楊亨利、林青峰（2018），『應用網路評價的功能商品推薦系統』，
中華民國資訊管理學報，第二十五卷，第三期，頁 335-362。

應用網路評價的功能商品推薦系統

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

隨著電子商務的發展，現在一個虛擬商場能夠查詢到的商品資訊常常超過人們所能負擔；如何從玲瓏滿目的商品中找到最符合需要，對使用者而言是非常重要的。傳統上的推薦系統是由個人資料、社會關係、購買或點閱記錄等資料來進行相似性的比對，進而完成推薦。本研究則採用網路上關於商品使用者評價文的意見傾向資料來進行推薦。因應消費者可能對商品功能屬性有不同的權重偏好，我們的系統考量到各個屬性構面的網路評價，並具有擴充彈性，會以代理人系統去網路上自動讀入可能的新商品與新屬性資料。在訓練好意見分類器之後，我們建立了一個推薦雛型系統，最後還進行了可用性及易用性的實驗。我們採用 SVM 分類器，在測試階段中，在巨觀平均上，我們的「非常負面、負面、中立、正面、非常正面」五類正面分類器的平均精準率、召回率、F1 值分別為 80.85%、79.81%、80.32%；五類負面分類器的平均精準率、召回率、F1 值分別為 81.66%、79.17%、80.39%。若如傳統一樣只分為「正面、中立、負面」三類的程度的話，其分類各項績效均為 80%以上、或甚至超過 90%。而在可用性、易用性實驗亦顯示這個系統在有用性、易用性與總評的各項目上，均可令使用者滿意。

關鍵詞：意見挖掘、情感分析、網路評論、推薦系統

* 本文通訊作者。電子郵件信箱：yanh@nccu.edu.tw
2017/09/24 投稿；2018/04/16 修訂；2018/06/27 接受

Yang,H.L. and Lin,Q.F. (2018), 'Functional product recommendation system based on sentiment mining from web reviews', *Journal of Information Management*, Vol. 25, No. 3, pp. 335-362.

Functional Product Recommendation System Based on Sentiment Mining from Web Reviews

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 a prototype system based on reviews opinion mining for functional product recommendation.

Design/methodology/approach－A multi-level SVM classifier was proposed and a prototype was also built. Finally, we designed an experiment for evaluating usability and ease of use.

Findings－In classifying five categories, i.e., {very negative, negative, neutral, positive, very positive}, our positive classifier had precision 80.85%, recall 79.81% and F1-score 80.32%; and negative classifier had precision 81.66%, recall 79.17% and F1-score 80.39%. If using traditional three categories, all the above performance were above 80% or even above 90%. The experiment also indicated that our system is easy to use and more accurate than searching information manually.

Research limitations/implications－This study ignored the posted date of product reviews. Future research may further explore the possibly opinions in the different stages of the life cycle of a product.

Practical implications－It is very important for people to find the products meeting their needs. Traditional recommendation system might use personal data, social

* Corresponding author. Email : yanh@nccu.edu.tw
2017/09/24 received; 2018/04/16 revised; 2018/06/27 accepted

relationship, purchase data or click-through records to calculate users' similarity for recommendation. This study tries to apply sentimental orientation of product reviews on each feature crawled from internet for recommendation. The proposed framework is also extensible since an automated agent would automatically process new products or new features.

Originality/value—We proposed a new hybrid-based recommendation system by applying web opinion mining to combine content-based and collaborative filtering.

Keywords: opinion mining, sentiment analysis, internet review, recommendation system

壹、緒論

隨著網路及 Web2.0 技術的發達，有越來越多的使用者會在網路上發表其對事物的心得評價提供網友參考；而參考網友評價來進行商品的決策也漸漸變成一個重要又便利的資訊來源。在這樣的背景下，針對使用者評價的各種應用如雨後春筍般紛紛出現，而在推薦系統中考量到網路評價資訊也變成了十分熱門的課題（如 Jing et al. 2018）。然而，這類的推薦系統對於商品的網路評價通常只以「整個商品」正負評價來看，並未關注對其各個屬性構面其實有不同網路評價，而消費者在下購買決策時，其實也有不同的屬性權重偏好。再者，也忽略網路評價的動態性：除了評論是每天增加外，商品會有新舊版本系列、屬性也會有擴增。因此，如何建立一個有效、能自動化收集、自動分析並判斷網路上對商品各個屬性評價傾向的整合推薦系統的研究，有其重要性。

若能有上述「有效且自動化的收集使用者在網路上對商品的意見傾向」的系統，一方面能夠協助廠商即時掌握到網路上對自家商品各功能屬性以及競爭者商品各功能屬性評價。另方面，也能協助消費者在作購買決策時，能依其對各商品功能屬性在意程度，有效參考網路資訊、同時解決資料過載的問題。

本研究選擇了行動電話做為試作標的，我們利用機器學習方法先建立一個多階層的屬性意見類別分類器來處理行動電話各屬性的意見傾向，並將分析收集到的傾向資訊整合到推薦系統，該系統具有自行擴增資料的能力。最後，針對推薦雛型系統我們進行了一個實驗以驗證系統有用性、易用性。

本文將在第貳節探討相關文獻、第參節提出研究方法、第肆節說明雛型系統的建置方法、第伍節介紹實驗設計與實驗結果、最後提出結論與建議。

貳、文獻探討

一、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，乃收集網路上或其他來源搜尋到的對某個商品或服務的評價文本資料，針對此商品或服務屬性的意見彙總找出正負評價程度（Dave et al. 2003; Pang & Lee 2008）。意見挖掘有很廣泛的應用領域，例如，分析新聞文本可有助於瞭解競爭者的運作模式（Ye et al. 2006）；而有更多關注的是分析網路商品使用者評價的文本，如電子產品（Turney & Littman 2003）、電影（Ye et al. 2006; Chaovalit & Zhou 2005）、餐廳評價（Yan et al. 2015）。近年隨著社群網路的發達，也有以微網誌為分析對象，來挖掘使用者間聯繫是否對品牌偏

好有影響 （Goudas et al. 2015; Mostafa 2013）。分析的文章也可能為是網路上的短文字心情發言（如狀態文或塗鴨牆），以期瞭解使用者當下的情緒或連續的情緒起伏（Li & Lu 2017）。例如，楊亨利與林青峰（2017）曾以新浪微博為例，提出微網誌短句的情感指數，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、喜樂、憤怒、悲傷、厭惡或恐懼等情感的指數。

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

（一）依評價的對象分類

Kempf（1999）將商品分為情感商品（Hedonic Product）及功能商品（Functional Product）二大類，藉此我們也可將評論的文章評價的對象區分為上述二大類別。情感商品包含有形的書本、繪畫、藝術品等、或如無形的電影戲劇、音樂等。使用者經由使用情感商品可直接激發出情緒；如喜劇的作者或觀眾期望經由這個作品帶給使用後開心情緒的感覺。這類商品依其內容表現可能激發使用者多種不同的情緒，如悲傷、歡喜、興奮、驚嚇、喜愛、噁心等（楊亨利 & 林青峰 2017）。

而相對地，一般的功能商品包含有形的家電設備、汽車、手機等、或無形的交通運輸服務或金融商品等，此類商品的設計是期待使用者經由操作此類事物，得到此類事物預先設計的功能服務。這類商品通常不會直接引發使用者多種的情感。例如在一般正常的狀況下，人們應該不太會覺得「一台汽車開起來很悲傷」或是覺得「這手機真是嚇死我了」。但經過使用這類事物的功能後，使用者還是會對其本體或各屬性有正負評價的感想。Zheng、Wang 與 Gao（2018）提出了一個針對行動電話中文網路評價的屬性萃取方法，但其只針對目前評論中出現的屬性進行萃取，並未考慮未來可能會新屬性的問題。Jin 等人（2016）則利用 Amazon 的評論來做競爭分析，他們鎖定特定商品，先分析其各屬性的重要意見，再進一步針對其競爭商品的評論做類似屬性的意見分析。

（二）分析資料的層級

依照不同的分析目標及不同的使用目的，可依分析資料的粗細程度，分成片語（Phase-level）、單句（Sentence-level）及通篇（Document-level）三種意見分析。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）。

本研究鎖定的研究目標為使用者發表在網路論壇上的商品評價。由於在一篇評價文中對討論商品的各屬性該使用者可能都會有不同的意見，對搜尋網路資訊者而言，每個屬性的評論都是重要的；所以本研究的分析單位為單句。

（三）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 et al. 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，以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。

