一點參考價值也沒有。」

→ 此筆語料原應以問號斷句，但斷句後其中“1987 年 ~” 句子過短且並無意義，因此本句不斷句。

11 中研院資訊所詞庫小組所開發的中文斷詞系統，詳細資料參見計畫網站 <http://ckipsvr.iis.sinica.edu.tw/>

12 中研院資訊所詞庫小組所完成的「中央研究院平衡語料庫詞集及詞頻統計」，詳細資料參見 http://www.aclclp.org.tw/doc/wlawf_abstract.pdf

▶ 斷句範例 2：

「有點失望 ... 去中興的圖書館竟然不得其門而入 ... 」

→此筆語料原應以連接號斷句，然“有點失望”該句雖有情緒但無情緒目標，意義不清楚，因此本句不斷句。

▶ 斷句範例 3：

「在中興大學圖書館唸書還挺爽的，好桌好椅，廁所也好乾淨，光線充足適中，又蠻安靜的，真是相識恨晚呀 !!! 」

→此筆語料所涵蓋的情緒目標涵蓋圖書館整體、圖書館設備以及圖書館空間與環境，若不斷句無法精確分類，因此該筆語料依情緒目標斷為 4 句：（1）在中興大學圖書館唸書還挺爽的；（2）好桌好椅；（3）廁所也好乾淨，光線充足適中，又蠻安靜的；（4）真是相識恨晚呀 !!!。

2. 斷詞

斷詞採用 CKIP 斷詞服務客戶程式（Console）可一次將多筆語料同時進行斷詞，然而 CKIP 斷詞後，除了原本語料的分句外，發現部分標點符號也會使句子斷開，如全形的逗號（，）、句號（。）與問號（？），以及半形逗號（,），必須人工將被斷開的句子連結起來。為增加效率，進行斷詞之前，先將上述標點符號以其他不會造成分句的標點符號取代，即可解決此問題。

▶ 斷詞範例：

圖書館 (Nc) 每 (Nes) 次 (Nf) 預約 (VC) 都 (D) 要 (D) 重新 (D) 輸入 (VC) 帳 (Na) 密 (VH) 真是 (D) 煩死 (VH) 了 (T) 啦 (T)

▶ 處理未知詞後之斷詞範例：

圖書館 (Nc) 每 (Nes) 次 (Nf) 預約 (VC) 都 (D) 要 (D) 重新 (D) 輸入 (VC) 帳密 (Nc) 真是 (D) 煩死 (VH) 了 (T) 啦 (T)

▶ CKIP 斷句之例外處理範例：

「是全南區的人全部都來中興的圖書館嗎？這麼滿是想嚇唬誰」

→語言處理時以標點符號做為句子與句子之間的分隔，若問號為全形問號，斷詞結果會從問號分成兩句。因此，若改成半形問號便不會分成兩句。

3. 合併詞彙

基於 CKIP 中文斷詞系統在處理詞典中未收錄之未知詞 (Out of Vocabulary) 時，無法區分出準確的斷詞結果，如人名、地名、組織名、書名、慣用複合詞等，皆會被斷成較短的單字或詞，而為便於後續特徵詞選取的處理，將被斷開的未知詞合併起來。與圖書館相關的專有名詞，包括提供的服務與設備，如電子期刊、流通櫃台、館際合作等等，同樣會予以合併。此外，在人工標記出研究所需之情緒詞、情緒搭配詞、程度詞、否定詞時，若詞彙有被斷開，則亦應將其合併。

▶ 未知詞合併範例：

中興 (Vi) 圖書館 (N) 居然 (ADV) 進 (Vt) 了 (ASP) 貞 (N)
本 (DET) 義行 (N) 畫冊 (N) !(EXCLAMATIONCATEGORY)
而且 (C) 看起來 (ADV) 還 (ADV) 有點 (ADV) 久 (Vi) 了 (T)
!(EXCLAMATIONCATEGORY)

→此筆語料斷詞結果將貞本義行斷開，但貞本義行為人名，所以應將其合併，此筆語料的斷詞應為：

中興 (Vi) 圖書館 (N) 居然 (ADV) 進 (Vt) 了 (ASP) 貞
本義行 (N) 畫冊 (N) !(EXCLAMATIONCATEGORY) 而且
(C) 看起來 (ADV) 還 (ADV) 有點 (ADV) 久 (Vi) 了 (T)
!(EXCLAMATIONCATEGORY)

▶ 圖書館相關專有名詞合併範例：

中興 (Vi) 圖書館 (N) 的 (T) 電子 (N) 期刊 (N) 真 (ADV)
他媽的 (T) 靠 (P) 杯 (N)

→此筆語料斷詞結果將「電子」與「期刊」斷開，但電子期刊為圖書館館藏，具有專指性，所以應將其合併；此外，「靠」和「杯」被斷開成為兩個單字，無法顯示其意涵，因此也應將其合併。此筆語料的斷詞應為：

中興 (Vi) 圖書館 (N) 的 (T) 電子期刊 (N) 真 (ADV) 他媽的
(T) 靠杯 (N)

▶ 情緒詞合併範例：

有點 (ADV) 失望 (Vi) ... (ETCCATEGORY) 去 (Vt) 中興 (Vi) 的 (T) 圖書館 (N) 竟然 (ADV) 不得 (ADV) 其 (DET) 門 (N) 而 (C) 入 (Vt) ... (ETCCATEGORY)

→此筆語料之情緒詞為失望與不得其門而入，然而不得其門而入被斷為多個詞，因此應予以合併為「不得其門而入 (Vt)」

▶ 程度詞合併範例：

中興 (Vi) 圖書館 (N) 真的 (ADV) 很 (ADV) 棒 (Vi) 耶 (T)

→此筆語料中情緒詞為棒，程度詞為真的很，但程度詞被斷開為「真的 (ADV)」與「很 (ADV)」，因此予以合併為「真的很 (ADV)」。

(二) 人工標記之特殊句型處理

前處理完成後，找 3 位標記者以人工方式將語料分成館員、館藏、服務、設備，以及空間與環境 5 大類，以多數決劃分語料類別，若語料屬於整體性觀點，則每類別皆放入該語料，其餘列入其他類。同時標記情緒詞、程度詞、否定詞與情緒搭配詞，若 3 人皆不相同則委請第 4 人加入討論，確定出共同認可的標記結果。3 位標記者中兩人為動物科學系四年級學生，另一人為外語系三年級學生，在進行實際人工標記之前，3 人皆經過簡單的語言句法認知以及標記的訓練測試。

情緒詞以 (S, +/-) 標記；程度詞以 (D, 3/2/1) 標記，並分成強 (3)、中 (2)、弱 (1) 3 種程度；否定詞以 (N) 標記；情緒搭配詞以 (C, 情緒詞) 標記。整個句子也標記帶有的情緒與強度，以 (S +/-, D 3/2/1) 表示之，單一句子中可能包含多個程度詞，整句情緒之程度取決於中庸、保守以及多數規則。舉例而言，若一個句子有 2 個情緒詞，程度分別為 D1 與 D2，則整句情緒程度依據保守規則取為 D1；若句子有 3 個情緒詞，程度分別為 D1、D2 與 D3，則整句情緒程度依據中庸規則取為 D2；若句子有多個情緒詞，程度分別為 D1、D1 與 D2，則整句情緒程度依據多數規則取為 D1。最後，將 4 種詞彙匯總建立情緒詞辭典、程度詞辭典、否定詞辭典與情緒搭配詞辭典。

然基於中文語法困難、字詞間組合豐富多變，尤其語義複雜，使得實驗語料中，發現部分語料雖然由人類可以判斷出語料中之主題、情緒目標與情緒詞等，然而，於人工標記上卻無法直接找出適合的情

緒詞進行標記，必須將語料加以變形或改寫。本研究實驗語料中，特殊句型有反諷句與比較句，分別有 9 句與 27 句，其他須改寫句子 2 句，無法判斷語境者 10 句。雖然須改寫情況的句子並不多，但因實驗語料總計僅 3633 句，倘若語料數量龐大，或許能呈現出改寫前後之成效差異。

1. 反諷句

反諷句是文句中採用修辭學的倒反法。倒反法意指言詞表面所呈現的意義，和作者內心實際的想法剛好相反，並藉以與陳述事實相反的言語來突顯事實的重要，可分為倒辭與反語，其中反語往往帶有強烈的幽默或嘲諷的意味（黃慶萱，2002）。實驗語料中的反諷句強調作者本身對圖書館服務之不滿意情緒，反諷句會導致句子情緒極性反轉，文句中的情緒詞為正向情緒詞，但整句情緒卻為負向，在人工標記上會產生矛盾，解決方法為利用本研究提出之情緒搭配詞與情緒詞作搭配，使情緒詞的正向情緒，因與情緒搭配詞連結後而轉變為負向情緒。

