東吳大學商學院

資訊管理學系碩士論文

指導教授：黃日鉦　博士

文本探勘與情緒分析於產品推薦之應用-以 PTT 電影版為例

Text Mining and Sentiment Analysis for the Application of the Product Recommendation-The Case of PTT Movie Board

研究生：　張傳珩 撰

中　華　民　國　一　〇　八　年　七　月

文本探勘與情緒分析漁產品推薦之應用-以 PTT 電影版為例

Text Mining and Sentiment Analysis for the Application of Product Recommendation-The Case of PTT Movie Board

研究生：張傳珩　　Student：Chuan-Heng, Zhang

指導教授：黃日鉦　　Advisor：Jih-Jeng, Huang

東吳大學商學院

資訊管理學系

碩士論文

A Thesis

Submitted to Department of Computer Science and Information Management

School of Business

Soochow University

July 2019

Taipei, Taiwan, Republic of China

中　華　民　國　一　〇　八　年　七　月

SOOCHOW
UNIVERSITY
UNTO A FULL GROWN MAN

誌謝

終於要畢業了，雖然因為身體關係而休學延誤了畢業時間，但是能畢業的感覺真是好，從當初進了東吳找了黃日鉦老師請他當指導教授，到現在口試完成準備離開學校，一轉眼好幾年也過去了，很幸運的碰上了幾個很熱心的同學，復學回來後，也很幸運的遇到了一個熱心的學弟幫忙處理事情，這一路走來很充實，也讓我學到了對於事情更好的做法及想法，成長的不少。

非常感謝黃日鉦老師對於我的細心指導，雖然論文進度因為事情關係而緩慢，但是有任何問題老師還是都非常細心的教導我，從一開始的手足無措，到現今的慢慢把事情解決，論文的撰寫也從題目的構想到有了現在的成果，真的非常感謝黃日鉦老師。

感謝同學及學長學弟的教導勉勵，給了我學術上的指導以及未來工作上的建議，對於其他跟學校無關的事情也是提供了不少建議，也幫助解決了不少論文上的問題，非常感謝你們。

最後要感謝我的父母，在研究所的過程中給予我支持與鼓勵，讓我能順利完成學業，謝謝弟弟的幫忙，也謝謝你們的幫助及鼓勵，也感謝耿華學長在我一開始入學給我的指導，也感謝陳劭與冠年，在碩士這階段認識你們真的是我的福氣，非常感謝你們，真的謝謝。

張傳珩 謹誌於

東吳大學資訊管理研究所

中華民國一〇八年七月

摘要

隨著網際網路資訊的進步以及智慧型裝置的普及化，網路上有著各種大量的資訊以及不同的社群網路平台，而消費者在購買物品時從以前的詢問消費者已經演變成至今會先上網搜尋相關資料以及評論，但是網路資訊非常龐大，消費者透過關鍵字在尋找評論時，需要閱讀大量的網頁以及文字，花費大量的時間，這對於消費者來說其實非常費神，本研究透過主題分析以及情緒分析後，提供消費者搜尋電影多面向之情緒分析結果，使消費者不需要再看完許多評論後，才能得知電影評價。

本研究蒐集 PTT 電影版半年的評論內容，透過情緒分析出形容詞的情緒分數，並且根據情緒分數推薦電影，接著使用主題分析得到的 Topic model 結合情緒分析的字詞分數，推薦更適合的電影給消費者。

關鍵字:文字探勘、情緒分析、電影推薦、主題模型。

ABSTRACT

Thanks to internet technology improvements and the smart devices popularized, we can find a huge variety of information and different kind of social media platforms. Nowadays people prefer to search for comments and information on the internet than ask others opinions before they make purchases. However, there is massive information around the internet world. When people use the keywords to search in the comments, they will have to read a lot of texts and pages, which will take a bunch of time. This is not an easy job for people. The research "Subject analysis" and "Emotional analysis" help people to search for the diversity of emotional analysis consequences from movies. People won't have to review many comments to understand the movie evaluation.

By collecting the half-year comments from PTT, this research has analyzed the adjective words to get the emotional score and use the score to build movie recommendations. After that, analyze the topics to get the topic models including the emotional score from analyzed words to give people the movie they prefer.

Keywords: Text Mining, Emotional Analysis, Movie Recommendation, Topic Model.

目錄

圖目錄

表目錄

第一章 緒論

1.1 研究背景與動機

大約在西元兩千年之前，當時網際網路還尚未普及，世人在做一件事情決定的時候，通人大部分的人是像身邊的親朋好友去詢問與事情有關的意見以及想法，隨著行動裝置以及社群網路的越來越發達及便利，人們在生活中可以接觸到網際網路的時間會越來越多，世人會開始使用網路來參考網上使用者的想法，而且越來越多的人會將使用者的心得以及相關意見 POST 上網，現今社會消費者如果要去購買一件商品，或是看一部電影，吃一間餐廳，通常都會上網去搜尋相關的心得或是評價當作一個參考，而且隨著網路技術的越來越發達，現在越來越多的平台可以讓消費者去上網分享心得，例如：mobile01、批踢踢實業坊、伊莉論壇…等。隨著平台越來越普及，上網分享心得的使用者也會來越越多。

目前台灣使用者人數最多的社群網站為台大批踢踢實業坊（telnet://ptt.cc），它的創站年齡以及使用人數皆為台灣各大社群網站之冠，以往在評論電影方面之相關的研究是使用評論系統，並且用情緒分析電影評論文章，但是只有情緒分系有時候並不是非常完善，所以本研究會加入主題分析並且結合情緒分析，使用兩種分析方法去得到讓消費者有更適合的電影類型推薦。

現今社會行動裝置之普及以及網路使用者人數越來越多，使用者面臨到資訊超載（information overload）之問題，相關資訊量非常的多，因為資訊量龐大，就算使用了多個關鍵字去上網搜尋，也非常難得到搜尋者滿意的答案，所以想要找到更準確且讓搜尋者覺得準確且滿意的答案的話，就要用其他的方法來輔助才行。

文字探勘與情緒分析的目的是使用所得到之相關主要的意見，取有用的內容，這個方法不需要花費大量的時間、金錢、以及人力，所需要的成本比普通的市場調查低了非常的多。

1.2 研究目的

電影的內容評論對於消費者來說是一件極為重要的事情，電影並沒有像是其他商品一樣的七天內退換貨，不滿意退費之類的服務，消費者在決定要不要觀看這部電影的時候，通常只能從電影公司剪接的預告片以及網路上消費者或是身邊親朋好友觀看之後的心得來當作參考。

本研究使用了文字探勘以及情緒分析消費者在台大批踢踢實業坊內之電影版上的評論內容，透過軟體分析之後取得資料，收集語料並且建立一個資料庫，把蒐集到之資料進行分析處理，建立相關的電影評論的情緒分析辭典，然後使用集群分析去分析各消費者在批踢踢實業坊之關係，並且將兩種分析所得到著結論去得到適合的電影類型並且推薦給消費者。

本研究目的可分為下列三樣:

1. 建立特定類型的特徵關鍵詞庫，提升情緒分析於電影主題的準確度。
2. 使用主題分析，將電影主題分類，得到推薦類型的主題模型。
3. 透過兩種實驗方式，方便消費者搜尋有幫助的目標資訊。

1.3 論文章節架構

本論文架構如下 ： 第一章為緒論，說明本論文之研究背景動機與目的。第二章則是介紹文本探勘、情緒分析與研究相關之文獻。第三章說明本研究所使用之方法。第四章則是將 BBS 留言板帶入本文所使用之研究方法。第五章則為研究結論以及未來發展。圖 1.1 為本研究流程圖。

