

楊亨利、林青峰（2020），『針對情感商品的推薦機制－以流行音樂為例』，*中華民國資訊管理學報*，第二十七卷，第二期，頁 175-204。

## 針對情感商品的推薦機制－以流行音樂為例

楊亨利\*

國立政治大學資訊管理學系

林青峰

國立政治大學資訊管理學系

### 摘要

情感商品，如音樂、電影等，與一般單純為了使用功能的功能商品有很大的不同。因為情感商品的評價與個人感受有關，情感商品在網路上通常會存在比較多主觀的評論；商品的效用也更與商品本身內容及通常能帶給使用者什麼感覺與情緒來的有關。傳統上，對於網路評論，我們通常只關注評論中所述及的商品屬性，主要在找正負傾向規則，而不會去企圖找出像是「聽了讓人感到很遺憾」這種引發人類情緒的情感商品規則。本研究以流行音樂這個情感商品為例，提出一個針對情感商品的推薦機制。首先我們先建立能了解網路評論狀況的情感標籤分類器，用於隨時了解某商品目前網路評論的情感傾向；另外也建立一個同時考慮到音樂歌詞及音質特性的音樂內容分類器，用於從音樂的內容特徵來得到某音樂商品可能音樂情感傾向。經過資料的收集、分析與訓練，網路評論分類器與音樂內容分類器的精準率、召回率與 F1 均達令人滿意程度，進而本研究以實驗分析在用戶悲傷情緒下應推薦的音樂來說明情感商品的推薦規則建立過程。

**關鍵詞：**情感分析、流行音樂、意見挖掘、網路評論、推薦規則

---

\* 本文通訊作者。電子郵件信箱：yanh@nccu.edu.tw  
2019/11/29 投稿；2020/01/19 修訂；2020/02/20 接受

Yang, H.L. and Lin, Q.F. (2020), 'Recommended mechanism for hedonic products-Taking pop music as an example', *Journal of Information Management*, Vol. 27, No. 2, pp. 175-204.

## Recommended Mechanism for Hedonic Products-Taking Pop Music as an Example

Heng-Li Yang\*

Department of Management Information Systems, National Cheng-Chi University

Qing-Feng Lin

Department of Management Information Systems, National Cheng-Chi University

### Abstract

**Purpose**—This study aims to propose a mechanism based on web reviews opinion mining and product contents (e.g., audio and lyrics in our case) for hedonic product recommendation.

**Design/methodology/approach** — The classifiers, web review SVM classifiers and music content SVM classifiers, were proposed and a prototype was also built. Finally, we designed an experiment for exemplifying the process of determining the recommended product when the user is in a particular mood.

**Findings**—The acceptable precision, recall, F1 ratio were obtained for the two classifiers. The experiment indicated the recommendation rule while users are in sad mood.

**Research limitations/implications**—We only take as an example of pop music. Other hedonic products (e.g., dancing) might be more complicated to analyze their contents owing to video.

**Practical implications**—Following our proposed mechanism, the suppliers of hedonic products would know how to recommend proper contents to users to invoke

---

\* Corresponding author. Email: [yanh@nccu.edu.tw](mailto:yanh@nccu.edu.tw)  
2019/11/29 received; 2020/01/19 revised; 2020/02/20 accepted

their desirable feelings.

**Originality/value**— The proposed mechanism is brand new. As we know, there is no such a recommended mechanism for hedonic product in literature.

**Keywords:** sentiment analysis, pop music, opinion mining, internet review, recommendation mechanism

## 壹、緒論

推薦系統在企業的應用已有多多年，不過，一般是用來推薦功能性商品。Chaudhuri 與 Holbrook (2001) 提出商品的價值可分為實用性價值 (utilitarian value) 與享樂性價值 (hedonic value)。實用性價值被認為是任務性及理性的，使用者滿足了其對商品本身的功能或效用上的需要；而享樂性價值被認為是較主觀且個人化的，使用者享有情感、美感或其它感官上的愉悅、幻想的感覺體驗。依上述二種價值的含量多寡，商品可區分成享樂價值高的情感商品 (hedonic product) 與實用價值高的功能商品 (functional product) 二類 (Kempf 1999)。人們購買情感商品目的是得到某種情感，這種情感的製造者就是人們自己，這類商品可能是有形的，也可能是無形的，如：小說、音樂、影片、戲劇、舞蹈、繪畫、藝術陳列品等。