▶ 反諷句標記範例 1：

「中興圖書館的書真的很 (D, 2) 妙 (S, -) ... 有下冊沒上冊 (C, 妙)」

→ 句子情緒標記為 (S-, D2)。此筆語料情緒詞為「妙」，查詢 NTUSD 情緒詞典「妙」的情緒極性為正向，但後句卻說有「下冊沒上冊」，顯示出作者認為圖書館館藏蒐集不完整。由於本句子為反諷句，若情緒詞僅標記「妙」，無法使標記結果顯現該句情緒，須同時標記情緒搭配詞，而此筆語料情緒詞「妙」因「有下冊沒上冊」轉為負向情緒，整句情緒為負向。

▶ 反諷句標記範例 2：

「中興的圖書館真是太 (D, 2) 讚 (S, -) 了嗎?? 九點半不到前面排了十個等候的人 (C, 讚)，一個半小時終於讓我等到，待了五個多小時離開時，外面又十幾個在等」

→ 句子情緒標記為 (S-, D2)。此筆語料情緒詞為「讚」，NTUSD 中「讚」為正向，然而之後句子皆在陳述作者排隊進館的情形，人很多且等很久，因此此筆語料情緒詞「讚」因後面陳述轉為負向情緒，整句情緒判定為負向。

2. 比較句

比較句是兩個（或多個）比較對象或者是同一個對象的不同方面，如程度、數量或性狀等，進行有差別或相同表現的句子，其主要的功能是用來表達事物之間經過比較後所產生的關係與結果（李靜枝，2009）。比較句的基本構成要素有比較主語、比較字、比較標準、比較結果等四個比較項目，例如「哥哥比弟弟高」，“哥哥”為比較主語，“比”為比較字，“弟弟”為比較標準，“高”為比較結果，但四個項目在同一句子中並非絕對存在，實際用法上偶爾會有所省略。比較句一般分為等比與差比，等比包含等同與近似，差比包括不及、勝過與極比（許國萍，2007），其句型結構多變且語義多樣，人類容易分辨比較句中之主語以及比較主語與比較標準之間的關係，然而恐有難以精確標記之虞，應事先予以改寫，但無法針對所有句型一一探討，本研究僅針對實驗語料中出現之比較句進行改寫。觀察實驗語料中之比較句，比較句之等比於人工標記並無大問題，因不論主題為比較主語或比較標準，兩者比較結果相同，皆可以相同情緒詞標記。須改寫的是比較句之差比且主題落於比較標準，標記尤其困難的是，比較句中無比較字，而是以描述型態呈現比較結果。

▶ 比較句之差比改寫範例：

「為什麼一間私立的南台科技大學圖書館設備會比中興好呢?！」

→此筆語料之比較主語為南台科技大學圖書館，比較標準為中興大學圖書館，針對圖書館設備在做比較，但中興圖書館為實驗語料主題，單純依照句中文字陳述標記，無法表達出本句作者針對中興圖書館欲傳達的語意。因此，建議改寫為：

「為什麼中興會比一間私立的南台科技大學圖書館設備不（N）好（S，+）呢?！」

→改寫準則為：

A. 比較主語與比較標準相互對調。

B. 比較字（如形容詞或動詞片語）改寫為其反義詞。

▶ 比較句之無比較字改寫範例：

「在使用完中興大學的圖書館及其設施後深深的感覺到，台北大

學還在鑽木取火，過著史前時代的生活…」

→此筆語料之比較主語為中興大學圖書館，比較標準為台北大學圖書館，但無明確比較字與比較結果。基於中興大學圖書館為實驗語料主題，本句以“台北大學還在鑽木取火，過著史前時代的生活”表達中興大學圖書館設施較為先進，因此，建議改寫為：

「在使用完中興大學的圖書館及其設施後深深的感覺到，不像 / 不若台北大學還在鑽木取火，過著史前時代的生活…」

→改寫準則則是：

將描述主題之陳述句“台北大學還在鑽木取火，過著史前時代的生活”加上否定詞不像或不若即可。

3. 其他

實驗語料中尚有其他無法標記出切合的情緒詞時，也應加以改寫。

▶ 範例 1：

「我要圖書館，就算只有中興三分之一大也可以」

→此筆語料主題為中興圖書館，但主語卻是某圖書館，其語意顯示中興圖書館較某圖書館大。首先，先將主語改為中興圖書館，再取三分之一的倒數為比較結果。因此，建議改寫為：

「中興圖書館有我們圖書館的三倍大」

▶ 範例 2：

「中興大學圖書館（推）」

→此筆語料情緒為正向，情緒詞為「推」，此情緒詞為動詞，單個字用法過多且極性為中性，判斷此為簡略詞，建議將簡略詞依照語意還原為「推薦」：

「中興大學圖書館（推薦）」

4. 無法判斷語境

實驗語料中，有部分語料其情緒無法由語境判斷，因此，無法正確標記出情緒詞與整句情緒，僅能將語料從實驗中去除。

▶ 範例 1：

「把中興圖書館變成彰師圖書館吧 !!!」

→ 此筆語料無論主題為中興圖書館或彰師圖書館，皆無法辨別其情緒極性，因此不納入實驗語料中。

▶ 範例 2：

「去中興逛了一圈 ~ 為啥圖書館跟宿舍品質差這麼多」

→ 此筆語料為比較句，針對圖書館與宿舍之品質相互比較，主題雖為中興大學圖書館，然因比較標準“差這麼多”無法辨別圖書館與宿舍的品質究竟是圖書館品質差於宿舍，或是宿舍品質差於圖書館，因此也不納入實驗語料中。

三、語料分類方法

利用 K- 最鄰近分類法 (KNN)、簡單貝氏分類法 (Naïve Bayes)、SVM 等方法進行分類 (曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯, 2006)，將測試語料分成館員、館藏、服務、設備以及空間與環境 5 大類。首先擷取每類的特徵詞，針對各類之所有詞彙計算 TF×IDF (Salton & Buckley, 1988)。如果將每一個句子視為一個檔案，TF (Term Frequency, 詞頻) 指的是每個詞在各句中出現的頻率；IDF (Inverse Document Frequency, 逆文件頻率) 指的是總句數 / 該詞出現句數的商。接著，擷取各類權重值大於某個閾值 (Threshold) 的詞彙或者 Top-N 詞彙作為特徵詞，形成各類的類別特徵向量。

(一) K- 最鄰近分類法 (KNN)

K- 最鄰近分類法是根據在空間中的同一類物件彼此距離較近為基礎進行預測未知資料的類別。KNN 的計算必須先計算未知資料與所有訓練資料之相似度，找出相似度最高的前 K 篇文件，再從這 K 篇文件的類別決定輸入未知資料的類別 (Cover & Hart, 1967)。

首先將測試語料與訓練語料之所有句子以向量方式表示，再計算測試語料句與訓練語料句之間的距離，接著以距離最接近的 K 個訓練語料句為範圍，根據範圍中訓練語料句最多之類別，將測試語料句分入該類。

(二) 簡單貝氏分類法 (Naïve Bayes)

Naïve Bayes 是一種簡單且實用的分類方法，且利用已知的事件發生之機率來推測未知資料的類別。基於機率概念計算資料特徵與類別的條件機率，以機率最大者作為該資料可能的分類結果。進行分類之前必須事先知道資料要分成的類別，使用訓練資料進行樣本學習，取得事物曾經發生過的機率值，才能有效地用來處理未分類的資料（McCallum & Nigam, 1998）。

先利用訓練資料的所有句子計算每個類別的事前機率，並以此計算所有特徵在各個類別的條件機率。當測試句子進行分類時，先分別計算每個特徵在各個類別中的條件機率，再將每個特徵在各個類別的條件機率相乘之後，可求得該測試句子屬於各個類別的機率，最後再依照機率最高的類別當作未知句子之類別。

（三）支援向量機 (SVM)

SVM 是一種很有效率的分類演算法，Vapnik and Cortes (1995) 根據統計學習理論提出的監督式機器學習方法，適用於小樣本、高維模式，也可解決非線性資料的問題（黃承龍、陳穆臻、王界人，2004）。簡單的 SVM 是將一群資料分成兩個不同類別的集合，希望找出一個可以將資料分隔開的超平面（Separating Hyperplane）。SVM 進階的做法是輸入一群以未知方式已分類好的資料，經由訓練後獲得一個模型，爾後，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。基於 SVM 實作容易使用、正確率高、相對快速的特性，加上即使資料量不足的時候，也僅需要幾個關鍵的 Support vector 就足以達到分類目標，經常被應用於自然語言處理或自動辨識分類的研究上（Hsu & Lin, 2002）。

