

國立政治大學數位內容碩士學位學程

碩士學位論文

一個考慮閱聽人體驗喜好的電子新聞推薦模型

An E-news Recommendation Model Based on Consumer's

Experience and Preference



指導教授：許志堅博士

副指導教授：廖峻鋒博士

研究生：朱為丞撰

中華民國 105 年 7 月

誌謝

首先很感謝聖三位和主讓我能夠考上並在政大數位內容所當中學習，過程雖然辛苦，但讓我學習到很多寶貴的事物，如何和人相處、如何團隊合作、如何做好一份研究，如何構想等等。

很感謝父母們總是不求回報的支持著我，感謝許志堅老師一步步帶著我教導我如何做一份研究，感謝很多雖然我不太熟悉但是卻熱心幫助我的老師們、感謝很多不認識的同學幫我填寫非常長的問卷、也感謝許多過程中幫助我的朋友以及忍受我不足的朋友們。

在過程當中深刻感受到整個學術領域的浩瀚，若沒有過去許多偉大的學者的貢獻，無法讓各個研究領域如此蓬勃的發展，也體會到要完成一份論文真的不是靠一個人就可以了，需要有許多研究人員的貢獻，也需要有很多人的幫助。期望自己能夠把所學到的所得到的，在未來更多的回饋給這個世界。

2017.1 月 朱為丞

摘要

本研究嘗試建立一個考慮使用者體驗喜好之電子新聞推薦模型。我們以 Schmitt 提出之策略體驗模組為基礎了解使用者對各體驗之重視程度，分析使用者對各種不同型式體驗之重視程度以作為 ID3 決策樹機器學習演算法的輸入屬性，並以消費者對於電子新聞的喜好與否作為目標屬性，利用決策樹演算法計算這些輸入屬性(使用者對各種不同型式體驗之喜好)與目標屬性(使用者對於電子新聞的選擇)之間的關聯式規則。接著利用這些規則來建構一個預測模型，以評估閱聽人對於未知電子新聞的接受程度，從而建立一個能有效符合使用者個人體驗喜好之新聞推薦模型。

關鍵詞：體驗行銷、決策樹、新聞推薦

目次

第一章 研究動機與目的.....	8
第一節 研究背景與動機.....	8
第二節 研究目的.....	10
第二章 文獻回顧.....	14
第一節 體驗行銷.....	14
一、 體驗的定義.....	14
二、 策略體驗模組.....	15
三、 體驗媒介.....	17
第二節 資料探勘(Data Mining).....	18
一、 資料探勘的操作步驟.....	19
二、 資料探勘的技術與方法.....	19
三、 決策樹演算法(Decision Tree).....	22
四、 ID3 演算法(Iterative Dichotomiser 3).....	22
第三節 李克特量表.....	25
第四節 軟性新聞.....	26
第三章 研究設計.....	32
第一節 使用者行為資料蒐集.....	35
一、 閱聽人體驗喜好之評估.....	36
二、 設計問卷.....	37
三、 問卷測試、修正及問卷調查.....	42
第二節 機器學習.....	43
一、 訓練階段.....	43
二、 執行階段.....	50
第四章 實驗結果與分析.....	55

第一節 問卷設計及結果分析.....	55
一、 問卷設計.....	55
二、 問卷測試及修正.....	56
三、 問卷信度分析.....	57
四、 問卷效度分析.....	59
第二節 推薦系統衡量指標及驗證方法.....	60
第三節 實驗結果分析.....	62
一、 演算法之參數調整.....	63
二、 實驗結果分析.....	73
第五章 結論與未來展望.....	78



表次

表 1、現有新聞推薦演算法優缺點比較.....	11
表 2、台灣紙本新聞分類.....	27
表 3、國際紙本新聞分類.....	28
表 4、網路新聞分類.....	29
表 5、體驗模組之於新聞內容的意義.....	36
表 6、各新聞體驗之重要關聯因素.....	38
表 7、各重要關聯因素之對應問句.....	39
表 8、各新聞類別所包含之新聞內容.....	42
表 9、好感得分配分表.....	45
表 10、使用者對各重要關鍵因素之喜好表.....	46
表 11、新聞喜好得分配分表.....	47
表 12、使用者對各新聞類別之喜好表.....	48
表 13、體驗重要關鍵因素及新聞類別種類.....	52
表 14、體驗重要關鍵因素信度.....	58
表 15、各新聞類別因素信度.....	58
表 16、CRONBACH A 係數與可信度高低之對照表.....	59
表 17、各問卷之 KMO 值及球型檢定值.....	60
表 18、混淆矩陣.....	61
表 19、決策樹停止分割純度門檻之變化影響表.....	64
表 20、決策樹停止分割純度門檻對不同各類別新聞之變化影響表.....	66
表 21、權重分配係數 ALPHA 值之變化影響.....	68
表 22、權重分配係數 ALPHA 值對各類別新聞之變化影響表.....	69
表 23、推薦門檻 MIN_SUGGEST_SCORE 值之變化影響表.....	71

表 24、推薦門檻 MIN_SUGGEST_SCORE 值對各類別新聞之變化影響表72

表 25、推薦模型對不同數量測試資料之推薦表現.....74

表 26、推薦模型對不同類別新聞之推薦表現表76



圖次

圖 1、美國成年人使用各種媒體的時間比例	8
圖 2、獲得新聞的主要來源比例	9
圖 3、超過 18 歲並使用網路的人口中，會觀看網路新聞的人數	10
圖 4、決策樹架構	24
圖 5、研究示意圖	32
圖 6、問卷設計流程	33
圖 7、機器學習架構圖	35
圖 8、訓練階段架構圖	44
圖 9、各新聞類別決策樹	49
圖 10、執行階段流程圖	51
圖 11、問卷設計流程	55
圖 12、問卷修正流程	56
圖 13、決策樹停止分割純度門檻之變化影響圖	64
圖 14、決策樹停止分割純度門檻對不同各類別新聞之變化影響圖	65
圖 15、權重分配係數 ALPHA 值之變化影響圖	67
圖 16、權重分配係數 ALPHA 值對各類別新聞之變化影響圖	69
圖 17、推薦門檻 MIN_SUGGEST_SCORE 值之變化影響圖	70
圖 18、推薦門檻 MIN_SUGGEST_SCORE 值對各類別新聞之變化影響圖	72
圖 19、推薦模型對不同數量測試資料之推薦表現圖	74
圖 20、推薦模型對不同類別新聞之推薦表現圖	76

第一章 研究動機與目的

第一節 研究背景與動機

近年來，由於網路和行動通訊的興起，使得傳播產業受到了很大的衝擊和影響，新媒介的發展改變了閱聽人獲取資訊的方式。使用者除了透過傳統的媒體獲取資訊之外，越來越多使用者透過 Facebook，Twitter，Youtube 等等新媒介平台來獲取最新資訊。而傳統的媒體公司為了因應這潮流也各自展開新媒介的佈局。美國知名市場研究機構 eMarketer 針對使用者主要獲取資訊的方式進行調查(圖 1)[1]，發現在 2013 年時，美國成年人使用新媒介的時間首次超越了觀看電視的時間，並且使用行動通訊的時間比率不斷的增加，到 2014 年已經有 23.3%。使用傳統媒體如電視，廣播，印刷品的時間則逐年的下降。

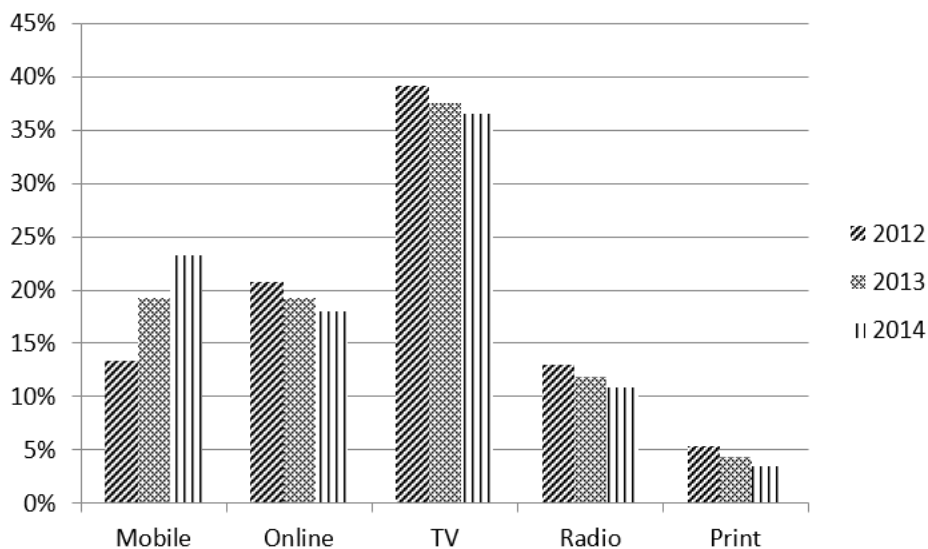


圖 1、美國成年人使用各種媒體的時間比例

在新聞領域當中，使用者比起透過紙本越來越習慣透過網路來閱讀新聞，美國皮尤研究中心(Pew Research Center)，2008 年時調查發現，使用者主要獲取新聞的來源，比例上電子新聞已經超越了紙本新聞(圖 2)[2]。而美國 comScore 網

站則針對十八歲以上會使用網路的美國人進行觀看網路新聞的狀況調查，發現在 2014 年一月到 2015 年一月，會觀看數位新聞的人數比例增加了 19%，從一億四千六百萬，成長到了一億七千三百萬人，會觀看網路新聞的佔其中 82%(圖三) [3]。而在大量的內容以及龐大數量的使用者當中，如何推薦使用者所喜歡的新聞內容，成為了重要的課題。以現有的新聞網站，常用使用的推薦方式，是按照點擊的人次做熱門新聞推薦，以及推薦和使用者所點擊的內容分類相同的新聞。以 New York Times 為例，在新聞之後會推薦和所點閱相同分類的數個新聞，以及推薦焦點新聞。而其他如 BBC，CNN 則會推薦熱門新聞，以及與觀看新聞內容中有相關的新聞，也會推薦和所點閱的內容相同分類的新聞。

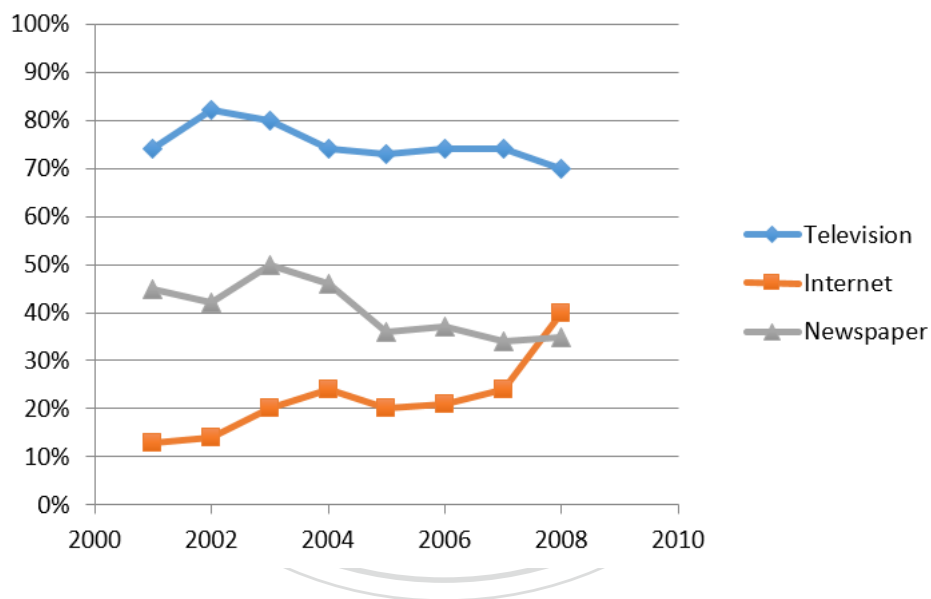


圖 2、獲得新聞的主要來源比例

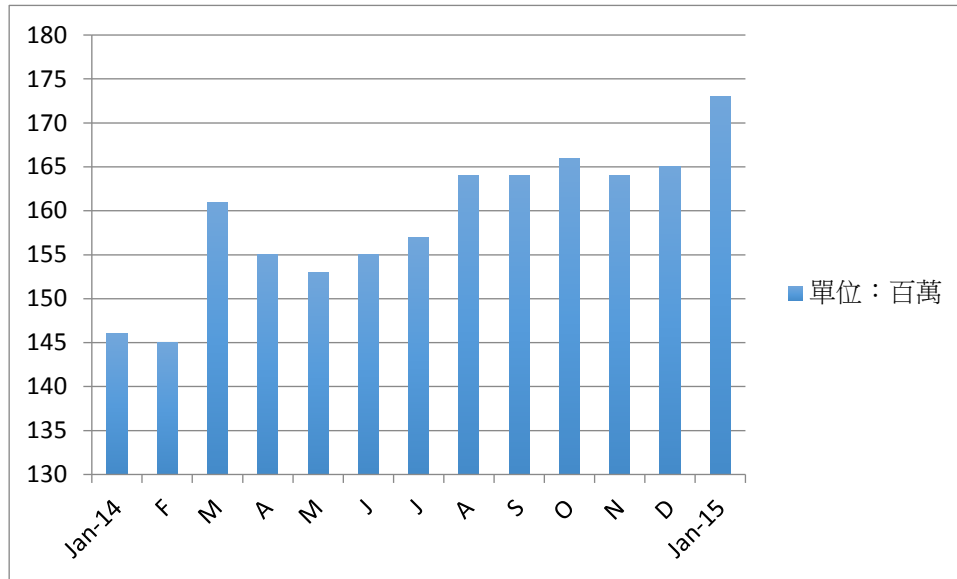


圖 3、超過 18 歲並使用網路的人口中，會觀看網路新聞的人數
(資料來源：comScore)

第二節 研究目的

如何推薦給使用者所想要觀看到的內容成為了學者們爭相研究的題目，許多學者提出使用機器學習的演算法，針對使用者進行分析並推薦新聞。其中演算法大致上可以分為兩種，一種是以內容為底進行的推薦 (content-based recommendation)，另外一種則是透過使用者群的演算法來推薦新聞內容 (collaborative filtering)，這種演算法就不針對新聞的內容分析[4]。

以內容為底進行的新聞推薦部分 (content-based recommendation)，Kompan, M.和 Bielíková, M.,提出了以餘弦相似度 (cosine-similarity) 演算法針對內容運算來推薦新聞[5]。也有學者提出 Hermes framework [6]和 Athena framework[7]，對內容進行知識本體 (Ontology) 的比對和分析之後推薦新聞。

另外，透過使用者群的演算法來推薦新聞內容（collaborative filtering）的部份，Joseph A. Kostas 等學者將 collaborative filtering 的演算法使用在 Usenet news 中[8]。Google 在 2007 年也提出了延展性高的 collaborative filtering 推薦[9]。

除了上述兩類型的演算法之外，google 在 2010 年提出另一種演算法[10]認為人想要觀看新聞，就是想要找尋一些新的事物，因此不會只提供相同類別的內容，反而會提供一些比較一般的內容，並針對使用者的點擊進行記錄以及分析，觀察在這期間當中，點擊內容的趨勢，並分析出使用者是不是有新的興趣產生，進而推薦相關的內容。

表 1、現有新聞推薦演算法優缺點比較

推薦方法	優點	缺點
依照點擊次數推薦	能夠推薦出大眾最關心的議題	無法針對使用者喜好進行推薦
content-based recommendation	可以提供讀者喜歡的領域的熱門文章	無法針對讀者個人喜歡的體驗進行分析推薦
collaborative filtering	能夠對同類型興趣的人推薦該類型的新聞	無法直接針對使用者喜好進行分析並推薦

本研究分析後發現，目前所使用的推薦演算法，最後的決策主要是依據新聞的內容類別作為決策來推薦，或者是透過擁有相同興趣使用者的記錄進行推薦，沒有辦法針對使用者本身的喜好體驗精準的進行分析並推薦。

阿里巴巴集團董事局主席兼執行長馬雲曾說：體驗很有意思，就是感受，上世紀講了很多服務，不斷地增加服務能力，其實客戶要的不是服務，而是體驗。近來對於消費者和使用者而言，消費者和使用者喜好的不只是訊息或服務的本身，而是整個服務所給人的體驗。像是買電子產品時，相較於過去產品只是擺設在產品架上，現在的商店都會擺設實機在店中供消費者體驗以及使用。經營書店時也不再只是顧慮到書種類的齊全，數量的多寡而已，如何營造出好的體驗成為了吸引人潮的關鍵。另外消費者也會願意花相當的金費去得到一個難忘的體驗，比如說去遊樂園遊玩，農場生活體驗，或是創造出一個生日的驚喜等等。

Pine and Gilmore(1998)在"體驗經濟(The Experience Economy)"[11]一書中就提到，過去從農業經濟時代、工業經濟時代、服務經濟時代，到現在進入了體驗經濟的時代，透過體驗可用來創造使用者難以忘懷的感受，而增加該服務所帶來的效益。另外 Schmitt 在 1997 年則提出了具體的體驗分類模組，可將體驗大致分成感官式(sense)、情感式(feel)、思考式(think)、行動式(act)以及關聯(related)式體驗[12]。

雖然依靠點擊率進行推薦可以推薦出最熱門的即時話題，但除了熱門的即時話題之外，尚有其他多樣化的新聞內容，其中較軟性的議題像是旅遊或是親子議題等等就較不能用點擊率作為推薦的依據，而是需要針對使用者喜好的體驗進行分析再加以推薦。因此本研究想要針對新聞的體驗構面，結合機器學習的演算法來建構一套推薦系統。

擁有巨大的資料量的時候，透過機器學習和資料探勘的演算法，可以對巨大的資料進行統計，比對，解析，針對特定的問題能夠產生出客觀的結果，這樣的

方法和技術被廣泛運用在電子商務，社群網路，醫療記錄，甚至是科學的分析。不少的公司會運用機器學習的技術來預測並推薦使用者感有興趣的內容。

機器學習中有許多的方法，像是集群法，決策樹，類神經網路，基因演算等方法，分別使用不同的演算方法對資料進行分析，其中決策樹是一種樹狀資料結構的預測模型，能夠產生歸納性原則也兼具直覺的表達優點。

本研究嘗試建立一個考慮使用者體驗喜好之電子新聞推薦模型。Schmitt 在 1997 年提出的體驗模組：感官式(sense) 體驗、情感式(feel) 體驗、思考式(think) 體驗、行動式(act) 體驗以及關聯式(related) 體驗。我們以此體驗模組為基礎了解使用者對各體驗之重視程度，並運用機器學習的決策樹演算法，分析使用者重視的體驗類型以及使用者喜好的新聞類別之關聯式規則，再結合使用者重視之體驗類型和關聯式規則建立新聞推薦模型。