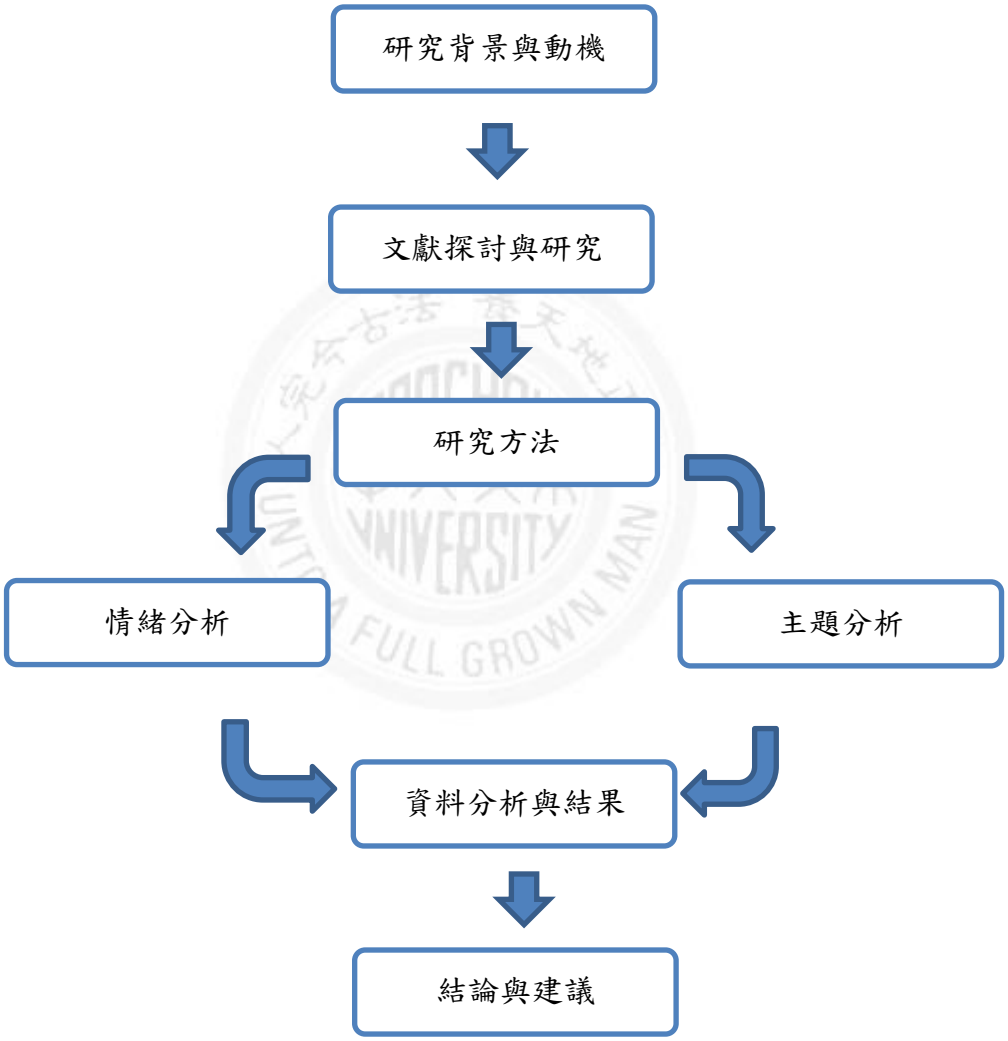


圖 1-1 研究流程圖

第二章 文獻探討

2.1 文字探勘

網際網路時代的來臨，讓使用者可以上網得到非常巨大的資訊，但是這些資訊量有部分是以文字來顯現出，這些龐大的資料皆可稱為大數據(Big Data) (Bollier, 2010) 。文字探勘的技術主要是來處理半結構化的文字資料，以得到這些文字資料中相關的結構規則。Tang＆Guo (2015)指出文字探勘包含的面向有非常多種，像是統計分析與資料庫的建立都是。

文字探勘(Text Mining)也可稱為文字知識挖掘(Knowledge Discovery from Text, KDT)是一種跨各種領域並且結合了資料探勘、資料萃取、以及語言處理的一種技術，此技術使用大量的文字資料經由軟體分析整理歸納。Sullivan（2001）定義文字探勘為「一種編輯、組織及分析大量文件的過程，為了要提供特定使用者特定的資訊，以及發現某些特徵及其間的關聯」。相較於傳統資料探勘，文字探勘需要加上一些額外的資料選擇處理程序，以及較為複雜的特徵萃取步驟。所以文字探勘其實是結合了多種傳統的資訊萃取技術，像是全文檢索以及關鍵字搜尋...等，圖 2-1 文字探勘技術金字塔顯示了目前常見的文字探勘技術及各技術間的層級關係。

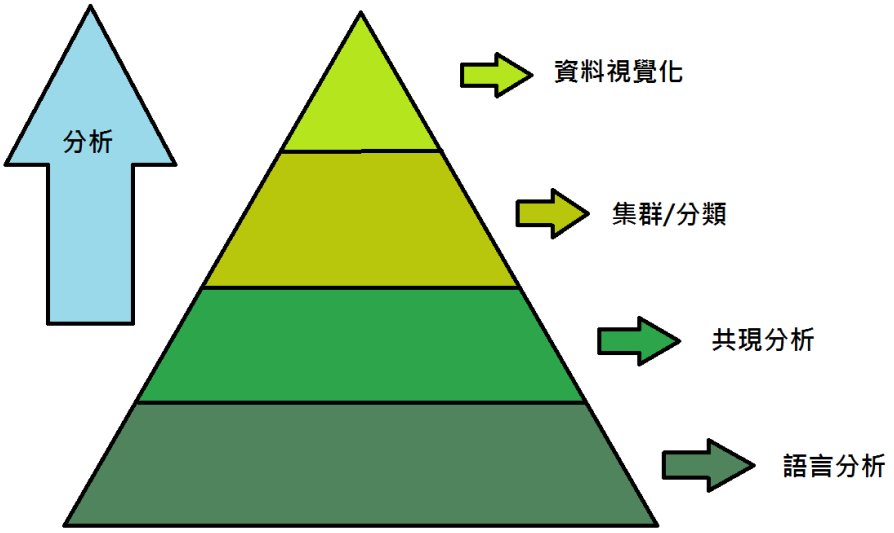


圖 2-1 文字探勘技術金字塔

(資料來源：Knowledge Management Systems－A Text Mining Perspective)

Brachman et al.(1996)的研究指出，利用資料探勘的演算法而找出有用資料的目的，整個資料挖掘的過程活動，就是得到解決問題的原因，而尹其言、楊建民(2010)的研究指出文字探勘需要先將文字特性量化，然後降低用字遣詞的差異性，接著是將把文字依據語言語義準確的辨識出來。因為文字並非全都是同性質的單位，所以需要量化後才能找出相關性，文字探勘在技術上結合了各個領域，如:文字關鍵字搜尋、人工智能、知識萃取…等，所以文字探勘與資料採礦技術相比會比較的費時費工。Han and Kamber (2000)的研究內容指出，資料探勘(Data Mining)為知識挖掘(Knowledge Discovery)中最重要的一步，其過程大致上可以分為以下幾個步驟，資料清理(Data Cleaning)、資料整合(Data Integration)、資料選擇(Data Selection)、資料轉

換(Data Transformation)、資料探勘(Data Mining)、樣式評估(Pattern Evaluation)、知識呈現(Knowledge Presentation)。

目前網路上的文章資料非常的巨大，並且這些文章通常都為非結構性文章，文字探勘就能處理這些非結構性的文章，並且得到文章中的重要資料，文字探勘在分析的時候要先擁有分析的文字語言資料庫(text corpus)，像是報告、信件…等。然後建立半結構化全文資料庫(text database)，之後得到結構化的資料術語矩陣（term-document matrix）。

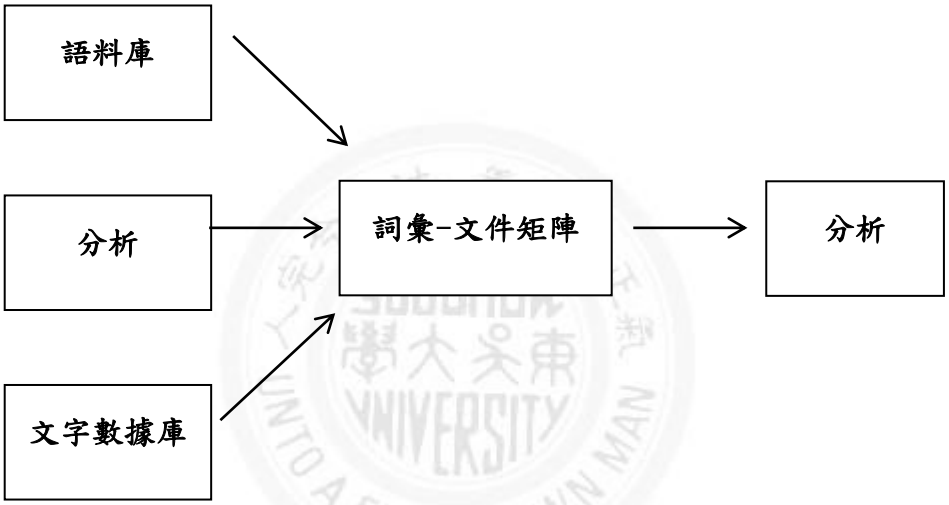


圖 2-2 文字探勘的處理流程

（資料來源：Text Mining in R（2012））

2.2 情緒分析

情緒分析是依據資料中的情緒以及使用者的心得將評論分類，可以得到文章作者在文章中的情緒評論，相較於傳統的評論分類，情緒分析更能有效率且快速的幫助分析者得到想要的資訊，目前情緒分析已經有效地應用於商品評論、企業之形象、音樂情緒辨識（Music Emotion Recognition）等領域之中，且人們在需要作出決定時，

會去尋求他人的意見，所以情緒分析也適用於個人以及團體組織(Zhang et al.,2018)。情緒分析可以分析出文章中所表達的情緒，並且分析出文章是正面還是負面的，情緒分析的分析方法為找出文章內經由人工標記的情緒類型，然後得到文字與情緒兩者之間的關聯性，所以如果使用者所蒐集的語料內容豐富，所分析出來知研究結果就越好。

情緒分析也是一種跨領域的技術，以確定文章中的情緒並竊決定是否為正向與負向(Milagros et al., 2016)。Day and Teng (2017)指出情緒分析也可以是作者的評價、意見或是情緒狀態。楊昌樺的研究內容為以文章中的表情符號來分析作者的情感，並且將文章中的表情符號做情緒分類，最後研究結果發現降情緒分為正負兩種然後做 SVM 的情緒分類器。

Liu（2012）指出情緒分析是分析者對於資料來做出意見判斷或是分析作者的情緒、觀點、評價…等，然後得出意見之極性，而資料可以是商品、服務、議題…等。Isidoros 與 Ioannis (2016)指出情緒分析的目的主要是為了瞭解人們透過不同媒介對不同主題所發表的意見極性，例如演講、文字資料、手語…等。李啟菁(2010)的研究指出情緒判斷的單位通常可以是字詞、語句、段落、或是其他相關的大量網路語料，所以在進行分析之前會先探討分析對象之是否有主觀性質。Thayer（1990)的研究指出，把情緒的特徵分為能量(Energy)與壓力(Stress）x 兩軸，能量軸(Energy)的兩端為平靜(Calm)與積極(Energetic)，壓力軸(Stress)的兩端則為快樂(Happy)與焦慮(Anxious)。

2.3 情緒辨識方法

情緒分析的辨識方法分為以下兩種，字典法（lexicon-based methods）以及機器學習法（machine learning approaches）(Maynard & Funk, 2011)，介紹如下：