四、情緒分析方法

參考既有文獻，本研究提出之情緒分析的方法有兩種：

（一）導入程度詞、否定詞與情緒詞權重之情緒分析

測試語料分類後，計算 5 大類句子中情緒詞分數，正向情緒詞為 1 分，負向情緒詞為 -1 分，此外利用程度詞和否定詞給予加權計算句子情緒分數，弱程度詞「稍微」乘以 1.5、中程度詞「很」乘以 2，強程度詞「超級」乘以 3，否定詞則乘上 -1，接著將句中各情緒詞分數加總，即可得到句子情緒分數。之後，將各類中句子情緒分數加總且平均，此為類別情緒分數，類別情緒分數加總代表圖書館的總評分。

如下示範例，其句情緒分數為 -2，表示該句情緒為負向情緒。

▶ 句子的情緒分數計算範例：

「圖書館資料的種類 (C, 多) 沒有 (N) 很 (D+) 多 (S, +) 耶！」

→ 句子情緒標記為 (S-, N, D2)

句子情緒分數為 $1 \times (-1) \times 2 = -2$ 。

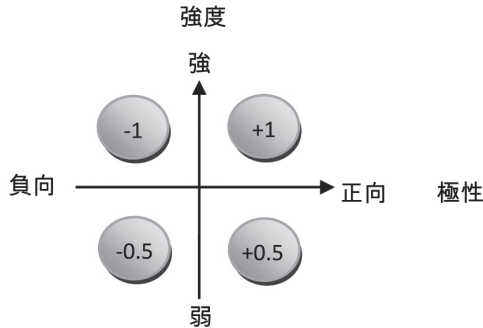


圖 2 情緒極性與程度類別之情緒分數

(二) 情緒極性與程度類別之情緒分析

測試語料分類後，分別計算 5 大類句子中情緒詞分數。以情緒極性與程度類別之情緒分數（如圖 2），將句子中情緒詞分數加總並求平均值，所得分數為句子的情緒分數，再將各類別中所有句子的情緒分數加總後求平均，得出每類別情緒分數，各類情緒分數相加後所得即為圖書館的總評分。

肆、研究結果與分析

一、辭典建置

由 3 位標記者 (A、B：動物科學系四年級；C：外文系三年級) 分別標記出每筆語料的類別、情緒詞、情緒搭配詞、程度詞、否定詞，以及整句情緒，3 人標記結果以多數決決定，若 3 人皆不相同則委請第 4 人加入討論，確定出共同認可的標記結果，再依據標記結果彙整出情緒詞辭典、情緒搭配詞辭典、程度詞辭典、否定詞辭典四本辭典，並刪除重複詞彙，其結果如表 2 所示。標記完成後，以統計方法針對每位

標記者的標記結果，利用 Kappa 檢定標記者之標記一致性，分別針對類別、極性以及整句情緒進行標記一致性檢定，其結果如表 3 所示。參考 Landis & Koch (1977) 提出的 Kappa 值一致性解釋 Kappa 檢定結果皆為 0.6 以上，可見三位標記者彼此的標記結果皆屬高度一致性，其中仍有少數不完全相同的情形發生。在極性標記的部分 Kappa 值皆高達 0.8 以上，屬於最強一致性，因此作為產生情緒辭典之依據非常可靠。

表 2 情緒分析辭典

辭典種類	詞彙個數	總計	圖資領域專屬	總計
情緒詞辭典	正向	481	29	82
	負向	1071	53	
情緒搭配詞辭典	正向	130	22	76
	負向	368	54	
程度詞辭典	弱程度	39		
	中程度	61	120	
	強程度	20		
否定詞辭典		56	56	

表 3 Kappa 一致性檢定結果

標記者	類別	極性	整句情緒
A*B	.693	.814	.687
A*C	.686	.824	.709
B*C	.705	.843	.724

二、語料分類

Weka (Mark et al., 2009) 為紐西蘭 Waikato 大學電腦科學系所發展出的資料探勘應用軟體，提供眾多資料探勘演算法，主要包含資料預處理、關聯規則、分類預測法、叢集分析法等，並具備視覺化的圖形介面。另一方面，由於 Weka 屬於共享軟體，本研究使用此軟體執行分類預測。為了提高分類品質，降低 Weka 執行所需要的時間，事先利用 Weka 中的篩選器 (Filter) 進行特徵挑選 (Attribute Selection) 大幅減少特徵值數量。

語料分類實驗以三種測試模式為基礎：一、採用測試模式 Training Set 的 Inside Test；二、採用測試模式 Cross Validation 的 Outside Test；三、採用測試模式 Test Set 的分類結果預測，分別使用 KNN、Naïve Bayes 以及 LibSVM 分類器進行語料分類，另結合七種語料來源組合 (Plurk、

自動化、論壇、Plurk+ 自動化、Plurk+ 論壇、自動化 + 論壇、全部語料) 訓練模型，並檢視其分類效能，三種分類器的參數皆採預設值，唯 LibSVM 的核心函數 (Kernel Function) 選用放射性核心函數 (Radial Basis Function, 簡稱 RBF) 以建立分類模型。Inside Test 和 Outside Test 使用如表 1 所示之 3633 句語料，其中館員類 26 句、館藏類 363 句、服務類 50 句、設備類 394 句、空間與環境類 301 句、整體性 247 句、其他 93 句、無分類 1521 句以及無關係 196 句。另外，測試分類實驗則使用中興大學 2010/8-2011/7 的 Plurk 語料共 226 句作為測試語料，其餘 Plurk 語料 1326 句為訓練語料。上述 Plurk 測試語料會扣除 96 句屬於無情緒者，實際以 130 句進行情緒分析，其中館員類 0 句、館藏類 12 句、服務類 38 句、設備類 10 句、空間與環境類 27 句、其他類 5 句，以及整體性 38 句。

如表 4 所示，無論是 Inside Test 或 Cross Validation 實驗，KNN 針對不同語料來源進行分類得到的正確率皆高於 NB 與 LibSVM。如表 5 所示，測試語料之正確預測類別者僅三成，而正確率最高的仍是使用 KNN 的分類結果，KNN 分類錯誤的語料，NB 與 LibSVM 也同樣分類錯誤。過去研究皆指出 SVM 的效能優於 KNN 與 NB，與本實驗呈現結果並不相同，是因為本實驗使用 LibSVM 時，核心函數 RBF 的兩個重要參數 C 與 γ 採用預設值，若能針對語料內容將參數加以調整並找出最佳參數值則對結果定能有所改善。

表 4 Inside Test & Cross Validation 語料分類正確率

Classifier	語料來源 %	Plurk	自動化	論壇	Plurk+ 自動化	Plurk+ 論壇	自動化 + 論壇	全
	NB	64.0	78.5	82.7	56.3	59.7	73.8	55.0
	LibSVM	34.7	67.3	38.8	50.9	33.5	64.9	50.5
Cross Validation	KNN(k=1)	32.1	69.6	43.9	55.3	29.6	70.0	53.8
	NB	30.4	67.7	66.3	49.1	28.5	66.8	50.2
	LibSVM	23.8	67.3	32.7	47.8	23.7	64.9	47.0

表 5 測試語料分類預測正確率

Classifier	正確率 %
KNN	37.2
NB	36.3
LibSVM	30.5

三、情緒分析

將所蒐集到的所有語料進行情緒分析之 Inside Test，根據三種不同語料來源（Plurk、自動化、論壇）相互搭配組合，再分別以導入程度詞、否定詞與情緒詞權重之情緒分析，以及情緒極性與程度類別之情緒分析兩種方法進行實驗，並嘗試利用量表將情緒分析所得情緒分數予以正規化，使兩種方法能在相同標準下與標準答案相互比較，找出最佳量表正規化分數區間與情緒分析方法。實驗分為四個部分：一、導入程度詞、否定詞與情緒詞權重之情緒分析；二、情緒極性與程度類別之情緒分析；三、程度詞權重值實驗；四、中興大學圖書館使用者滿意度之情緒分析。

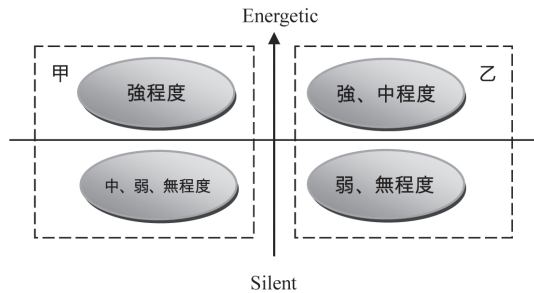


圖 3 程度類別甲、乙

實驗標準答案是人工標記的整句情緒，依據圖 3 情緒極性與程度類別給予情緒分數，圖中象限程度僅分強與弱，實際人工標記分為強、中、弱與無程度詞，為能相互對應將程度又分成 2 類：第一種（甲）為強程度屬於象限程度中的強程度，其他為弱程度；第二種（乙）為強程度與中程度屬於象限程度中的強程度，弱程度與無程度為弱程度。由甲和乙產生兩個標準答案 I 和 II。基於情緒分析實驗乃針對含有情緒之句子計算情緒分數，因此本實驗語料扣除全部語料中為無分類與無關係之句子進行情緒分析，總共 1925 句，Plurk 有 1008 句、自動化 836 句、論壇 81 句。