過去對功能商品如日常用品、3C 用品、教科書等推薦是鎖定在與屬性相關的正負傾向意見上，如「iphone8 的畫質普遍評價為佳」。但對以享樂性價值為主的情感商品來說，只找出與屬性相關正負評價（如「此電影的男主角演技很好」）這樣是不夠的。情感商品的使用者會比使用功能商品更重視商品本身的內容所能帶來他的感受。而情感商品的網路評論也通常隨著評論者自己的偏好、生活體驗等原因而呈現出其個人主觀感受。是以，想要有更好的支援此類商品的推薦，需要有針對情緒更精細的處理方法。本研究以華語流行音樂為例，提出一個支援情感商品選擇特性的商品推薦機制，這機制會同時用到商品內容的資訊（以我們的例子而言，是音樂音訊、歌詞）以及網路評價來建立分類器。而相對功能商品，我們的分類器希望能分出該商品設計時作者希望傳達給使用者的感受，也就是激起其情緒變化，同時也希望掌握網路評價者在寫評論時，其情緒變化。最後，我們也設計一個實驗來測試當使用者處於某情緒（如悲傷）的情境下時，我們的機制是否能適當的推薦情感商品（如音樂）給他們，以此例來說明對情感商品較恰當的推薦作法。

過去研究者（如 Hu & Downie 2010; Hu & Yang 2017; Wang & Yang 2019）已在音樂所蘊含的情感辨識方法上，有不少突破。本研究的重點不在其辨識技術方法的改進，我們仍是採用傳統的方法，而是提出一個完整的推薦機制，同時考量網路評論、音訊與歌詞三部分，也就是強調情感商品的情感辨識應同時考量到商品內容的資訊（以我們的例子而言，是音樂音訊、歌詞）以及網路評價，以及其推薦程序應先經過一連串的情緒實驗，方能有合適的推薦。

接下來，本文將在第貳節探討相關文獻、第參節提出我們建議的推薦機制、第肆節說明該機制下的情感標籤分類器的建置、第伍節描述我們透過情緒音樂實

驗來找出悲傷的情境下，適當的音樂推薦，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) 的研究是收集對某個商品或服務的評價文本資料，針對此商品或服務的使用意見利用機器學習的方法找出正負評價或感覺激發狀況的分類方法 (Pang & Lee 2008)。現今意見挖掘技術已有很廣泛的應用，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網友評價的文本，如電子產品 (Turney & Littman 2003)、電影 (Ye et al. 2006)、餐廳評價 (Yan et al. 2015)、手機 (楊亨利 & 林青峰 2018)。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 (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。

Esuli 與 Sebastiani (2006) 依分析產出將情感分析的相關研究區分成三類：(1)決定文字主客觀的分類、(2)決定文字正負傾向的分類、及(3)決定文字正負傾向強度的分類。進一步 Pang 與 Lee (2008) 更依分析的產出更細的區分出以下主要類別，分別是：(1)情感極度及強度的分類：不一定是正負，特定感受的有無，可能或不可能的分類研究都算此類，例如從選舉討論區中將意見區分成可能會贏或不可能會贏二個類別。(2)主客觀偵測及意見辨視、(3)關聯性的分類：例如說從醫學文本中的資料試著去分類出新病患的可能結果、(4)評論星級的推論、(5)相同意見監測的分類：從二個文本當中，找出相同或不同意見的分類研究、(6)通篇文章的主題偵測、(7)看法與觀點分類研究：將觀點分群，例如說將某個議題的觀點區分為保守派或激進派兩群。(8)類別分類研究：例如對某個文本，分類器可依內容將其分類為公告或是廣告。及(9)出處分類研究：依文本的內容，分析其出自何人之手。本研究則屬利用進行情感極度與強度的分類。