2.3.1 字典法（lexicon-based methods）

字典法的情緒分析辨識方法是利用軟體內已經建立好的字典進行情緒分析（徐筱雁，2014），目前僅有 NTUSD(台灣大學資訊工程研究所所建置之中文情緒詞辭典)與 HowNet(廣義知網）。

2.3.2.1 監督式機器學習法(Supervised)

監督式機器學習法通常用在文件分類上，透過文件量化的方法來訓練機器完成學習的工作，耗費人力進行類別標註，利用標註後的文件來訓練機器，並且每一個文件有一個正確答案，使用特徵向量與標籤給機器學習。Agarwal et al. (2011)的研究指出機器學習法的基本問題使從文本資料提取複雜的特徵並且找出其中相關性。

2.3.2.2 非監督式機器學習法(Non-Supervised)

非監督式機器學習法並不需要像是監督式學習法那樣要先耗費大量人力進行人工標註，而是利用詞庫內帶有情緒的詞而把資料進行分類，將相似度高的資料歸類在同一類別，分類在同一個群內。Chaovalit and Zhou(2005) 的研究指出非監督式的學習方法雖然有快速方便的優勢，但是整體來說準確度遠遠不及於監督式機器學習法

以下是本文所找到之過去幾年的有關於情緒分析的研究，李啟菁（2010）所研究之數位相機之評論為例子，此研究是因為要了解 BLOG 文章撰寫者的情緒，然後分析文章中有主觀性之句子，此研究方法為字典法，分析文章中的程度與意見詞，得到文章的整體評價，謝鎮宇（2010）的研究為飯店的評價系統，分析飯店評論中的意見詞，並且用意見探勘技術得到研究內容。

2.4 中文斷詞與情感分析用詞語集

我們在分析文章的之前，必須要先分辨文章中的詞，才能做程式分析處理，中文文章通常都是用段詞處理來分解出字詞，英文的話兩個詞之間都會有空格，斷詞方法有兩種，第一個為統計式斷詞法（statistical based approach），第二個是辭典式斷詞法（dictionary based approach），辭典斷詞法事前必須要先建立斷詞法專用的詞庫，詞庫如果資料越大，斷詞的效果越優，反之詞庫如果資料小，所得到之效果就不好，

林孟翰（2011）指出中文斷詞法通常會碰到未知詞的問題，所以在使用斷詞系統的時候，讓人工標記的量減少，斷出正確的字詞為發展重點，中研院所判定的斷詞規則為長辭優先、詞長標準差小者優先、附著語素最小者優先、定量複合詞自數最少者優先、一字詞詞頻最高者優先、總詞頻最高者優先。

目前來說知網(HowNet)是最多人使用的中文斷詞系統，知網(HowNet)於 1988 年由董振東教授所創立，董振東教授在 2003 年時語中研院詞庫小組把中研院詞庫小組詞典(CKIP Chinese Lexical Knowledge Base)與知網(HowNet)兩樣作為連結並且修改，得到了廣義知網(Extended-HowNet, E-HowNet），董振東指出，在語言處理之前，並須要有一個資料充足巨大的知識庫，知識庫內必須要有屬性與概念之間的關係，下表為知網概念相關連結，以此表說明

表 2-1 知網各概念相關連結

編號	概念關係	範例
1.	上下位關係	由概念的主要特徵體展現
2.	同義關係	（可通過《同義、反義以及對義組的形成》獲得）
3.	反義關係	（可通過《同義、反義以及對義組的形成》獲得）
4.	對義關係	（可通過《同義、反義以及對義組的形成》獲得）
5.	部件－整體關係	心、CPU
6.	屬性－宿主關係	顏色、速度
7.	材料－成品	布、麵粉
8.	施事／經驗者／關係主體－事件關係	司機、雇主
9.	受事／內容／領屬物等－事件關係	患者、雇員
10.	工具－事件關係	手錶、計算機
11.	場所－事件關係	銀行、醫院
12.	時間－事件關係	假日、孕期
13.	值－屬性關係	藍、慢
14.	實體－值關係	傻子、傻瓜
15.	事件－角色關係	購物、盜墓
16.	相關關係	穀物、煤田

（資料來源：知網（http://www.keenage.com/zhiwang/c_zhiwang.html））

知網(HowNet)並不只是一個詞典，有點名符其實的算是一個真正的網，主要是把概念的共性與個性所反映出來，董振東以下圖 2-3 為例子，「醫生」與「病患」，「人」就是這兩者中間的共性，「醫療」這個動作為醫生個性，「病患」的個性就是本身「病患」的經驗，而對於「富翁」與「窮人」，「帥哥美女」與「醜人」來說，「人」就是這四者之間的共性，而富有程度以及美不美觀就是這四者之間的個性。

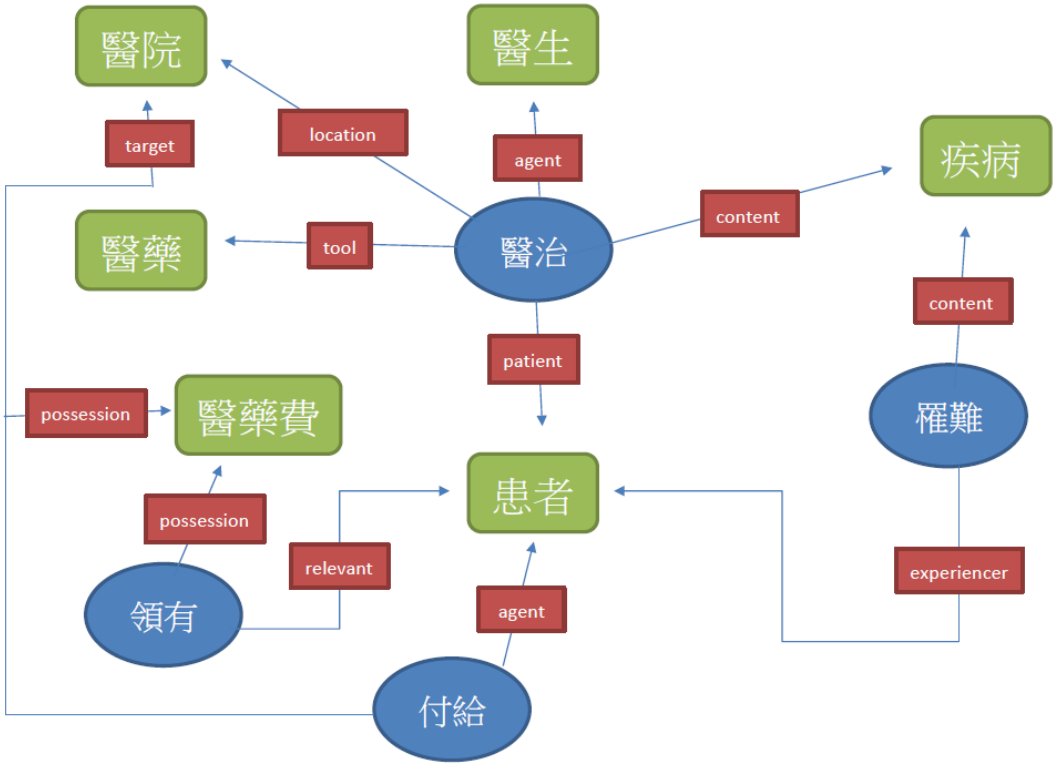


圖 2-3 知網概念系統圖

（資料來源：知網 http://www.keenage.com/zhiwang/c_zhiwang.html）

第三章 研究方法

3.1 研究架構

本論文之研究架構圖如下，首先抓取資料並且建構語料庫，然後進行文字處理，去除數字與符號等…，接著找出所有的形容詞並且建立文字雲，分析之後得到字詞情緒分數 ，接著進行主題分析並且得到 TOPIC 模型，之後進行情緒分析得到 THETA 分數再探討分析結果是否準確。圖 3-1 為本研究架構圖。

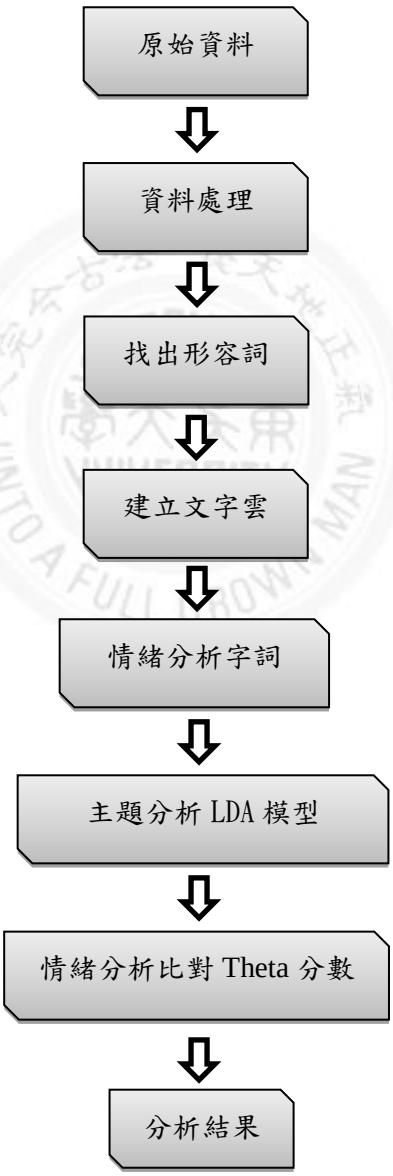


圖 3-1 研究架構圖

1. 收集資料: 利用 Python 抓取 PTT 電影版這半年來的文章，並且作為本研究的實驗資料。

2. 實驗分析：實驗資料進行處理 : 將從 PTT 電影版抓取的這半年文章進行清理，刪除數字與符號，並且進行中文斷詞處理，找出資料中的形容詞並且建立文字雲。

情緒分析 : 針對建立的文字雲進行情緒分析處理，分析出自詞的情緒分數。

主題分析 : 此步驟使用 LDA 模型處理，分析評論者主題中的特徵詞。

情緒分析 : 根據主題模型與字詞的情緒分數分析之後得到 Theta 分數作為分析本研究的最終結果並且探討分析成果。

3.2 論文實驗語料

本研究所使用之實驗語料來自 BBS 站台大批踢踢實業坊的 MOVIE 板，PTT 實業坊歷史悠久，也是台灣最大的網路論壇，每日熱門時間約有 14 萬人同時在線上使用，且 PTT 站的討論內容非常廣泛，本文所研究的 MOVIE 板便是用來討論電影的內容，使用者也可以發表對於電影內容之相關心感想。