（一）導入程度詞、否定詞與情緒詞權重之情緒分析

測試語料分類後，計算 5 大類句子中情緒詞分數，正向情緒詞為 1 分，負向情緒詞為 -1 分，此外利用程度詞和否定詞給予加權計算句

子情緒分數，弱程度詞「稍微」乘以 1.5、中程度詞「很」乘以 2，強程度詞「超級」乘以 3，否定詞則乘上 -1，接著將句中各情緒詞分數加總，即可得到句子情緒分數。之後，將各類中句子情緒分數加總且平均，此為類別情緒分數，類別情緒分數加總代表圖書館的總評分。

利用情緒詞、程度詞和否定詞給予加權計算出句子原始情緒分數，弱、中、強程度分別加權 1.5、2、3，情緒分數經過加權介於於 -3 ~ +3，利用四種量表（A ~ D，如圖 4 ~ 圖 7）將情緒分數正規化，量表 A 與 B 基於加權值而設定，採用程度類別甲之實際原始情緒分數分布情形後，設定出量表 C。採用程度類別乙之實際情緒分數分布情形，設定出量表 D，之後再計算正規化後之句子情緒分數與標準答案之正確率，其結果如表 6。

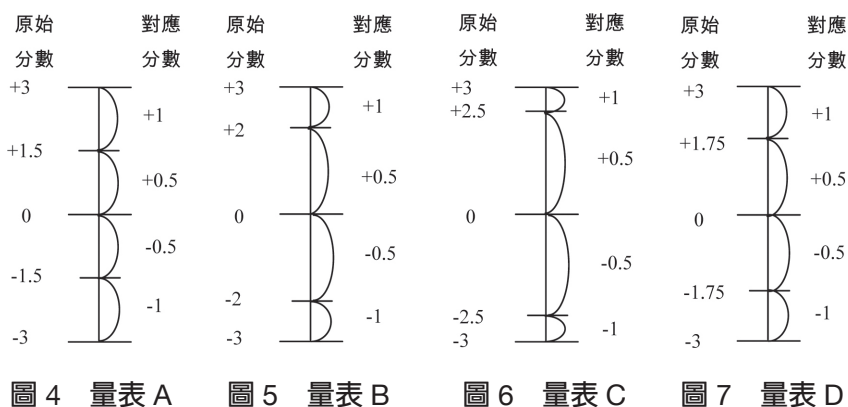


表 6 導入程度詞、否定詞與情緒詞權重之情緒分析正確率

語料來源 量表	Plurk	自動化	論壇	Plurk+ 自動化	Plurk+ 論壇	自動化 + 論壇	全
1- 甲 -A	0.7113	0.8971	0.8765	0.7956	0.7236	0.8953	0.7984
1- 甲 -B	0.9940	0.9928	1.0000	0.9935	0.9945	0.9935	0.9932
1- 甲 -C	0.9960	0.9952	1.0000	0.9957	0.9963	0.9956	0.9953
1- 乙 -A	0.9663	0.9486	0.9383	0.9582	0.9642	0.9477	0.9574
1- 乙 -B	0.7371	0.9222	0.9383	0.8210	0.7521	0.9237	0.8260
1- 乙 -D	0.9722	0.9498	0.9753	0.9620	0.9725	0.9520	0.9626

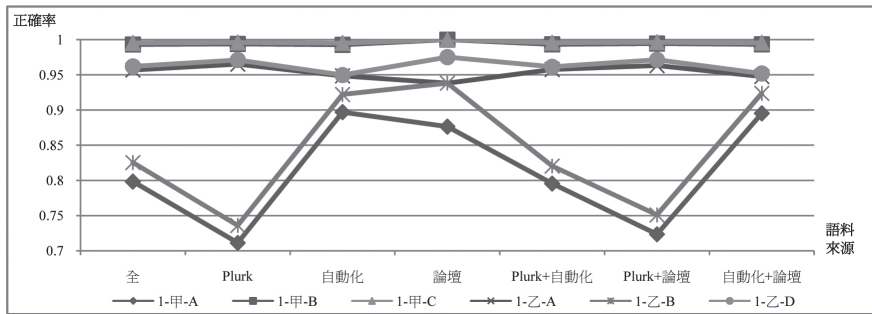


圖 8 導入程度詞、否定詞與情緒詞權重之情緒分析折線圖

從表 6 和圖 8 中可知，1-甲-C 獲得結果最佳，數字 1 代表採用方法一進行情緒分析的第一種實驗，甲代表強弱的分類是採用程度類別甲（即使用標準答案 I），字母 C 代表採用量表 C 進行情緒分數正規化。由於程度類別甲將中、弱、無程度歸於同一類，如果使用等距量表 A 時，會高估情緒分數。另一方面，若使用量表 B 時，由於門檻提高消除了高估的現象。因此，1-甲-A 的正確率低於 1-甲-B。同理，由於類別乙（即使用標準答案 II）將強、中程度歸於一類，若使用量表 B 時，會低估情緒分數。而若使用量表 A，由於門檻降低，可以消除低估的現象。因此，1-乙-A 的正確率高於於 1-乙-B。程度類別乙（即使用標準答案 II）搭配量表 A 的正確率又高於程度類別甲搭配量表 A 的正確率。以量表來看，不同語料來源使用量表 A 與標準答案 I 相比，以及使用量表 B 與標準答案 II 相比，結果皆不盡理想，而使用依據語料情緒分數實際分布情形所設定之量表 C 和 D，則會得到相對佳的結果。

(二) 情緒極性與程度類別之情緒分析

測試語料分類後，以情緒極性與程度類別之情緒分析分別計算 5 大類句子情緒分數，將句子中情緒詞分數加總並求平均值，所得分數為句子的情緒分數，再將各類別中所有句子的情緒分數加總後求平均，得出每類別情緒分數，各類情緒分數相加後所得即為圖書館的總評分。

情緒極性與程度類別給予情緒分數，利用三種量表（E ~ G，如圖 9 ~ 圖 11）將情緒分數正規化，量表 E 是基於象限分數而設定，採用實際原始情緒分數分布情形後，設定出量表 F 與 G，再與兩種標準答案相比較並計算正確率。然因象限分數皆介於 +1 ~ -1 之間，因此程度類別甲與乙可使用相同量表進行正規化。方法二情緒分析正確率如表 7。

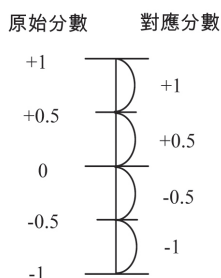


圖 9 量表 E

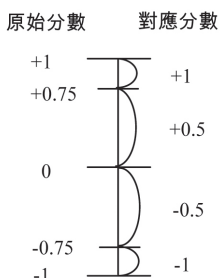


圖 10 量表 F

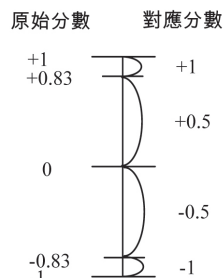


圖 11 量表 G

表 7 情緒極性與程度類別之情緒分析正確率

語料來源 量表	Plurk	自動化	論壇	Plurk+ 自動化	Plurk+ 論壇	自動化 + 論壇	全
2-甲-E	0.9782	0.9617	0.9136	0.9707	0.9734	0.9575	0.9683
2-甲-F	0.9990	0.9976	1.0000	0.9984	0.9991	0.9978	0.9984
2-甲-G	0.9990	0.9976	1.0000	0.9984	0.9991	0.9978	0.9984
2-乙-E	0.9058	0.8110	0.7407	0.8628	0.8935	0.8048	0.8577
2-乙-F	0.9931	0.9641	0.9877	0.9799	0.9927	0.9662	0.9803
2-乙-G	0.9921	0.9641	0.9877	0.9794	0.9917	0.9662	0.9797

由表 7 中顯示，無論語料來源組合、程度類別以及量表，情緒分析結果大多皆有高正確率。在相同語料來源與相同量表的情形下，程度類別甲的正確率高於乙，表示在象限中程度詞之強或弱程度的分類上，中程度的程度詞應分類為弱程度，僅將強程度的程度詞視為強程度，可以得到較高的情緒分析結果。一般使用者在網際網路上以文字陳述對圖書館的讚賞或批評，使用的情緒詞以強程度的程度詞加以修飾，則顯示使用者欲表達對圖書館是極度滿意或極度不滿意，圖書館則須特別注意、儘快改善。

針對單一語料來源而言，在程度類別甲中，基於論壇語料所屬句數少且無強程度，分數區間落於 +0.5 ~ -0.5 之間，經過量表正規化，情緒分數非 +0.5 即 -0.5，正確率自然提高，甚至完全正確；相對地，Plurk 句數多、涵蓋率高，以 Plurk 進行情緒分析所得正確率可信度較高，證明單以 Plurk 語料進行情緒分析之可用性與有效性。語料來源組合 Plurk+ 論壇若加入自動化以全部語料會導致正確率降低，由圖 12 可更清楚得知自動化語料中包含雜訊，會影響情緒分析正確率。