第二章 文獻回顧

第一節 體驗行銷

體驗行銷是現在行銷的風潮，舉凡大小的公司都會透過體驗來去宣傳自己的商品。例如販售電子產品的公司，像是蘋果，HTC 等等，會設置體驗專區使消費者可以親身體驗使用的感覺，亦或是到校園當中舉辦體驗活動，使消費者對於產品產生美好的體驗。而體驗行銷的概念，Schmitt 在 1999 年透過 *Experiential Marketing* 這本著作已經提出。在 Schmitt 的著作中也預示了體驗行銷將席捲 20 世紀的商場。

一、體驗的定義

目前成為主流的體驗行銷，Schmitt (1999) 以著三種角度闡述體驗的定義，包括哲學 (philosophical)、神經生物學 (neurobiological)、心理學與社會學 (psychology & sociology) 等觀點。透過這些觀點說明體驗和消費者的連結。

I. 哲學的觀點：

人對平常生活當中所經歷的事情和接受的刺激會產生一些感覺和回應，我們稱之為體驗。其中體驗包含的層面非常廣，會透過各種媒介得到不同的體驗。例如在平常透過眼睛看見的，耳朵聽到的，鼻子聞到的，四肢去碰觸到的，甚至是虛擬的感覺，這些都是體驗。而提供體驗的主體則可以透過各種媒介包裝所想要傳遞的訊息或是想販售的商品，提供體驗引人迴響。

II. 神經生物學觀點：

當人受到刺激時，透過大腦不同的神經區塊的互動反應會產生出體驗。更細部的說明，人的感官，例如眼睛，耳朵，鼻子，嘴巴，皮膚等等接受刺激之後，大腦中的『知覺與感覺（sensation）系統』會處理這些訊息，接著傳送到掌管思考和創造的『認知（cognition）系統』，最後則進入『自覺情感（affect）系統』進行記憶處理。而透過這三個系統彼此的互動會產生體驗。

III. 心理學與社會學觀點：

以社會學角度來說，是因為社會團體以及文化的互相影響而產生出體驗。而以心理學的角度，因著人長期的習慣，以及生活的模式從小到大累積的經驗，成為了體驗的根源。

二、策略體驗模組

Schmitt 認為體驗是經由數種互相獨立的體驗形式所組合而成。Schmitt 將體驗大致分成五形式：感官式體驗（sense），情感式體驗（feel），思考式體驗（think），行動式體驗（act），關聯式體驗（relate）。體驗策略模組則是以著這些形式為基礎，針對每個形式可以再進行進一步的設計，透過這些形式的整合達成想要傳達給消費者的體驗。

I. 感官式體驗：

感官式體驗主要是針對人基本的五個感官提供的體驗。透過提供感官上的刺激使得消費者對於產品的感受更加提升，進而使消費者對於產品更

有好感。像是書店當中提供舒適的音樂，透過聽覺使人放鬆的在書店中享受。或是食物商品透過產品的試吃，讓消費者可以感受味覺的刺激進而增強想購買的意願。

II. 情感式體驗：

情感式體驗則是藉由一些方式使消費者對於公司的品牌或是產品產生情感。透過這些方式使得消費者和產品產生連結超越只是需求的層次。常見的手法像是創造一些標語，為產品加上一些個性，例如 NBA 季後賽的廣告當中的標語『Where Amazing Happens』引發人想到過去在球場上發生的偉大事件，進而期待再度觀看球賽的時候可以見證奇蹟。

III. 思考式體驗：

思考是體驗主要是要讓消費者產生思考的動作。透過消費者感興趣的事物誘導他進入更深入的思考當中，這過程當中將產品和消費者之間做連結，如此營造出思考式的體驗。一些電影或是書本的宣傳當中常會提出一些問題引發人思考，進而使消費者對那主題產生興趣。

IV. 行動式體驗：

行動式體驗主要是使消費者透過親身去實踐一些項目而產生感受如此得到體驗。例如宣傳活動時常常擺放可以合照的人型立牌，使得參觀者不只是觀看而是能夠透過行動進行互動。

V. 關聯式體驗：

關聯式體驗是透過個人體驗和他人事物產生關聯，例如在論壇上面的討論，透過討論的時候不只是自己一個人的體驗而是和他人一起共創體驗。

或是在遊戲的進行當中，透過玩家間彼此的互動而產生不一樣的效果和樂趣。

三、體驗媒介

Schmitt 認為上述的五種體驗需要透過體驗媒介來進行傳達，透過這些體驗媒介，消費者和使用者可以得到 體驗傳達者所要傳遞的體驗。

I. 溝通體驗媒介：

例如新聞網站中，不同讀者的留言和討論，或是透過客服詢問瞭解的部分。

II. 視覺口語識別體驗媒介：

此媒介泛指的一些企業商標、設計或是產品的標語等可以區別企業本身和其他企業的象徵。例如 NIKE 的--Just Do It、Adidas—Impossible is nothing, 東森新聞雲--樂在分享，愛在雲端，聯合新聞網--觸動未來新勢力。

III. 產品呈現體驗媒介：

包括書店當中擺放書的方式，服飾店當中擺放衣服的方式以及新聞的排版等。

IV. 空間環境體驗媒介：

包括商場當中的佈置設計。

V. 網站體驗媒介：

現今網路已經成為了不可或缺的媒介，透過網站或是粉絲團的經營，使消費者以及民眾更加瞭解企業和產品的精神，其中包含官方網站、粉絲團和各個通訊平台的官方帳號。

VI. 共建品牌體驗媒介：

包括任何有共同合作的案件或是場合，例如發表會、企業聯盟、贊助商、共同合作活動案等等。

VII. 人員體驗媒介：

包含員工，客服人員，銷售員，以及任何會跟品牌、企業或商品連結的人。

第二節 資料探勘(Data Mining)

資料探勘的技術是結合人工智慧、機器學習、統計學、以及資料庫系統，透過這些領域的理論的演算法，應用在資料中而得到資料和資料間有價值的特殊規則。不過對於找出潛在規則這樣的行為其實不是一個新的技術，人類從過去就會研究動物的遷徙而找到適合的時機去打獵，也會找出穀物的生長週期而進行農事的規劃，政治上也會研究如何可以獲得更多支持者，這些都是相關的應用。但是近年來，由於電腦，行動通訊的普及，使得資料的儲存變得非常容易，因此產生出龐大的資料量。對於現在巨量的資料量，使用過去的方法無法精準的分析出其中隱藏的規則和關連性。因此學者們發展出許多的演算法，針對不同性質的資料，使用適合的演算法進行運算並產出分析結果，所產出的結果能夠對各領域的研究或是趨勢能夠有很大的助益。資料探勘目前已經廣泛的使用在各個領域，像是手寫識別、語音識別、生物特徵識別、醫學診斷、商業分析等等領域。在之後的幾個章節將會介紹資料探勘的步驟和常使用的方法，以及各種方法的特性。

一、資料探勘的操作步驟

為了從龐大的資料當中取得有價值的分析，必須要對原始資料進行處理並選擇適合的演算法來做運算，最後以著合適的方式呈現。在 Pang-Ning Tan et al.所作的 Introduction to Data Mining 書中把這過程分為三個階段[13]: 資料前處理，資料探勘，後處理。而 Han & Kamber(2001)[14]則把資料探勘的過程細分成七個步驟：

Step1：資料清理(Data Cleaning)：將雜亂和與資料及不相同的資料刪除

Step2：資料整合(Data integration)：將多個來源的資料整合

Step3：資料選擇(Data selection)：在資料庫中選擇適當的資料。

Step4：資料轉換(Data transformation)：轉換資料使其得以更方便地被使用

Step5：資料探勘(Data mining)：運用資料探勘的演算法來進行分析

Step6：模式評估(Pattern evaluation)：評估適合的模式加以使用

Step7：知識表達(Knowledge presentation)：使用適合的呈現方式表達資料探勘所得到的結果

二、資料探勘的技術與方法

對於原始資料進行初步的處理之後，需要依照資料的性質選擇最適合的演算法進行運算。本研究根據 Ian H.Witten et al.所著的”Data Mining”一書中的分類列出較常使用的演算法：

I. 1-R 規則法(1-rule)

是一種極為簡單直觀的演算法，其做法是在資料的每個屬性所延伸出來的結果，找出錯誤率最低的一組屬性而得到分析的結果。此演算法雖然簡單，但是實際操作上卻出乎意料的好，比起其他的演算法的準確率只有少零點多個百分點。這種演算法很適合用在讓人了解整個資料及的整體狀況，因為他的效率，也可以使用在很大的資料集當中。

II. 統計建模法(Statistical Modeling)

此演算法是在建立在 Bayes' 規則上，透過機率的運算產生出分析的結果，同樣的，這也是一個非常簡單的演算法，但是卻會產生出令人驚艷的結果，甚至表現的比一些複雜的演算法還要出色。但是統計建模法必須要在一個屬性影響的權重是平均散佈狀況下運算，結果才會較為精準。

III. 線性模組(Linear Models)

相較於上述的決策和規則方式的演算法，線性模組對於處理數字行的資料是更適合的。雖然說統計及資料探勘為兩門各自獨立的科目，但統計學與資料探勘的預測仍是有很大的相關性，資料探勘常會使用統計的方法對資料進行分析，其中包括一般迴歸(regression)、logistic regression、區別分析(discriminal analysis)等。

IV. 支援向量機

主要的操作方法是運用特徵值以及數學之計算來訓練資料，透過計算特徵值得到一個理想的分界，即可利用此超分界將資料進行分類。其中分類邊界若距離最近的訓練資料點越遠，則其分類之誤差會越小。分類邊界

和最近資料點之距離稱為間隔(margin)，支持向量機演算法期望找到一個最大間格分類邊界，此邊界和最近之訓練資料點擁有最大的距離。

V. 分群演算法(Clustering)

分群演算法是利用物件之間的關係資訊來做分群，其中物件彼此相關的話會被分到同一個群當中。當資訊當中各個群內相似程度大而群和群之間相差不大的話，這樣的狀況之下會有最好的結果。分群演算法當中最常見的有 K-nearest neighbor 演算法。

VI. 類神經網路(Neural Network)

類神經網路一開始的發想是想要模擬生物的神經系統。大腦是由很多神經元細胞組成，神經元彼此透過神經軸和突觸互相連結。神經學家發現人的大腦可以透過突觸受到的重複刺激進行學習。而學者運用這概念在資料探勘的領域，透過訓練和學習的方式發現資料彼此的關係。此技術適用於線性和非線性的狀況。

VII. 基因演算法(Genetic Algorithm)

基因演算法（或稱遺傳演算法）是由 John Holland 於 1975 年提出。此演算法是借鏡自然生態中的演化規律『適者生存』的概念加以應用在資料探勘的操作當中。透過模擬大自然的基因操作（例如：交配 cross-over, 反轉 inversion, 突變 mutation）對資料進行處理和分析。

VIII. 決策樹(decision tree)

決策樹演算法將資料透過分類判斷而建立成一個樹狀的模型。透過這種樹狀結構可以清楚地表達資訊的結構和相互關係。因為其清楚的結構和呈現常常可以提供很明確的分析，決策樹也成為了最受歡迎的資料探勘技術之一。而常見的決策樹演算法有 ID3,C4.5,CART 等。

三、決策樹演算法(Decision Tree)

決策樹演算法會將資料建構成樹狀的資料結構，將相同的，藉此分析出資料之間互相的關聯特性，其產出的結構十分易懂及視覺化，在各個領域都廣泛被使用。在決策樹演算法中 ID3 (Iterative Dichotomiser 3)和 C4.5 為普遍被使用的決策樹演算法，ID3 是由 Quinlan 於 1986 提出的[17][18]，C4.5 則是在 1993 提出。在本研究當中，由於樹狀的結構可以最清楚的表現出使用者喜歡的體驗類型以及使用者喜歡的新聞類別的關聯式關係，因此本研究決定使用以決策樹為基礎之演算法---ID3 決策樹演算法對資料集進行分析。

四、ID3 演算法(Iterative Dichotomiser 3)

ID3 演算法為決策樹演算法的一種，決策樹演算法有許多種類，其主要差別是在於用何種理論決定分類屬性的判斷。ID3 演算法在建構決策樹時，計算資訊獲利(Informational Gain)來決定分類屬性的判斷依據。本研究會首先介紹 ID3 演算法演算法所使用的資訊獲利，接下來依序介紹 ID3 決策樹架構、ID3 演算法的建構。

I. 資訊獲利

美國數學家 Claude Shannon 於 1948 年 10 月發表的 A mathematical Theory of Communication 提出了透過計算熵(Entropy)來進行資訊的量化。而在 ID3 決策樹演算法中將經過處理後的熵值和經過處理前的熵值相減，得到的數值稱為資訊獲利，ID3 演算法會比較各項資訊獲利的數值來決定該進行何種操作。

Claude Shannon 提出的熵值的計算方法，假設所有事件 x 的集合為 A ，而事件 x 發生的機率為 $P(x)$ ，而該資訊的熵值 $E(c)$ 為 A 中所有事件 x 的發生機率 $P(x)$ 乘上 $\log_2 P(x)$ 的總和負值(式 1)。

$$E(c) = - \sum_{x \in A} P(x) \log_2 P(x) \quad (1)$$

而資訊獲利 Information_gain 的計算方法是將經過處理前的資訊熵值 Entropy_before 減去經過處理後的資訊熵值 Entropy_after(式 2)

$$\text{Information_gain} = \text{Entropy_before} - \text{Entropy_after} \quad (2)$$

II. ID3 決策樹架構

ID3 演算法會建立一個決策樹分類模型，此決策樹的是透過遞迴(Recursive)和分治法(Divide-and-conquer)建立的，其架構是由上而下(Top-down)的樹狀結構(圖 4)，詳細介紹如下：

1. 根節點：在決策樹的建構過程中，一開始所有的資料都儲存於根節點中，在建立決策樹的過程當中會將資料分配到不同的節點。

2. 分支線：表示節點資訊的分枝路徑
3. 子節點：為根結點以下的之節點，亦可稱為內部節點
4. 葉節點：決策樹演算法進行至最後所停止之節點，在每個葉節點當中之資料具有類似的特徵。

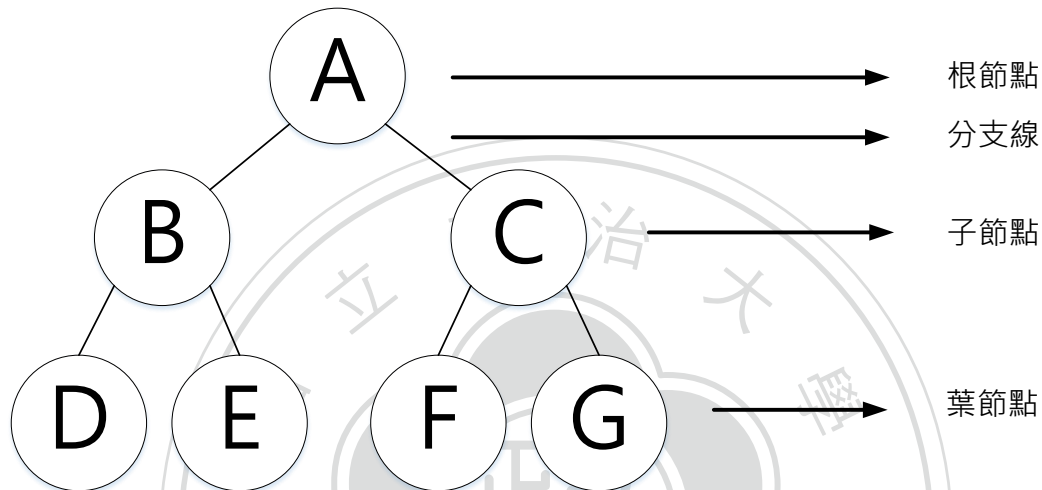


圖 4、決策樹架構

III. ID3 決策樹演算法運算流程：

本研究將所有的資料樣本集合以 T 來表示，若一共有 n 筆樣本資料，則 $T = (T_1, \dots, T_n)$ 。每一筆樣本資料皆有一組特徵屬性，我們將其表示為 $F = (F_1, \dots, F_n)$ ，每個特徵屬性有兩個特徵值 $f = (f_1, f_2)$ 。在執行 ID3 決策樹演算法時，會選定一項特徵屬性 F_n 作為目標屬性 C ，目標屬性 C 會有 n 個特徵值 $c = (c_1, \dots, c_n)$ ，除了目標屬性之外其餘的特徵屬性則為關鍵屬性。以下為 ID3 決策樹執行步驟：

Step1: 起初，所有資料物件皆屬於根節點 A 。

Step2: 分析目前節點中所有資料物件的目標屬性 C，若目標屬性 C 的特徵值 c 皆相同則停止運算，若非全部相同，則繼續進行步驟 3。

Step3: 若尚有關鍵屬性未被選取則進行步驟 4；倘若所有關鍵屬性皆被選取過，則計算節點中各關鍵屬性的類別出現次數，並選擇最多出現次數之關鍵屬性類別作為該節點之類別。

Step4: 計算目前節點 k 之熵值 $E(k)$ ，並計算所有尚未選擇過的關鍵屬性 F 之資訊獲利值 Information_gain，比較所有關鍵屬性之資訊獲利值 Information_gain 選取具有最大資訊獲利值之關鍵屬性 F，依據關鍵屬性的特徵值 f 將該節點的資料樣本分入兩個子節點。

Step5: 對 Step4 所產生的兩個子節點，從 Step2 開始重新執行。

第三節 李克特量表

在本研究的問卷設計是以李克特量表(Likert Scale)為基礎進行設計。李克特量表為心理學家 Rensis Likert 於 1932 所提出，透過回答李克特量表的題目，受測者會具體的回答出其認同程度。此量表是目前在調查研究中最常被使用的量表。

通常典型李克特量表給予受測者回答的認同程度選項為五個等級：非常不同意、不同意、無意見、同意、非常同意。不過也有研究者使用七個或九個等級的回答選項，也有些研究者將中間(無意見)項目的選項去除，只剩下四個等級的回答選項。

李克特量表的設計程序需進行以下步驟:

Step1: 針對研究議題設計出與了解受測者對議題態度的陳述，每一項的陳述皆予以設定相同等級數目的答案。

Step2: 根據針對每個陳述回答的偏向程度給予其不同的分數配置，若該項陳述為正面的陳述，則答案越正向分數越高；反之若該項陳述為負向陳述，則答案負向則分數越高。

Step3: 尋找些許受測者進行前測，根據受測者的回答進行對陳述句的調整。若陳述句無法具有鑑別力則需要對陳述句進行修改，避免受測者對每一陳述皆回覆相同答案。

Step4: 將調整好的陳述整合成正式量表(若無調整即直接進行整合)。

設計完量表後即可將此量表發放給受測者，所得到的資訊則為受測者對於該項議題之態度分數。而所有受測者的態度分數總合則為此群受測者對此議題的態度偏好狀況。

第四節 軟性新聞

在本研究中，主要進行推薦分析的新聞類型是屬於軟性新聞的範疇，Thomas E. Patterson 在描述軟性新聞的定義時說道，硬性新聞是較具時效性的新聞，像是報導重大事件，頭條，以及會影響甚至打亂日常生活的事件報導，而軟性新聞則是相對較無時效性，較以個人為中心，較感性的類型的新聞[19]。由於硬性新聞是屬於每位使用者必須要及時了解的新聞，因此此類新聞的推薦可以單純以熱門的程度進行推薦即可得到良好之果效。但軟性新聞則須分析每個使用者之喜好進行推薦，方可推薦出使用者所喜歡之新聞。下列表格為本研究蒐集之國內外各大報紙及網站之新聞分類方式。