二、推薦系統

推薦系統是經由分析使用者曾經買過或有興趣的項目或是去了解使用者的社會關係（如屬於什麼社團）來去預測使用者可能的喜好商品（Goldberg et al. 1992; Resnick & Varian 1997）。推薦系統產生推薦的方式通常可分為協同過濾（Collaborative Filtering）及內容推薦（Content-based Recommendation）兩大類。

協同過濾主要是經由收集、分析大量使用者的行為、屬性及有興趣的商品後，將使用者分為數個使用者相關程度非常高的社群，作為合作推薦參考，一個新的使用者到來之後，系統就可以基於其已有社群的相似性來推薦這個使用者可能會喜歡的商品（Linden et al. 2003）。協同過濾依分析的資料不同，可分為分析使用者行為、屬性等資料的使用者導向（User-based）的方法，以及分析商品項目資料的項目導向（Item-based）二類方法。一般的協同推薦並不需要去「了解」推薦項目是什麼，這點跟內容推薦很不同，它是以使用者為主的分析，重點是要去了解使用者與使用者之間的關係，找出與其最相近的群組。

協同過濾因為只考慮使用者間的相似性，並不考慮商品的內容，所以較常能找到令人出乎意料但有效的跨類別推薦名單；另外協同過濾因為不考慮商品的內容，所以也免去了對於內容的理解，不需要苦心設想應該如何表示商品的內容特

徵。但傳統協同過濾在實務上有三個較大的問題（Cho & Kim 2004），分別是(1)稀疏性（Sparsity）問題：通常會發生有很多的項目是沒有評分的，導致可分析的資料太過少造成稀疏矩陣。(2)冷啟始（Cold-start）問題：不容易一開始就會擁有使用者資料，而沒有可分析的資料，也就無法分析出使用者群組。最後是(3)系統的擴展性（Scalability）問題：當使用者數量成長時，計算量會大量成長可能會造成系統再也無法快速的回應。

內容導向推薦系統則不從使用者之間相似性來考慮推薦，而是對商品的內容特徵先進行分析及分類，在掌握了商品的內容特徵分類關係後，內容導向的推薦系統對使用者會推薦與他們之前喜歡物品在內容上相似的其它商品（Wright 2002）。若「商品內容的相似性」能構成使用者喜好的基礎，所產生出的推薦就會有不錯的效果。但因為並不是所有商品的內容特徵，都可以用一組特徵表示的方式來表達，倘若所處理的項目，其內容不易表示時，也就不容易使用這種推薦方法。此外內容導向的推薦方式，也需要系統設計者對於內容的資料特性，甚至是所應用的領域特性有更深入的了解，才能設計出夠好的特徵表示方式。

本研究的推薦系統是尊重使用者對商品各屬性的重視程度，並使用意見分析技術針對商品內容的評價找出評論者對該商品各屬性的意見值，以找出使用者可能會喜歡的商品。本質上是屬於內容導向推薦系統，但參考了其它使用者的意見，也是協同概念；應算是混合式推薦系統。同時，其它使用者的意見來自於網路，所以，上述傳統協同過濾會產生的問題，在我們的研究裡會大幅減輕。

三、利用意見分析的推薦系統

Zhang 等人（2007）利用使用者的意見建立一個電影推薦系統。Poirier 等人（2010）也分析電影部落格文章，先建立意見分類器來完成使用者－商品評價矩陣，再利用協同推薦技術來進行商品的推薦。同樣地，國內的陳仕堯（2017）提出了一個利用分析網路電影資料庫 IMDB 使用者評價進行電影的推薦系統； 孫育辰（2017）也利用 Tripadvisor.com 的遊客評論來作旅館推薦。可是他們對商品的傾向評分都只有針對總評分數並沒有深入討論到在決策過程很重要的屬性分類，也沒有考慮到意見正負的強度。

Priyanka 等人（2015）提出了一個利用意見分析個人化推薦書的系統，他們利用意見分析的方法去分析書的心得評論，判斷出針對一本書的評論所談到的屬性正負意見值。然後依據有談到「這本書」這個屬性的意見值數量來比較產出每一本書每個屬性的正負傾向值，當正意見大於負意見，那這個屬性就是正的，反之亦反；而當正負意見數量相當時，這個屬性就認為是中立的。取得每一本書的每個屬性的正負傾向之後，再配合從使用者資料找出的喜愛偏好後，就可以找到

合適的推薦書目給使用者。但是，Priyanka 等人（2015）的研究中，並沒有考慮意見強度的問題，也忽略了經常可能會出現的一篇評論談到二種以上商品比較的問題，且不能如本研究可處理新商品、新規格的延展性。而 Jing 等人（2018）則試圖證明利用網路評價的推薦系統績效會比傳統的推薦系統好。不過，他們鎖定京東商城 29 種已知筆電的客戶回應，不考慮商品在評論中的別名問題；文字探勘用的是傳統的 TFIDF，未驗證正確率；在推薦時，並未真正考量到消費者屬性權重；而且只能推薦給有在京東商城評論過的客戶；系統也不具自動化擴充的能力。

參、研究流程

本研究分成四階段來進行，在第一階段中，我們針對「行動電話」這類在網路上的商品評價進行資料收集及領域知識的分析，目的是為了得到評論者會從那些「屬性」及「屬性的分類」來去評價這類的商品。每類物品都會擁有不同的屬性，像汽車擁有「顏色」、「廠牌」、「排氣量」等屬性；而行動電話就沒有「排氣量」，相反有汽車沒有的「螢幕大小」、「鏡頭」、「握感」等屬性。在這個階段中我們關心的是「評價者會去評價的屬性」，而不是所有的屬性；對意見分析來說，不會被拿出來評價的屬性之重要性是低的。取得商品會被評價的屬性後，我們會將屬性進行分群，得到「屬性的分類」；如果屬性的分類還是過多，就需進一步簡化。

第二階段則是多階層屬性類別分類器的最佳化，在這個階段中我們會利用第一階段中收集及分析的成果來建立與最佳化各屬性類別層及商品總評的意見分類器。本研究採用多階層屬性類別分類器的設計，而非採用一般作法將各屬性直接輸入訓練的作法；這主要有二個原因，一是這樣的設計是為了模擬人類在多屬性類別商品的比較時會將「屬性細節」的優劣歸類成「屬性類別」優劣的現象；第二是這樣的設計我們可以得到各類別屬性的意見評價分數；這些分數對於想要了解總評產生的原因及對系統的展示及使用上是比較有利的。

第三階段是評價知識的自動化擴充與使用，其中包含一個比較網站雛型界面的建立、以及建立能每日處理舊商品新評論、沒有新屬性的新商品評論以及有新屬性的新商品評論能力的自動化系統。本研究中最後階段，我們設計了一個有用性/易用性實驗，提供使用者進行試用，並評估本系統是否對幫助決策是有用的。

肆、系統的建置

本研究之整體系統架構圖如圖 1，整個系統可以分成「資料前處理系統」與較為核心的「意見分析整合推薦系統」。資料前處理系統主要負責網路意見文章

的取得、清洗、標記與資料格式化等功能。意見分析整合推薦系統則是負責分析前處理系統已格式化的資料，訓練出商品各屬性分類與總評的分類器，訓練好分類器之後，最後則是整合商品意見的推薦功能與易用性實驗。