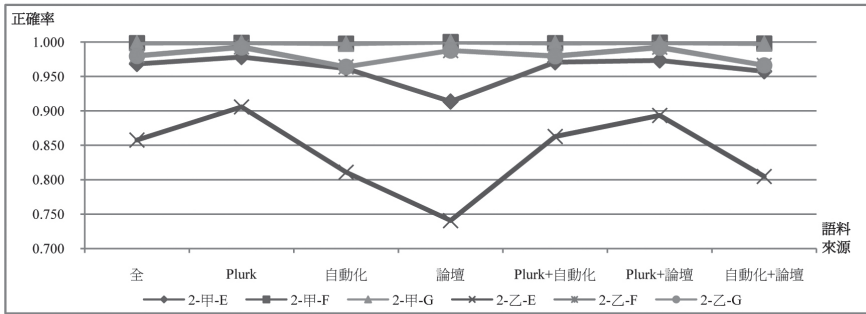


圖 12 情緒極性與程度類別之情緒分析折線圖

綜合而論，將方法一與方法二計算情緒分析後所得正確率最高者相比較，並依照程度類別對應。程度類別甲中，方法一結果最佳者為 1-甲-C，方法二為 2-甲-F 與 2-甲-G，方法二正確率高於方法一；針對程度類別乙，方法一結果最佳者為 1-乙-D，方法二則是 2-乙-F，同樣是方法二正確率高於方法一，因此可以了解進行情緒分析，利用本研究提出之情緒極性與程度類別之情緒分析較佳。

(三) 程度詞權重值實驗

為瞭解本研究提出的程度加權值是否能有效提升情緒分析正確率，因此另以弱、中、強程度加權值較低與較高分別進行實驗，較低加權值（1.2、1.5、2）為方法三，較高的加權值（2、3、4）為方法四，再各別以量表進行正規化，方法三的量表為如圖 13～圖 16 所示之量表 H～K，方法四則採用如圖 17～圖 20 所示之量表 L～O，實驗結果如表 8 所示。依據前述方法得知，利用實際分布量表之結果最具可信度，因此三種不同程度詞權重使用常態分布量表相較於其他量表所得正確率也最高，若以觀察實際語料分布所設定的量表（C、D、J、K、N、O）相比，本研究提出之權重值加權（1.5、2、3）後可獲得最佳結果。

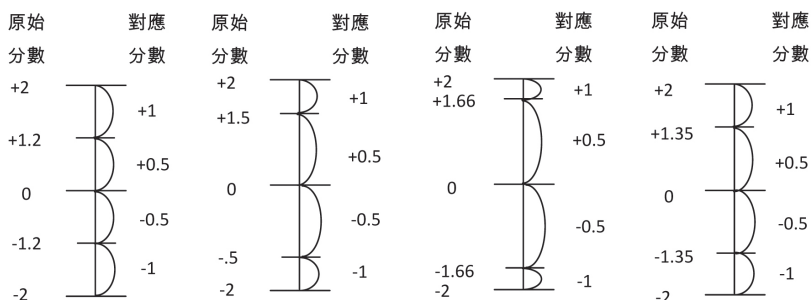


圖 13 量表 H

圖 14 量表 I

圖 15 量表 J

圖 16 量表 K

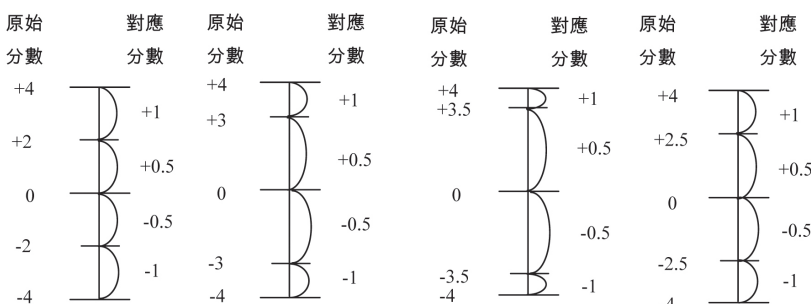


圖 17 量表 L

圖 18 量表 M

圖 19 量表 N

圖 20 量表 O

(四) 中興大學圖書館使用者滿意度之情緒分析

為瞭解中興大學圖書館 99 年度使用者之滿意度，擷取測試語料計算情緒分數，語料分為館員、館藏、服務、設備、空間與環境，以及其他等六大類，測試語料共有 226 句，然扣除 96 句屬於無情緒者，實際以 130 句進行情緒分析，館員類 0 句、館藏類 12 句、服務類 38 句、設備類 10 句、空間與環境類 27 句、其他類 5 句，以及整體性 38 句。首先，使用人工標記的整句情緒（程度副詞只有強和弱兩種），參照圖 2 所示的情緒分數計算句情緒分數，再將各類別中所有句子的情緒分數加總後求平均，得出每類別情緒分數，稱之為句子原始分數（可視為標準答案）。接著，分別以方法一與方法二進行情緒分析，方法二依程度類別可分兩種，因先前實驗程度類別甲的結果較佳，因而僅照程度類別甲的方式進行情緒分析。此外，為了增加各類語料的數量，三種方式皆計算導入整體性語料（亦即將與各類都相關的整體語料置

表 8 不同程度詞權重之情緒分析正確率

量表	語料來源 Plurk	自動化	論壇	Plurk+ 自動化	Plurk+ 論壇	自動化 + 論壇	全
1- 甲 -A	0.7113	0.8971	0.8765	0.7956	0.7236	0.8953	0.7984
1- 甲 -B	0.9940	0.9928	1.0000	0.9935	0.9945	0.9935	0.9932
1- 甲 -C	0.9960	0.9952	1.0000	0.9957	0.9963	0.9956	0.9953
1- 乙 -A	0.9663	0.9486	0.9383	0.9582	0.9642	0.9477	0.9574
1- 乙 -B	0.7371	0.9222	0.9383	0.8210	0.7521	0.9237	0.8260
1- 乙 -D	0.9722	0.9498	0.9753	0.9620	0.9725	0.9520	0.9626
3- 甲 -H	0.6538	0.8182	0.8148	0.7283	0.6657	0.8179	0.7319
3- 甲 -I	0.9950	0.9952	1.0000	0.9951	0.9954	0.9956	0.9953
3- 甲 -J	0.9950	0.9976	1.0000	0.9962	0.9954	0.9978	0.9964
3- 乙 -H	0.9187	0.8768	0.8765	0.8997	0.9155	0.8768	0.8987
3- 乙 -I	0.7381	0.9246	0.9383	0.8227	0.7530	0.9258	0.8275
3- 乙 -K	0.9663	0.9510	0.9630	0.9593	0.9660	0.9520	0.9647
4- 甲 -L	0.7113	0.9043	0.9136	0.7988	0.7264	0.9051	0.8036
4- 甲 -M	0.9911	0.9916	1.0000	0.9913	0.9917	0.9924	0.9917
4- 甲 -N	0.9950	0.9940	1.0000	0.9946	0.9954	0.9945	0.9948
4- 乙 -L	0.9663	0.9510	0.9753	0.9593	0.9669	0.9531	0.9600
4- 乙 -M	0.7341	0.9211	0.9383	0.8189	0.7493	0.9226	0.8239
4- 乙 -O	0.9831	0.9545	0.9753	0.9702	0.9826	0.9564	0.9704

入各類中) 與否之情緒分數, 之後再將分析結果以圖形化方式呈現。依照未導入整體性語料之情緒分數畫出長條圖, 由於原始句情緒分數含有負數, 而雷達圖乃適用大於 0 的數值, 因此導入整體性語料後之情緒分數則得以畫於雷達圖之上。

採用原始分數以及利用方法一、方法二進行情緒分析之結果如表 9。其分別將未導入整體性語料計算所得情緒分數以長條圖方式呈現情緒分析結果(如圖 21 ~ 圖 23), 另以雷達圖方式呈現導入整體性語料之情緒分數。由長條圖中可明顯看出, 館藏類與服務類無論是原始分數、利用方法一或方法二皆是使用者對圖書館不滿意的項目, 而設備則呈現滿意結果。較不同的是其他類的情緒分數, 方法二與原始分數是不滿意, 而方法一則顯示為滿意, 觀察 5 句其他類語料發現, 其中 1 筆語料之極性為正向, 具有 1 個情緒詞且其程度標記為 D2, 因方法一的中程度會加權 2, 使得情緒分數放大, 而方法二以象限方式差距最大為 1, 該句情緒分數為 0.5, 相比之下兩種方法計算的情緒分數已相差 1.5, 因此可能使分數正負數改變。另一方面, 若從導入整體性