表 2、台灣紙本新聞分類

自由時報	焦點新聞·政治新聞·生活新聞·財經新聞·國際新聞·新北都會新聞·新北都會生活·自由廣場·影視名人·專輯·國際娛樂·消費新聞·文化·藝術·自由副刊·旅遊·旅遊專輯·家庭親子·兩性異言堂·健康醫療·社會焦點·大社會·社會新聞·體壇焦點·運動場上·證券新聞·投資理財
中國時報	要聞·焦點新聞·政治綜合·政治新聞·財金新聞·生活新聞·社會新聞·國際新聞·時論廣場·文化新聞·消費·人間副刊
聯合報	焦點·要聞·話題·生活·社會·文化·綜合·國際·民意論壇·兩岸·產業·策略·投資理財·證券行情·健康·聯合副刊·繽紛·新北市新聞·運動·基隆新聞·教育·星話題·星火線·星 blog·星品味·消費·旅遊休閒
蘋果	頭條·要聞·政治·社會·法庭·論壇·全球焦點·國際新聞·運動焦點新聞·尬輸贏·籃球火·運動 bar·運彩王·教育·求職王·好屋·靚車王·財經·暖流·娛樂名人·電影表·木瓜霞吐槽·國際娛樂·名人時尚·頭條新聞·消費高手·讓我更美·健康醫療·3C 通訊·車市快遞·流行尖端·美食ㄟㄟ看·廚房 DIY·蘋果遊戲王·看漫畫·一日遊·每日運勢·蘋果遊戲王解答
經濟日報	頭版·話題·焦點·國際·兩岸·陸港股市·金融·產業·產業·稅務法務·經營管理·化學論壇·自動化周報·基金理財要聞·基金·基金走廊·理財·證券行情表·證券產業要聞·市場脈動·市場焦點·上市櫃公司·期貨·權證
工商時報	火線焦點·財經要聞·綜合要聞·政經八百·國際經濟·國際投資·國際產業·兩岸焦點·陸港股市·金融市場·科技要聞·產業商業·稅務法務·地方要聞·產業動態·房地產·健康照護·證券頭版·法人看市·上市櫃·集中市場·櫃買市場·期貨·權證·證券表板·財富頭版·金融商品·證券投資·證券表板·投資頭版·理財百寶箱·金融理財·證券表板·流通頭版·活動資訊·經營知識
旺報	焦點·話題·綜合·文教·財經要聞·產業·理財·地方新聞·社會·影視·兩岸史話·民意·論壇·兩岸徵文·社論
都會時報	頭版·評朝論野·全國公告·公益彩卷·都會偶像巨星
眾聲日報	焦點新聞·國會要聞·新聞大追擊·財經焦點·產業證券·財稅專刊·醫藥生活·金融匯市·全國教育·直轄市·北部地方新聞·直轄市·南部地方新聞·全國版公告·深度報導

表 3、國際紙本新聞分類

<p>International New York Times</p>	<ol style="list-style-type: none"> 1.First Page 2.World News : Europe, Asia, United 3.States 4.Style 5.Opinion 6.Culture : Art, Television, Books 7.Sports 8.Business (Dealbook) <p style="text-align: right;">2016/7/21</p>
<p>The New York Times</p>	<ol style="list-style-type: none"> 1.First Page 2.International 3.National 4.Weather Report 5.New York 6.Editorials/letters 7.Business Day: dealbook, automobiles, market gauges 8.SportsFriday 9.WeekendArts I : Movies Performances, The Listings 10.WeekendArts II : Fine Arts Leisure <p style="text-align: right;">2016/7/15</p>
<p>the guardian 英國 衛報</p>	<ol style="list-style-type: none"> 1.News 2.National :Special report 3.International 4.Financial 5.journal : the long readlopinionreviews 6.obituaries 7.education 8.theguardianjobs 9.weather 10.sport 11.shortcuts 12.women 13.style Q&A 14.TV & radios 15.puzzles & crossword <p style="text-align: right;">2016/7/19</p>

The Straights Times 新加坡海峽時報	1.Top of the News 2.World 3.Opinion 4.life weekend : happenings trlrvision buzz 5.Travel 6.Home 7.Business 8.sport <div style="text-align: right;">2016/7/15</div>
---------------------------------------	--

表 4、網路新聞分類

奇摩新聞	<ol style="list-style-type: none"> 1.首頁(圖片集錦、熱門新聞、心情新聞、新聞總覽、縮時新聞、雜誌專區、YAHOO 民調中心、人氣、Y 頭腦、懶人新聞卡、強震重創台南) 2.政治(政治熱門、政治影音、日本留學、外勞仲介、遊學團) 3.財經(股市匯市、產業動態、經貿財稅、房地產、理財就業、國際財經、財經熱門、公共消息、財經影音、投資聚焦) 4.影劇(藝人動態、電視廣播、音樂、電影、日韓、影劇熱門、名人娛樂、影劇影音、微博、線上音樂、卸妝油、美白產品、春季保養、爆紅內褲) 5.運動(棒球、籃球、高爾夫、網球、綜合、運動熱門、運動影音、2014 世足、塑身衣、飛輪、按摩椅) 6.社會(社會熱門、公共消息、綜合影音、汽車玻璃、鑽戒、濾水器) 7.地方(大台北、北台灣、中部離島、南臺灣、東台灣、地方熱門、公共消息、通馬桶、裝潢、桶裝水) 8.國際(亞澳、中港、美洲、歐非、國際財經、國際熱門、國際影音、遊學、美國遊學、海外婚禮) 9.生活(消費、交通、寵物、美食、生活綜合、生活熱門、一分鐘報氣象、生活影音、開胃料理、圓仔日記、檜木桶、外燴、成分簡單、黃牛包) 10.文教(教育、藝文、文教熱門、看新聞學英文、英文怎麼說、情緒教育、遊學代辦、英語學習、學英文) 11.健康(醫療衛生、美容保健、健康熱門、蘋果肌、外籍看護、褐藻糖膠) 12.科技(資訊 3C、科學發展、自然環境、科技熱門、3C 大事記、淨水器、網頁製作、硬碟救援) 13.旅遊(美食、旅遊熱門、達人帶你玩、香港住宿、訂房系統、東京飯店) 14.氣象(新頁面) 15.新奇(酷搜圖聞、FUN 影音、影音懶人包、澳門住宿、美白牙齒、氣墊粉)
------	--

	<p>餅)</p> <p>16.影音(影劇、綜合、生活、政治、財經、運動、國際、YAHOO 獨家、FUN 影音、凱蒂酷瑞克、2016 選舉、民視新聞、音波拉皮)</p> <p>17.專欄(社會觀察、休閒風尚、哈燒趨勢、Y!oung 觀點、給年輕人的信)</p>
PCHOME 新聞	<ol style="list-style-type: none"> 1. 首頁(氣象、新聞人物、對獎專區、影音新聞、一分鐘新聞、名人專欄) 2. 政治(政治焦點、最新政治、政治專題、政治照片) 3. 社會(社會焦點、最新社會、社會專題、社會照片) 4. 財經(財經焦點、最新財經、財經照片、股市新聞、基金新聞) 5. 科技(科技焦點、最新科技、科技專題、科技照片、遊戲新聞) 6. 國際(國際焦點、最新國際、國際專題、國際照片) 7. 大陸(大陸焦點、最新大陸、大陸照片) 8. 健康(健康焦點、最新健康) 9. 娛樂(新頁面) 10. 體育(體育焦點、最新體育、體育專題、體育照片) 11. 生活(生活焦點、最新生活、生活專題、生活照片) 12. 消費(消費焦點、最新消費) 13. 旅遊(旅遊焦點、最新旅遊) 14. 房產(房屋首頁、房產新聞、特別企劃、實用工具) 15. 專題(精選專題、政治、社會、財經、娛樂、體育、生活、科技、國際) 16. 雜誌(雜誌首頁、合作媒體) 17. 星座(牡羊、金牛、雙子、巨蟹、獅子、處女、天秤、先蠍、射手、魔羯、水瓶、雙魚) 18. 汽車(汽車焦點、最新汽車) 19. 政治(政府消息焦點、最新政府消息、行政院、中央政府、台北市、新北市、高雄市) 20. 民意(最新議題、熱門議題、熱門回應) 21. 通知
NOWnews 新聞	<ol style="list-style-type: none"> 1. 新聞(人氣總覽、名家論壇、公益行善、圖集總覽、新聞影音、今日頭條、快樂志工、星座算命) 2. 政治(政治看板、軍事新聞、NOW 評論、網友評論、親青爆報、政治最速報) 3. 財經(財經看板、經貿報一下、復華金管家、錢進東協、房市、理財、職場、台北房市、財經最速報) 4. 生活(生活看板、健康、校園、媽媽保健、命理、勾勾好事、愛鮮 iFRESH、保健新知、生活最速報) 5. 地方(台北、基隆、新竹、台中、南投、嘉義、高雄、宜蘭、金門、新北、桃園、苗栗、彰化、雲林、台南、屏東、花東)

	<ol style="list-style-type: none"> 6. 社會(社會看板、社會志工、社會最速報) 7. 運動(運動看板、台灣之光、綜合、棒球、籃球、足球、運動最速報) 8. 娛樂(娛樂看板、日韓流行線、電視、電影、音樂、八卦、色區) 9. 國際(國際看板、軍武大觀、國際最速報) 10. 大陸(大陸看板、花生網、熱話題、名人堂、兩岸最速報) 11. 新奇(新奇看板、新奇圖片、寵物、新奇最速報) 12. 消費(消費看板、流行時尚、民生消費、美食料理、汽車鑑賞、NOW 無 GAME) 13. 旅遊(旅遊看板、走遍台灣、環遊世界、旅行玩家、活動好康、劈腿女孩 Yaya) 14. 科技(科技看板、行動通訊、電玩動漫、雲端電視、網頁遊戲、Swapub) 15. 健康(健康養生、醫美減重、兩性關係、醫藥衛生、癌症新知、達人部落格、活動情報、診所地圖、健康專題)
蘋果即時	<p>最新、焦點、熱門、爆社、動物、搜奇、3C、影片、正妹、體育、圖解、媒陣、娛樂、時尚、生活、社會、國際、財經、地產、政治、論壇</p>
eNews	<ol style="list-style-type: none"> 1. 首頁 2. 名人娛樂 3. 即時新聞(其他、地方新聞、社會新聞、奇聞軼事、社會新聞、國際新聞、政治新聞、美通社、運動、今日頭條) 4. 休閒趣聞(寵物資訊、親子話題、生活知識、星座命理、搞笑趣聞、勵志感人、恐怖驚悚) 5. 健康生活(健康醫療、養生食譜、減肥塑身、美妝時尚、時尚穿搭) 6. 情感家庭(兩性關係、親子話題) 7. 其他(上班甘苦談、財經資訊、行銷創業、雜誌、經營管理、名人觀點、房市快報、手機&APP、電玩遊戲、電腦資訊、3C 配件) 8. 吃喝玩樂(活動展覽、美食特搜、料理食譜、旅遊資訊、達人分享) 9. 暖新聞

第三章 研究設計

本研究是從使用者的體驗之中分析其喜好來作為行為分析的基礎。我們運用機器學習當中的決策樹資料探勘演算法(decision tree data mining algorithm)來計算使用者行為的關聯式規則：我們分析使用者對各種不同型式體驗之重視程度以作為決策樹資料探勘的輸入屬性，並以消費者對於電子新聞的喜好與否作為目標屬性，利用決策樹演算法計算這些輸入屬性(使用者對各種不同型式體驗之喜好)與目標屬性(使用者對於電子新聞的選擇)之間的關聯式規則。接著利用這些規則來建構一個預測模型，以評估閱聽人對於未知電子新聞的接受程度，從而建立一個能有效符合使用者個人體驗喜好之新聞推薦模型：從眾多電子新聞之中經由預測模型的資料分析來預測並挑選出使用者可能感興趣的新聞，推薦給使用者(圖 5)。

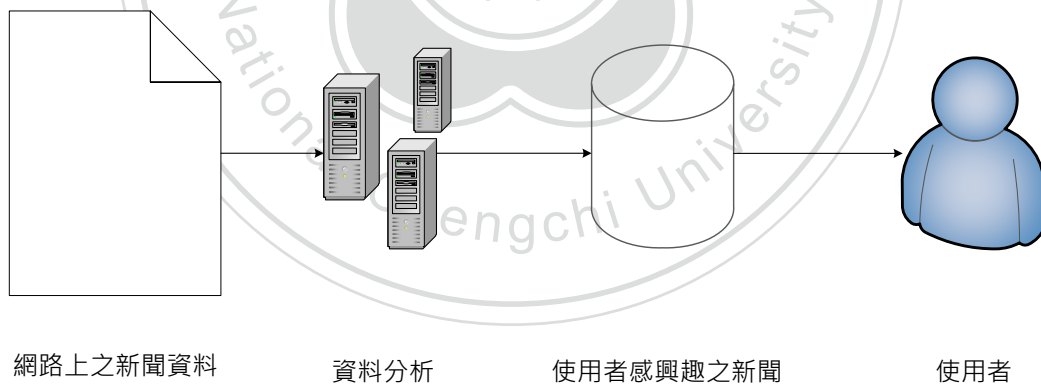


圖 5、研究示意圖

本研究的架構主要包含兩部分：(一)使用者行為資料蒐集、(二)透過機器學習建立新聞推薦模型。

其中，使用者行為資料蒐集的主要目的是分析使用者源自其經驗所產生的喜好狀況。為了要了解使用者對於不同的體驗類型偏好的程度以預測使用者會喜歡

的新聞，首先需要得到使用者對於不同類型體驗的偏好程度資料，同時收集使用者對於各種新聞類型的偏好程度之資料；這些使用者的喜好資料，本研究預定將透過問卷的方式來取得。問卷設計的流程則如圖 6 所示，首先是進行閱聽人體驗喜好之評估，接著根據評估結果設計問卷。而在實際給受測者填寫之前先進行前測以及修正，修正完畢後即進行問卷調查。

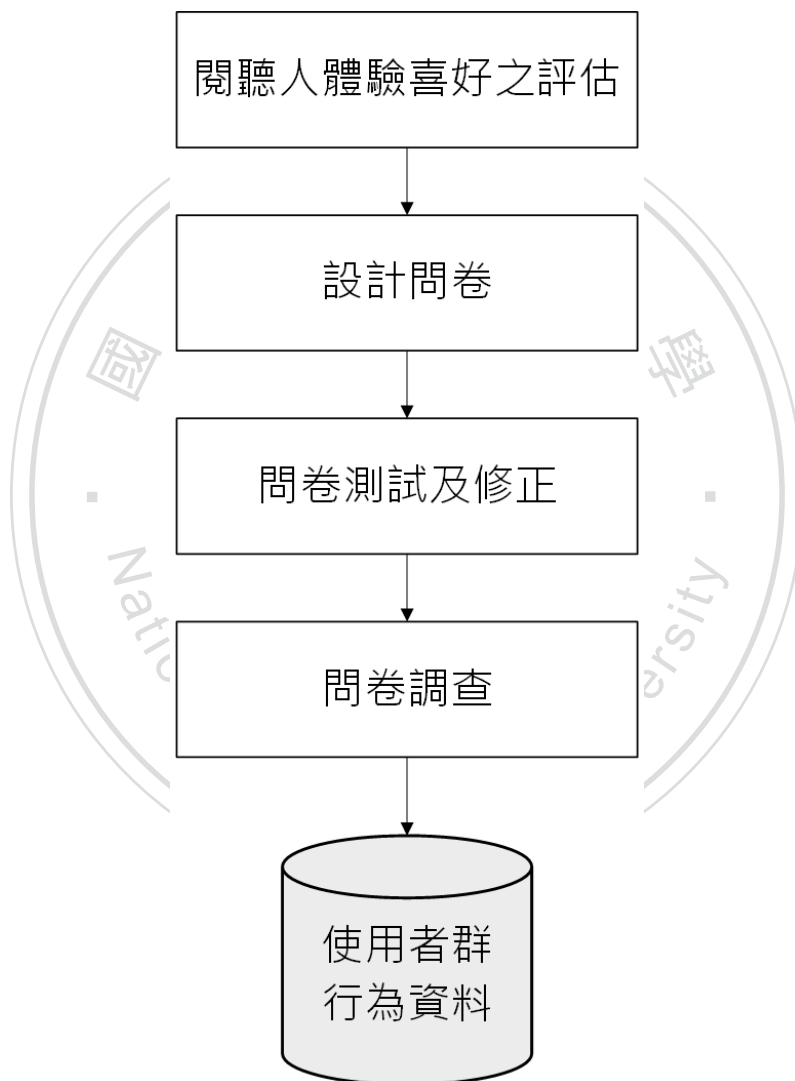


圖 6、問卷設計流程

經由第一部份的使用者行為資料蒐集階段而得到使用者對於不同類型體驗的喜好資訊之後，在第二部分：機器學習，本研究會以這些資訊為輸入，採用機

器學習的方式，利用資料探勘的演算法，來建立符合使用者喜好體驗之新聞推薦模型(圖 7)。

機器學習這個部分又可以再細分成為「訓練階段」和「執行階段」，主要目的是進行電子新聞的分析與推薦計算。在訓練階段，本研究分析前述問卷所得到關於使用者喜好的體驗資料以及使用者對於不同類型新聞的喜好度，以這些喜好資料作為輸入屬性，透過決策樹演算法進行資料探勘的計算，以分析出輸入屬性(代表使用者對於各類體驗的重視程度)與目標屬性(使用者對於各類電子新聞的偏好狀況)之間的關聯式規則。之後，以這些規則作為推薦計分之基礎，相關規則資料將儲存到資料庫當中以提供執行階段作為推薦決策之使用。而在執行階段，首先將對未知的電子新聞進行分析，再進一步由本研究的機器學習推薦模組結合訓練階段所得到的關聯式規則資料進行分析計算，以此判斷是否推薦此未知新聞給使用者。