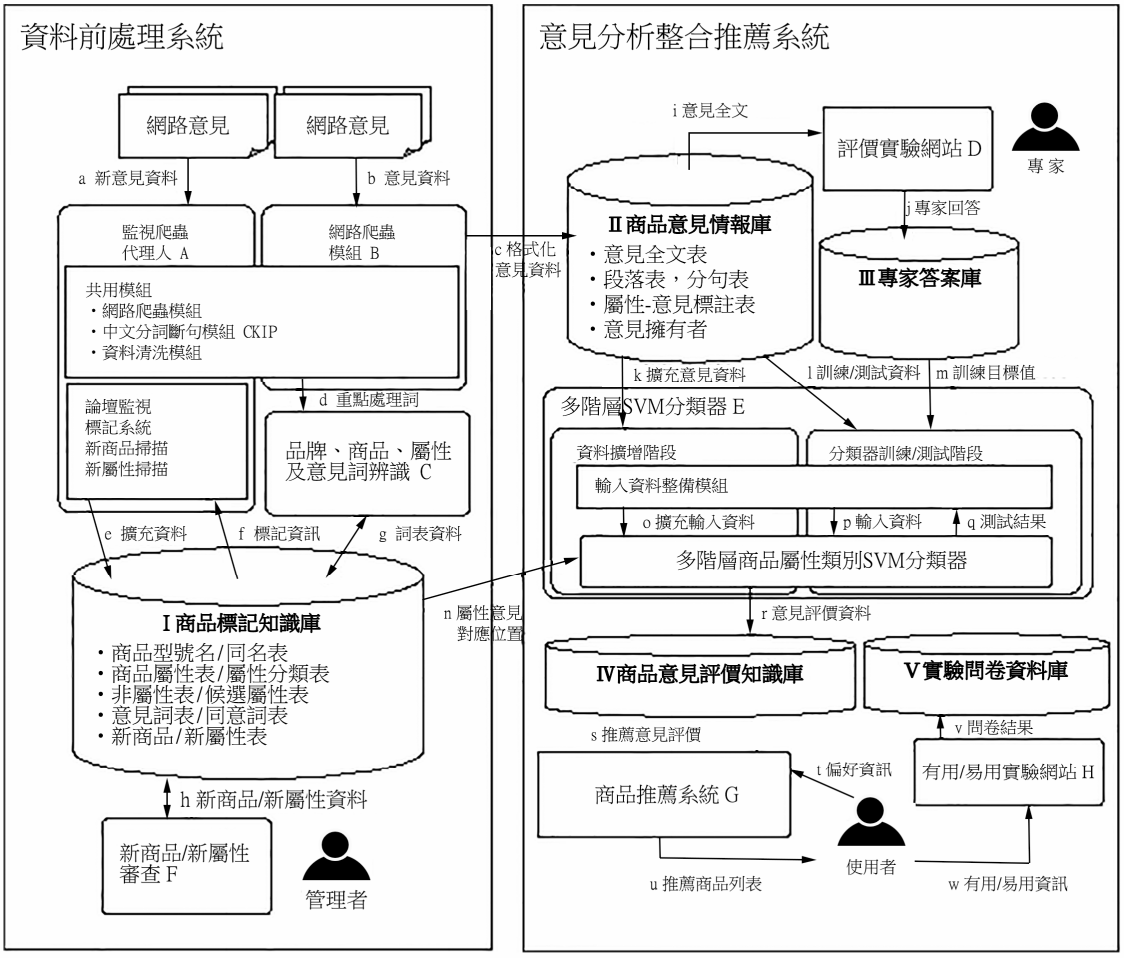

圖 1：本研究的系統架構圖

一、資料前處理系統：目標商品評論的領域知識分析

（一）原始資料的收集及分詞斷句

我們利用自行開發的網路爬蟲程式（圖 1 之模組 A, B），從台灣 PTT 電子佈告欄系統，針對作者會分享「行動電話」心得的「MobileComm」、「Android」、「iPhone」與「WindowsPhone」等四個討論版，收集從 2010 年 1 月 1 日至 2014 年 12 月 31 日止在版上所發表的行動電話心得分享相關評價文章。得到共 3825

篇的原始評論的資料，再將每篇文章資料依「文章標題」、「文章內容」、「作者」、「發表時間」等資料欄位，將其存放在商品意見情報資料庫（圖 1 之資料庫 II）的「意見全文」資料表中。原文資料的收集上我們採用不修改即儲存的策略，這樣雖然比直接清洗資料來的浪費空間，但在遇到問題資料或是因分析項目改變還要再調用原始資料時，此策略可以避免對相同文章的重覆爬蟲工作。在完成原始文章的儲存後，我們會針對標題及文章內容二欄位進行中文分詞斷句的作業，我們利用中央研究院的分詞斷句系統 SINICA CKIP 將內文完成分詞斷句工作。在我們的研究中，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。另外，為了盡量避免 SINICA CKIP 系統因為網路用語或特殊的書寫格式而發生重大的標記錯誤，研究者對於分詞斷句結果是有做再快速人工確認的動作。當發現明顯的顏文字（即表情符號）或是特殊書寫格式而發生的錯誤時，我們會用人工修正輸入文句的工作。在本研究中發生這種現象的句子，在 72972 句中共有 442 句，在調整格式後再送 CKIP 進行分詞斷句皆可順利完成分詞斷句。

接下來，我們會將完成標記的詞、句，依文章的段落、句出現的先後順序結構分別儲存在資料庫的單詞及單句資料表中，使得每一篇文章中的每一句評論都有個別的單句代碼，單句資料表還記錄了該句是出現在那篇評論文章裡以及該句出現在文章中的第幾段落、出現在該段的第幾句等位置相關的資料。類似的，單詞資料表也記錄了獨一無二的單詞代碼，及其出現在那一句裡、是該句的第幾個出現的單詞及該詞是什麼詞性。如此的資料庫設計對任一單詞都可以快又容易的找到靠近另一詞的相對位置；這除了可以方便研究者找到屬性及其情感搭配詞外，知道每句話在文章中的位置，對研究者想要找到全文評價上會有直接的幫助。這樣的設計對後續處理作業上是非常重要的。

（二）品牌、商品、屬性及意見詞的辨識

本研究為了增加未來系統的可擴增性，不採用一般從已知型錄屬性的方法來去建立商品品牌、商品、屬性等資料。若直接從網路上取得相關公司的行動電話商品型號列表，雖有助於商品品牌與型號的辨識，但是卻無法處理網路論壇中使用者對同一商品常常有大量的別稱或是非正規用語的問題：如有無空白的別稱，如：iphone 5（有空白的）、iphone5（無空白的）；簡寫的別稱，如：i5s、ip5s 或 5s（皆代表同一種商品 iphone5s）；不同語言的別稱，如：butterfly 與蝴蝶機（二者皆為官方名稱）；發音或逐字翻譯的別稱，如：唉鳳 5（iphone5）、新一（New One）。為了因應如上幾類大量別稱的狀況，我們採用可能型號字詞辨識功能及例外規則二個方法來建立別稱與商品公司、型號的對應表。以手機商品為例，可能是型號的字詞通常是英數符號混合字，如：iphone5, ip5, m2 等。所以我們可以針

對 CKIP 標註英文的字詞以及出現頻率，進行可能型號的辨識與取得。之後我們會針對各個不同的別稱建立如「5s 對應到蘋果公司所出產的 iphone 5s」的對應表；也會配合利用官方型號列表及人工加入例外規則的方式來完成不容易自動辨識的中文、發音及逐字翻譯的別稱（圖 1 之模組 C），完成如「新一對應到 htc 公司的 New One」的對應表。經過型號分析及人工對應標註之後，我們共得到由 18 間公司（Acer、Apple、Asus、Google、hTC、Huawei、InFocus、LG、Moto、Nokia、OPPO、Philips、Samsung、Sharp、Sony、TWM、小米及長江）所出產之共 138 種對應的手機型號及 172 種可能的字串與產品型號的對應表。完成商品及品牌的對應表之後，我們將此對應表中所列的公司列為代號為 M 的詞性標註；將商品中所列的商品列為代號為 P 的詞性標註，並針對已完成分詞斷句的單詞資料表進行單詞標註的調整及替換。在詞性替換的同時，我們也維護一個每篇文章曾提到的品牌與商品的資料表。並將以上各資料，儲存在「商品標記知識庫」（圖 1 之資料庫I）之對應資料表中。經過此調整作業後，可很快找出某篇文章中所有提及的商品，同時也可以修正商品品名常被 CKIP 歸類成英文或其它不正確標記的錯誤。

這個階段中我們另一個要做的重要工作是找出所有被評價的屬性詞及其意見詞列表。通常一般的屬性詞是以名詞出現在文中，為了找出人們最常用來評價行動電話商品所會重視的屬性，我們將 3825 篇文章中（共 72972 單句、722948 個被標記的單詞），挑選出被標記為名詞的單詞，依出現數量排序進行列表。在排除掉在中文中常用但沒意義的名詞如「我」、「你」、「地方」等及不常出現的字詞（出現文章數 20 次以下）後，我們得到了 177 個第一階段的屬性詞表。

在得到第一階段的屬性詞表之後，我們把所有出現屬性詞的句子先挑選出來，進一步的對這些含有商品屬性的句子（2618 句）裡，所有的副詞、程度詞、否定詞及形容詞等可能的意見詞及其副詞的位置標記出來，並計算其與屬性詞相距的字詞距離。在本研究中，我們認為單句內屬性詞前後三個字詞以內的意見詞才有可能是與其直接相關的[1]。而意見詞相關的副詞、程度詞、否定詞則通常會出現在意見詞的前方。利用這樣的規則，再經過人工過濾掉不是屬性的詞，我們可以找出共 721 個第一階段的意見詞表。在整個找尋屬性及其情感搭配詞的過程中，我們會一直維護「非屬性詞」及「候選屬性詞」的資料表，對於沒意義的中文名詞或是人工過濾的非屬性名詞，我們利用非屬性詞表來去排除，而因為次數不足的規則（出現文章數 20 次以下）我們則會將該詞及目前出現的評論文章數記錄在「候選屬性詞」表中，這樣的設計對未來進一步維護屬性詞表及後續的新