圖 3-2 為批踢踢畫面截圖：

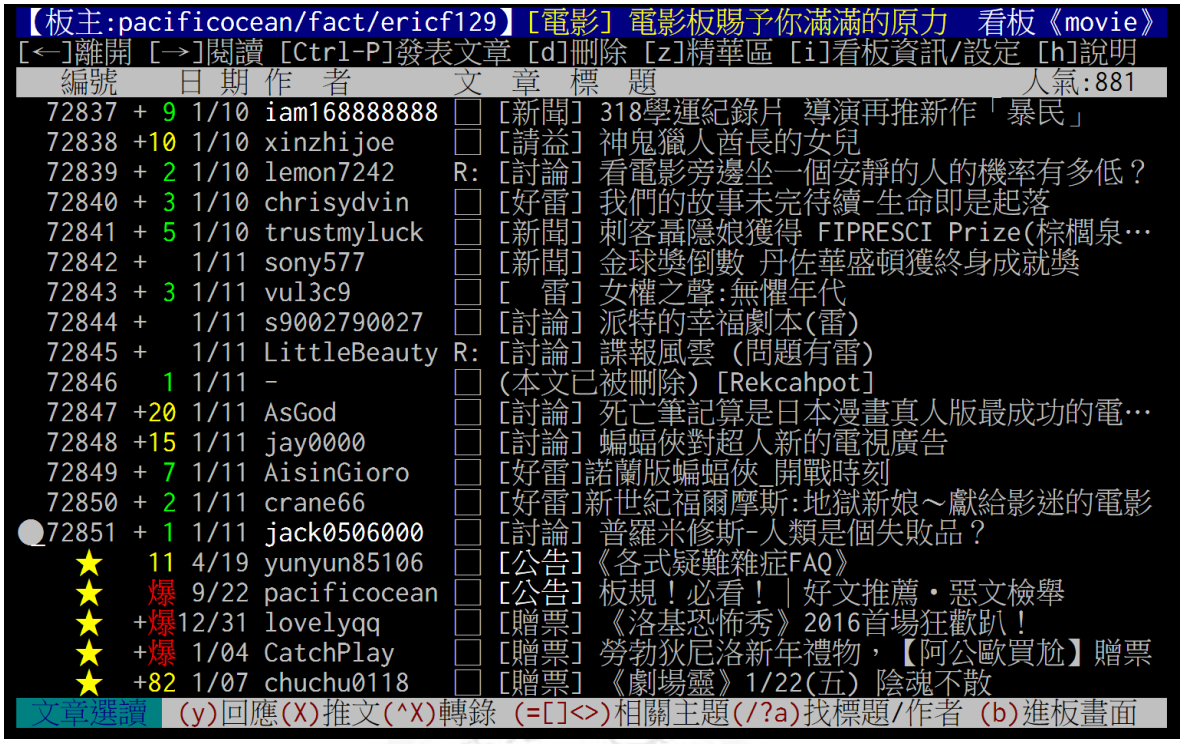


圖 3-2 批踢踢電影版截圖

本研究會先擷取電影評論文章，文章從 2019 年 1 月-2019 年 6 月的文章中，抓取這半年來的文章資料，並且使用 Python 從 MOVIE 板上抓取這 3788 部電影的評論文章。取出評論文章之後，以程式方法檢查每篇評論文章，如有內文明顯與評論主題無關的，予以刪除，評論文章內有大量重複文字、特殊符號與不具文章性質之評論內容也同樣予以刪除，剩下的擷取文章就為本研究之實驗語料。

3.3 實驗語料處理

在開始實驗分析之前，本研究進行情緒分析與主題分析都需要先將抓取下來的 PTT 電影版資料做前置處理，這步驟會清楚不需要的語料，並且人工整理刪除不需要的資料以供後續情緒分析與主題分析計算與使用。

3.3.1 刪除無意義文字

本研究資料來源為台大批踢踢實業坊之電影版(MOVIE），圖 3-3 為電影版之評論文章內容之文字截圖，截圖內容包含文章評論內容、作者名稱、看板名稱、文章日期、標題名稱、作者 IP、文章網址、下方使用者之推文與虛文、下方推噓文使用者的 IP 與使用者 ID。

圖 3-3 PTT 電影版評論文章內容之文字截圖

本研究將上圖內容如作者 IP、推噓文者 IP、文章日期、文章網址…等文字對於實驗無任何幫助，本階段將會刪除這些無意義文字，並且將文章內容標題人工刪除修改成只有電影名稱，無其他無意義之文字如圖 3-4 所示：

	A	B	C	D	E	F	G	H	I	J
1		content	title							
2	老大人	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:老大人 2.觀影時間:4/6 3.觀影地點:中影屏東景								
3	護墊俠	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:護墊俠 2.觀影時間:晚上 3.觀影地點:pass 4.觀								
4	大冒險家	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:2.觀影時間:拒答 3.觀影地點:美麗華影城 4.觀								
5	侵密室友	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:侵密室友 2.觀影時間:4/1 3.觀影地點:京華城 4								
6	雞不可失	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:雞不可失 2.觀影時間:2/20 3.觀影地點:大直美								
7	一個巨星的誕生	昨天看完一個巨星的誕生 覺得女神卡卡絕對是未來的影后 演得太精彩了 從第一次上台唱歌								
8	一個巨星的誕生	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:一個巨星的誕生 2.觀影時間:/2月28日10:35 3月								
9	比悲傷更悲傷的故事	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:比悲傷更悲傷的事 2.觀影時間:/ 拒答 3.觀影地								
10	比悲傷更悲傷的故事	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:比悲傷更悲傷的故事 2.觀影時間:/ 12/01 3.觀影								
11	我們	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:我們 2.觀影時間:11:10 3/22 3.觀影地點:板橋								
12	復仇者聯盟4	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:復仇者四 終局之戰 2.觀影時間:/拒答 3.觀影地								
13	喜歡綠皮書	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:幸福綠皮書 2.觀影時間:3.觀影地點:長春國賓								
14	怪獸與葛林戴華德的罪行	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱:2.觀影時間:11/16 3.觀影地點:大千 4.觀影方								
15	花樣奶奶說英文/I Can Speak	https://i.imgur.com/gDADRo7.jpg 本來是衝著李帝勳看的 電影的第一幕他搭公車 擠到常常差點								

圖 3-4 標題修改

3.3.2 文章資料中文斷詞處理

中文句子因為沒有空白符號，跟英文句子相比，在取得文中的詞會花費比較多的時間。本文利用中央研究院的中文斷詞服務系統，將 2019 年 1 月-6 月的文章資料(PTT 的 MOVIE 版內的文章與底下使用者討論之內容)使用斷詞系統之後，取得語料庫。語料庫內的句子需要做正面或反面的情緒判定，不過因為大多數的字詞都跟情緒無關，所以判定句子情緒這步驟如下圖所示。