語料實驗結果之雷達圖來看可發現，三種方法雖然座標數值不同，但呈現出的圖形卻極為相似，因此可以說導入整體性語料實驗產生的結果是可信的，更進一步驗證導入整體性語料的必要性。

表 9 中興圖書館測試語料結合程度加權與象限之情緒分析分數表

類別	原始分數		方法一（程度加權）		方法二（象限）	
	未導入整體性語料	導入整體性語料	未導入整體性語料	導入整體性語料	未導入整體性語料	導入整體性語料
館員	0.00	0.43	0.00	1.42	0.00	0.41
館藏	-0.13	0.30	-0.33	1.00	-0.13	0.28
服務	-0.21	0.11	-0.65	0.38	-0.22	0.09
設備	0.20	0.39	0.50	1.23	0.18	0.36
空間與環境	-0.02	0.25	0.06	0.86	0.00	0.24
其他	-0.10	0.37	0.20	1.28	-0.05	0.35

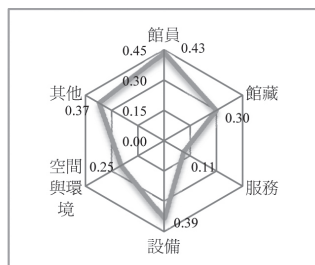
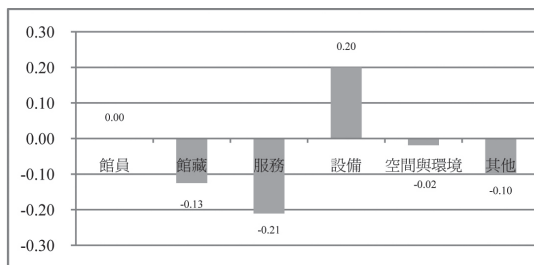


圖 21 原始分數未導入整體性語料之情緒分析結果長條圖與雷達圖

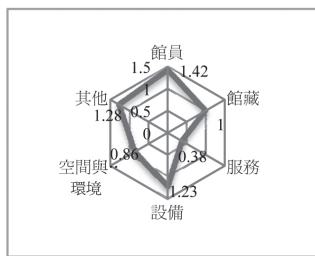
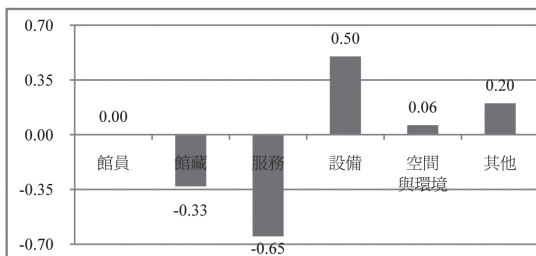


圖 22 方法一未導入整體性語料之情緒分析結果長條圖與雷達圖

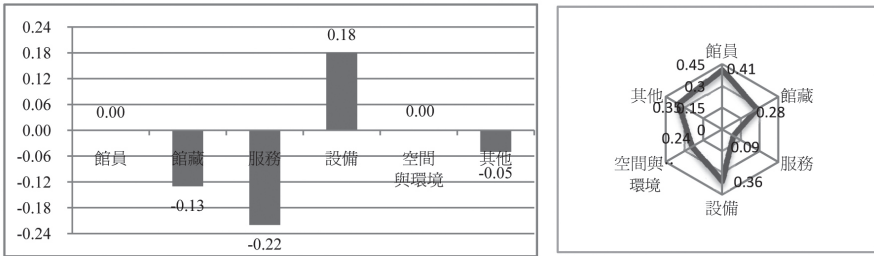


圖 23 方法二未導入整體性語料之情緒分析結果長條圖與雷達圖

伍、結論與未來研究

提出本研究重要結論，以及使用情緒分析於圖書館使用者滿意度評估之後續研究建議，以作為圖書館應用以及後續研究人員研究之參考。

一、建構一本圖書資訊學領域專屬的情緒分析辭典，提供大學圖書館利用並實施使用者滿意度之評估。

基於情緒分析辭典是進行情緒分析不可或缺的工具，目前雖已有研究建置出情緒分析辭典，但為了將情緒分析應用於圖書資訊學領域使用者滿意度的評估，需要一本適用於圖書館的情緒分析辭典，所以須從人工標記出的情緒詞加以彙整。由於有些情緒詞的極性會隨著所指的情緒目標而改變，因而發展出情緒搭配詞辭典，加上從人工標記過程中發現，人類常用句法中之反諷句，因句子前後情緒反轉，與情緒搭配詞特性雷同，因此可藉此彌補一般情緒詞無法呈現意見反轉的情況。未來期能以自動化方式抽取情緒詞，但中文詞句組合多樣，許多用法人類習以為常，而電腦程式無法輕易判斷特殊句型，比較句即是一例。建議系統開發者在開發成是以自動化方式抽取情緒詞的同時，參考本研究對電腦針對語意不清的比較句擷取錯誤詞彙，所提出之語料改寫方法與規則的建議，先行改寫。最常見的問題是主語以及比較標準位置對調，其改寫方法是將彼此復位並將比較字改為反義詞即可。

二、本研究擷取之圖書館使用者滿意度評估之情緒分析辭典具可用性與可信度。

情緒分析辭典是由三位標記者分別標記，再以多數決方式決定最

終標記結果，為確保情緒分析辭典之可用性，因此利用統計方法 Kappa 檢定標記者之標記一致性。在類別、極性與整句情緒之 Kappa 值皆大於 0.6，三位標記者之標記結果屬於高度一致性，尤其是極性部分，皆高於 0.8 屬於最強一致性，因此可得知雖然三位標記者來自不同系所，有不同背景與對語言的認知，依據標記結果所擷取出的情緒詞辭典與情緒搭配詞辭典皆具可用性且是可信的。

三、KNN、Naïve Bayes 與 SVM 三種語料分類方法，以 KNN 成效最佳。

除用三種方法外，語料分類分別進行 Inside Test、Outside Test (Cross Validation) 與分類預測三種測試模式，利用 KNN 分類不同語料組合計算所得正確率皆高於其他兩種方法，根據相關實驗指出，LibSVM 的分類效能相當不錯，然而本實驗結果卻是最差的，KNN 之 Inside Test 正確率最高可達 100%，最低也有 90%，Cross Validation 的結果雖大幅降低，但 KNN 仍維持最佳結果。分類結果預測僅近三成左右正確，此次並沒考慮到最佳參數設定，若有則能提升 LibSVM 分類效能。

四、利用情緒極性與類別象限計算所得的情緒分數相較於程度詞加權，與人工標記之標準答案相比正確率較高。

利用程度詞加權情緒分析方法與情緒極性及類別象限情緒分析方法分別計算句子情緒分數，分數區間不相同，必須以量表將分數正規化，使兩方法結果可處於同一比較基準上。本研究嘗試多個量表，最好的量表區間是依據語料內容不同觀察後而定。建議圖書館可以情緒分析評估使用者滿意度，或者以問卷與情緒分析並行，除每年例行的滿意度調查外，平時以情緒分析瞭解圖書館使用者近期不滿意的項目，使使用者與圖書館間的溝通對話的距離縮短。同時，將 Plurk、自動化系統與論壇三種語料來源相互結合，構成八種語料來源組合，對於語料來源組合去計算情緒分析分數，利用類別象限情緒析方法與 Plurk 語料可得到令人滿意的結果。

以中興大學圖書館為例，實際取測試語料，依不同類別計算情緒分數總分並求平均，未導入整體性語料之情緒分數會有正負數，可以長條圖呈現正向與負向類別，以突顯圖書館所需改進與留意之處。而針對可能類別數量為 0 時，可採用導入整體性語料，即每種類別皆算分，情緒分數便增加可以雷達圖畫出。情緒分析兩種方法所畫出的雷

達圖，雖然分數大小區間不同，但圖形卻極為相似，尤其情緒極性與類別象限與以原始分數計算所得分數與圖形皆相去不遠，更可證明情緒極性與類別象限之成效。

本研究雖致力於情緒分析辭典的建置、語料分類及情緒分析方法進行探討，但囿於能力及時間等因素，仍有未盡周延之處。謹對於本研究不足之處，提出下列進一步研究之建議，以供對本研究議題同感興趣之後續研究者參考。

- 一、情緒分析結果可採用統計方式檢定情緒分析與問卷調查使用者滿意度之效度。
- 二、情緒分析辭典之呈現方式可改以 Topic Map 或 Semantic Web，建立情緒詞關聯性，精緻化情緒分析辭典。
- 三、除使用長條圖或雷達圖外，可探究圖書館使用者滿意度如何以圖形化組合方式呈現。
- 四、基於情緒詞本身也有強弱程度的區別，可透過情緒詞的程度，進一步檢驗情緒分析結果。
- 五、以臺灣大學林智仁老師開發的 easy.py 程式，找出 LibSVM 的核心函數 RFB 之最佳 C 與 γ ，探討 LibSVM 分類效能。
- 六、利用本研究擷取出的情緒分析辭典，實際針對任一圖書館採用情緒分析進行使用者滿意度評估，以瞭解情緒辭典之實用性，並持續增加情緒詞擴大詞彙涵蓋度。
- 七、提升圖書館滿意度評估系統的效能，導入規則式專家系統 (Rule-based Expert System) 執行語料改寫與否的判斷與自動改寫作業。