意見挖掘的文獻中常用到的計算方法，大概可區分為字典法、知識本體法、與機器學習法等演算方法。字典法是利用專家事先已定義好的情感字典，經由分析句子或文章與這些情感字的關係，如在語類庫中共同出現的次數，來決定評價的意見傾向 (Wiebe et al. 2005)。與字典法類似的，知識本體法同樣也是利用專家事先建立好的領域知識本體，如 ConceptNet、HowNet、SenticNet，來進行節點

(字)與情感字之間關係的計算,但與字典法不同的是本體法二個節點的關係計算通常是用知識本體的推論引擎計算而來的(Liu & Singh 2004)。另外有一些研究則採用統計方式來計算出字詞可能的意見傾向,如 Turney 與 Littman (2003)先定義一組正負情緒字詞(如 good, bad, unfortunate),再利用 PMI (pointwise mutual information) 和 LSA (latent semantic analysis) 的統計方法計算字與正反面詞之間的關係來決定該字的可能意見傾向。

而機器學習式的分析法則不需要預先定義的字典;經由輸入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(Li & Wu 2010)。機器學習式法還可區分有監督式(supervised)及非監督式(unsupervised)二類的方法。非監督式的學習法的訓練資料並不用經由專家評等分類,而是讓演算法自行摸索出資料之間的規律,如集群(clustering)演算法。而監督式的學習法是由已經被專家標記好結果的訓練資料開始進行訓練,雖然這種方法需要較大量的資源來處理訓練資料,但也因為有訓練目標通常訓練完成的分類器準確度較高(Kontopoulos et al. 2013)。機器學習式的意見分析分類器因為是訓練出來的,分類器常常能準確的說出分類應是什麼,但對於為什麼如此分類,並沒有一個較明顯的規則可以做為證據。機器學習法應用在意見分析上較常見的核心演算法有支援向量機(support vector machine; SVM)、簡易貝式法(naïve Bayes; NB)、最大熵值法(maximum entropy)、深度類神經網路(deep neural networks; DNN)、自適應增強法(adaptive boosting)、邏輯迴歸法(logistic regression; LR)、最近鄰居法(k-nearest neighbors; KNN)等方法。

本研究在兩分類器的資料匯總過程採用了 ConceptNet 的技術去計算情緒關鍵字與基本情緒之間關係強度來降低維度。而我們分類器的訓練則是採用監督式 SVM 機器學習法。

## 二、音樂的意見挖掘

Hu 等(2005)利用簡易貝式法去分析網路音樂評論,從評論文字中找出音樂的分類及預測評論者給此音樂的星級。Zhuang 等(2006)利用自然語言處理法去分析電影評論,考慮如劇情、視覺效果、特效、導演、演員等電影屬性,可以找出如「劇情很簡單」、「演員很棒」此類的規則。Oramas 等(2016)利用自然語言處理的方法分析大量的音樂網路評論資料,試著利用評論來判斷音樂類別分類,他們發現音樂評論的極性與時間是會不斷的變化的。這些研究都不能如本研究找出網路評價的隱含情緒。

Kumar 與 Minz (2013) 試著利用 SentiWordNet 計算已被標記成喜悅、生

氣、愛或難過四類的音樂歌詞中單詞的正負面相關分數再匯總成單首歌的特徵分數，利用特徵分數再去進行各種技術的分類器的訓練與正確率比較。Corona 與 O'Mahony (2015) 利用向量空間模式來表達每首歌歌詞中出現的情緒字特性，再利用 Last.fm 的網路電台和音樂社群中的情緒標註資料來進行各種分類方法的比較。Sharma 等 (2016) 將歌詞進行詞性標註後，再利用 SentiWordNet 計算關聯強度最後利用支援向量機的方法來分類音樂是否適合聽眾。這些容或考慮了音樂中填詞者對歌詞所想傳達的情感，但並未同時考量作曲者付諸音訊內容的情感。另方面，Gómez 與 Cáceres (2017) 則是只考慮音訊內容，而且只侷限於 16 種音色 (Timbre) 的音訊特徵做為輸入資料來作音樂分類，反而沒考慮歌詞所傳達的情感，無法處理同曲不同詞。