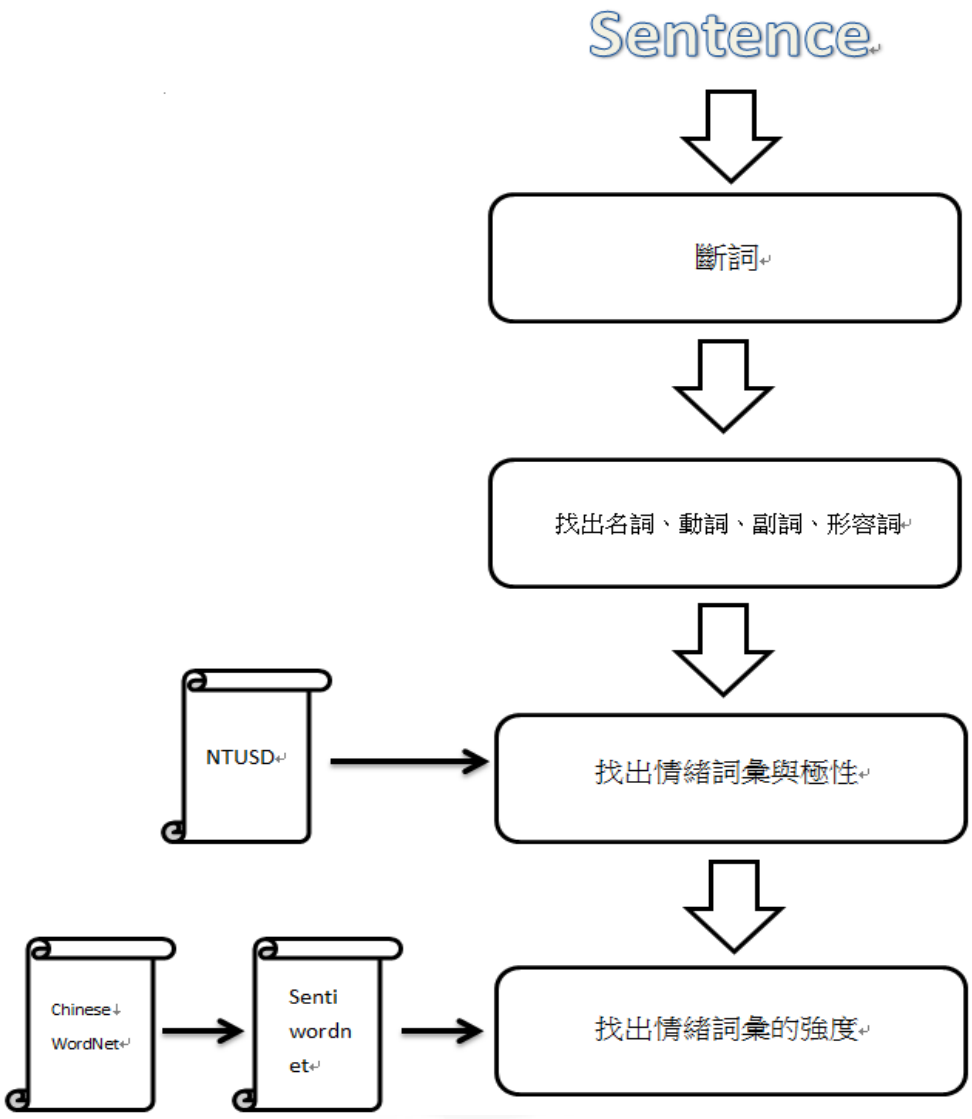


圖 3-5 語料庫情緒判定步驟

步驟 1. 斷詞

本研究的資料內容來源為台大批踢踢實業坊內之 MOVIE 電影版內的電影評論文章與底下其他使用者的相關推文，並且將抓取下來的文章經由中研院斷詞系統來分析。

步驟 2.找出名詞、動詞、副詞、形容詞

名詞、動詞、副詞、形容詞、是使用頻率最高的字詞，所以這個步驟會將這些有磁性的字詞擷取出來。

步驟 3.找出情緒詞彙以及極性

從步驟 2 挑選出來這些詞之後，判定這些字詞是否為情緒字詞，並且參考 NTUSD(台大情緒字典），如果在 NTUSD 裡面有找到相同的字詞，則給予該字詞標記極性。

步驟 4.找出情緒詞彙的強度

本步驟則參考中文辭彙網路(Chinese Wordnet)來找出辭彙的情緒強度，先使用中文辭彙網路找出情緒字詞的意義，如果有的話則會出現意義號碼，並且本研究再使用 Sentowordnet 比照使用中文辭彙網路所找出意義帶碼兩個相對比較，就可以找出情緒強度 。

下圖 3-6 與 圖 3-7 為本研究進行中文斷詞前後的範例

作者 czqaz（最近很 ok~ˆˆ） 看板 movie

標題 ［選片］復仇者 4 上古尊者的問題

時間 Fri May 17 00:13:33 2019

請問

復仇者 4

上古尊者聽到史傳奇交出石頭後

說 也許是他自己錯了，這是什麼意思？

什麼東西錯了？為什麼錯了，好糾結啊。

為何上古尊者覺得自己錯了，然後也交出石頭？？

煩請厲害的各位幫忙解答～～

圖 3-6 評論文章斷詞前

請問

復仇｜者 ｜4

上古｜尊者｜聽到｜史｜傳奇｜交出｜石頭｜後

說｜也許｜是｜他｜自己｜錯｜了｜這｜是｜什麼｜意思

圖 3-7 評論文章斷詞後

3.4 情緒分析

本步驟將用 R 語言撰寫程式碼進行情緒分析，將電影名稱全部獨立分開並且分析出各個電影內的字詞情緒分數， 本研究根據抓取資料所整理出的字詞表如表 3-1 所示

表 3-1 形容詞字詞表

巧妙	好奇	年輕	有趣	自由	完整	快樂	幸福	明顯
勇敢	幽默	很大	很棒	流暢	美好	重新	容易	恐怖
特效	強大	強烈	清楚	細膩	連結	單純	悲傷	最大
最好	殘酷	傷害	意外	瘋狂	緊張	輕鬆	厲害	複雜
遺憾	優秀	尷尬	簡單	豐富	驚奇	驚悚	驚喜	

字詞情緒分析分數計算方式有兩種，第一是利用情緒辭典本身就有的分數，第二是根據情緒字詞與特徵關鍵詞兩者間的距離給予分數，本研究根據程式所計算出各個電影的所對應情緒字詞分數如下圖 3-8

title	巧妙	好奇	年輕	有趣	自由	完整	快樂	幸福	明顯	勇敢	幽默	很大	很棒	流暢	美好	重新	容易	恐怖
人面魚 平均值	0	0	0	0	0	0.43872	0	0	0	0	0	0	0	0	0	0	0	0.526608
十二個想死的少年 平均值	0	0	0	0.845308	0	0	0	0	0	0	0	0	0	0	0	0	0	0
地獄怪客血后的崛起 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
花樣奶奶說英文/I Can Speak 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
限時好友 平均值	0	0	0	0	0	0	0	0	0	0	0.260114	0	0.235728	0	0	0	0	0
借來的一百天，借屍還魂新玩法 平均值	0	0.576203	0	0	0	0	0	0	0	0	0	0	0	0	0	0.445037	0.894915	0
原本以為只是手機掉了 平均值	1.053499	0	0	0	0	0	0	0	0	0	0	0	0	0	0.806517	0	0	0
哥吉拉：哪噩者 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
旅貓日記 影評 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
真竊 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
祝我忌日快樂 平均值	0	0	0	0	0	0	4.213995	0	0	0	0	0	0	0	0	0	0	0
屠宰場守則 Slaughterhouse Rulez 平均值	0.093644	0	0	0	0	0.068245	0	0	0	0.095856	0	0	0	0	0	0	0	0.163834
移動城市 平均值	0	0	0	0	0.651905	0	0	0	0	0	0	0	0	0	0	0	0	0
窒息 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.819168
復仇者聯盟4—些小小私心 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
搖擺男孩 平均值	0	0	0	0.230538	1.066754	0	0	0	0	0	0	0.349524	0.321448	0	0.293279	0	0	0
新喜劇之王 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
跳痛先生 平均值	0	0	0	0.298344	0	0	0	0	0	0	0	0	0	0	0	0	0.184247	0
福爾圖螺之瞳 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
誰先愛上他 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.5MM 平均值	0	0	0.436448	0	0	0	0.183217	0	0	0	0	0	0	0	0	0	0	0
1分54秒 平均值	0	0	0	0	0	0	0	0.162104	0	0	0	0	0	0	0	0.141603	0	0
28天毀滅倒數—末日的孤獨與考驗 平均值	0	0	0	0	0	0	0	0	0.131773	0	0	0	0	0	0	0	0	0.294901
303 之旅 平均值	0	0	0	0	0	0.139593	0	0	0	0	0	0	0	0	0	0	0	0
7月22日重生 平均值	0	0	0	0	0	0.085307	0	0.297191	0	0	0	0	0	0	0	0.086535	0	0.248676
90分鐘末日倒數 平均值	0.063368	0	0	0.036318	0	0.03656	0	0	0	0	0	0	0.042094	0	0.057608	0.050861	0.059661	0.297289
A star is born. 一個巨星的誕生 平均值	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.522034	0

圖 3-8 電影名稱對應字詞之情緒分數

所有資料依據關鍵字詞與電影名稱進行情緒分析後，可得到各個電影與字詞所對應的情緒分數，情緒分數越大則代表越屬於這個面向，相對的情緒分數越小，則該面向越低。

3.5 主題分析(LDA 模型)

根據資料來源批踢踢實業坊電影版內版友討論的文章內容，此步驟利用 R 語言 Topicmodels 套件進行 LDA 主題分析並且得到 LDA 主題候選詞，LDA 是非監督式學習(Unsupervised Learing)，所以在分析之前並需要足夠的主題數量才有辦法進行分析，LDA 分析最重要的數值為 Topic 數，所以本研究會根據上面情緒分析所得到之字詞來設定 Topic 數，本步驟將會採用 K 層交叉分析法(K-fold Cross Validation)來進行分析，K 層交叉分析法就是將資料數分成 K 份，並且把一個子集與剩下的子集來做測試數據，之後經過 K 次測試就能得到一個數據，本研究用 15 層 K 交叉分析法來得到所需要 Topic 數 ，下圖 3-19 為 K 層交叉分析法圖。