1 我們會選擇 3 這個距離，是因為發現在資料庫中的分布，中文直接有關的屬性詞與意見詞距離最常見就是±3，已佔 59.09%（1547 句），而其他分布：±5，佔 62.03%（1624 句）；±8，佔 77.92%（2040 句）；±10，佔 86.25%（2258 句）。為避免太大的錯誤，我們較為保守採用 3。後續研究者若願意忍受較多的錯誤，也可採用較大的距離。

屬性自動化處理會有幫助。

得到第一階段的意見詞表後，參考 Liu 等人（2013）的作法，我們認為反過來可以進一步從包含意見詞的句子裡，找出一些可能被自動過濾但是可能的屬性詞。利用這個方法及人工比對，我們在第二階段額外得到 63 個可能的屬性詞。得到第二階段的屬性詞後，也可以再進一步的找尋第二階段的意見詞表。總計二個階段共得到 240 個屬性詞、981 個意見詞表。在經過另一位專家審查後，我們認為在第二階段得到的屬性表及意見詞表已堪用。以後類似研究若覺得不足，也可以用同樣的方法進一步找尋下一階段的屬性詞及意見詞。

（三）決定屬性類別

本研究嘗試過陳宗義等人（2015）的方法，採用 K-Mean 的分群演算法，企圖將「有相同意見詞的屬性詞」歸類成同一類。首先我們針對每個屬性字建立一個有搭配過的意見詞矩陣表（有搭配為 1，沒搭配過為 0）。計算二個屬性字距離的方法為此二屬性字搭配意見詞組的餘弦相似度（Cosine Similarity）。在陳宗義等人的研究中，屬性「食材」可以利用「香味」、「食感」等食材的意見詞，進行完全的屬性自動歸類。可是，我們發現這樣的結論在行動電話商品的屬性中，是無法自動分類的。我們猜測會有如此差異的可能原因有二，其一是因為「行動電話商品」這類功能商品本身特性，這種功能商品中對於不應該在同屬性分類下的二個屬性，也很有可能會搭配很類似的意見詞組。例如二個屬性字[外殼]（手機外觀）與[盒子]（外包裝）完全不同；但與[外殼]常搭配的意見詞有{大、小、漂亮、有質感}，而與[盒子]常搭配的意見詞也有類似的{大、小、漂亮}。因為外殼與盒子有如此相近的意見詞，進行自動分類會將此二種屬性分在同一類。另一個可能的原因則是綜合意見詞的使用數量，在本研究中，使用者常用可以搭配幾乎所有屬性詞的意見詞，如「好」、「差」、「不錯」等；這類意見詞如果頻繁出現的話，也會有很大的機率干擾自動分群的結果。這可供後緒研究者做為參考。

在本研究中，雖然利用分類法不能完全自動歸類屬性，但有些屬性的自動分類結果還是可參考的，所以我們採用先做分群法後再利用人工去調整屬性分類的方法，以期對每一個屬性，都可以找到其所屬的唯一主分類及唯一次分類。在完成主分類（4 類）及次分類（14 類）的初步分類後，為了進一步要確定這些分類是否為目標客群在選擇手機時所重視的，我們利用台灣共 54 名常使用該論壇的大專學生進行調查。首先我們將所有的初步次分類及其下的重要屬性列表出來，要求學生將他們認為會是他們選擇行動電話的重要次分類打勾。經過調查整理，最後在徵詢領域專家後，本研究將專注在表 1 所列八種主、次要屬性分類（服務、軟體、外觀、效能、相機、畫面、其它硬體[2]、價格）和總體評價的評論。

2 「其它硬體」這一類的產生是由前導實驗中較少被學生選到的原「硬體」次分類的集合，會建立此分類是為了屬性分類的周延性原因。

表 1：屬性字標註的主分類與其次分類

主分類	次分類
服務	服務類：如客服、維修、保固、服務等
軟體	軟體類：如相容性、功能性、操作性、作業系統、市集等
硬體	外觀類：如外觀、外殼等 效能類：如效能、作業能力等 相機類：如相機、像素、光圈等 畫面類：如畫面、螢幕等 其它類：如包裝、周邊、按鍵、訊號網路、散熱力、儲存容量、聲音、擴充插槽、續航力、手感等
價格	價格類：如定價、報價、價錢等
總評	總評類：如整體、品質、定位等

而對於意見字詞，為了有效的降低屬性－意見字詞的組合數，我們的系統也會將意見相似的意見字詞聚成一類；表 2 列出了關於「靈敏」的意見字詞，在這些意見字詞找尋出來時，我們同時也標註了該意見的「程度」與是否為「否定」。目前我們決定意見字詞分類的方式是利用系統排序意見字詞後，再用人工進行兩類別的聚合。利用如表 2 之「分類」、「程度」與「否定」三個欄位，我們可以將表列的這五個意見字詞，分成有相同「等級」和「極向」的三類意見詞（「靈敏」、「太靈敏」、「不靈敏」）。這樣的方法可以降低意見字詞的數量又不影響其要表達的內涵，經過這樣的處理之後，意見字詞從原本的 981 個，降低到 642 個分類。

表 2：意見字詞與意見詞分類

意見 ID	意見字詞	出現句數	程度詞	否定詞	分類	意見詞分類編號
336	不靈活	7		不	靈敏	301
102	不靈敏	21		不	靈敏	301
466	太靈敏	5	太、很		靈敏	302
572	很靈敏	4	太、很		靈敏	302
119	靈敏	18			靈敏	303

使用上述這二個類別概念，理論上的「屬性與情感詞對組合」最大可能組合數會從 235440（240 屬性*981 意見）種屬性意見詞對組合數目下降到只有 5136

（8 屬性類*642 意見類）種詞對類別；而在本研究中，我們實際掃描到的文章資料可以得出組合數目，則由原本 12704 種屬性意見詞對組合，下降到了 2783 種詞對類別組合，這對增加訓練系統的效能很有幫助。

接下來，我們需要判別「意見擁有者」。對於商品評論文本中任一組「屬性意見詞對組合」，都需要判別此屬性意見指的是那一個商品。在全文（含標題）只有提到一個商品的狀況下，其下的所有「屬性意見詞對組合」當然就只能針對該項商品討論。但是當全文（含標題）有提到一個以上的商品時，我們決定某屬性意見應該在談那個商品的作法是採取由屬性意見所屬的該句位置開始向前找尋的第一個商品名稱來判定為該意見的擁有者。如該句話完全沒有商品名稱，且前文及文章標題也無商品名稱時，該句的意見則為無商品名稱的意見。無擁有者的意見，在本系統中目前無法被歸類，將被捨棄。

二、多階層的屬性類別分類器的最佳化

我們雖然已經從大量資料中找出常被人談論的商品屬性及搭配的意見詞，但我們還需要擁有讓機器能夠判斷什麼樣的評論會導向正面或負面評價的能力。本研究將先針對評論文章各屬性類別及總評的可能評價值進行實驗，在取得各篇屬性類別的評價值之後，我們會進行多階層分類器的最佳化訓練。

（一）專家對評論文章的評價實驗

我們要對每一篇要加入訓練的評論文章，我們得先取得讀者對於這篇評論中所提到的屬性類別及總評的建議的評價值；也就是說，對於每一篇文章，我們要取得一般讀者對其服務、軟體、價格、外觀、效能、相機、畫面、其它硬體，加上總合評價，一共九個的評價傾向的分數。

因為資源限制的問題，我們採用人工挑選出共 387 篇較適合訓練的中長篇評論文章來進行專家傾向值的收集。其中文章全文只有提及一種功能商品的評論有 232 篇，另外同篇提及二種以上商品的評論則有 155 篇。為此我們建立了一個評分網站，對於全文只有提及一種功能商品文章的評分，我們邀請了台灣北中南四所大學共 328 位的大專學生進行評分，得到 2343 份評論的傾向分數，經過清洗後，共得到 1821 份可用評論（平均一篇文章有 7.84 個有效回答）。對於提及二種以上商品的較複雜文章，則是由 5 位已有熟練答題經驗的大專學生擔任全部 155 篇文章的回答，此部分我們一共得到 1125 份評論的傾向分數：雖然此類文章內文中曾提及二種以上的商品，但經過 5 位專家個別評分及討論會議討論的結過，其中 111 篇文章還是認定為單一商品的評論；真正的多商品評論，在一篇文章內提到兩商品有 42 篇、提到三樣商品的有 6 篇、四樣商品則只有 3 篇。