以下章節分別就各個階段進行細部的介紹。

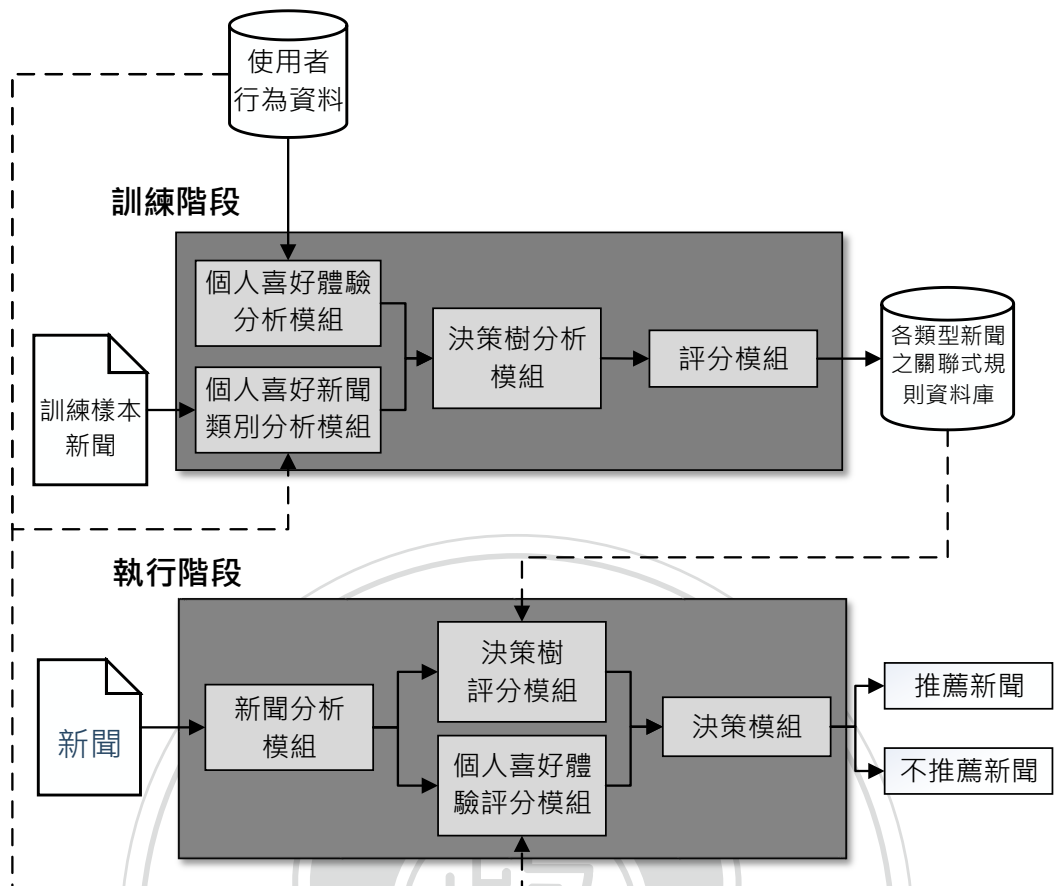


圖 7、機器學習架構圖

第一節 使用者行為資料蒐集

在進行機器學習時需要取得使用者行為資料而對其進行分析運算，因此在此階段本研究會以問卷的方式取得使用者所喜歡的新聞體驗類型資訊以及使用者喜歡的新聞類型資訊，問卷設計流程如圖 6 所示。流程共分為五個步驟：(1) 閱聽人體驗喜好之評估；(2)設計問卷；(3)問卷測試以及修正；(4)問卷調查；(5)儲存使用者群行為資料。以下分別就每個步驟進行說明。

一、閱聽人體驗喜好之評估

本研究的問卷設計主要分為兩個部分：(I)使用者體驗喜好調查，以及(II)使用者新聞類別喜好調查；分別取得使用者對於各類型體驗的偏好資訊以及取得使用者喜歡的新聞類別之資訊。為了要能夠設計出有效的問卷，本研究在文獻資料研究及分析階段會針對體驗模組理論以及現有的實體和網路新聞類別進行研究。

在使用者體驗喜好調查的問卷設計中，本研究以 schmitt 在 Experiential marketing 一書所提出的策略體驗模組作為基礎（見第二章），將體驗類型分為五個類型：感官式體驗、情感式體驗、思考式體驗、行動式體驗以及關聯式體驗。原本 schmitt 提出此策略體驗模組時主要是針對行銷應用，在本研究中，我們將取其原始意義並擴充這些體驗在新聞領域當中的相對應意涵(如表 5)。在問卷設計階段，我們將會以這些定義設計相對應的問題來取得使用者對五種體驗的喜好程度。

表 5、體驗模組之於新聞內容的意義

體驗類型	意義
感官式體驗	讀者觀看新聞時會感受到關於五感的刺激或想像，透過新聞的文字可以感受到新聞所描述之實體感覺。
情感式體驗	讀者閱讀新聞內容時會產生柔性情感之共鳴，內心會產生許多感性的情緒。
思考式體驗	讀者因為新聞的關鍵字、內容而激發自身的挑戰性或者好奇心而願意深入思索，或是報導內容具有深度、涉及重大的議題而引發讀者願意深思、腦力激盪。
行動式體驗	讀者閱讀新聞時會產生想要身體力行去從事某種行為、或願意參與某互動的想法。

關聯式體驗	讀者閱讀新聞報導時，透過其中的文字敘述或是報導可以得知其他相關的議題資訊或是與自身相關的利害資訊
-------	--

在使用者新聞類別喜好調查的問卷中，本研究分析推薦的新聞主要以軟性新聞為主，將各大報紙及網站之分類方法整合，選擇若干種新聞類型提供使用者，以確定使用者的喜好類型。這部分的結果將會切割為兩部分，分別應用於兩處：其一是作為本研究之機器學習模組在進行訓練階段時之輸入，以之作為決策樹資料探勘計算之目標屬性(使用者所喜歡的新聞類型)，藉此分析出使用者各類型體驗類型的喜好狀況對於目標屬性(各新聞類別的喜好狀況)的影響，找出之間的關聯式規則；另一處則是在本研究之機器學習模組進行執行階段時，作為系統之計算結果之對照印證，以確認正確性。系統將依據使用者的體驗喜好，以訓練階段所得到的關聯式規則進行分析計算，決定推薦或不推薦各個新聞；問卷事先詢問使用者之結果可以直接用於判斷本研究之機器學習模組之推薦準確程度。

為能夠找出最適合進行研究的軟性新聞分類方法，因此本研究在文獻資料研究及分析階段蒐集了紙本新聞以及較具代表性的網路新聞分類作為參考(見第二章)，重新規劃出較適合本研究進行分析的新聞分類(文化藝術類別/設計時尚類別/體育類別/健康類別/影視娛樂類別/旅遊類別/親子類別/美食類別)，以利問卷設計之進行。

二、設計問卷

1. 使用者體驗喜好調查：

如同上一節所描述的一般，本研究問卷的設計上主要分成兩個部分。在使用者體驗喜好調查的問卷，本研究會對上一節所提出的體驗類型再進一步設計，針對每一種體驗類型找出會讓使用者產生該體驗的重要關鍵因

素，例如會讓使用者產生感官式體驗的重要關鍵因素分別有嗅覺味覺要素、視聽要素、觸感要素。詳細的每一種體驗類型的重要關鍵因素列於表 6。

表 6、各新聞體驗之重要關聯因素

體驗類型	意義	重要關鍵因素	
感官式 體驗	讀者觀看新聞時會感受到關於五感的刺激或想像，透過新聞的文字可以感受到新聞所描述之實體感覺。	色香味感受	閱讀新聞內容時，讀者會被引發產生對於嗅覺與味覺的愉悅感受(e.g., 美食新聞)
		視聽愉悅	新聞內容或關鍵字能引發讀者視覺與聽覺的愉悅感受、共鳴。(視聽類型新聞)
		美感要素	讀者可以透過新聞內容觸發其對美感方面的感受
情感式 體驗	讀者閱讀新聞內容時會產生柔性情感之共鳴，內心會產生許多感性的情緒。	同情關懷	新聞報導的內容以及題材引起讀者的關懷及同情
		煽情感受	讀者透過新聞報導內容或關鍵字可以感受到催淚、煽情的感受。
		共鳴感受	新聞報導內容或關鍵字引發讀者美好的想像或憧憬
思考式 體驗	讀者因為新聞的關鍵字、內容而激發自身的挑戰性或者好奇心而願意深入思索，或是報導內容具有深度、涉及重大的議題而引發讀者願意深思、腦力激盪。	震撼性	新聞報導內容或關鍵字具有聳動、驚人的意涵。
		重大意義	新聞報導內容或關鍵字陳述一個重要的公共議題或者拋出一個重大的新問題
		反省檢討	讀者透過新聞內容可以對社會或某事件進行更深入的思考 (例如，論壇)
行動式 體驗	讀者閱讀新聞時會產生想要身體力行去從事某種行為、或願意參與某互動的想法。	聲援呼應	讀者閱讀新聞之報導時會激發想要行動的情緒(例如“滅頂”的呼籲)
		活動宣傳	報導內容或關鍵字大力地報導某個具有渲染力的重大活動(例如反課網活動)
		持續關注	報導內容或關鍵字陳述某個讓讀者願意持續關注的重大事件
關聯式 體驗	讀者閱讀新聞報導時，透過其中的文字敘述或是報導者的觀	議題相關性	新聞報導內容或關鍵字攸關群眾權益

	點，引發讀者想要去關懷的心情；或是報導是關於讀者本身利益的相關議題	利害相關	新聞報導內容或關鍵字挑動讀者去了解、比較
--	-----------------------------------	------	----------------------

找出影響每一個體驗類型的重要關鍵因素之後。接下來要針對體驗類型設計五等級之李克特量表(見第二章)，本研究針對每一個重要關鍵因素設計三種不同面相之陳述(表 7)，受測者可以針對每個陳述選擇對其之同意程度(非常同意、很同意、普通、不同意、非常不同意)。一旦回收了問卷，本研究便可以從使用者對以上問題所做出的回答了解他們對每一種體驗類型的喜好程度以及對每一個重要關鍵因素的敏感程度。

表 7、各重要關聯因素之對應問句

體驗類型	重要關鍵因素	問題
感官式體驗	色香味感受	1. 在閱讀新聞時，我喜歡內容當中有描述食物或飲品的味道的報導
		2. 如果閱讀到仔細描述食品香味和口味的報導，我會很有興趣
		3. 在新聞當中如果有出現很仔細描寫食品的滋味和嗅味的內容，我會去想像那感覺並享受在其中
	視聽愉悅	1. 我覺得在新聞當中有出現關於視覺衝擊和音效衝擊的內容會引發我的興趣
		2. 閱讀有視覺和聽覺內容的新聞的時候，我會沉浸在其中
		3. 我喜歡吸收有關視覺和聽覺內容的新聞，閱讀時會讓我產生愉悅的感受
	美感要素	1. 我喜歡閱讀描述或展示物品質感和雅致的新聞
		2. 閱讀有關於描述物品的質感的新聞時，我會花心思去思考並感受那新聞
		3. 我會花心思描述物品或商品形狀樣式的新聞，透過欣賞那物品，我會覺得開心

情感式體驗	同情關懷	1. 我喜歡閱讀到一些資訊可以讓我知道有哪些地方或人事物是需要我去關懷的
		2. 閱讀到一些新聞在描述某些人事物是需要被關懷的時候，我會很重視並且想要去了解
		3. 我很想要去知道有哪些人事物是亟需要被幫助獲支持的，我很希望在新聞當中吸收到這些資訊
	煽情感受	1. 閱讀新聞時，我喜歡閱讀到關於有關煽情的內容
		2. 閱讀新聞時，我會喜歡看到有催淚感受的內容
		3. 我會喜歡讀到可以給予自己一些刺激情感的內容，像是催淚或是煽情等情感
	共鳴感受	1. 我喜歡閱讀到一些新聞是可以讓我對生活或對未來是可以產生美好的憧憬
		2. 當閱讀到一些可以讓我產生美好的生活想像的新聞，我會覺得很開心
		3. 我會希望透過閱讀一些可以帶給自己正面想像的新聞，透過這些新聞得到正面的力量
思考式體驗	震撼性	1. 我喜歡閱讀到擁有聳動標題或是內容的新聞
		2. 當閱讀到有聳動的標題或內容的新聞會讓我重視一些議題，所以我很喜歡
		3. 我希望可以瀏覽到一些報導驚人意涵的新聞，讓我可以更仔細去思考一些議題
	重大意義	1. 我喜歡閱讀報導重大公眾議題或是提出重要新問題的新聞
		2. 我會希望在新聞當中能夠看到報導重大議題的新聞，這讓我可以持續關注這社會
		3. 當新聞報導到關於重大議題，我會留意並閱讀，因為想要了解在社會當中發生的事情，也希望能夠出一分力
	反省檢討	1. 我喜歡閱讀一些對重大事件進行更深入的思考的新聞
		2. 我會喜歡一些新聞報導讓我們對重大議題或是現況做深入的思考，因為這會讓我更深入看清楚真實社會現況
		3. 我喜歡看關於深入探討某件議題的新聞，因為這會讓我從不同方面去看待並對待事情
行動式體驗	聲援呼應	1. 我喜歡閱讀到一些新聞是會激發我們想要去參加活動的心理
		2. 我會希望從新聞當中夠得到一些想要一同參與某些重大的活動的刺激，並了解其重要性
		3. 我會希望閱讀到新聞報導呼籲名眾去參與一些活動或是行動，因為我喜歡去參加以及感受許多事物
	活動宣傳	1. 我喜歡閱讀報導重要活動的新聞

		2. 我希望可以從新聞知道一些資訊讓我可以去參與一些實際活動
		3. 我希望從新聞當中了解重大活動的意義以及來龍去脈
		持續關注
		1. 我喜歡閱讀到針對一個行動或活動持續的追蹤報導
		2. 我會希望可以得知一個重要事件或是行動的持續報導，因為我也想要一同參與
		3. 我想要持續關注一個重要行動或事件的狀態和發展狀況，所以很希望從新聞當中得到這些資訊
關聯式體驗	議題相關性	1. 我會想要閱讀到一些新聞報導的內容引導我可以連結到其他相關議題的內容
		2. 因為想要了解並關注其他地區遇到的相關的議題，所以會想要閱讀到一些可以引導我得知更多相關類似議題的新聞內容
		3. 我喜歡吸收並了解資訊，所以我會想要閱讀到一些引導我會去查詢相關資訊的新聞內容
	利害相關	1. 我會喜歡閱讀跟自身或是社會利益相關的新聞內容
		2. 我會從新聞當中得知和自身利益相關的資訊，讓我可以更掌握可以幫助自己的資訊
		3. 我很關注和自己利益相關的資訊，因為這對來說很重要

以上為關於使用者體驗喜好調查的問卷內容，第二部分的問卷是關於使用者新聞類別喜好調查，透過這部分的問卷，可以得到使用者喜歡的新聞類別，本研究可以透過第一部分和第二部分的結果作為資料集來對使用者行為進行分析。

II. 使用者新聞類別喜好調查：

在使用者新聞類別喜好調查的問卷中，本研究整理了紙本新聞和網路新聞的分類類別之後，找出適合本研究的方式來推薦的軟性新聞類型，分別為：文化藝術類別、設計時尚類別、體育類別、健康類別、影視娛樂類別、旅遊類別、親子類別、美食類別，表 8 詳細介紹各種內容的新聞歸屬於何種新聞類別。針對每個類別，本研究分別找出七則新聞放在問卷中供

使用者閱讀，並設計李克特量表以取得使用者對此新聞之喜好程度。其中李克特量表採的五個等級的回答選項設計：非常喜歡，很喜歡，尚可，不喜歡，非常不喜歡。

表 8、各新聞類別所包含之新聞內容

新聞類別名稱	所包含新聞之新聞內容
文化藝術類別	文化新聞、藝術新聞等
設計時尚類別	時尚新聞、精品新聞
體育類別	體育新聞、運動賽事新聞等
健康類別	健康相關新聞、醫療新聞等
影視娛樂類別	娛樂新聞、名人偶像新聞、國際娛樂新聞、八卦新聞等
旅遊類別	休閒相關新聞、旅遊新聞等
親子類別	教育新聞、親子相關新聞等
美食類別	美食新聞

三、問卷測試、修正及問卷調查

設計完問卷之後，將問卷的樣式排版進行微調之後，即進行問卷調查。本研究鎖定的研究族群為大學生，針對五百名大學生進行問卷調查。回收問卷之後本研究會將問卷的結果輸入至使用者行為資料庫，以供機器學習使用。完成使用者行為資料蒐集之後，本研究將透過此部分所蒐集到的使用者所喜歡的體驗類型以及使用者喜歡的類別的資料進行機器學習，藉此分析出兩者的關聯性規則。

第二節 機器學習

經過第一階段問卷的調查得到使用者喜歡的新聞體驗類型以及喜歡的新聞類別之後，這個階段要運用機器學習的演算法首先分析出兩者的關聯式規則，並運用此關聯式規則來建立能有效符合使用者個人體驗喜好之新聞推薦模型。在機器學習段中分為訓練階段和執行階段，訓練階段顧名思義就是要對系統進行訓練而讓系統有能夠推薦出適合的新聞，在本研究的訓練階段會透過問卷的結果分析出使用者所喜歡的體驗和使用者喜歡的新聞類別的關聯式規則，例如說喜歡思考式體驗的人有較大的機率會喜歡社會類的新聞等等。並且將這些關聯式規則依照影響程度的大小加以配分，所產生的配分資料會進一步使用在執行階段。而在執行階段則會實際操作推薦新聞的動作，本研究會輸入未知樣本新聞資料，機器學習模組將會分析此新聞所含有的體驗類型以及文件內容屬於何種分類，之後將這些資訊分別和訓練階段所得到的使用者資訊做分析運算，最後判斷是否推薦該新聞。以下分別對訓練階段以及執行階段更進一步作解說。

一、訓練階段

在訓練階段本研究首先輸入訓練樣本新聞，訓練樣本新聞是在問卷階段已經由受測者閱讀過的新聞資料，本研究系統當中的個人喜好新聞類別會從問卷得到的使用者行為資料當中得到使用者對於該新聞的喜好度，並且也會將新聞進行分類。

而個人喜好體驗分析模組則會從使用者行為資料分析出使用者喜歡的體驗類型資訊。接下來決策樹分析模組會對以上資訊進行決策樹演算法的分析，獲取使用者喜好的體驗資料以及使用者對於各新聞類別的喜好度的關聯式規則。之後評分模組會對所得到的關聯式規則進行權重的配分，並將此結果儲存至各類型新聞之關聯式規則資料庫。