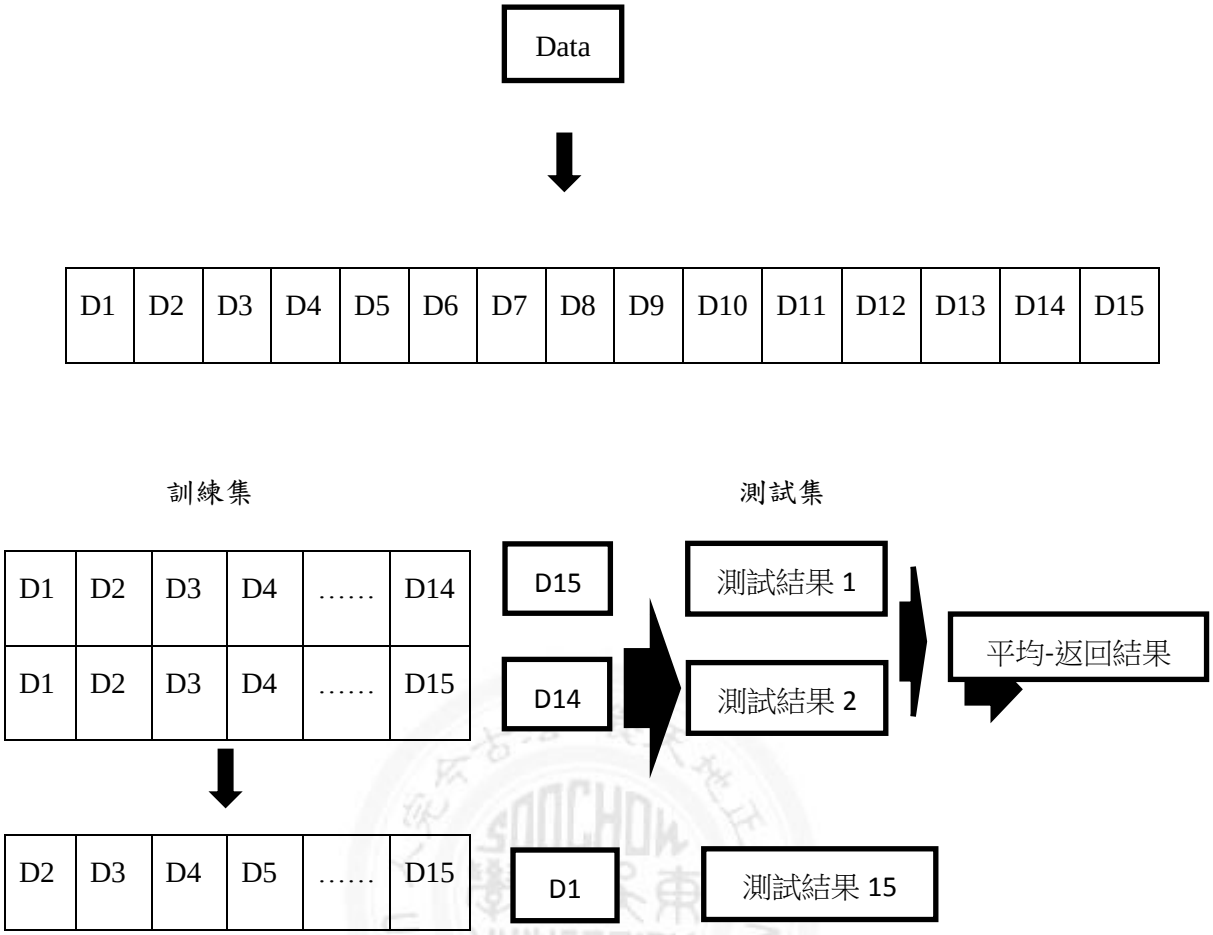


圖 3-9 K 層交叉分析法圖示

由圖示方法得到 Topic 數之後，將 TWord 數值設定為 15，所以每個 Topic 分類皆得到 15 個字詞作為候選詞。

第四章 實驗結果

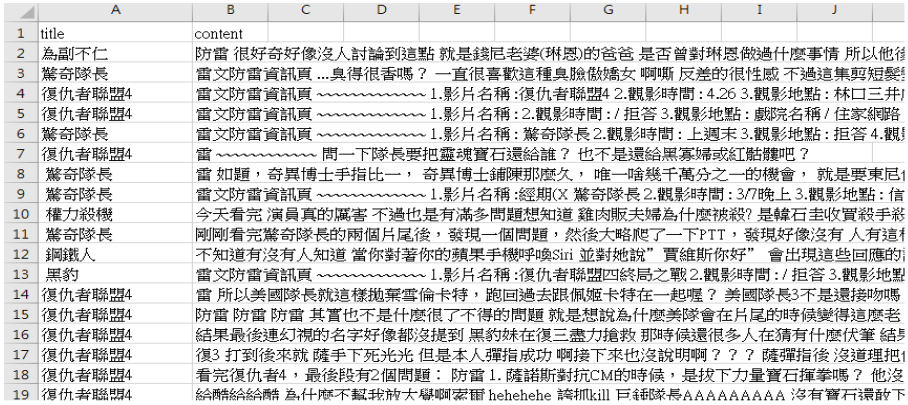
4.1 實驗方法

本研究採用台大批踢踢實業坊(Ptt. cc)論壇中電影版(Movie)內消費者對於電影評論內容做為實驗資料，並且依據研究方法章節之流程說明實驗結果，第一節為實驗語料處理，本節針對電影評論文章內容進行分類，以方便後續進行情緒分析，評斷電影評論內容是屬於哪一類方面，第二節字詞庫做情緒分析之結果，本節會把資料內所有電影個別做情緒分析判斷屬於哪一類電影，第三節為主題分析結果，本節採用 LDA 模型進行分析，並且根據結果提出 TOPIC 模型，第四節為 Theta 分數判斷，根據分數高低判斷出電影是屬於哪一類主題。

4.2 實驗語料處理

本研究實驗資料來源於台大 PTT 實業坊(Ptt. cc) 內其電影版(MOVIE)，抓取了半年來的電影評論內容文章，共有 3788 篇文章。

文字預先處理:把抓取下來的文章做資料預先處理，刪除數字與符號或是其他跟後續無相關之資料，此階段用程式抓取 3788 篇文章後，用 EXCEL 打開資料並且人工刪除數字與符號，以圖 4-1 與 4-2 所示。



	A	B	C	D	E	F	G	H	I	J
1	title	content								
2	為副不仁	防雷 很好奇好像沒人討論到這點 就是錢尼老婆(琳恩)的爸爸 是否曾對琳恩做過什麼事情 所以他後								
3	驚奇隊長	雷文防雷資訊頁 ...臭得很香嗎？ 一直很喜歡這種臭臉傲嬌女 啊嘶 反差的很性感 不過這集剪短髮								
4	復仇者聯盟4	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱 :復仇者聯盟4 2.觀影時間 : 4.26 3.觀影地點 : 林口三井								
5	復仇者聯盟4	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱 : 2.觀影時間 : / 拒答 3.觀影地點 : 戲院名稱 / 住家網路								
6	驚奇隊長	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱 : 驚奇隊長 2.觀影時間 : 上週末 3.觀影地點 : 拒答 4.觀								
7	復仇者聯盟4	雷 ~~~~~~~~~~~~ 問一下隊長要把靈魂寶石還給誰？ 也不是還給黑寡婦或紅骷髏吧？								
8	驚奇隊長	雷 如題，奇異博士手指比一， 奇異博士鋪陳那麼久， 唯一啥幾千萬分之一的機會， 就是要東尼								
9	驚奇隊長	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱 :經期(X 驚奇隊長 2.觀影時間 : 3/7晚上 3.觀影地點 : 信								
10	權力殺機	今天看完 演員真的厲害 不過也是有滿多問題想知道 雞肉販夫婦為什麼被殺? 是韓石圭收買殺手殺								
11	驚奇隊長	剛剛看完驚奇隊長的兩個片尾後，發現一個問題，然後大略爬了一下PTT，發現好像沒有 人有這								
12	鋼鐵人	不知道有沒有人知道 當你對著你的蘋果手機呼喚Siri 並對她說”賈維斯你好” 會出現這些回應的								
13	黑豹	雷文防雷資訊頁 ~~~~~~~~~~~~~~ 1.影片名稱 :復仇者聯盟四終局之戰 2.觀影時間 : / 拒答 3.觀影地點								
14	復仇者聯盟4	雷 所以美國隊長就這樣拋棄雪倫卡特，跑回過去跟佩姬卡特在一起嘿？ 美國隊長3不是還接吻嗎								
15	復仇者聯盟4	防雷 防雷 防雷 其實也不是什麼很了不得的問題 就是想說為什麼美隊會在片尾的時候變得這麼老								
16	復仇者聯盟4	結果最後連幻視的名字好像都沒提到 黑豹妹在復三盡力搶救 那時候還很多人在猜有什麼伏筆 結								
17	復仇者聯盟4	復3 打到後來就 薩手下死光光 但是本人彈指成功 啊接下來也沒說明啊？？？ 薩彈指後 沒道理把								
18	復仇者聯盟4	看完復仇者4，最後段有2個問題： 防雷 1. 薩諾斯對抗CM的時候，是拔下力量寶石揮拳嗎？ 他沒								
19	復仇者聯盟4	給酷給給酷 為什麼不幫我放大學啊索爾 hehehehe 跨抓kill 巨鋼隊長AAAAAAAAA 沒有寶石還敢下								