在清洗策略上，單一商品文章和二種以上商品的文章我們採用不同的策略。

單一商品文章，每一位連線到本問卷系統的大專學生可以選擇回答一篇文章以上的傾向問題。在他閱讀完文章之後，他可針對這篇文章提到的商品，回答服務、軟體、價格、外觀、效能、相機、畫面、其它硬體與總合評價九題的的傾向分數（採單選[未提及（null）、很差(-2)、差(-1)、中立(0)、好(1)、很好(2)]）。此自行撰寫的問卷系統擁有自動平均答題數、避免回答重覆回答及回答時間記錄等功能。在單一商品文章中，我們認為對正常人來說 20 秒鐘根本不夠看完一篇此類的評論文章（平均每篇文章字數為 1054），所以，當一題的總回答時間不足 20 秒鐘時，我們的清洗策略是認為該回答是應該被捨棄的。如果一位答題者答題總數的 1/2 以上都被認為有問題被捨棄的話，我們採用較為嚴格的認定，該答題者的所有答題我們都認為是不可靠的，應該被全部捨棄。另一方面，二種以上商品的文章因為較為複雜，我們採用已有答題經驗，也願意作滿所有問題的 5 位訓練過的專家讀者來收取意見資料。對於專家讀者答案不一致的狀況，我們採取的原則如下：如果 5 位答題者的平均分數加減一倍標準差的區間範圍中，沒有 4 位以上的答題者的話；或是未提及（null）的人數及有評分的人數相差不足 3 的時候，該篇文章就得列入委員會進一步當面討論出一致的結果；委員會實際共討論了 9 篇的文章內容。對於單一商品或二種以上商品的文章專家傾向值都採平均計算。

（二）意見分析整合推薦系統

在完成資料收集之後，我們共有 387 篇文章產生出的可供使用的 457 個評論資料集。這 457 筆評論資料都已完成擁有者判定、屬性情感摘要、並也從專家方取得了九個分類題目的傾向值。

本研究採用分層訓練的方法來訓練支援向量機（Support Vector Machine; SVM）分類器；SVM 分類器架構圖如圖 2。本研究並不像一般傳統情感分析的作法去評斷每一單句句子的正負傾向，而是從推薦商品的角度出發，以一篇評論為單位，找到討論商品的各個屬性意見傾向。由於 Yang and Chao（2015）的研究中，已證明搭配詞（Collocation）作法績效優於事先定義情感辭典法。所以，在計算正負傾向的演算法上，我們是採用搭配詞作法。利用整篇文章中所有提到的搭配屬性意見詞當成輸入值來訓練 SVM 分類器，得到訓練後的 SVM 分類器之後，可以用來決定未知評論的各屬性類別意見傾向。比較特別的是，我們利用各屬性類別分類器的輸出與「總評」相關的屬性意見來決定這篇文章總評的意見傾向分數。

在訓練階段中當此篇心得有提到某個屬性－意見詞對的時候，該筆訓練輸入資料的屬性意見對應位置的值即設為 1；沒有提及這個屬性意見詞對即設為 0。以服務類的屬性類別分類器來舉例，其中共包含如：[服務－好]、[維修－不爽]、[官網－資料少]、[客服－好]等 78 個屬性意見詞對的位置，當有某篇心得提到

「服務－好」時，其輸入資料的對應位置即為 1。我們利用 10 折交叉驗證（10 Fold Cross Validation）的方法來進行正確率的測試（Kohavi 1995）。首先我們將所有 457 筆資料亂數平分成 10 個資料集，每次選 1 個資料集出來當作訓練分類器的資料，其它的 9 個資料集當成測試資料，來進行除了總評之外的八個 SVM 分類器的訓練與測試。而如圖 2 最下方所示，總評分類器的輸入資料除了有自己對應的屬性意見詞對之外，還包含其它八個類別屬性的輸出值。依此方法將所有文章一一完成輸入資料的整備後，再加上專家所提供的各文章各類別評分當成訓練答案，即完成輸入資料的準備。接著我們會先訓練除了總評之外的八個屬性類別分類器，並記錄每篇文章各屬性類別分類器的訓練輸出資料。再來去進行總評的訓練。最後完成這九個分類器的訓練。

在測試階段中，我們也依上述的方法先整備好輸入的資料，再一一輸入到各分類器中，再來比對分類器分類出的結果與專家評價的結果是否正確。如果整體的正確率還不滿意，則可以修正屬性意見列表後，再重新開始訓練；如果測試的正確率已可接受，則進行後資料增強階段。

在我們的設計中，後續資料增強階段可以被認為是新資料的導入。當監視爬蟲代理人發現論壇有一篇新的關於某手機的文章，系統會利用屬性、意見列表先去分析文中所有有提到屬性意見的詞對，例如評論內提到：「服務－很好」、「電池容量－差」、「維修－快」、「外觀－美」等，再加上已建立好的屬性與屬性類別對應庫，就可以把有提到的屬性意見詞對標在各對應位置成為輸入資料。把這個輸入資料丟到各類的分類器，就可以得到各屬性類別分類結果。再將此類別分類結果及總評和關的屬性意見輸入總評分類器中，就可得到總評的結果。這樣系統就能得到這篇作者對於某手機的意見傾向資料，系統會將其儲存起來，做為後續推薦系統使用。

以一個實例來說，一篇名為「LG-G2 入手心得」內含 10 句：「(1)本來在等其它手機，等到很不爽，……(5)螢幕很正點，窄邊框很美。(6)相機我覺得就普普通通。(7)家人下個月想換手機 這會讓我真的考慮勸他換 G2……(9)電池真的蠻省電…..」10 句話中，其實只有三句(5)(6)(9)有內含真正的商品屬性意見，經過我們以 CKIP 處理後，我們會得到屬性意見詞對：「螢幕（很,正點）」「窄邊框（很,美）」、「相機（普普通通）」、「電池（蠻,省電）」，並標註其程度詞或否定詞。整篇分析完之後，即可轉換成格式化的屬性意見陣列，當作 SVM 的輸入值。