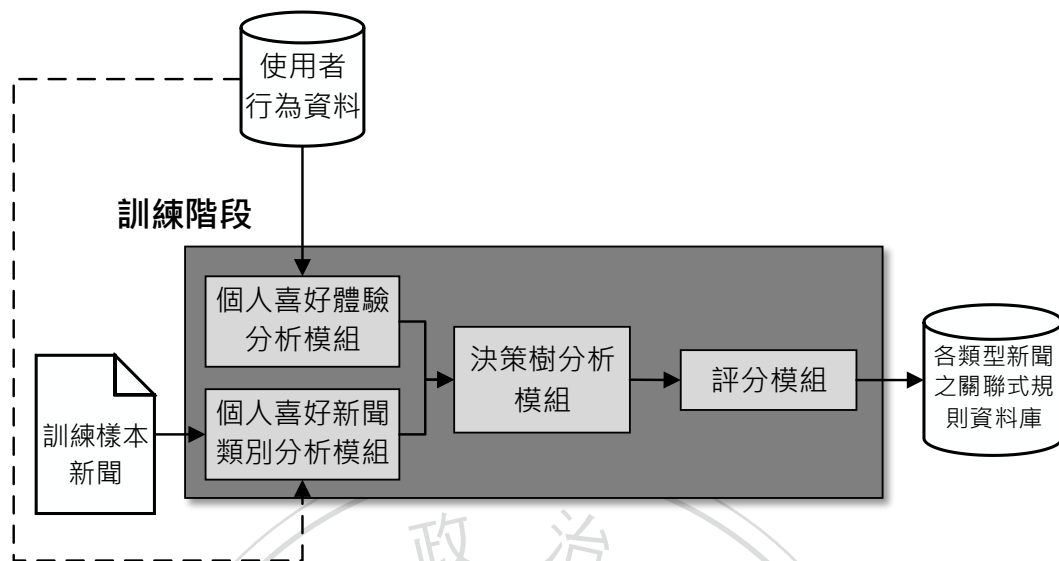


圖 8、訓練階段架構圖

I. 個人喜好體驗分析模組

在個人喜好體驗分析模組中首先輸入問卷當中所得到的使用者行為資料。在第一階段新聞體驗溝面的問卷設計中，本研究針對每一種重要關鍵因素Factor_i分別設計了三個重要關鍵要素問題(j)，在此模組中本研究針對問卷的結果進行運算，分析出受測者對於各個重要關鍵因素Factor_i的好感程度。

Step1：將每個受測者針對每一個重要關鍵要素問題(j)不同的回答進行配分，得到不同的體驗好感得分Exp_preference_point(j)，當中的配分如表 9。

表 9、好感得分配分表

回答	Exp_preference_point(j)
非常同意	100
同意	75
普通	50
不同意	25
非常不同意	0

Step2：由於問卷的題目是針對新聞體驗的重要關鍵因素所設計的，因此透過計算問卷回覆的得分可以得知測試者所喜歡的體驗關鍵因素。算式(1)將每個重要關鍵因素Factor_i回答的答案(j)所得的體驗好感得分 Exp_preference_point(j)相加並除上每個重要關鍵因素的問題數量 Question_amount(Factor_i)，得到體驗好感分數 Exp_preference_score(Factor_i)。若是體驗好感分數 Exp_preference_score(Factor_i)的值大於 50，則本研究定義該名測試者喜歡該重要關鍵因素，若是體驗好感分數 Exp_preference_score(Factor_i)小於或等於 50，則本研究定義該名測試者不喜歡該重要關鍵因素。其結果可以表示如表 10。

$$\text{Exp_prefernce_score}(\text{Factor}_i) = \frac{\sum_{j=1}^n \text{Exp_preference_point}(j)}{\text{Question_amount}(\text{Factor}_i)} \quad (1)$$

表 10、使用者對各重要關鍵因素之喜好表

重要關鍵因素	使用者喜好
色香味感受	喜歡/不喜歡
視聽愉悅	喜歡/不喜歡
美感要素	喜歡/不喜歡
同情關懷	喜歡/不喜歡
煽情感受	喜歡/不喜歡
共鳴感受	喜歡/不喜歡
震撼性	喜歡/不喜歡
重大意義	喜歡/不喜歡
反省檢討	喜歡/不喜歡
聲援呼應	喜歡/不喜歡
活動宣傳	喜歡/不喜歡
持續關注	喜歡/不喜歡
議題相關性	喜歡/不喜歡
利害相關	喜歡/不喜歡

II. 個人喜好新聞類別分析模組

在此模組中首先將被輸入的訓練樣本新聞和使用者行為資料中新聞類別構面問卷的回答進行比對，在前面章節有提到訓練樣本新聞是在問卷階段已經由受測者閱讀過的新聞資料，因此可以在使用者行為資料中對應

到相同的新聞以及該新聞所屬的新聞類別，並取得每個受測者對訓練樣本新聞中各種類別新聞的喜好程度。

Step1：將每個受測者針對每一個新聞喜好問題(k)不同的回答進行配分，得到不同的新聞喜好得分 $News_preference_point(k)$ ，當中的配分如表 11。

表 11、新聞喜好得分配分表

回答	News_preference_point(k)
非常喜歡	100
喜歡	75
尚可	50
不喜歡	25
非常不喜歡	0

Step2：透過計算問卷回覆的得分可以得知測試者對各種類型的電子新聞之喜好程度。算式(2)將每種新聞類別 $Type_i$ 中針對每個新聞喜好問題(k)回覆的答案所得的新聞喜好得分 $Exp_preference_point(k)$ 相加並除上屬於每個新聞類別 $Type_i$ 的新聞數量 $News_amount(Type_i)$ ，得到新聞喜好分數 $News_preference_score(Type_i)$ 。若是新聞喜好分數 $News_preference_score(Type_i)$ 的值大於 50，則本研究定義該名測試者喜歡該新聞類別，若是新聞喜好分數 $News_preference_score(Type_i)$ 小於或等於 50，則本研究定義該名測試者不喜歡該新聞類別。其結果可以表示如表 12。

$$\text{News_preference_score}(\text{Type}_i) = \frac{\sum_{k=1}^n \text{Exp_preference_point}(k)}{\text{News_amount}(\text{Type}_i)} \quad (2)$$

表 12、使用者對各新聞類別之喜好表

新聞類別	使用者喜好
藝術與文化類別	喜歡/不喜歡
影視娛樂八卦類別	喜歡/不喜歡
美食消費類別	喜歡/不喜歡
運動與競賽類別	喜歡/不喜歡
醫療健康類別	喜歡/不喜歡
休閒旅遊類別	喜歡/不喜歡
家庭綜合事務類別	喜歡/不喜歡

III. 決策樹分析模組

得到測試者對於每個新聞體驗當中重要關鍵因素的喜好程度以及測試者對各種新聞類別喜好的程度之後，接下來將這些資訊送入至決策樹分析模組進行 ID3 決策樹演算法的分析運算。本研究將新聞類別當作目標屬性，新聞體驗的重要關鍵因素作為決策屬性，分別對每個新聞類別建立其決策樹(圖 9)。經由各個決策樹本研究可以得到測試者喜歡的重要關鍵因素和測試者喜歡的新聞類別之關聯式規則。

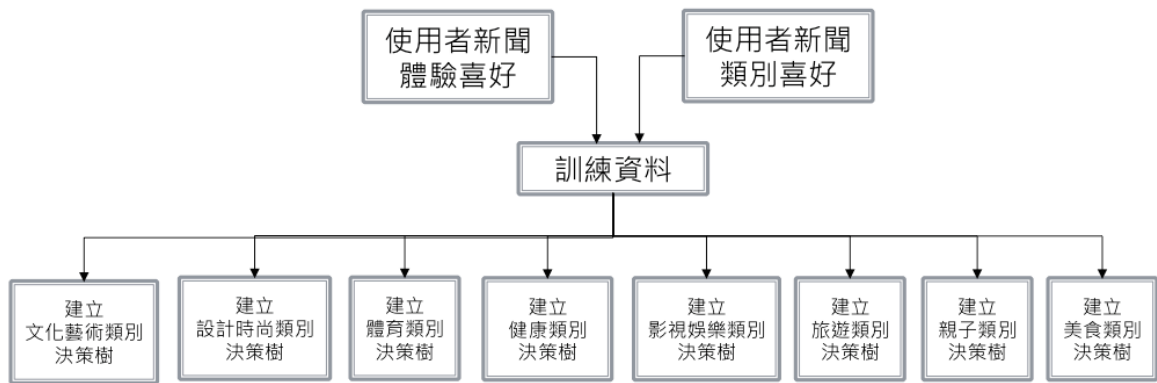


圖 9、各新聞類別決策樹

IV. 評分模組

建立完決策樹之後，本研究的配分模組會將決策樹所產生的每一個關聯規則加以配分後並儲存至關聯規則資料庫，供系統在執行階段的決策模組判斷時使用。

Step1: 在算式(3)中本研究對節點 p 取得該節點的支持度 $Support(p)$ 及節點純度 $Purity(p)$ 之乘積，並計算所有節點當中 $Value$ 的最大值 $Value_max$ 以及最小值 $Value_min$ 。

Step2: 計算出權重 $Weight(p)$ (算式 4)

Step3: 透過權重 $Weight(p)$ 以及 $Value_min$ 調整所有 $Value(p)$ 值之分布，使 $Value(p)$ 值介於 0 和 50 之間，其調整過後之值為 $Adjust_value(p)$ (算式 5)

Step4: 若節點 p 之屬性為不喜歡該決策樹之新聞類別，則節點分數為 $50 - Adjust_value(p)$ ，反之則為 $50 + Adjust_value(p)$ (算式 6)

$$Value(p) = Purity(p) \times Support(p) \quad (3)$$

$$Weight(p) = \frac{50}{Value_max - Value_min} \quad (4)$$

$$Adjust_value(p) = (Value(p) - Value_min) \times Weight(p) \quad (5)$$

$$Score(p) = 50 \pm Adjust_value(p) \quad (6)$$

計算完各節點的配分之後將每個節點所對應的關聯式規則 R 以及配分資訊儲存至各類別規則資料庫中，在執行階段將會以這些資訊為依據決定是否推薦新聞。

二、執行階段

執行階段運用訓練階段所分析出的關聯式規則資料庫以及使用者對各體驗類型的喜好程度來推薦新聞。本研究在此階段輸入未知樣本新聞，首先新聞分析模組會對此新聞進行分析，並將分析完之資訊分別送入決策評分模組以及個人喜好體驗評分模組進行評分。在決策樹評分模組中會依照訓練階段得到的各類型新聞的關聯式規則對新聞進行評分。而在個人喜好體驗評分模組則會透過使用者喜好的體驗資訊為基礎對新聞進行評分。完成這兩部分的評分之後會將結果皆送至決策模組整合並判斷是否要推薦該未知新聞。以下分別對各模組進行介紹。

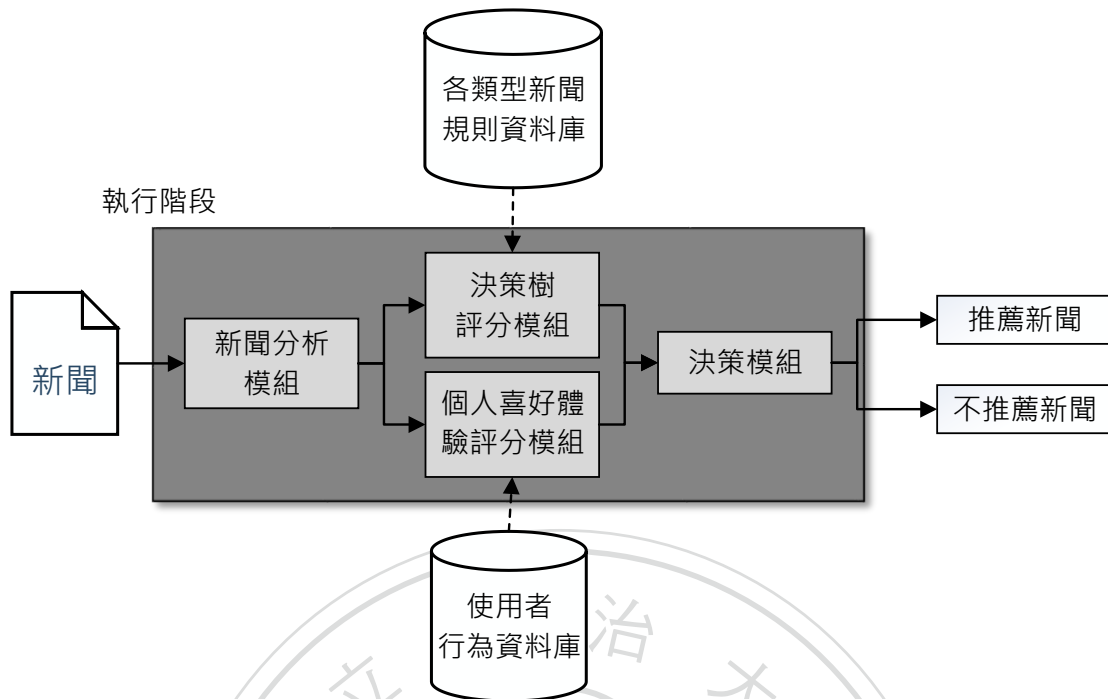


圖 10、執行階段流程圖

1.新聞分析模組

新聞的在此模組中，系統將未知新聞進行兩個層面的分析：新聞體驗層面以及新聞類別層面。本研究會以監督式學習的方式對新聞擁有的體驗重要關鍵因素以及新聞的類別進行分析。

- a. 新聞體驗層面：分析新聞當中具有的體驗重要關鍵因素，其結果可能不單只有一種因素。
- b. 新聞類別層面：分析新聞是屬於哪一種類別的新聞，而在本研究當中每一則新聞只會屬於一種新聞類別。

表 13、體驗重要關鍵因素及新聞類別種類

體驗重要關鍵因素	色香味感受/視聽愉悅/美感要素/同情關懷/煽情感受/共鳴感受/震撼性/重大意義/反省檢討/聲援呼應/活動宣傳/持續關注/議題相關性/利害相關
新聞類別	文化藝術類別/設計時尚類別/體育類別/健康類別/影視娛樂類別/旅遊類別/親子類別/美食類別

II. 決策樹評分模組

在決策樹評分模組中本研究運用訓練階段計算出的各類型新聞之關聯式規則資訊對新聞進行評分。

Step1 : 從新聞分析模組取得新聞的類別資訊

Step2 : 從使用者行為資料庫當中取得使用者對於各重要關鍵因素($Factor_i$)的體驗好感分數 $Exp_preference_score(Factor_i)$

Step3 : 從各類型新聞之關聯式規則資料庫中取得此新聞類別的關聯式規則 R_i

Step4 : 配對每條關聯式規則 R_i ，若關聯規則條件成立則得到該規則的規則得分 $Score(R_i)$

Step5 : 計算算式(7)，將優化係數(c) 乘上所有規則得分 $Score(R_i)$ 的和得到決策樹得分($Desiciontree_score$)

$$Desiciontree_score = c \times \sum Score(R_i) \quad (7)$$

III. 個人喜好體驗評分模組

Step1：將新聞分析模組中分析出來的新聞體驗層面資訊進行配分，如果某項重要關鍵因素 $Factor_i$ 沒有出現在新聞 t 當中，則設定那一項重要關鍵因素的權重為 0，如果有出現則按照下列式子算出該重要關鍵因素 $Factor_i$ 之配分權重 $Exp_weight(Factor_i)$ 。算式(8)為配分權重 $Exp_weight(Factor_i)$ 之計算方式，其中 t 代表新聞， $Factor_amount(t)$ 代表該新聞總共擁有重要關鍵因素之總數，配分權重 $Exp_weight(Factor_i)$ 等於 1 除以 $amount(t)$ 。

$$Exp_weight(Factor_i) = \frac{1}{Factor_amount(t)} \quad (8)$$

Step2：計算算式(9)，式中之 $Factor_i$ 為重要關鍵因素，從使用者行為資料庫中取得使用者對每個重要關鍵因素之體驗好感分數 $Exp_preference_score(Factor_i)$ ，將各個重要關鍵因素之體驗配分權重 $Exp_weight(Factor_i)$ 和體驗好感分數 $Exp_preference_score(Factor_i)$ 相乘之後累加起來，得到對此新聞的喜好體驗分數 Exp_score 。

$$Exp_score = \sum_{i=1}^n Exp_weight(Factor_i) \times Preference_score(Factor_i) \quad (9)$$

IV. 決策模組

Step1：取得決策樹評分模組中算式(7)計算出的決策樹得分 $Desiciontree_score$ 和個人喜好體驗評分模組中算式(9)計算出的喜好體驗分數 Exp_score 。

Step2 : 計算算式(8)，算式中 α 為權重分配係數， α 介於 0 和 1 之間。將決策樹得分 $Desiciontree_score$ 和喜好體驗分數 Exp_score 分別乘上 $1-\alpha$ 以及 α 之後相加，得到最終推薦分數 $Final_score$ (算式 10)。

$$Final_score = \alpha \times Exp_score + (1 - \alpha) \times Desiciontree_score \quad (10)$$

Step3 : 若最終推薦分數 $Final_score$ 大於或等於最小推薦分數 $Min_suggest_score$ 則推薦此新聞，反之，若推薦分數 $Suggest_score$ 小於最小推薦分數 $Min_suggest_score$ 則不推薦此新聞。



第四章 實驗結果與分析

第一節 問卷設計及結果分析

本節共分為三部分，以下分別介紹本研究如何取得使用者之行為資料以及取得結果之後問卷的信度及效度分析情形。

一、問卷設計

由於目前無法取得使用者實際於網路上點擊各新聞之資料情形，因此本研究使用問卷之方式取得使用者對於各種體驗的重視程度以及對各種類型新聞的偏好狀況。問卷共分為兩個部分，(1) 使用者體驗喜好調查、(2) 使用者新聞類別喜好調查。問卷的設計流程如圖 11，共分為文獻資料研究及分析、閱聽人體驗喜好之評估、設計問卷、問卷測試及修正以及問卷調查共五個階段。

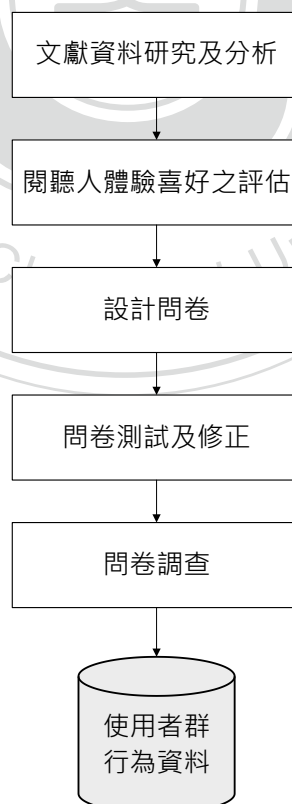


圖 11、問卷設計流程

i. 使用者體驗喜好調查

本研究以 schmitt 提出的體驗模組中的五種體驗(感官式體驗、情感式體驗、思考式體驗、行動式體驗、關聯式體驗)作為基礎，針對每一種體驗類型找出會讓使用者產生該體驗的重要關鍵因素。為了取得受測者對每個體驗重要關鍵因素的喜好程度，我們對每個體驗重要關鍵因素設計三個問題，製作李克特量表，用以分析使用者喜好之體驗。

ii. 使用者新聞類別喜好調查

本研究蒐集分析數個國際知名新聞之分類，並將所有非熱門類型，較屬於軟性新聞之新聞類別彙整，共整理成八種新聞類別。我們針對每個不同類別的新聞，蒐集各十篇網路上該類別的新聞作為問卷之題項列於問卷中。