謝辭

感謝國科會計畫 NSC 100-2221-E-005-087 給予本研究部份經費之補助和支持。

參考文獻

- 孔誠志 (1998)。形象公關：實務操演手冊。台北縣：科技圖書。
- 吳明德 (2003, 12 月)。圖書館評鑑。國立成功大學圖書館館刊，11，1-8。
- 吳蕙欣 (2011)。結合多辭典與常識網路的情緒分析系統。未出版之碩士論文，國立台灣大學資訊工程學系，台北市。
- 李孟潔 (2009)。利用機器學習作法之中文意見分析。未出版之碩士

- 論文，國立清華大學資訊工程學系，新竹市。
- 李政儒（2011）。應用廣義知網以支援情緒分析之研究。未出版之碩士論文，國立台灣大學資訊工程學系，台北市。
- 李啟菁（2010）。中文部落格文章之意見分析。未出版之碩士論文，國立台北科技大學資訊工程學系，台北市。
- 李靜枝（2009）。現代漢語有標記的比較句之教學排序 - 以母語為西班牙語的學習者為例。未出版之碩士論文，國立台灣師範大學華語文教學研究所，台北市。
- 孫瑛澤、陳建良、劉峻杰、劉昭麟、蘇豐文（2010，9月）。中文短句之情緒分類。在中華民國計算語言學學會、國立國際暨南大學資工系與電機系主辦，第廿二屆自然語言與語音處理研討會，南投縣。
- 張郁蔚（2007）。以訪談法探討「全國檔案目錄查詢網」使用者之滿意情形。大學圖書館，11（1），139-165。
- 許國萍（2007）。現代漢語差比範疇研究。上海：學林出版社。
- 創市際（2011）。ARO 網路測量研究。上網日期：2011年11月20日，檢自 <http://www.iama.org.tw/upload/ResourceReport/20110830105124836.pdf>
- 彭金堂、鄭姿均、卓欣姿、陳幸慈、程瑞琪（2010）。圖書館服務品質滿意度之研究 - 以某科技大學圖書館為例。績效與策略研究，7（1），53-73。
- 黃承龍、陳穆臻、王界人（2004）。支援向量機於信用評等之應用。計量管理期刊，1（2），155-172。
- 黃慶萱（2002）。修辭學。台北：三民。
- 曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯（2006）。資料探勘。台北：旗標。
- 楊昌樺、高虹安、陳信希（2007，9月）。以部落格語料進行情緒趨勢分析。在中華民國計算語言學學會主辦，第十九屆自然語言與語音處理研討會，台北市。
- 楊昌樺、陳信希（2006，9月）。以部落格進行情緒分類之研究。在中華民國計算語言學學會主辦，第十八屆自然語言與語音處理研討會，新竹市。
- 楊國樞、文崇一、吳聰賢、李亦園（1990）。社會及行為科學研究法。台北：東華書局。

- 劉峻霖 (2000)。電腦化印前從業人員工作滿意度之研究。未出版之碩士論文，國立台灣師範大學工業教育學系，台北市。
- 盧秀菊 (2003, 12月)。學術圖書館之績效評估。中國圖書館學會會報，71，1-19。
- Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.
- Drake, M. A. (1994). Technological Innovation and Organizational Change. *Journal of Library Administration*, 19(3-4), 39-49.
- Goodall, D. L. (1988). Performance Measurement: A Historical Perspective. *Journal of Librarianship and Information Science*, 20(2), 128-144.
- Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 415-425.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Retrieved October 14, 2011, from <http://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf>
- Ku, L. W., & Chen, H. H. (2007). Mining Opinions from the Web: Beyond Relevance Retrieval. *Journal of American Society for Information Science and Technology*, 58(12), 1838-1850.
- Landis, J. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.
- Lu, Y., Kong, X., Quan, X., Liu, W., & Xu, Y. (2010). Exploring the Sentiment Strength of User Reviews. *Lecture Notes in Computer Science*, 6184, 471-482.
- Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Mark H., Eibe F., Geoffrey H. Bernhard P., Peter R., & Ian H. W. (2009). The WEKA data mining software: an update. *Sigkdd Explorations*, 11(1), 10-18.
- McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naïve Bayes Text Classification. In *Proceedings of AAAI/ICML-98 Workshop on Learning for Text Categorization*(pp.41-48). California: AAAI Press.

- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Information Retrieval*, 2(1-2), 1-135.
- Salton, G., & Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing & Management*, 24(5), 513-523.
- Vapnik, V., & Cortes, C. (1995). Support Vector Networks. *Machine Learning*, 20(3), 273-297.
- Wilson, T., Wiebe, J., & Hwa, R. (2004). Just How Mad Are You? Finding Strong and Weak Opinion Clauses. In *Proceedings of The National Conference On Artificial Intelligence*. Retrieved October 14, 2011, from <http://www.cs.pitt.edu/~hwa/aaai04.pdf>

A Study on Library Users' Satisfaction Evaluation Using Sentimental Analysis

June-Jei Kuo

Assistant Professor
Graduate Institute of Library and Information Science
National Chung Hsing University

Yu-Jung Zhang

Administrative Assistant
Library, Asia University

Introduction

Along with the rapid development of computer technology and the Internet, traditional library services and library users' use habits have also changed. Libraries have realized the need to conduct library service satisfaction research from the user's perspective, provide customized services and understand users' real needs and opinions, which help achieve the ultimate goals of service users, and achieve the ideal of optimizing library collections and concomitant services. Traditional library users' satisfaction assessment relied on questionnaire surveys, interviews or focus groups, to understand the satisfaction of library users, and needs for improvement. However, there are some restrictions as to implementation of the traditional methods. The need to design a questionnaire with reliability and validity requires expert reviews and pretest findings. The process from distributing questionnaires to disseminating usually takes several months. For interviews and focus groups it is not easy to find representative subjects. Plus, the process from preparing the interview template to transcribing audio files, to analyzing the results is time-consuming and laborious. From the aforementioned, one can see how the traditional method of implementing

users' satisfaction evaluation involves excessive overhead, and cannot immediately identify library problems. Nor can it really identify the user's dissatisfaction issues, or permit timely improvements.

The rapid development of MicroBlogging has helped people network online by sharing texts, pictures, videos and other media to express their feelings, to share some opinions of lived experience, and express their views that may be positive or negative. Therefore, in order to solve the traditional library satisfaction evaluation problems, this study used the Internet to retrieve opinions made about libraries. The classification prediction was used to classify opinions into different library dimensions, and then the sentimental analysis was employed to elucidate sentiments for each dimension. Besides, an overall sentiment score for library users' satisfaction was provided.

Research Method

This study was divided into two processes, that is, the process of feature extraction and establishing the sentiment analysis dictionary, and the process of sentiment analysis. During the process of feature extraction and establishing the sentiment analysis dictionary, initially the relevant corpus was collected from the web and segmented by the Chinese word segmentation tool. Feature dictionaries used for classification prediction and sentimental analysis dictionaries used for sentimental analysis can then be extracted automatically or manually. During the process of sentiment analysis, the test corpus was processed by the Chinese word segmentation tool and classified into the five categories of librarians, collections, services, equipment, and space and environment, using classifiers and the related feature dictionaries. Then, the corpus of each category was analyzed using the proposed sentimental analysis algorithms to obtain the sentimental score (positive or negative). Moreover, the composite sentiment score was provided to be the library satisfaction score.

This study used Plurk as the corpus source. By referring to the "Aim for the Top" university grants list in 2010 from Ministry of Education, 14 university libraries were chosen to collect their complete Plurk corpus. The corpus collection data extended to December 31, 2011, for a corpus

of 1557 entries. In addition, in order to increase the corpus data types and quantity, both the automation system and the questions & answers forum system of National Chung Hsing University were employed to retrieve the library users' relevant opinion corpus as well, for 1994 and 98 entries, respectively, and a total of 2092 entries. However, some sentiment corpus cannot be judged by the context, or this study did not deal with English corpus, which was outside the scope of this study. There was a total of 16 entries, and an actual corpus of 3633 entries was used. After completing the corpus collection, the entries were segmented into Chinese words and cut into sentences. Moreover, all the stop words were removed. However, in order to avoid short sentences or sentence semantic ambiguity, flexibility in combining sentences was permitted. In addition, if one sentence has two or more associated sentimental targets, then the sentence was to be segmented into multiple sentences. Next, for each sentence manual tagging was conducted to annotate sentiment words, adverbs, negation words, collocation words and sentence sentiment. Then the annotation results are compiled into four dictionaries consisting of a sentiment dictionary, sentiment collocations dictionary, adverbs dictionary, and negation word dictionary, and duplicate words were removed.