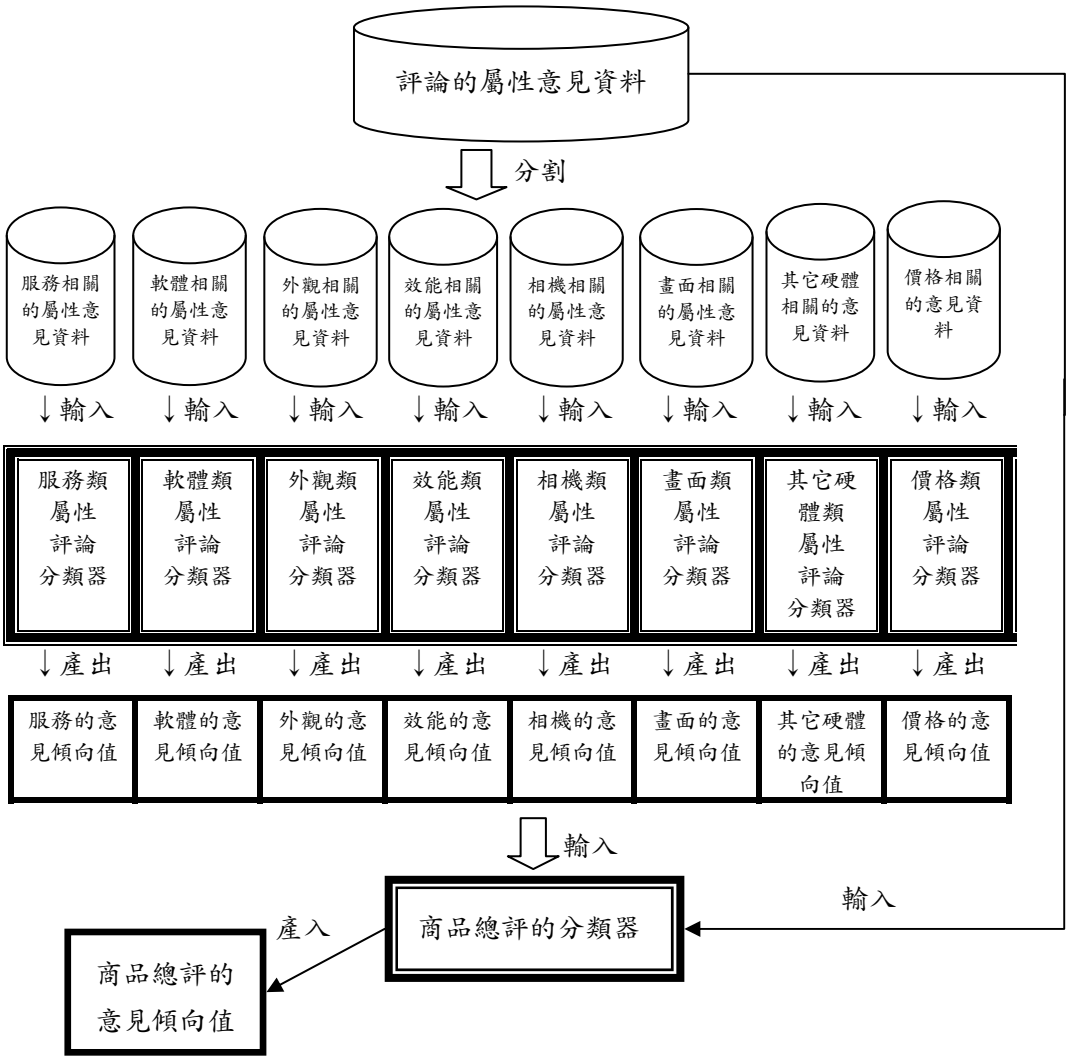

圖 2：SVM 分類器架構圖

在實際操作時 SVM 訓練參數，我們採用 cost 值為 2048，gamma 值為 0.00048828125 進行訓練[3]。由於是屬於多分類問題，在分類績效報導上我們參考 Sokolova 與 Lapalme（2009）的建議，分別提供精準率（Precision）、召回率（Recall）、F1 值的巨觀平均（Macro-Averaging）與微觀平均（Micro-Averaging）兩種數值。

傳統分類績效只分為「正面、中立、負面」三類，而我們則可有五類「非常正面、正面、中立、負面、非常負面」。由表 3 與表 4 可見，我們的 10 折交叉驗

3 依資料集的不同，需先計算出適合的參數再進行訓練（Chang & Lin 2011），本研究是利用 libsvm 中所提供的工具 grip.py 程式算出針對這個實驗資料的最佳 cost 及 gamma 參數。

證分類器的平均績效結果，不管是訓練階段或測試階段，在五類都已不錯；若依傳統三類，更是升至接近或超過 90%以上。從文獻來看，目前意見分析研究情感分類績效在 60%~80%間已是可以接受，如 Liu 與 Zheng 整理不同電影類別的分類器的分類績效是落在 66%~84%之間（Liu & Zheng 2012）；Zhang 等人在 2009 年提出不同計算方式與不同類別商品的意見分析分類績效亦是落在 64%~83.8%之間（Zhang et al. 2009），因此，本研究的分類績效應已達令人滿意水準。

表 3：本實驗總評分類器的績效整理表（五類別）

	訓練階段			測試階段		
	Precision	Recall	F1	Precision	Recall	F1
正面分類	95.18%	87.03%	90.92%	88.88%	85.76%	87.29%
負面分類	90.39%	91.88%	91.13%	91.74%	83.78%	87.58%

表 4：本實驗總評分類器的績效整理表（三類別）

		訓練階段			測試階段		
		Precision	Recall	F1	Precision	Recall	F1
正面分類	巨觀	82.09%	82.05%	82.07%	80.85%	79.81%	80.32%
	微觀	89.04%	89.04%	89.04%	85.15%	85.15%	85.15%
負面分類	巨觀	74.07%	74.98%	74.52%	81.66%	79.17%	80.39%
	微觀	83.59%	83.59%	83.59%	84.49%	84.49%	84.49%
整體	巨觀	64.59%	68.61%	66.54%	72.57%	76.50%	74.48%
	微觀	75.25%	75.25%	75.25%	74.42%	74.42%	74.42%

三、評論知識的自動化擴充與使用

（一）商品評價的擴充：監視爬蟲代理人

在完成分類器的訓練之後，本研究還設計了後續資料擴充機制。也就是說，我們希望這個架構能處理以下狀況：(1)舊商品關於舊屬性的新評論（如對 iPhone7 關於螢幕的新意見，螢幕是系統原已知的屬性）、(2)舊商品關於新屬性的新評論（如對 iPhone7 關於新功能指紋辨示的意見，此屬性是系統原不知）、(3)新商品舊屬性的新評論（如對 iPhone8 關於螢幕的意見，螢幕是系統原已知的屬性，但商品 iPhone8 是新的）、與(4)新商品新屬性的評論（如對 iPhone8 關於無線充電功能的意見）。其中(1)舊商品舊屬性在前一小節中已有說明。但要有能處理

(2)(3)(4)這三種狀況，我們首先得有判斷一個未知的文字字詞是否可能為新商品或新屬性的能力。

當在同一論壇中有一評論出現時，本研究用到進一步所開發的「監視爬蟲代理人」(圖 1 中的模組 A)，以自動化的進行商品評價的彙整。這代理人會首先將新增加的評論文章進行分詞斷句及分段、編碼及重要關鍵詞標註等前處理。商品名稱因為新商品的推陳出新，相較於屬性字與意見字來說，是比較容易出現新資料的。我們利用可能型號辨識功能來處理可能的新商品名稱，如前文所述，此功能的開發原則是由之前的商品名稱整理出比對規則，以行動電話為例，首先是去分析目前我們找出來的型號字串的長度，我們發現目前市面上行動電話的型號長度，通常是 2 到 9 個字元的長度；過短或是超過 10 個字元的過長英數混合字串，是較少被拿來當型號的。所以我們的可能型號辨識功能會配合上述的商品可能命名規則，並透過網路查詢疑似型號的字串來自動確定可能的新商品名稱，如果一但確定此字串可能為新商品，代理人會先在標記知識庫的新商品名稱表中新增資料。但為了其它種的別名及系統完整性，我們的設計目前有提供新商品審查的功能（模組 F)，管理者可以依情況選擇新商品需審查或是不審查自動新增。通過審查或是自動新增新商品名稱會在新增在商品名稱表中。如此一來，上述(3)新商品舊屬性的情況也可以解決。

接下來要解決(2)(4)新屬性的問題，可以分成二階段，首先我們得先解決「這個字詞是不是標記資料庫裡沒有的新屬性？」，之後我們還要解決「這個新屬性應該要歸到那一屬性類別？」我們認為一個不在屬性詞庫裡的新屬性必須滿足三個條件才有可能是新屬性，第一是它必然是有意見詞的名詞；第二是該名詞在討論商品的官方相關文件（促銷、廣告、商品情報）必須要同時出現；第三是全部文章討論該屬性的次數必須大於一個門檻值（如新屬性詞須出現在至少 20 篇意見文本內)。當系統在掃描期間發現有新詞同時滿足第一、第二條件時，系統會將該詞加到候選屬性詞表中，並在每次掃描到該屬性詞時更新之後出現在文本內的總次數。當總出現次數已超過門檻值後，我們得進一步去決定這個新詞的主、次屬性分類。在這裡我們假設商品屬性字之搭配的意見詞在同一個論壇中的評論是較不容易改變的，意見詞相近的二個詞大多屬於同一屬性類別。如此一來，我們可以比對這個新屬性目前所有的意見搭配詞，與目前商品屬性資料庫裡的那一個屬性詞最為接近，並找出最可能的屬性分類。當也確定好可能的屬性分類之後，與新商品品名的狀況一樣，系統會將這個新屬性詞先存在新屬性詞表中，管理者可以自由選擇是否須要審查後再新增到屬性表中。不過，如果涉及出現新屬性的情況，因為需要重新訓練新的該屬性及總評 SVM 分類器，所以管理者介入的新屬性管理方式可能還是必要的。