二、問卷測試及修正

在問卷的測試與修正階段，本研究將設計好的問卷首先進行初步的測試，選定 50 位大學生作為測試者進行測試，回收問卷後，我們觀察是否有影響甚至干擾受測者作答之編排方式或是說明敘述，並加以改善之後發放正式問卷 (圖 12)。本研究發放 429 份問卷作為樣本資料。

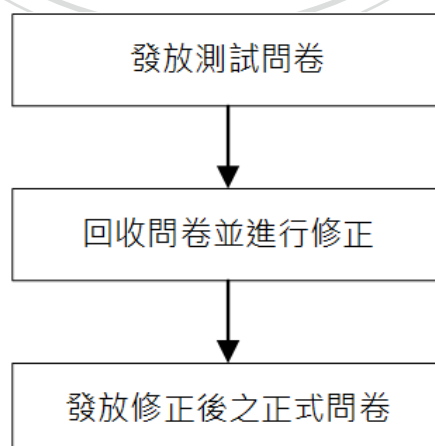


圖 12、問卷修正流程

三、問卷信度分析

信度分析的用意是在於測量受測者回答結果的一致性、可靠性以及穩定性。若計算出來真信度係數越高，則表示此問卷之回答結果有較高的一致性。信度衡量的標準有數種方法，以下列出較常使用之方法

I. 再測信度

讓同樣之受測者，在一定時間內重複填寫相同之問卷，以兩次填寫的結果計算出其相關係數，相關係數越高，則表示其回答有較高之一致性，若相關係數較低，則表示回答的結果較沒有一致性

II. 折半信度

折半信度是將受測的題目分成兩半，計算受測者在這兩部分題目各自的得分，再將其以「斯—布公式」(Spearman-Brown Formula)進行運算，以求得其相關程度

III. Cronbach' s α 信度

在 1951 年由 Lee J. Cronbach 提出，計算 α 值以評估問卷結果之信度指標。

本實驗採用實務上應用最廣之 Cronbach' s α 信度作為衡量信度之標準，在本研究中使用者體驗喜好調查問卷以及新聞類別喜好調查問卷之回答結果的 Cronbach' s α 分數列於下表，根據學者吳宗正&吳育東(2000)建議信度參考範圍(表 16)，在本實驗兩個問卷的各項問題組合的回答結果之信度(表 14、表 15)皆超過 0.7，代表使用者回答的結果具有相當的一致性，有很高的可信度。

表 14、體驗重要關鍵因素信度

體驗重要關鍵因素	信度
色香味感受	0.849
視聽愉悅	0.818
美感要素	0.876
同情關懷	0.888
煽情感受	0.912
共鳴感受	0.879
震撼性	0.834
重大意義	0.889
反省檢討	0.882
聲援呼應	0.894
活動宣傳	0.842
持續關注	0.845
議題相關性	0.871
利害相關	0.905

表 15、各新聞類別因素信度

新聞類別	信度
文化藝術	0.721
設計時尚	0.774
運動	0.842
健康	0.786
影視娛樂	0.735
旅遊	0.790
親子	0.833
美食	0.882

表 16、Cronbach α 係數與可信度高低之對照表

可信度	Cronbach α 係數
不可信	Cronbach α 係數 < 0.3
勉強可信	$0.3 \leq$ Cronbach α 係數 < 0.4
可信	$0.4 \leq$ Cronbach α 係數 < 0.5
很可信 (最常見)	$0.5 \leq$ Cronbach α 係數 < 0.7
很可信 (次常見)	$0.7 \leq$ Cronbach α 係數 < 0.9
十分可信	$0.9 \leq$ Cronbach α 係數

四、問卷效度分析

檢測效度的意義在於判斷問卷是否確實取得實驗所欲取得之資訊。本研究使用因素分析進行效度的檢測。因素分析的作用是用來檢驗問卷各個問項之『建構效度』，透過抽取各個問項之間之共同因素(common factor)，計算各個問項之因素負荷量用以評估解釋各個構面。進行因素分析須考慮到以下幾點：KMO 值、Barlett 球型檢定。

KMO 值是以相關分析之形式，顯示出實驗所取得的樣本資料量是否適宜。KMO 值介於 0 到 1 之間，若值越大則代表資料集具有越高的效度，反之則反。

而 Barlett 球型檢定是對變項之間的相關矩陣進行的球型檢定，若變項彼此之間的相關係數越高，則表示資料集具有較高的效度。

本研究的問卷共分為兩部分，使用者體驗喜好調卷以及使用類別喜好調查問卷，表 17 為兩問卷之 KMO 值和 Barlett 球型檢定值。

表 17、各問卷之 KMO 值及球形檢定值

	Kaiser-Meyer-Olkin 取樣適切性量數	Bartlett 的球形檢定 近似卡方分配	顯著性
使用者體驗 喜好調查問卷	0.827	11761.077	0.000
使用者新聞類別 喜好調查問卷	0.851	7123.942	0.000

由表 17 可得知兩份問卷之 KMO 值皆大於 0.6，兩份問卷亦有較高的 Bartlett 的球形檢定值，顯示此二份問卷皆有良好的效度[20]。

第二節 推薦系統衡量指標及驗證方法

建立推薦系統在目前是一個熱門的研究領域，而如何評估眾多推薦系統的準確度亦是一個重要的議題，許多學者提出不同的方式以評估推薦系統的推薦準確度。Guy Shani 和 Asela Gunawardana 指出通常在判斷推薦系統是否推薦出合適的項目時，常使用的衡量方法為計算為 precision 及 recall，使用這些數值互相對照來衡量推薦的準確程度。另外也有許多學者使用 f-measure 來綜合 precision 以及 recall 的結果進行衡量。以下分別介紹各種衡量的方法。[21][22]

I. 混淆矩陣

推薦系統在推薦每一個項目時，其結果共可分為四種：

- a. True Positive, TP : 預測為 Positive，結果為正確
- b. False Positive, FP : 預測為 Positive，結果為錯誤
- c. True Negative, TN : 預測為 Negative，結果為正確
- d. False Negative, FN : 預測為 Negative，結果為錯誤

表 18、混淆矩陣

		實際	
		Positive	Negative
預測	Positive	TP, True Positive	FP, False Positive
	Negative	FN, False Negative	TN, True Negative

我們可以將所有之推薦結果累積而得到推薦系統結果之混淆矩陣，作為衡量推薦準確度運算的依據，許多推薦系統評估之數值也是以混淆矩陣為基礎進行運算。

II. precision

precision 為推薦之精確率，即在推薦出的項目當中有多少是準確的。計算的方法是將推薦系統推薦結果之 True Positive 的總量除上 True Positive 數量及 False Positive 數量的和。

$$\text{precision} = \frac{TP}{TP + FP}$$

III. recall

recall 為召回率，系統實際推薦出使用者喜歡的項目占有所有使用者喜歡的內容的比例。計算的方法是將推薦系統推薦結果之 True Positive 的總量除上 True Positive 數量及 False Negative 數量的和。

$$\text{recall} = \frac{TP}{TP + FN}$$

IV. f-measure

f-measure，亦稱為 F1 score 或 F-score。precision 以及 recall 兩個數據必須要互相衡量對照來取得一個平衡點，由於若推薦的項目多的時 recall 有很大的機率會提高，但 precision 相對有可能會降低，反之，若系統推薦的項目偏少，則 precision 相對會提高，但是 recall 的數值則會減少。因此許多學者使用 f-measure 來評估 precision 以及 recall 之平衡。

f-measure 之計算方式：

$$f\text{-measure} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

V. K 次交叉驗證

交叉驗證是在統計學當中將樣本分割成較小子集的方法，可將一部份的樣本做為分析以及建立模型使用，另一部份的樣本則做為驗證此模型之測試資料。而 K 次交叉驗證(K-fold cross validation)是將資料集隨機分為 K 個子資料集，一共做 K 輪的測試，每輪輪流取一份資料集作為測試資料，而存留的 K-1 個資料集則作為訓練的資料集，如此進行 K 輪之後求其均值。一般來說 K 值會設定為 10。

第三節 實驗結果分析

本研究共取得 429 位大學以及碩士學生族群作為樣本資料，其中我們將使用者資料分成兩部分，一部分作為訓練資料集，另一部分則作為測試資料集，在訓練階段，本研究使用 K 次交叉驗證(K-fold cross-validation)對訓練資料即進行分

析，將 K 值設定為 10，對資料集進行 10 次的交叉驗證後對每個新聞類別各自調整出最好之參數值。並以得到之決策樹規則及各類別決策樹參數值作為推薦模組之設定對測試資料集進行推薦。

一、演算法之參數調整

本研究回收問卷取得樣本資料後，便將樣本資料輸入至使用者行為資料庫，並運用之進行機器學習(見 3.2 節)。在進行機器學習過程當中調整各個參數使此推薦模型能夠得到最高的 f-measure 表現。

I. 決策樹節點停止分割之純度門檻 Min_node_purity

在決策樹模組建立各類別新聞之決策樹時，須設定停止決策樹節點繼續進行分割之條件，其中一個條件為最小純度之設定，當該節點的純度大於中止運算最小純度之標準時，該節點則不繼續進行分割，成為葉節點，不同的停止條件會使決策樹產出不同的規則，從而影響推薦模組的推薦準確程度。圖 13 及表 19 顯示最小純度設定之下，推薦模型之推薦準確度。(其他可變參數之設定， $\text{Min_suggest_score} = 60$ ， $\alpha = 0.2$)

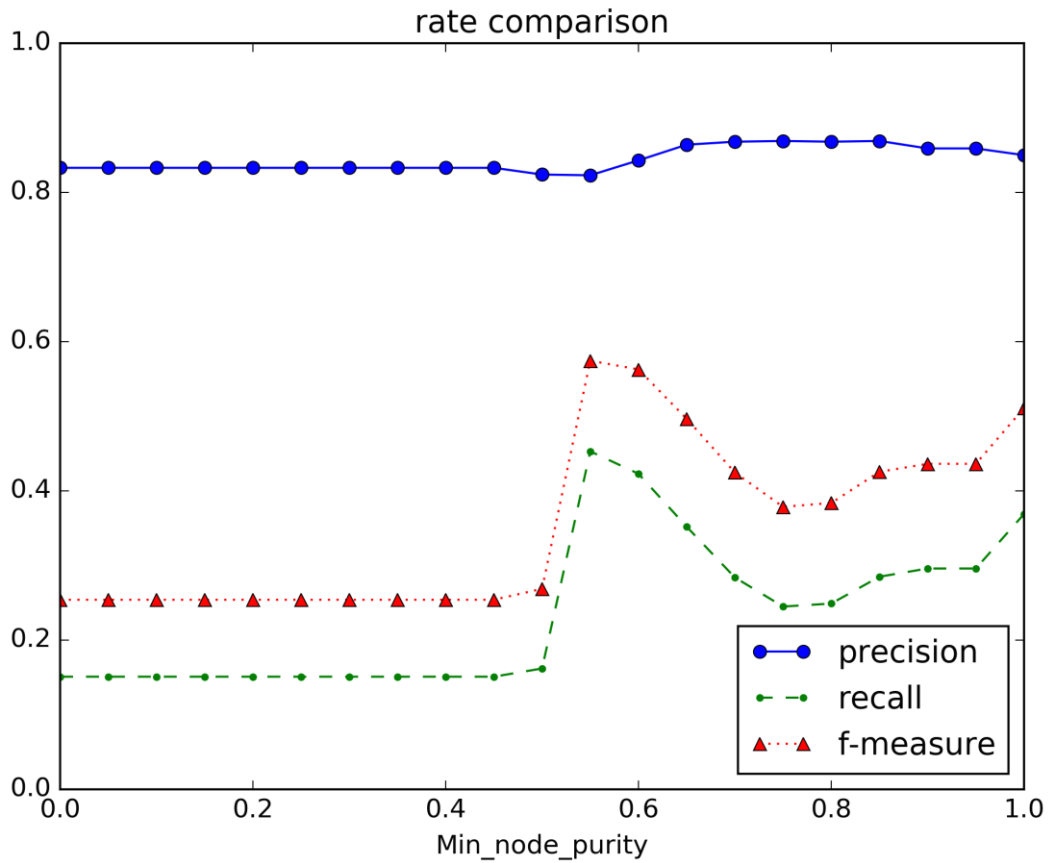


圖 13、決策樹停止分割純度門檻之變化影響圖

表 19、決策樹停止分割純度門檻之變化影響表

	precision	recall	f-measure
0.0	0.833	0.151	0.2541
0.05	0.833	0.151	0.2541
0.1	0.833	0.151	0.2541
0.15	0.833	0.151	0.2541
0.2	0.833	0.151	0.2541
0.25	0.833	0.151	0.2541
0.3	0.833	0.151	0.2541
0.35	0.833	0.151	0.2541
0.4	0.833	0.151	0.2541
0.45	0.833	0.151	0.2541
0.5	0.824	0.162	0.2688
0.55	0.823	0.453	0.5745
0.6	0.843	0.423	0.5628
0.65	0.864	0.352	0.4967
0.7	0.868	0.284	0.4247
0.75	0.869	0.245	0.3787
0.8	0.868	0.249	0.3838
0.85	0.869	0.285	0.4256
0.9	0.859	0.296	0.4363
0.95	0.859	0.296	0.4363
1.0	0.85	0.369	0.5111

由圖 13 可知，當中止節點分割的純度門檻超過 0.5 時，推薦模型之 precision 會略為下降，但 recall 之數值會開始大幅的提升，f-measure 在純度門檻為 0.55 時達到最高值，之後 f-measure 開始下降直到純度門檻為 0.75 為止，純度門檻值大於 0.75 之後 f-measure 又開始提升。為了調整每個新聞類別的決策樹使之皆有良好之表現，我們針對每個新聞類別之決策樹找出最好類別之節點分割的純度門檻值(圖 14、表 20)，使每個類別之新聞能擁有最佳之 f-measure 值。

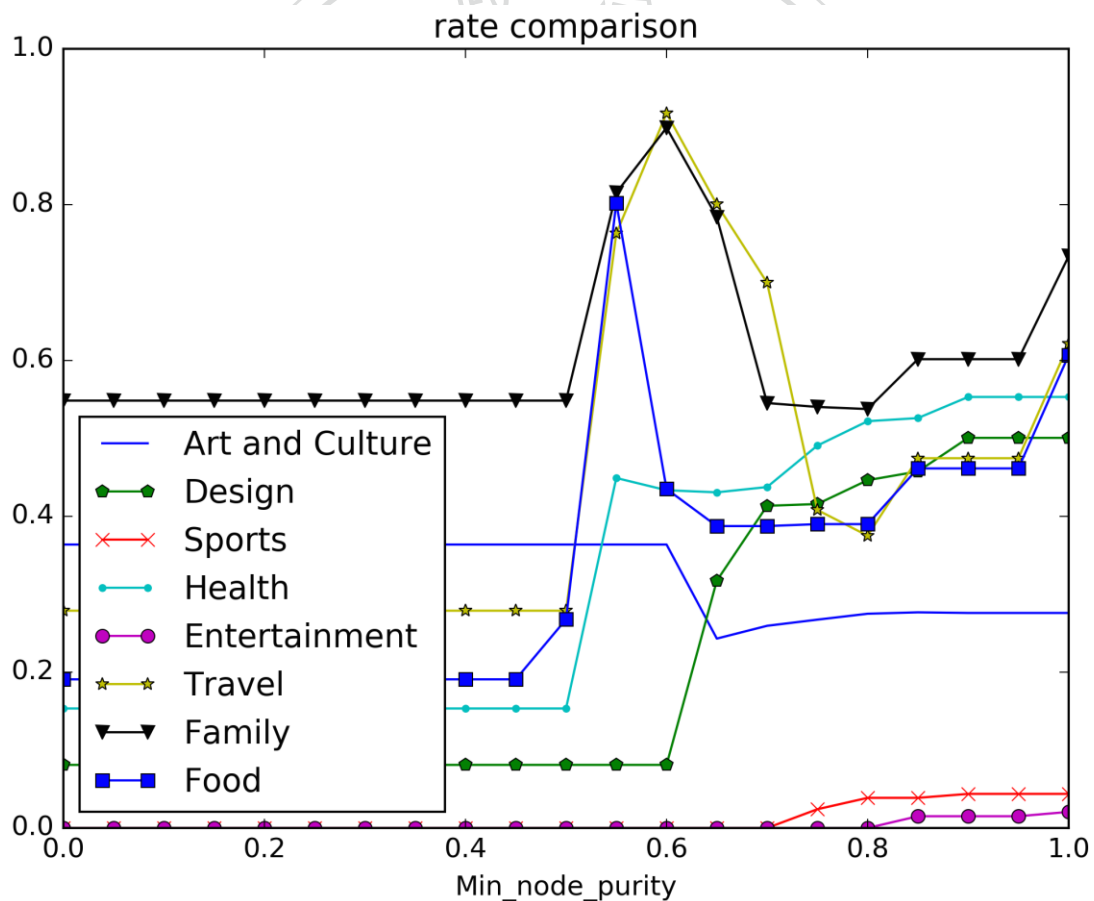


圖 14、決策樹停止分割純度門檻對不同各類別新聞之變化影響圖

表 20、決策樹停止分割純度門檻對不同各類別新聞之變化影響表

	Art and Culture	Design	Sports	Health	Entertainment	Travel	Family	Food
0.0	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.05	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.1	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.15	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.2	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.25	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.3	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.35	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.4	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.45	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.1908
0.5	0.3638	0.081	0.0	0.1534	0.0	0.2789	0.5485	0.2678
0.55	0.3638	0.081	0.0	0.4492	0.0	0.7629	0.8152	0.8023
0.6	0.3638	0.081	0.0	0.4335	0.0	0.9173	0.8988	0.4351
0.65	0.2431	0.3174	0.0	0.4308	0.0	0.8004	0.784	0.3875
0.7	0.2595	0.4135	0.0	0.4376	0.0	0.6999	0.5454	0.3875
0.75	0.2674	0.4158	0.0241	0.4908	0.0	0.4082	0.5404	0.39
0.8	0.275	0.4464	0.0387	0.5221	0.0	0.3751	0.5377	0.39
0.85	0.2768	0.4576	0.0387	0.5261	0.0151	0.4746	0.6017	0.4615
0.9	0.276	0.5007	0.0438	0.5532	0.0151	0.4746	0.6017	0.4615
0.95	0.276	0.5007	0.0438	0.5532	0.0151	0.4746	0.6017	0.4615
1.0	0.276	0.5007	0.0438	0.5532	0.0207	0.6212	0.7339	0.6069