圖 4-1 無意義資料刪除前

為副不仁	防雷 很好奇好像沒人討論到這點 就是錢尼老婆(琳恩)的爸爸 是否曾對琳恩做過什麼事情 所以他後來對兩個孫女親近時 錢尼才
驚奇隊長	雷文防雷資訊頁 臭得很香嗎？ 一直很喜歡這種臭臉傲嬌女 啊嘶 反差的很性感 不過這集剪短髮變得很像女T 就沒硬了
復仇者聯盟4	雷文防雷資訊頁 ~影片名稱 復仇者聯盟4 觀影時間 .26 觀影地點 林口三井威秀.觀影方式 自行購票 其他防雷說明(非必要)
復仇者聯盟4	雷文防雷資訊頁 ~影片名稱 觀影時間 / 拒答 觀影地點 戲院名稱 / 住家網路 / 其他等地點 (單選 自行刪除選項) / 拒答 .觀影方
驚奇隊長	雷文防雷資訊頁 ~影片名稱 驚奇隊長 觀影時間 上週末 觀影地點 拒答 .觀影方式 自行購票 ~此區為發文防雷頁 可選擇性
復仇者聯盟4	雷 問一下隊長要把靈魂寶石還給誰？ 也不是還給黑寡婦或紅骷髏吧？
驚奇隊長	雷 如題 奇異博士手指比一 奇異博士鋪陳那麼久 唯一啥幾千萬分之一的機會 就是要東尼你英雄式犧牲 這點我覺得ok。 無限手
驚奇隊長	雷文防雷資訊頁 ~影片名稱 經期(X 驚奇隊長 觀影時間 3/7晚上 觀影地點 信義威秀.觀影方式 ㄟ屁屁訂票 ~此區為發文防
權力殺機	今天看完 演員真的厲害 不過也是有滿多問題想知道 雞肉販夫婦為什麼被殺? 是韓石圭收買殺手殺人? 韓石圭倒車撞死的殺手和
驚奇隊長	剛剛看完驚奇隊長的兩個片尾後 發現一個問題 然後大略爬了一下PTT 發現好像沒有 人有這樣的想法 所以想問問看 有沒有這樣
鋼鐵人	不知道有沒有人知道 當你對著你的蘋果手機呼喚Siri 並對她說”賈維斯你好” 會出現這些回應的話 https://i.imgur.com/fjuIGoa.jpg
黑豹	雷文防雷資訊頁 ~影片名稱 復仇者聯盟四終局之戰 觀影時間 / 拒答 觀影地點 戲院名稱 / 住家網路 / 其他等地點 (單選 自行
復仇者聯盟4	雷 所以美國隊長就這樣拋棄雪倫卡特 跑回過去跟佩姬卡特在一起嘿？ 美國隊長3不是還接吻嗎？ 還是這條線就沒有要收的意思
復仇者聯盟4	防雷 防雷 防雷 其實也不是什麼很了不得的問題 就是想說為什麼美隊會在片尾的時候變得這麼老 美隊被改造成強化人類的肉體
復仇者聯盟4	結果最後連幻視的名字好像都沒提到 黑豹妹在復三盡力搶救 那時候還很多人在猜有什麼伏筆 結果這一集好像沒出現相關的劇情
復仇者聯盟4	復3 打到後來就 薩手下死光光 但是本人彈指成功 啊接下來也沒說明啊？？？ 薩彈指後 沒道理把他自己的軍隊全滅啊 就算 一根

圖 4-2 資料修改後

此步驟要找出文章內所有形容詞，根據抓取下來的 3788 篇文章，用程式找出文章內的所有形容詞，並且建立文字雲圖，本研究所建立之字雲圖有兩個，分別為正面向字雲圖以及負面向字雲圖，正面向字雲圖形容詞像是有趣、幸福、簡單之類的正面向形容詞，所以本研究透過字雲圖上的關鍵形容詞判斷此字圖雲屬於正面向類別，如圖 4-3 所示。

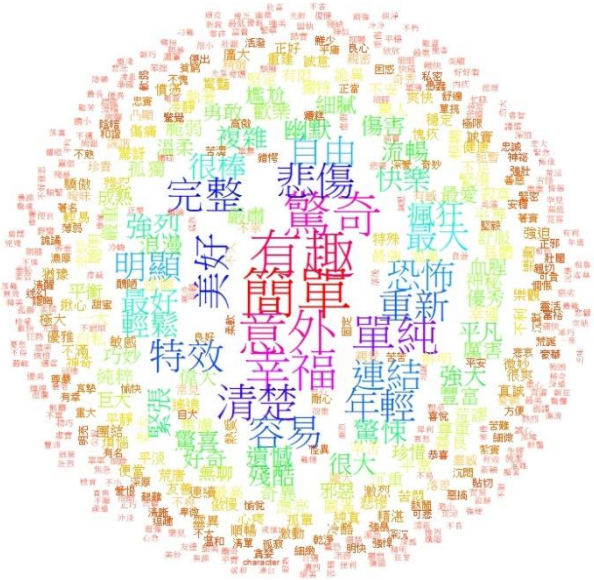


圖 4-3 正面向形容詞字雲圖

另一張字雲圖從圖上可以看出關鍵形容詞有悲傷、恐怖、尷尬、無聊…等，所以根據字雲圖上關鍵形容詞判斷出來此字雲圖屬於負面項類別，如圖 4-4 所示。

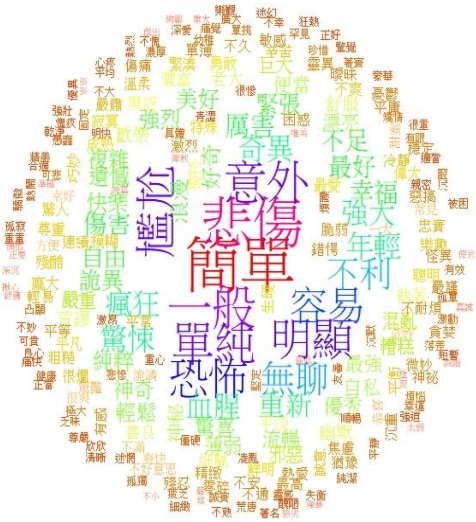


圖 4-4 負面向形容詞字雲圖

4.3 字詞情緒分析分數

本階段將資料內 690 部電影之情緒詞做情緒分析並且得到情緒分析分數，本階段提出情緒分析之方法為分別計算情緒詞之分數如下圖 4-5 所示，並且從電影名稱與形容詞分數可以選擇是否推薦哪一類的電影給消費者。

title	巧妙	好奇	年輕	有趣	自由	完整	快樂	幸福
人面魚 平均值	0	0	0	0	0	0.43872	0	0
十二個想死的少年 平均值	0	0	0	0.845308	0	0	0	0
地獄怪客血后的崛起 平均	0	0	0	0	0	0	0	0
花樣奶奶說英文/I Can Spea	0	0	0	0	0	0	0	0
限時好友 平均值	0	0	0	0	0	0	0	0
借來的一百天，借屍還魂	0	0.576203	0	0	0	0	0	0
原本以為只是手機掉了 平	1.053499	0	0	0	0	0	0	0
哥吉拉：噬星者 平均值	0	0	0	0	0	0	0	0
旅貓日記 影評 平均值	0	0	0	0	0	0	0	0
真寵 平均值	0	0	0	0	0	0	0	0
祝我忌日快樂 平均值	0	0	0	0	0	0	4.213995	0
屠宰場守則 Slaughterhouse	0.093644	0	0	0	0	0.068245	0	0
移動城市 平均值	0	0	0	0	0.651905	0	0	0
窒息 平均值	0	0	0	0	0	0	0	0
復仇者聯盟4一些小小私心	0	0	0	0	0	0	0	0
搖擺男孩 平均值	0	0	0	0.230538	1.066754	0	0	0
新喜劇之王 平均值	0	0	0	0	0	0	0	0
跳痛先生 平均值	0	0	0	0.298344	0	0	0	0
福爾圖娜之瞳 平均值	0	0	0	0	0	0	0	0
誰先愛上他 平均值	0	0	0	0	0	0	0	0
0.5MM 平均值	0	0	0.436448	0	0	0	0.183217	0
1分54秒 平均值	0	0	0	0	0	0	0	0.162104
28天毀滅倒數一末日的孤獨	0	0	0	0	0	0	0	0
303 之旅 平均值	0	0	0	0	0	0.139593	0	0
7月22日重生 平均值	0	0	0	0	0	0.085307	0	0.297191

圖 4-5 形容詞的情緒分析分數

之後從上述形容詞選出，如果消費者想要看形容詞-“有趣的”電影 ，就可以篩選出下圖電影並且推薦給消費者,如圖 4-6 所示,如果消費者想要看”有趣”加上”自由”的電影，也可篩選出如圖 4-7 所示

1	title	巧妙	好奇	年輕	有趣	自
3	十二個想死的少年 平均值	0	0	0	0.845308	
42	John Wick3裡的軍隊 平均值	0	0	0	0.633981	
64	七罪追緝令 平均值	0	0	0	0.845308	
95	大都會Metropolis 為何默片	0	0	0	0.633981	
96	大釣哥：有趣 平均值	0	0	0	1.267962	
107	小鬼當家 平均值	0	0	0	2.535923	
127	天做凶殺案 平均值	1.053499	0	0	0.633981	
183	亦正亦邪 平均值	0	0	0	0.633981	
207	至少沒失望的雞不可失 平均	0	0	0	0.633981	
334	阿馬爾菲：女神的報酬 平	0	0	0	0.845308	
484	尋找柏格曼 平均值	0	0	1.115366	0.845308	
605	綠色奇蹟 平均值	0	0	1.115366	0.845308	
625	樂高玩電影1與2 平均值	0	0	0	1.521554	
635	醉拳 平均值	0	0	0	2.535923	

圖 4-6 有趣的電影名稱排名

	title	巧妙	好奇	年輕	有趣	自由
	搖擺男孩 平均值	0	0	0	0.230538	1.066754
4	阿馬爾菲：女神的報酬 平	0	0	0	0.845308	1.303811
5	黑鏡 平均值	0	0.025209	0.059752	0.01585	0.928965

圖 4-7 有趣與自由的電影名稱

4.4 主題分析(LDA 模型)

本節採用 R 語言 Topicmodels 套件來做 LDA 主題模型分析研究，本研究將電影名稱進行主題分析，可以更方便快速的讓使用者了解電影的特性，並且整理出 TOPIC 代表詞讓消費者了解電影內的資訊，Topic model 表 4-1 如下所示:

表 4-1 LDA 主題分析 Topic Model

主題 1	主題 2	主題 3	主題 4	主題 5
驚奇	幸福	年輕	意外	完整
有趣	悲傷	連結	單純	恐怖
特效	最好	自由	美好	最大
很棒	快樂	瘋狂	容易	明顯
強大	傷害	清楚	重新	強烈
幽默	平凡	輕鬆	遺憾	驚悚
流暢	勇敢	複雜	歡樂	緊張
驚喜	偉大	浪漫	成熟	有趣
尷尬	殘酷	嚴肅	珍惜	血腥
最愛	嚴重	神秘	無聊	厲害
奇異	孤獨	好奇	巨大	純粹
很大	寂寞	舒服	脆弱	邪惡
漂亮	細膩	焦慮	幸運	優秀
老大	沉重	平等	清楚	荒謬
最強	善良	平常	溫柔	巧妙