（二）手機推薦系統

本研究進一步應用所收集到的商品意見評價資料，設計了一個商品的推薦雛型系統。使用者經由使用這個推薦系統可以有效的依其重視的屬性類別找到綜合滿意度最高的商品。

使用者連線到我們的雛型系統之後會先導到一個說明引導頁面。使用者在點擊設定「關心的屬性類別與權重」之後，就會到達喜好收集頁面，使用者可以依照其喜好不同來尋找最適合他喜好傾向的行動電話。經由此頁，我們可以收集到使用者對商品各屬性的在意程度：本研究中我們設定「非常在乎」的權重為 4、「很在乎」3、「有點在乎」2、「沒太大關係」1，而「完全不在乎」則為權重 0。我們的雛型收集頁面目前的設計是請使用者一一針對各類別以「按鈕」式設計勾選其在意程度，當然也可以採取排序方式來排此八個屬性，或只要求使用者只排出最在意的前三名，其餘均視為相同的第四名。再者，因為目前並未與真實經營的商業網站合作，所以在我們的雛型系統裡是由使用者直接表達其個人偏好排序。未來的實務應用上，網站經營者可從多方的使用者行為與基本資料，如其歷史觀看記錄、參與的社群網站活動、評價資訊、年齡、收入等資訊，來分析取得使用者喜好程度資料。

接下來我們採用加權計分法去計算資料庫中每個商品歷史評價資訊對此使用者的加權偏好總分。每一種商品的八個屬性分類的平均歷史評價資訊可以計算並列成一個 1x8 的矩陣，商品評價分數若為 0 是非常負面、1 是負面、2 是中立或未提及、3 是正面、4 是非常正面。假設兩商品平均歷史評價資訊如下：甲商品[3.2, 2, 1.2, 0.8, 2.4, 3.1, 3.8, 1]及乙商品[1.3, 2, 2, 3.8, 3, 2, 2.6, 3.2]。而依對應的屬性類別，使用者的喜好程度也可以列成如下的 8x1 矩陣：

$$\text{userA } [4, 0, 1, 1, 2, 4, 1, 1]^T$$

$$\text{userB } [1, 2, 3, 2, 0, 0, 0, 4]^T$$

所以我們就可以將此二個矩陣相乘，得到商品評價加權分。以上數例子來說，對 userA，甲商品的加權分為 36.8 分[4]。再將這個加權分除於 userA 所有喜好程度值的總和（4+0+1+1+2+4+1+1 為 14）得到 2.629 的推薦數值。這樣的推薦數值因為符合非常負面（0）到非常正面（4）的強度數值比例，所以對研究人員的解讀與使用來說均屬容易。同樣的算法，也可以算出對 userA 來說乙商品的推薦數值為 2.200；對 userB 也可以算出其甲商品的推薦數值為 1.367、乙商品的推薦數值則為 2.642。由此可見把甲商品推薦給 userA，推薦乙商品給 userB 較為合適。我們的系統目前作法是對每一個使用者推薦數值最高的前 10 個商品給使用

4 即 3.2*4+2*0+1.2*1+0.8*1+2.4*2+3.1*4+3.8*1+1*1=36.8

者參考。圖 3 所示即為當使用者點選了某一商品（iPhone 4S）後的網路評價摘要。

圖 3：雛型系統對某特定商品推薦的客製評價摘要

除了推薦數值外，此客製評價摘要系統為了讓一般使用者也能很容易上手使用，所以會對推薦的每一個商品給一個以 100 分為滿分的推薦指數，因為推薦數值是[0,4]之間的小數，我們現在的雛型系統先採用最簡單放大 25 倍取整數的方法，來計算推薦指數。但從實作後的資料來看，推薦數值總分大於 3 以上的商品通常很少，所以換算目前換算出來的推薦指數通常 80 分以上就可算是非常推薦了。本系統也會依使用者對屬性喜好的重要程度依序去呈現摘要，使用者越在乎的屬性放在越前面、不在乎的就完全不列出。在每一類屬性下，也會列出在資料庫中該類屬性中最常被提及的屬性意見，使用者可用來做為選用商品的參考。如假設某使用者對於手機「外觀」最在意，圖 3 中對於 iPhone 4S 第一個顯示的就是「外觀」屬性，網路評價非常正面（在 0 至 4 分中，取得 3.2 分），該屬性有四篇評論提及「外形很滿意」、三篇提及「外觀很好看」、兩篇提及「外殼多樣」。如果使用者還要進一步看更仔細的資訊，我們也提供閱讀各評論連結的功能，讓使用者可以進一步去看更詳細的原始評論資料。此外，在此介面中，我們將所推

薦的 10 個商品列在介面右側的浮動視窗中，供使用者方便切換觀看不同的商品。另外，在圖 3 中，「商品名稱」與「推薦指數」間的「功能規格」按鈕，也會提供如一般 3C 網站都有的該商品基本的詳細規格資訊供使用者必要時查考。

四、系統可用性及易用性的實驗

（一）實驗設計

為了檢定本雛型系統對使用者推薦商品的可用性與易用性，我們進行了一個簡單的實驗。從資料庫選出從實驗施測的前六個月（2014 年 10 月到 2015 年 3 月）間推出的 15 款各廠牌手機，並且將其相關的屬性及評價資料，導入實驗的雛型系統裡。我們的實驗對象是大專學生；他們雖是學生，但是為手機商品更新相當快的使用者，也偏好在網路上尋找「鄉民」的意見。

我們設計了一個實驗網站來進行實驗，實驗總共時間約 60 分鐘。施測人員會根據介面帶領受測者逐項了解任務的目的及相關的注意事項。(1)首先，受測者需輸入他們的喜好資訊。他們被要求依自己個人主觀的喜好進行各屬性類別的在乎程度，例如說某人可能輸入了，外觀－非常在乎；服務－在乎；價格－不在乎等。(2)在收集完所有受測者的個人喜好資訊後，施測者會要求受測者自我針對實驗系統給定的 15 支行動電話，根據上述所自述的喜好，進行 45 分鐘的網路評價資料搜集。在搜集的時候，實驗系統會展現此 15 支行動電話商品的基本資訊，並提供受測者筆記用紙，用以實際記錄其在網路上看到的評論狀況；在這個階段，施測人員也會提醒受測者接下來會要求他們從中選出與排列 5 支手機，以期他們會做較深度的網路評價搜集。(3)等到收集資料時間結束後，實驗介面會要求受測者從表列的 15 支手機中，選出並排列他們最喜歡的 5 支手機。上述過程是欲模擬消費者在沒有本研究所提供的雛型系統時，自我在網路搜集相關商品資訊。

在受測者送出資料之後，實驗介面會依其最先所輸入的喜好資訊，利用本研究的推薦雛型系統列出我們系統對他個人的喜好推薦，並依屬性列出我們搜集到的相關網路評價資料，如前述之圖 4，供使用者比對推薦雛型系統的結果與其自我人工在網路搜集結果的差異。最後，我們請受測者填寫對雛型系統可用性及易用性的問卷，該問卷乃參考一般的科技接受模式（Technology Acceptance Model）（Davis 1989）的問項，如表 5。

（二）實驗結果

表 5 顯示此問卷的內部一致性信度（Cronbach’s α 值）是可被接受的。共有 38 名台灣大專院校對手機有使用需求及在網路上找尋評價資料經驗的學生一同參與我們的實驗，但有 1 位學生在系統上並未有使用記錄故不採計其問卷，另有 1

位同學並未完成問卷，所以我們實際有 36 名受測者的問卷資料。所有問項在經過統計檢測之後，在 α=0.05 的信心水準下，皆顯著大於李克特尺度 1-5 的中間值 3。這顯示這個系統在有用性、易用性與總評均可讓使用者滿意。

表 5：問卷信度分析表

構面	問項		個數	平均數	信度 α 值
有用性	A1	我覺得這個系統能有效的提供決策所需要的資訊	36	3.92	0.70
	A2	這系統可提供我選購手機時需要的網路評價資訊	36	3.44	
	A3	我覺得這個系統對選手機商品很有用	36	3.67	
易用性	B1	我想大部份的人可以很快就學會使用這個系統	36	3.83	0.79
	B2	這系統的操作介面對使用者很親和	36	3.5.6	
	B3	我覺得這個系統很容易使用	36	3.53	
總評	T	整體上來說我覺得這個系統是好的	36	4.03	-
構面分析			個數	平均數	信度 α 值
A	有用性		36	3.68	0.84
B	易用性		36	3.64	