II. 最終推薦分數之權重分配係數 α (alpha)

權重分配係數 α (alpha) 是在執行階段的決策模組中，決定新聞是否被推薦的決策分數。在決策模組中會計算出最終推薦分數 Final_score，其組成是由兩個得分依照比例所結合而成，決策樹評分模組中計算出的決策樹得分 Desiciontree_point 和個人喜好體驗評分模組中計算出的喜好體驗分數 Exp_score。其運算方式是將決策樹得分 Desiciontree_point 和喜好體驗分數 Exp_score 分別乘上 $1-\alpha$ 以及 α 之後相加而成。其中 α 介於 0 到 1 之間，用以調整最終推薦分數之來源比例。

$$\text{Final_score} = \alpha \times \text{Exp_score} + (1 - \alpha) \times \text{Desiciontree_score}$$

圖 15 及表 21 分別顯示設定不同 α 值之下系統之推薦準確程度。(其他可變參數，Min_suggest_score = 50)

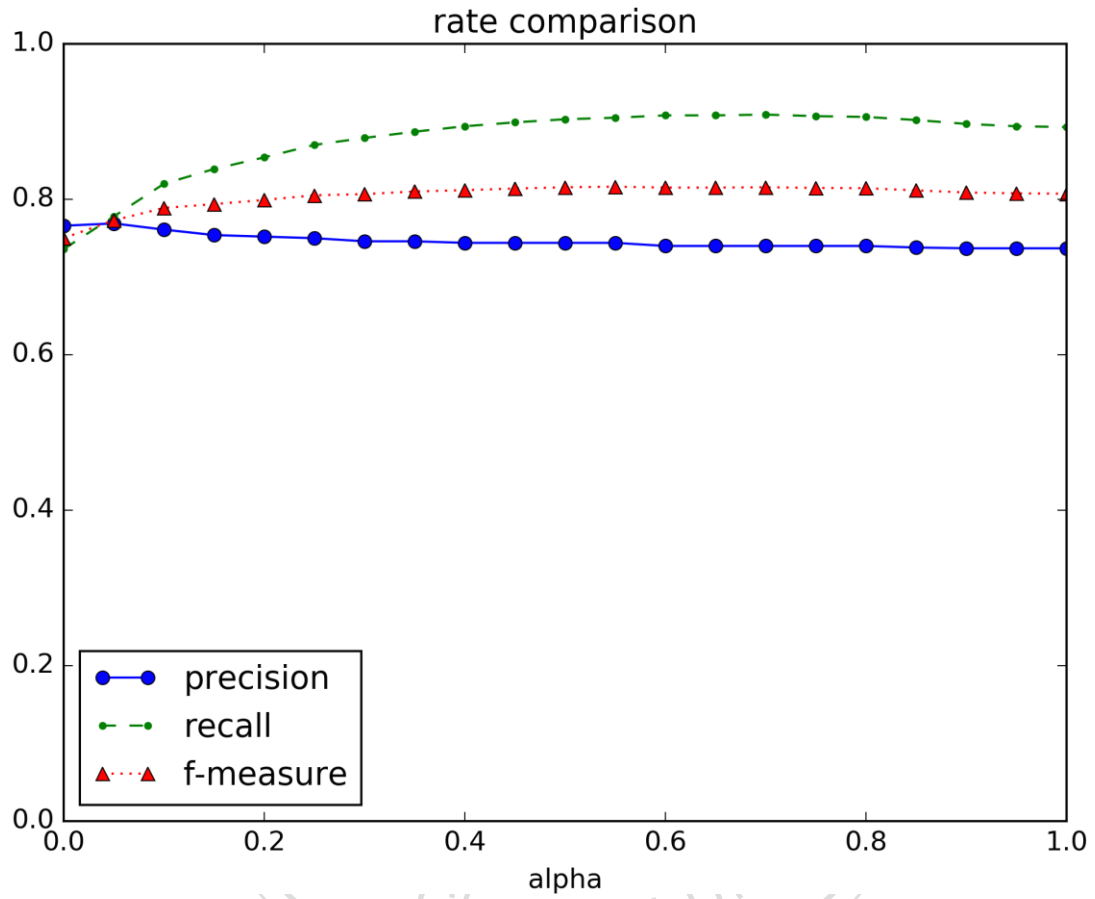


圖 15、權重分配係數 alpha 值之變化影響圖

表 21、權重分配係數 alpha 值之變化影響

	precision	recall	f-measure
0.0	0.766	0.736	0.7502
0.05	0.769	0.778	0.773
0.1	0.761	0.82	0.7889
0.15	0.754	0.839	0.7938
0.2	0.752	0.854	0.7992
0.25	0.75	0.87	0.8052
0.3	0.746	0.879	0.8067
0.35	0.746	0.887	0.8101
0.4	0.744	0.894	0.8118
0.45	0.744	0.899	0.8139
0.5	0.744	0.903	0.8155
0.55	0.744	0.905	0.8163
0.6	0.74	0.908	0.8151
0.65	0.74	0.908	0.8151
0.7	0.74	0.909	0.8155
0.75	0.74	0.907	0.8147
0.8	0.74	0.906	0.8143
0.85	0.738	0.902	0.8115
0.9	0.737	0.897	0.8088
0.95	0.737	0.894	0.8077
1.0	0.737	0.893	0.8072

透過圖 15 可以得知 α 值越大對於推薦結果之 precision 有降低之趨勢，對 recall、f-measure 則有提升之趨勢，當 α 值等於 0.5 時，本推薦模組可以取得最高之 f-measure 值。此數據顯示當推薦分數的分數組成比例，若是決策樹得分 Desiciontree_score 較高的話，本推薦模型推薦出的錯誤率會降低，但相對會有些使用者喜歡的新聞較無法被推薦出來。推薦量也會相對降低可推薦出較多使用者喜歡之新聞，但相對也會推薦較多錯誤的新聞。反之，若使用者喜好體驗分數 Exp_score 的比例較高的話，推薦模型可推薦出較多使用者喜歡之新聞，但相對也會產生較多錯誤推薦的狀況。以下之圖表(圖 16、表 22)為各個類別決策樹於不同 α 值之表現情形。

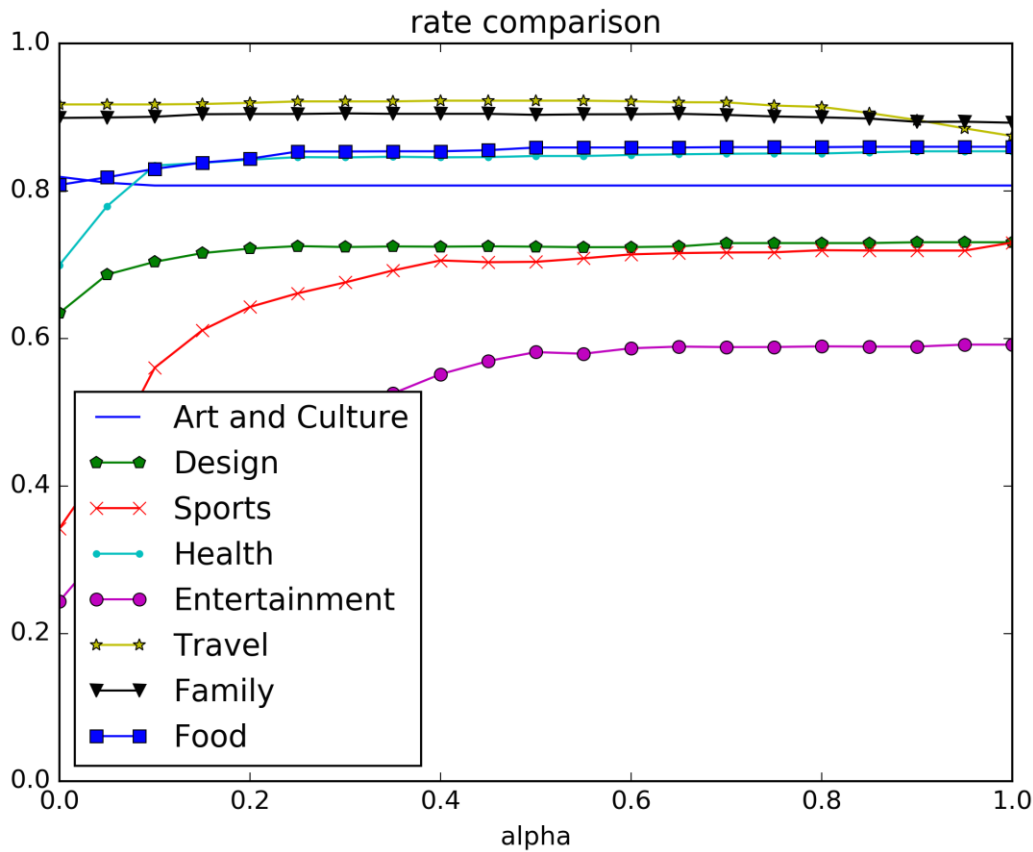


圖 16、權重分配係數 alpha 值對各類別新聞之變化影響圖

表 22、權重分配係數 alpha 值對各類別新聞之變化影響表

	Art and Culture	Design	Sports	Health	Entertainment	Travel	Family	Food
0.0	0.8192	0.6346	0.3427	0.6996	0.244	0.9173	0.899	0.8083
0.05	0.8112	0.6867	0.4364	0.7794	0.3199	0.9173	0.8996	0.8185
0.1	0.8074	0.704	0.5604	0.8345	0.3791	0.9173	0.9006	0.83
0.15	0.8074	0.7158	0.6113	0.8381	0.3946	0.9179	0.904	0.8385
0.2	0.8074	0.7219	0.6428	0.8424	0.4319	0.9194	0.9044	0.8437
0.25	0.8074	0.7253	0.6613	0.8459	0.4753	0.9214	0.9044	0.8535
0.3	0.8074	0.724	0.6761	0.8455	0.4989	0.9214	0.9052	0.8535
0.35	0.8074	0.7248	0.6923	0.8464	0.5256	0.9214	0.9046	0.8538
0.4	0.8074	0.7245	0.7059	0.8456	0.5517	0.9224	0.9046	0.8538
0.45	0.8074	0.725	0.7035	0.8459	0.5695	0.9224	0.9046	0.8553
0.5	0.8074	0.7245	0.704	0.8474	0.5816	0.9224	0.9032	0.8589
0.55	0.8074	0.7238	0.7087	0.8474	0.5794	0.9224	0.9039	0.8589
0.6	0.8074	0.7239	0.7141	0.8488	0.5868	0.9216	0.904	0.8589
0.65	0.8074	0.7248	0.7158	0.8497	0.5891	0.9203	0.9046	0.8589
0.7	0.8074	0.7293	0.7166	0.8504	0.5884	0.9201	0.9031	0.8594
0.75	0.8074	0.7294	0.7169	0.8508	0.5885	0.9158	0.901	0.8594
0.8	0.8074	0.7294	0.7197	0.8508	0.5894	0.9139	0.8998	0.8594
0.85	0.8074	0.7294	0.7193	0.8524	0.5891	0.9057	0.8982	0.8599
0.9	0.8074	0.7305	0.7192	0.8538	0.5891	0.8961	0.8937	0.8599
0.95	0.8074	0.7305	0.7192	0.8538	0.5918	0.8849	0.8939	0.8599
1.0	0.8074	0.7305	0.7301	0.8538	0.5917	0.8745	0.8925	0.8599

III. 推薦門檻 Min_suggest_score

前段提到最終推薦分數 Final_score 構成成分比例之調整，本推薦模組在計算出最終推薦分數 Final_score 之後，會比對推薦門檻 Min_suggest_score，若最終推薦分數 Final_score 大於推薦分門檻 Min_suggest_score，則推薦該則新聞，反之則反。在此，我們分析最小推薦分數 Min_suggest_score 的數值對推薦模型的推薦結果之影響。圖 17 及表 23 顯示不同的 Min_suggest_score 值對於推薦模型的 f-measure 之影響。

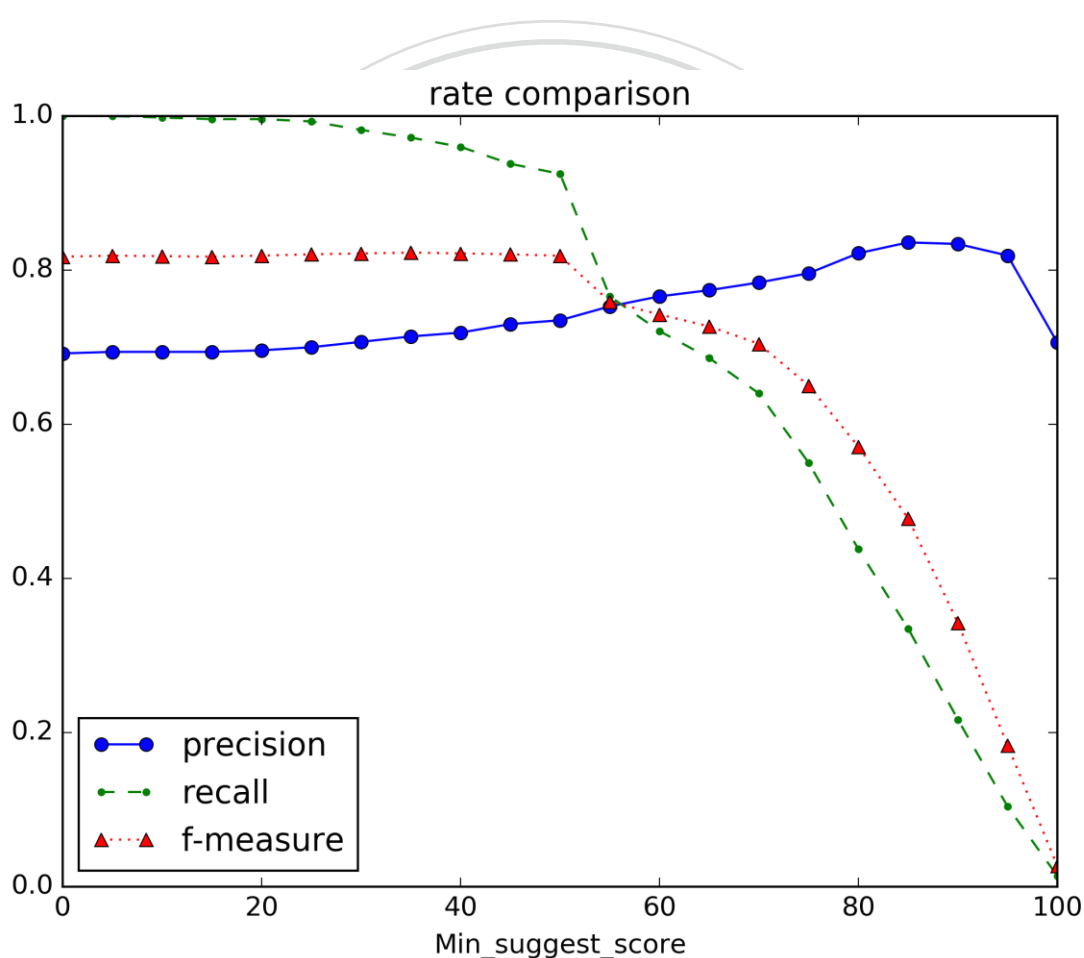


圖 17、推薦門檻 Min_suggest_score 值之變化影響圖

表 23、推薦門檻 Min_suggest_score 值之變化影響表

	precision	recall	f-measure
0.0	0.692	1.0	0.8175
5.0	0.694	1.0	0.8188
10.0	0.694	0.998	0.8182
15.0	0.694	0.996	0.8175
20.0	0.696	0.996	0.8189
25.0	0.7	0.993	0.8206
30.0	0.707	0.982	0.8217
35.0	0.714	0.972	0.8228
40.0	0.719	0.96	0.8218
45.0	0.73	0.938	0.8206
50.0	0.735	0.925	0.8187
55.0	0.753	0.766	0.7591
60.0	0.766	0.721	0.7425
65.0	0.774	0.686	0.7271
70.0	0.784	0.64	0.7045
75.0	0.796	0.55	0.6501
80.0	0.822	0.438	0.5711
85.0	0.836	0.335	0.4778
90.0	0.834	0.217	0.3426
95.0	0.819	0.104	0.1834
100.0	0.706	0.014	0.0271

在最小推薦分數 Min_suggest_score 小於 85 的狀況之下，數值越高時，推薦模型會有越高的 precision，但相對也會降低 recall 的值。實際運用此模型進行推薦時可以評估需求來調整最小推薦分數 Min_suggest_score，若是想要盡可能的推薦新聞且推薦錯誤較不會影響使用者的情況之下，可以將最小推薦分數 Min_suggest_score 設定較低的數值，反之，若是更重視推薦準確度以及希望降低錯誤推薦率，則可以將其設為較高的數值。本實驗重視 precision 及 recall 之平衡，使推薦之結果具有一定的準確性同時亦可推薦出大部分使用者會喜歡之內容，因此選擇有較佳 f-measure 表現之推薦門檻，圖 18 及表 24 顯示各個決策樹於不同推薦門檻值之 f-measure 表現。