First, for all category types the weights of all the words were computed, then for each category words with weights greater than a threshold or Top-N characteristic words were captured, to form feature word vectors. The K-nearest neighbor classification, Naïve Bayesian classification and SVM were employed to classify the sentences of each category into the five categories of librarians, collections, services, equipment, and space and the environment. In addition, with reference to the existing literature, this study proposes two methods of sentiment analysis: adverbs weighting sentimental analysis, as well as polarity and quadrant sentimental analysis, respectively. Finally, we used graphical representation to present the sentiment analysis results.

Research Result

In order to conduct the annotator consistency test, Kappa test was introduced. As all the Kappa test results were above 0.6, the results

showed high tagging/marker consistency. Then the Weka data mining tool was used to conduct the sentence classification. In terms of inside or outside test, the results showed KNN had a better result. In addition, for various combinations of corpus and categorization of degree adverbs, the abovementioned two sentimental analysis methods were deployed. In order to permit a comparison between the two different methods, multiple sentimental transformation scales were proposed. The results showed that the medium, weak, and no degree adverbs attributed to the same class, and polarity and quadrant sentimental analysis provide a better assessment of library user satisfaction results. In addition, as to the weight values for degree adverbs, strongest: 3, center: 2 and the weakest: 1.5, which is the best. Furthermore, the satisfaction evaluation using test corpus also confirmed the reliability analysis of the sentiment dictionaries. Concomitantly, for knowing every facet of user satisfaction the use of visualization methods, i.e., radar chart or bar chart, can illustrate better integrated library users' satisfaction.

Conclusions

This study not only explored Chinese corpus pre-processing and related technologies, but also constructed exclusive sentimental analysis dictionaries for library and information science. Meanwhile, sentimental analysis confirmed these dictionaries have both usability and credibility. In the corpus classification, when using only the default parameters for the classification tools the KNN results were better. In addition, compared with the adverbs weighting sentimental analysis, the polarity and quadrant sentimental analysis had a higher correct rate. Furthermore, combining the three kinds of corpus sources, Plurk, automation systems and forums constituted eight variety of corpus mix of sources, respectively. For different combinations of source corpus the sentimental scores were computed and compared. Among them, using the Plurk corpus and the polarity and quadrant sentimental analysis showed satisfactory results. Finally, the use of radar charts and bar charts can simultaneously present each category's outward sentimental tendencies. Besides, sentimental analysis result graphs are very similar to the standard answer graphs, while further enhance the effectiveness of

this study. For the future, in addition to continuing collecting the relevant corpus material, we will also continue enriching sentimental dictionaries and studying Chinese text pre-processing technologies. In addition, we will conduct a comparison study into the annual library user satisfaction evaluation results to understand the usefulness of this research and analysis to adjust the parameters to improve the performance of sentiment analysis. Furthermore, in order to timely provide libraries with accurate library user satisfaction, an important research topic will be how to deploy visualization for sentiment analysis results.

References

- Chang, Y. W. (2007). Applying Interview Approach to Study the Users' Satisfaction of "National Electronic Archival Retrieval System". *University Library Quarterly*, 11(1), 139-165. [Text in Chinese].
- Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.
- Drake, M. A. (1994). Technological Innovation and Organizational Change. *Journal of Library Administration*, 19(3-4), 39-49.
- Goodall, D. L. (1988). Performance Measurement: A Historical Perspective. *Journal of Librarianship and Information Science*, 20(2), 128-144.
- Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 415-425.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Retrieved October 14, 2011, from <http://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf>
- Huang, C. L., Chen, M.C., & Wang, C. J. (2004). Credit Card Scoring by Support Vector Machine. *Journal of Quantitative Management*, 1(2), 155-172. [Text in Chinese].
- Huang, Q. X. (2002). *Xiu ci xue*. Taipei: San Ming. [Text in Chinese].
- InsightXplorer Limited (2011). *Access Rating Online*. Retrieved November, 20, 2011, from <http://www.iama.org.tw/upload/ResourceReport/20110830105124836.pdf> [Text in Chinese].

- Kong, C. Z. (1998). *Corporate Image and Public Relations*. Taipei: Scientific and Publishing. [Text in Chinese].
- Ku, L. W., & Chen, H. H. (2007). Mining Opinions from the Web: Beyond Relevance Retrieval. *Journal of American Society for Information Science and Technology*, 58(12), 1838-1850.
- Landis, J. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.
- Lee, C. C. (2010). *Opinion Analysis of Chinese Blog Posts*. (Unpublished master's thesis). National Taipei University of Technology, Taiwan. [Text in Chinese].
- Lee, C. R. (2011). *Enhancing E-HowNet for Sentiment Analysis*. (Unpublished master's thesis). National Taiwan University, Taiwan. [Text in Chinese].
- Lee, J. J. (2009). *Xian dai han yu you biao ji de bi jiao ju zhi jiao xue pai xu-yi mu yu wei xi ban ya yu de xue xi zhe wei li*. (Unpublished master's thesis). National Taiwan Normal University, Taiwan. [Text in Chinese].
- Lee, M. C. (2009). *Opinion Analysis of Chinese Text using Machine Learning*. (Unpublished master's thesis). National Tsing Hua University, Taiwan. [Text in Chinese].
- Liu, C. L. (2000). *A Study of Job Satisfaction for Computerized Pre-press Workers*. (Unpublished master's thesis). National Taiwan Normal University, Taiwan. [Text in Chinese].
- Lu, S. J. (2003, December). Performance Measurement for Academic Libraries. *Bulletin of the Library Association of China*, 71, 1-19. [Text in Chinese].
- Lu, Y., Kong, X., Quan, X., Liu, W., & Xu, Y. (2010). Exploring the Sentiment Strength of User Reviews. *Lecture Notes in Computer Science*, 6184, 471-482.
- Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Mark H., Eibe F., Geoffrey H., Bernhard P., Peter R., & Ian H. W. (2009). The WEKA data mining software: an update. *Sigkdd Explorations*, 11(1), 10-18.
- McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for

- Naïve Bayes Text Classification. In *Proceedings of AAAI/ICML-98 Workshop on Learning for Text Categorization*(pp.41-48). California: AAAI Press.
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Information Retrieval*, 2(1-2), 1-135.
- Peng, J. T., Cheng, T. C., Cho, H. T., Chen, H. T., & Cheng, J. C. (2010). The study of Service Quality and User Satisfaction in Library-An Empirical Study of University Library. *Journal of Performance and Strategy Research*, 7(1), 53-73. [Text in Chinese]
- Salton, G., & Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing & Management*, 24(5), 513-523.
- Sun, Y.T., Chen, C.L., Liu, C. C., Liu, C. L., & Su, V. W. (2010). Sentiment Classification of Short Chinese Sentences. In *Proceedings of the 22th Conference on Computational Linguistics and Speech Processing*, Taiwan. [Text in Chinese].
- Tseng, S. S., Cai, X. M., Su, D. X., Tseng, Q. R., & Wang, Q. Y. (2006). *Data Mining*. Taipei: Flag. [Text in Chinese].
- Vapnik, V., & Cortes, C. (1995). Support Vector Networks. *Machine Learning*, 20(3), 273-297.
- Wilson, T., Wiebe, J., & Hwa, R. (2004). Just How Mad Are You? Finding Strong and Weak Opinion Clauses. In *Proceedings of the National Conference on Artificial Intelligence*. Retrieved October 14, 2011, from <http://www.cs.pitt.edu/~hwa/aaai04.pdf>
- Wu, H. H. (2011). *Sentiment Analysis Using Multi-dictionary and Commonsense Knowledgebase*. (Unpublished master's thesis). National Taiwan University, Taiwan. [Text in Chinese].
- Wu, M. D. (2003, December). Library Evaluation. *National Cheng Kung University Library Journal*, 11, 1-8. [Text in Chinese].
- Xu, G. P. (2007). *Studies of Comparative Category in Mandarin Chinese*. Shanghai: Academia Press. [Text in Chinese].
- Yang, C. H., & Chen, H. H. (2006). Emotion Classification Using Web Blog Corpora. In *Proceedings of the 18th Association for Computational Linguistics and Chinese Language*, Taiwan. [Text in Chinese].

- Yang, C. H., Kao, H. A., & Chen, H. H. (2007). Yi bu luo ge yu liao jin xing
ging xu fen xi. In *Proceedings of the 19th Association for Computational
Linguistics and Chinese Language*, Taiwan. [Text in Chinese].
- Yang, K. S., Wen, C. I., Wu, T. S., & Lee, Y. R. (1990). *Research Methods in
Social and Behavioral Sciences*. Taipei: Tunghua. [Text in Chinese].