本研究可透過 Topic Model 中的各種主題詞快速幫助消費者了解電影的主題性，以 Topic 2 來舉例，主題代表詞 1 至 5 為「幸福」、「悲傷」、「最好」、「快樂」、「傷害」，可以經由主題詞來推論出齊集群特色喜愛為「負面愛情類電影」，而以 Topic 5 的主題代表詞來推論出，此集群特色可能為「恐怖類電影」，以 Topic 4 的主題代表詞來推論出，此集群特色可能為「文藝類電影」

4.5 實驗結果推薦

根據主題分析(LDA 模型)的結果，本研究將上述 Topic 主題詞 1-5 各自重新命名，Topic 1 為「喜劇動作類型」，Topic 2 為「負面愛情類型」，Topic 3 為「劇情懸疑類型」，Topic 4 為「文藝類型」，Topic 5 為「恐怖類型」，並且根據程式分析結果得到電影對應 Topic 模型 5 種類型所計算出 Theta 分數，以電影「驚奇隊長」來舉例，此部電影的 Theta 分數在 Topic 1-5 為 0.33333 、0.174603 、 0.174603 、 0.15873 、 0.15873 ，根據分數高的來選擇，所以「驚奇隊長」此部電影屬於 Topic 1 喜劇動作類型，以下圖 4-8 所示 ：

新喜劇之王	0.211538	0.192308	0.192308	0.192308	0.211538
驚奇隊長	0.33333	0.174603	0.174603	0.15873	0.15873
誰先愛上他	0.207547	0.207547	0.207547	0.188679	0.188679

圖 4-8 「驚奇隊長」的 Theta 分數

以電影「幸福騙局」來舉例，此部電影最高的 Theta 分數在 Topic 2，所以根據 Theta 分數高的來選擇出「幸福騙局」為 Topic 2 「負面愛情類型」電影，如圖 4-9 所示 ：

地獄怪客血后的崛起	0.2068966	0.2068966	0.1724138	0.1896552	0.2241379
血派之樹	0.1466667	0.1466667	0.32	0.2	0.1866667
幸福騙局	0.1176471	0.4019608	0.2156863	0.1568627	0.1078431
限時好友	0.2096774	0.2258065	0.1774194	0.1935484	0.1935484
原本以為只是手機掉了	0.1851852	0.1851852	0.2037037	0.2037037	0.2222222

圖 4-9 「幸福騙局」的 Theta 分數

第五章 研究結論與建議

5-1 研究結論

本論文的研究目的在於使用兩種方法來推薦消費者更適合的電影類型主題，根據從批踢踢社群網站所抓取之資料分析出電影類型，以程式分析所抓取的語料進行分類，之後透過情緒分析出字詞的情緒分數，找出情緒詞姓，並且透過主題分析找出消費者的評論主離，將電影主題分類，依據上述方法可以根據消費者的喜好並且推薦適合的電影給消費者。

以往情緒分析在研究不同領域建立特徵詞的方式，幾乎都是透過大量人工的方法從文章資料中根據事先所設定好的研究議題篩選出來，並且不同時間範圍內的討論主題與內容會有所不同，所以抓取的語料也會產生很多不相同的主題，因此本研究會在情緒分析之後再透過 LDA 主題模型進行分析，這樣可以更容易抓取適合的消費者憑論內容，並且依據 LDA 分析出來的特徵關鍵詞，與詞庫進行語料分類。

5-2 研究貢獻

過往的情緒分析研究中，比少關於產品屬型方面的情緒分析，因此本研究以消費者觀點來研究，找出各種電影的不同屬性類型之情緒分析應用，透過這些面向進行消費者的語料情緒分析，本研究第一階段使用情緒分析實驗後，從分析結果可看出各個特徵關鍵詞所對應電影之情緒分數，研究第二階段透過主題分析電影評論文章，其中所有電影對應五個主題模型特徵關鍵詞，綜合上述兩種方法可以提供消費者搜尋出電影評論對應主題關鍵詞後有實用性的產品屬性多面向評價情緒，並且不用花費太多時間去瀏覽電影評論文章心得， 根據上述研究結論，本研究貢獻可歸納為以半自動化方式用程式建立產品屬性多面向特徵詞，並且使用兩個分析方法來推薦消費者更適合的電影主題類型。

5-3 研究限制與未來研究議題

此外本研究的不足之處，之後可以從實驗資料源頭多方蒐集，抓取不同領域的評論內容或是不同來源的資料並且進行分析，這樣可以得知是否不同領域的資料是否會導致有不同的結果，並且是否會影響推薦電影類型，此外，針對 LDA 分析之參數除了找出最高 Topic 數值之外，也可以另外探討 Tword 值，藉由不同的 Topic 與 Tword 值，去除重複的 Tword，這樣可以得到更精準的特徵關鍵詞，並且若能進一步的探討情緒程度，將情緒詞分成三種高中低的情緒，可以更精準的得到消費者的情緒表徵，此外納入語意分析，以加強情緒分析之準確度，由於本研究並沒有語意分析，所以無法得知消費者的評論內容是否有疑問或是有反串的意思，無法非常精準的判斷並且歸納到正確的情緒極性類別，在未來的研究上，文件的情緒分類，若評論文章有講述的重點，則該重點的情緒字詞分數應該要有所提高。

中文文獻

李啟菁.(2010).中文部落格文章之意見分析.(碩士),國立台北科技大學,台北市.

謝鎮宇.(2010).意見探勘在中文評鑑語料之應用.(碩專),交通大學,新竹市.

徐筱雁.(2014).情感分析中屬性詞與情感詞的關係之探討-以牛肉麵食評為例.(碩士),聯合大學,苗栗縣.

尹其言、楊建民(2010).應用文件分群與文字探勘技術於機器學習領域趨勢分析以 SSCI 資料庫為例,長榮大學學報，頁 1-16

林孟翰.(2011).基於中文斷詞技術之新聞網頁分類系統.(碩士),淡江大學,台北市.

顏國偉、譚慧敏.(1999).基於知網的常識知識標注，中文計算語言期刊 vol 4.no.2，頁 39-86

李淑惠.(2014).運用文字探勘技術與口碑分析之研究.(碩士),東吳大學,台北市.

邱鴻達.(2011).意見探勘在中文電影評論之研究.(碩士),交通大學,新竹市.

林國仲.(2017).運用情緒分析結合產品多面向自動分類於消費者評價之研究.(碩士),台南大學,台南市

張日威.(2014).應用LDA進行Plurk主題分類及使用者情緒分析.(碩士),雲林科技大學,雲林縣.

張莊平.(2012).中文文法剖析應用於電影評論之意見情感分類.(碩士),臺灣師範大學,台北市.

張育蓉.(2012).使用情緒分析於圖書館使用者滿意度評估之研究.(碩士),中興大學,台中市.

廖惠敏.(2015).網路美食評論情緒分析之研究.(碩士),高雄餐旅大學,高雄市.

趙玉娟.(2015).政治網路口碑的情感分析 ： 語意關聯性之觀點(碩士),交通大學,新竹市.

英文文獻

Agarwal, A., Xie, B.,Vovsha, I., Rambow., & Passinneau, R. (2011). Sentiment analysis of twitter data. *In proceedings of the workshop in languages in social media,30-38,*Association for Computational Linguistics.

Bing, Liu., (2012).Sentiment Analysis and Opinion Mining, *Morgan & Claypool Pulishers*.

Bollier, D., (2010). The Promise and Peril of Big Data, *The Aspen Institute*.

Brachman, R.J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G., & Simoudis, E., (1996). Mining Business Databases. *Communications of the ACM,* 39.(11),pp.47-48.

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 Sciences*, pp.112c- 112c.

Day, M. Y., and Teng, H. C., (2017). A Study of Deep learning to Sentiment Analysis on Word of Mouth of Smart Bracelet. *In Proceedings of the 2017 IEEE/ACM international Conference on Advances in Social Networks Analysis and Mining 2017*, pp.763-770.

Davis, F. D.,(1989). Perceived usefulness, Perceived ease of use, and user acceptance of information technology. *MIS Quarterly.13*(3), pp.319-340.

D. Sullivan.,(2001). Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales. John Wiley & Sons, Inc., New York.

Fernandez-Gavilance, M., Alvarez-Lopez, T., Juncal-Martinez, J., Costa-Montenegro, E., & Gonzalez-Castano, F. J.,(2016). *Unsupervised method for sentiment analysis in online texts.*Expert System With Application,58,57-75.

Han, J. and Kamber, M.,(2000). Data Mining: Concepts and Techniques. *Morgan Kaufmann,* San Francisco.

Li, G. and Liu, F.,(2012).Application of a clustering method on sentiment analysis. *Journal of Information Science* (38:2), pp 127-139.

Maynard, D. and Funk, A.,(2011).Automatic detection of political opinions in tweets. *InExtended Sentiment Web Conference*,pp.88-99.

Perikos, I., and Hatzilygeroudis, I., (2016). Recogninzing emotions in text using ensemble of classifiers. *Engineering Applications of Artificial Intelligence*, 51, 191-201.

R. E. Thayer.,(1990). The Biopsychology of Mood and Arousal.*Oxford University Press*,New York.

Sullivan, D,.(2001).Document warehousing and texe mining:techniques for improving business operations, marketing, and sales, *New York: John Wiley & Sons lnc*.

Tang, C. amd Guo, L.,(2015).Digging for gold with a simple tool:Validating text mining in studying electronic word-of-mouth(eWom) communication.*Marketing Letters*,26(1),67-80.

Zhang, L., Wang, S., and Liu, B.(2018).,Deep learning for sentiment analysis:A survey.*Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*,8(4).