伍、結論

本研究針對功能商品提出一個兼具內容推薦與協同過濾的混合式推薦系統，其各分類器的正確率、F1 皆已達令人滿意的程度，雜型系統亦有其可用性及易用性。此系統的核心是由網路評論自動化意見傾向分析所彙總而來，相對以往文獻所提出的推薦系統對於商品的網路評價只考慮「整個商品」正負評價，本研究進一步關注各個屬性構面的不同網路評價，這對消費者的購買決策可斟酌不同的屬性權重偏好以及廠商瞭解自家商品及競爭商品均有很大的幫助。例如，在分析手機的使用者評論後發現：若不分商品，評論者最關心的屬性前三名依次為螢幕、電池、相機。若細看特定商品，以最被關注的 iPhone5 來說，評論者最關心的屬性依次為：外殼、電池、電源鍵；對該商品，評價最差的屬性前三名為「價格（太貴）」、「home 鍵（不好、沒反應）」、「電源鍵（不好按）」；另方面評價最好的屬性則前三名則為「充電（快）」、「軟體（順暢）」、「拍照（漂亮、快）」；如能在即時得到這些訊息，將會對使用者選擇商品或廠商制定策略很有幫助。

再者，本研究提出了一個其它系統少見的自動化監測增添新商品、新屬性延展性的意見分析系統架構。利用本研究所開發出來的推薦系統，將可自動的監控

與分析消費者評論的傾向。在學術上，我們提出了一個多層次的 SVM 分類器架構，較特別的是其中總評分類器是需要上層分類器的輸出值做為來源。後續研究也許可以去討論這種多層次分類器架構中低層與高層分類的正確率相互影響。

本研究的技術架構將能順利的套用其它純功能商品（如金融、汽機車或家電等）的分析上，所可能的應用不只是推薦，還包含網路輿情監控、促銷方案效益評估與及時調整、分析消費者意見的長期趨勢、關注除了主流意見外的少數消費者多元觀點等。不過，類似的架構並不認為能直接套用在情感商品（如電影、閱讀或音樂等），這是因為是情感商品的評價並不見得有特定功能屬性，如「這電影讓我心跳不止」；而且情感商品會觸發的不止是意見正負傾向，而更可能觸動使用者的基本情感，如「令人感覺悲傷」；未來研究可開發情感商品的推薦系統。另一個研究限制，則是本研究的對使用者商品屬性偏好是直接問使用者，未來若能實際運用在社群或市集網站，將可直接由使用者的實際消費行為資料取得其商品屬性偏好（Jing et al. 2018）。

誌謝

本文接受科技部研究計畫（MOST 104-2410-H-004-135-MY3）補助，特此致謝。

參考文獻

孫育辰（2017），『Developing a Hybrid Collaborative Filtering-based Hotel Recommender System Based on Review Sentiment and Context Information』，未出版碩士論文，中正大學資訊管理系研究所，嘉義縣。

陳仕堯（2017），『結合意見探勘之電影推薦系統的研究』，未出版碩士論文，淡江大學資訊管理學系碩士班，台北市。

陳宗義、巫啟豪、陳裕民（2015），『馬斯洛需求理論為基之網路評論產品定位分析機制』，*電子商務學報*，第十七卷，第四期，頁 393-422。

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』，*資訊管理學報*，第二十四卷，第一期，頁 1-28。

Chaovalit, P. and Zhou, L. (2005), ‘Movie review mining: A comparison between supervised and unsupervised classification approaches,’ *Proceeding of the 38th Annual Hawaii International Conference on System Sciences* (HICSS'05), Hawaii, USA.

Chang, C.C. and Lin, C.J. (2011), ‘LIBSVM: a library for support vector machines,’ *ACM transactions on intelligent systems and technology (TIST)*, Vol. 2, No. 3, 27.

Cho, Y.H. and Kim, J.K. (2004), 'Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce,' *Expert Systems with Applications*, Vol. 26, No. 2, pp. 233-246.

Dave, K., Lawrence, S. and Pennock, D.M. (2003), 'Mining the peanut gallery: Opinion extraction and semantic classification of product reviews,' *Proceedings of the 12th International Conference on World Wide Web*, Budapest, Hungary, pp. 519-528

Davis, F.D. (1989). 'Perceived usefulness, perceived ease of use, and user acceptance of information technology,' *MIS Quarterly*, Vol. 13, No. 3, pp. 319-340.

Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992), 'Using collaborative filtering to weave an information tapestry,' *Communications of the ACM*, Vol. 35, No. 12, pp. 61-70.

Goudas, T., Louizos, C., Petasis, G. and Karkaletsis, V. (2015), 'Argument extraction from news, blogs, and the social web,' *International Journal on Artificial Intelligence Tools*, Vol. 24, No. 5, 1540024. https://doi.org/10.1142/S0218213015400242

Jin, J., Ji, P. and Gu, R. (2016), 'Identifying comparative customer requirements from product online reviews for competitor analysis,' *Engineering Applications of Artificial Intelligence*, Vol. 49, pp. 61-73.

Jing, N., Jiang, T., Du, J. and Sugumaran, V. (2018), 'Personalized recommendation based on customer preference mining and sentiment assessment from a Chinese e-commerce website', *Electronic Commerce Research*, Vol. 18, No. 1, pp. 159-179.

Kempf, D.S. (1999), 'Attitude formation from product trial: Distinct roles of cognition and affect for hedonic and functional products,' *Psychology & Marketing*, Vol. 16, No. 1, pp. 35-50.

Kohavi, R. (1995), 'A study of cross-validation and bootstrap for accuracy estimation and model selection,' *Ijcai*, Vol. 14, No. 2, pp. 1137-1145.

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, H. and Lu, W. (2017), 'Learning Latent Sentiment Scopes for Entity-Level Sentiment Analysis,' *AAAI*, 2007, pp. 3482-3489.

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.

Linden, G., Smith, B. and York, J. (2003), 'Amazon. com recommendations: Item-to-item collaborative filtering,' *IEEE Internet computing*, Vol. 7, No. 1, pp. 76-80.

Liu, H., He, J., Wang, T., Song, W. and Du, X. (2013), 'Combining User Preferences and User Opinions for Accurate Recommendation,' *Electronic Commerce Research and Applications*, Vol. 12, No. 1, pp. 14-23.

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.

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.

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

Pang, B., Lee, L. and Vaithyanathan, S. (2002), 'Thumbs up?: sentiment classification using machine learning techniques,' *Proceeding of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, Vol. 10, pp. 79-86, Pennsylvania, USA

Poirier, D., Fessant, F. and Tellier, I. (2010), 'Reducing the cold-start problem in content recommendation through opinion classification,' *Proceeding of the Web Intelligence and Intelligent Agent Technology* (WI-IAT), No. 1, pp. 204-207, IEEE. Toronto, Canada.

Priyanka, K., Tewari, A.S. and Barman, A.G. (2015, April), 'Personalized book recommendation system based on opinion mining technique,' *Proceeding of the Communication Technologies* (GCCT), pp. 285-289, IEEE. Thuckalay, India.

Resnick, P. and Varian, H.R. (1997), 'Recommender systems,' *Communications of the ACM*, Vol. 40, No. 3, pp. 56-58.

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.

Turney, P.D. and Littman, M.L. (2003), 'Measuring praise and criticism: Inference of semantic orientation from association,' *ACM Transactions on Information Systems* (TOIS), Vol. 21, No. 4, pp. 315-346.

Wright, S. (2002), 'Personalization: How a computer can know you better than yourself,' *Proceeding of the Multimedia Systems Conference*. UK: Hampshire.

Wiebe, J., Wilson, T. and Cardie, C. (2005), 'Annotating expressions of opinions and emotions in language,' *Language Resources and Evaluation*, Vol. 39, No. 2-3, pp. 165-210.

Yan, X., Wang, J. and Chau, M. (2015), 'Customer revisit intention to restaurants: Evidence from online reviews,' *Information Systems Frontiers*, Vol. 17, No. 3, pp. 645-657.

Yang, H.L. and Chao, A.F.Y. (2015), 'Sentiment analysis for Chinese reviews of movies in multi-genre-Based on morpheme-based features and collocations,' *Information Systems Frontiers*, Vol. 17, No. 6, pp. 1335-1352

Ye, Q., Shi, W. and Li, Y. (2006), 'Sentiment classification for movie reviews in Chinese by improved semantic oriented approach,' *Proceeding of the 39th Annual Hawaii International Conference on System Sciences* (HICSS'06), Hawaii, USA.

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, Y.C., Medo, M., Ren, J., Zhou, T., Li, T. and Yang, F. (2007), 'Recommendation model based on opinion diffusion,' *EPL(Europhysics Letters)*, Vol. 80, No. 6, 68003.

Zheng, L., Wang, H. and Gao, S. (2018), 'Sentimental feature selection for sentiment analysis of Chinese online reviews,' *International journal of machine learning and cybernetics*, Vol. 9. No. 1, pp.75-84.