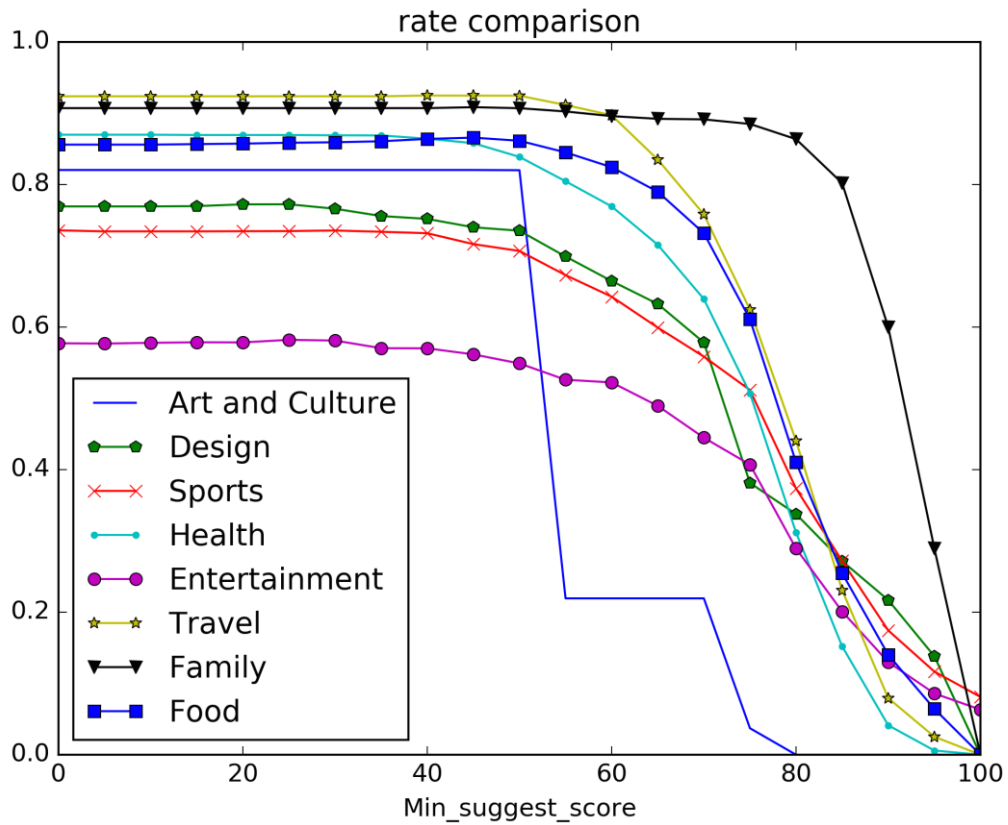


圖 18、推薦門檻 Min_suggest_score 值對各類別新聞之變化影響圖

表 24、推薦門檻 Min_suggest_score 值對各類別新聞之變化影響表

	Art and Culture	Design	Sports	Health	Entertainment	Travel	Family	Food
0.0	0.8204	0.7693	0.7357	0.8699	0.5773	0.9238	0.9072	0.856
5.0	0.8204	0.7693	0.7343	0.8699	0.5769	0.9238	0.9072	0.856
10.0	0.8204	0.7693	0.7343	0.8699	0.5779	0.9238	0.9072	0.856
15.0	0.8204	0.7697	0.7343	0.8695	0.5787	0.9238	0.9072	0.8567
20.0	0.8204	0.7722	0.7345	0.8695	0.5785	0.9238	0.9072	0.8574
25.0	0.8204	0.7723	0.7347	0.8695	0.5821	0.9238	0.9072	0.8586
30.0	0.8204	0.766	0.7355	0.8691	0.5811	0.9238	0.9072	0.8593
35.0	0.8204	0.7558	0.7337	0.8689	0.5704	0.9238	0.9072	0.8607
40.0	0.8204	0.7518	0.7317	0.8643	0.5702	0.9249	0.9072	0.864
45.0	0.8204	0.74	0.7164	0.858	0.5617	0.9247	0.9085	0.8658
50.0	0.8201	0.7352	0.7068	0.8388	0.549	0.9246	0.9071	0.8614
55.0	0.2194	0.699	0.6729	0.8046	0.5261	0.9116	0.9025	0.8451
60.0	0.2194	0.6646	0.6427	0.7693	0.5224	0.8972	0.896	0.8243
65.0	0.2194	0.6325	0.5993	0.7155	0.4895	0.835	0.8922	0.7899
70.0	0.2194	0.5783	0.5586	0.6394	0.4448	0.7582	0.8914	0.7317
75.0	0.0372	0.3813	0.5117	0.5071	0.4071	0.6246	0.885	0.611
80.0	0.0	0.3377	0.3735	0.3118	0.2893	0.4404	0.8637	0.4103
85.0	0.0	0.2717	0.2719	0.1521	0.201	0.2309	0.802	0.2548
90.0	0.0	0.217	0.1744	0.0413	0.13	0.0794	0.5998	0.1402
95.0	0.0	0.1382	0.1167	0.0059	0.086	0.0253	0.2896	0.0644
100.0	0.0	0.0	0.0811	0.0	0.0633	0.0	0.0	0.0

本研究希望推薦模型能同時兼具精確率以及召回率的數值，如此讓使用者能夠同時取得大部分其有興趣之內容，亦可以不收到過多不喜歡的內容。為達此目標，我們使用 f-measure 作為調整參數值之標準。

另外，本研究在實驗中發現，各類別新聞彼此之間對於參數的配置並無一致性，同樣的參數配置並不會對所有類別的推薦都有最佳之 f-measure 值。為了使推薦模組能夠有最好的推薦結果，本實驗對所有類別之停止分割純度門檻、Min_node_purity 最終推薦分數之權重分配係數 α 、推薦門檻 Min_suggest_score，分別取得最佳 f-measure 值之組合。以此組合作為實驗階段之參數組合(參數 1、2、3)。

$$\text{Min_node_purity}_i, 1 \leq i \leq 8 \quad (1)$$

$$\alpha_j, 1 \leq j \leq 8 \quad (2)$$

$$\text{Min_suggest_score}_k, 1 \leq k \leq 8 \quad (3)$$

二、實驗結果分析

I. 使用者資料樣本數之影響

本研究欲測試在分析不同的使用者行為資料樣本數的情況下，推薦模型所推薦出的結果。我們以程式隨機抽取各個使用者的資料集成四個資料集，每當完成一個資料集之後，被抽取到的資料完之後再放回有資料當中，即每個資料集本身不會有重複的使用者資料，但資料集彼此之間可能會有重覆資料，因可重複使用。四個資料集之的使用者行為資料數目分別為 50, 100, 200, 300, 432，建立完資料集之後，針對每個新聞類別本實驗皆設定上一節所求得之最佳參數組合。

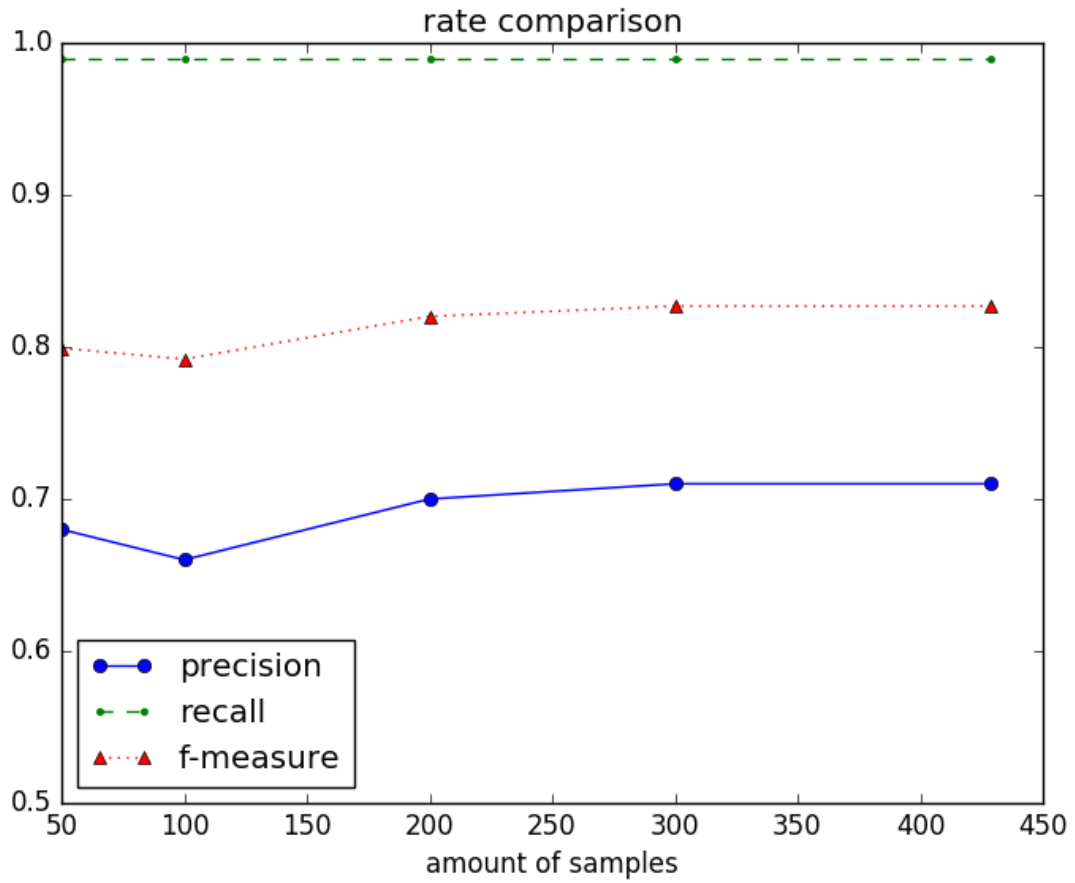


圖 19、推薦模型對不同數量測試資料之推薦表現圖

表 25、推薦模型對不同數量測試資料之推薦表現

	precision	recall	f-measure
50	0.68	0.99	0.7992
100	0.66	0.99	0.792
200	0.7	0.99	0.8201
300	0.71	0.99	0.8269
429	0.71	0.99	0.8269

由圖 19 可得知，雖然本實驗建構之推薦模型在對於各數量之樣本進行推薦時，推薦結果之 f-measure 略有不同，但 f-measure 均可達到大約 0.8。代表本實驗正確推薦比率以及召回率可以同時達到一個不錯比例，能同時對使用者進行準確的推薦並推薦出大部分使用者會喜歡之新聞內容。此數據結果也顯示本推薦模型有足夠的穩定度以及推薦的準確性。

II. 各決策樹之推薦表現

在進行實驗過程當中，本研究發現在推薦不同類別的新聞時，每種新聞類別有其各自最適合的分數配置比例，各類別彼此的參數配置不盡相同。舉例來說，在藝術與文化(Art and Culture)類別當中，透過決策樹結果進行推薦可以有較好的 f-measure 值，然而在設計(Design)類別當中，單純透過新聞的體驗進行推薦則會有較佳的 f-measure 值。由此可知，雖是同樣的推薦模型但是對各個新聞類別之表現會略有不同。圖 20 及表 26 顯示本實驗針對不同的新聞類別進行不同的參數配置，所得的各類別新聞推薦結果。

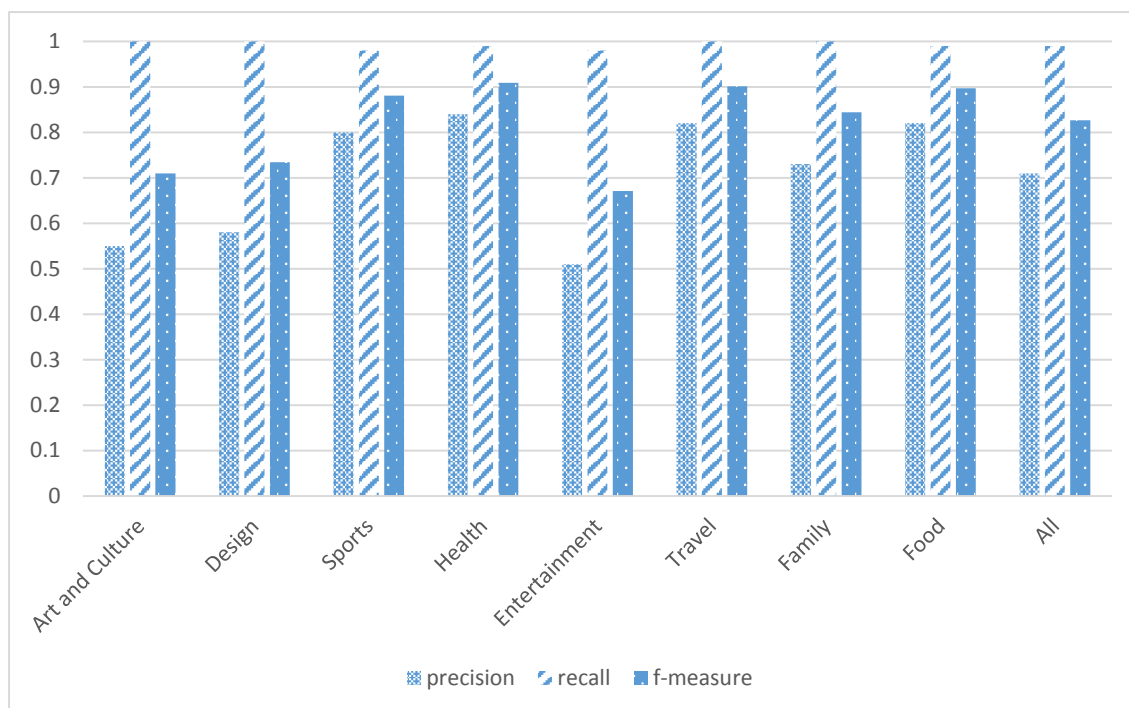


圖 20、推薦模型對不同類別新聞之推薦表現圖

表 26、推薦模型對不同類別新聞之推薦表現表

	Art and Culture	Design	Sports	Health	Entertainment	Travel	Family	Food	All
precision	0.55	0.58	0.8	0.84	0.51	0.82	0.73	0.82	0.71
recall	1	1	0.98	0.99	0.98	1	1	0.99	0.99
f-measure	0.7097	0.7342	0.8809	0.9089	0.6709	0.9011	0.8439	0.897	0.8269

在八種新聞類別的推薦中，本實驗模型在體育(Sports)、健康(Health)、旅遊(Travel)、親子(Family)、以及美食(Food)類別的新聞推薦皆能得到超過 0.84 的 f-measure 值，藉此結果我們可了解到以使用者喜歡之體驗作為推薦的依據，能夠有效的推薦合適的新聞給予使用者。

而在文化藝術(Art and Culture)、設計時尚(Design)、影視娛樂(Entertainment)類別本推薦模組則相對表現得較不像上述五類之新聞卓越，本推薦模型推薦此三類新聞的 f-measure 值分別為 0.7097、0.7315 以及 0.6577。本研究推測，在這三類

的新聞當中，在影響使用者喜歡新聞與否之因素中，體驗占一定的因素，但除了新聞的體驗，新聞主題層面的影響較其他類別之新聞更大。本研究希望往後之就加入主題層面的推薦參數以強化此部分之推薦。



第五章 結論與未來展望

由於過去以來電子新聞之推薦方法大多是透過使用者點擊過的內容推薦類似領域的電子新聞，或是透過協同篩選(collaborative filtering)推薦相同興趣使用者也觀看過之內容，並無演算法根據使用者喜歡的體驗類型進行推薦。

本研究建立一個考慮使用者體驗喜好之電子新聞推薦模型。我們以 Schmitt 提出之策略體驗模組為基礎了解使用者對各體驗之重視程度，分析使用者對各種不同型式體驗之重視程度以作為決策樹資料探勘的輸入屬性，並以消費者對於電子新聞的喜好與否作為目標屬性，利用決策樹演算法計算這些輸入屬性(使用者對各種不同型式體驗之喜好)與目標屬性(使用者對於電子新聞的選擇)之間的關聯式規則。接著利用這些規則來建構一個預測模型，以評估閱聽人對於未知電子新聞的接受程度，從而建立一個能有效符合使用者個人體驗喜好之新聞推薦模型。

新聞的推薦可以從很多個面向進行分析並推薦，本研究是以體驗為出發點進行推薦，然而在實驗的樣本當中，有些受測者的資料反應出，其喜歡新聞與否並非完全只和新聞的體驗有相關，或許是與其內容本身有相關而影響受測者之喜好程度，在未來，可以延續本研究之議題加入不同的新聞分析面向以優化此模組之推薦準確度。如同上述，本研究提出之推薦模型尚可以朝一些方向延續本研究之研究議題：

I. 新聞分類判斷自動化

目前本研究所提出之模型針對新聞類別以及新聞含有之體驗的判斷方法為監督式學習，未來希望完全使用機器進行判斷。目前有許多相關的文字探勘技術，可以透過關鍵字之方法進行判斷，亦可透過 clustering 之方法進行類似類別或類似體驗之新聞比對。

II. 不同年齡層不同族群之模型建立

目前本研究鎖定之族群為大學生以及研究生，未來可以針對不同年齡層，不同職業，不同背景等等面向都建立推薦之模型

III. 加入不同之分析元素

本研究主要以體驗為出發點進行推薦，在未來，判斷推薦與否的分數組成可以加入其他的面向進行整合，例如以內容為基礎進行推薦(content based recommendation)或是協同過濾等等之方法以提升推薦模型之準確度以及適用的廣泛程度。



參考文獻

- [1] M. Brandt. 2014 ,Apr 23. *Mobile is the New Media Star* [Online]. Available:<http://www.statista.com/chart/2168/share-of-time-spent-per-day-with-major-media/>
- [2] *Internet Overtakes Newspaper As Newsoutlet* [Online]. Available:<http://www.people-press.org/2008/12/23/internet-overtakes-newspapers-as-news-outlet/>
- [3] *173 Million Adults Engaged with Newspaper Digital Content in January* [Online]. Available:<http://mediamanagersclub.org/173-million-adults-engaged-newspaper-digital-content-january-naa>
- [4] Resnick, P., N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In Proceedings of the 1994 Computer Supported Cooperative Work Conference, 1994.
- [5] Kompan, M., Bieliková, M., 2010. Content-Based News Recommendation. In Proc. of the 11th Conf. EC-WEB, Springer, 61-72.
- [6] F. Frasinca, J. Borsje, and L. Levering. A Semantic Web-Based Approach for Building Personalized News Services. International Journal of E-Business Research, 5(3):35–53, 2009.
- [7] Intema W, Goossen F, Frasinca F, Hogenboom F. Ontology-based news recommendation. In: Proc 2010 EDBT/ICDT Workshops, Lausanne, Switzerland; 2010. pp 1–6.
- [8] Konstan, J. A., B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. GroupLens: Applying collaborative filtering to Usenet news. Communications of the ACM, 40(3):77-87, 1997.
- [9] A. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable

online collaborative filtering. In Proc. of the 16th International World Wide Web Conf., 2007.

[10] Liu, J., Dolan, P., Pedersen, E.R.: Personalized news recommendation based on click behavior. In Rich et al., eds.: Proc. of 14th Int. Conf. on Intelligent User Interfaces (IUI), ACM (2010) 31–40

[11] Pine, B.J., and Gilmore, J.H. The Experience Economy. Boston: Harvard Business School Press, 2011.

[12] Alex Simonson, Bernd H. Schmitt. Marketing Aesthetics: The Strategic Management of Brands, Identity and Image[M]. Free Press, August 30, 1997

[13] Schmitt, B.H. (1999). Experiential Marketing: How to Get Customers to Sense, Feel, Think, Act, and Relate to Your Company and Brands. New York: Free Press.

[14] Tan, P.-N., Steinbach, M., and Kumar, V. 2005. Introduction to Data Mining. Addison-Wesley.

[15] Han, Jiawei, Kamber, Micheline, 2000. Data Mining: Concepts and Techniques. Morgan Kaufmann.

[16] I.H. Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, 2011.

[17] Quinlan, J.R. (1986). Induction of decision trees. Machine Learning, 1, 81-106.

[18] Clair, D. C. St., Sabharwal, C. L. and Hacke, K. R. ,“Formation of clusters and resolution of ordinal attributes in ID3 classification trees”,Proc. of ACM/SIGAPP Symposium on Applied Computing: Technological Challenges of the 1990’s, pp.590-597, 1992.

[19] Patterson,T.E.2000. Doing Well and Doing Good:How Soft News and Critical

Journalism Are Shrinking the News Audience and Weakening Democracy—And What News Outlets Can Do about It. Cambridge, MA: The Joan Shorenstein Center for Press, Politics, and Public Policy at Harvard University

[20] Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39,31-36.

[21] F. Ricci, L. Rokach, B. Shapira, P. B. Kantor , *Recommender Systems Handbook*, Springer, 2011

[22] Gunawardana, A., Shani, G.: A survey of accuracy evaluation metrics of recommendation tasks. *J. Mach. Learn. Res.* 10, 2935–2962 (2009)