Hu 與 Downie (2010) 提出了一個同時考慮音訊內容及歌詞的 SVM 分類系統來同時兼顧音訊與歌詞的音樂內容情感分類。Hu 與 Yang (2017) 探索華語音樂情感以分類模式與二維 (正負評價、喚醒強度) 維度模式表達間的關係，以及作華語音樂與西方音樂的情感分布空間的比較，不過其資料只考量音訊的部份。近年來，Wang 與 Yang (2019) 則提出了一個深度學習法去處理中文音樂歌詞上的情感分類，並顯示其相較與一般機器學習方法及類神經網路法有較好的成果，不過其仍只專注於歌詞部分。相較這些研究，本研究則同時考量網路評論、音訊與歌詞三個部分。

### 三、音訊內容特徵

經過多年的發展，已有許多的音樂特徵被作為音樂類別分類之用。如 Tzanetakis 等 (2002) 利用音譜中心度 (spectral centroid)、音譜滾降率 (spectral rolloff)、音譜通量 (spectral flux)、過零率 (ZCR)、梅爾倒頻譜係數 (MFCC)、低強度率 (low energy rate) 等音樂特徵來進行音樂分類。Xu 等 (2003) 利用萃取網路上收集到的音樂中節奏譜 (beat spectrum)、過零率、短時間強度和梅爾倒頻譜係數等特徵，使用 SVM 的方法訓練一個多層次的分類器。Patel 與 Trivedi (2017) 則是使用了如方均根強度 (RMS)、平均強度、低強度率、節拍數、最強節拍強度、擾動值、快速傅立葉轉換係數 (FFTC)、梅爾倒頻譜係數、過零率等特徵來進行音樂分類。

本研究在選擇要萃取的音樂特徵做為分類參數時，是計算參考文獻中各音樂特徵的被選用次數，再從每個特徵分類中選取幾個較多學者採用的特徵做為本研究要萃取的音樂特徵。

#### 四、音樂推薦系統

早期學者們（如 Lee et al. 2010）應用協同推薦技術在音樂推薦上，但因常被諸如「新歌不會被推薦」、「沒有偏好的人不知如何推薦」的冷啟動問題影響（Schein et al. 2002），近期的文獻較常出現內容導向推薦系統或是混合式推薦系統。如 Chen 與 Chen（2001）使用二階段的方法來進行推薦：第一階段先對音樂進行分群與標籤標註，第二階段利用用戶過去的偏好資料。

Rho 等（2009）利用支援向量迴歸（support vector regression）來作心情分類器，同時利用協同過濾及知識本體的技術來分析用戶的心情之後再進行推薦。Van den Oord 等（2013）使用潛在因子模型（latent factor model）及深度學習法來改善新歌或已不受歡迎的歌不會被推薦的冷啟動現象。Wang 與 Wang（2014）也提出了一個利用深度信念網路（deep belief networks）及概率圖模型的音樂推薦系統，試圖改良傳統的二階段內容導向音樂推薦作法。

Celma（2006）利用朋友的朋友們（friends of friend）對朋友的留言與描述做為該朋友的側寫，並同時從音樂相關的 RSS 資訊及音檔本身的特性來進行音樂推薦。另外 Bu 等（2010）利用傳統的音檔特徵資料及使用超圖（hyper graph）來處理 Last.fm 中諸如朋友關係、會員關係、收聽關係、標籤關係等音樂社群資料來進行音樂推薦。他們分析的社群資料主要是為了進行用戶相似度計算，與本研究將網路評論做為音樂網路評論標籤的來源不同。

而 Nanopoulos 等（2010）也試著從 Last.fm 的社群標籤（social tagging）中得到用戶對音樂所加上的如音樂類別、風格、心情、用戶意見、樂器等多面向的標籤資訊；並且建議採用用戶、標籤及音訊特徵三維度的資料進行模型化，找出三個維度間的關係後再進行音樂推薦。其只考慮音訊特徵，並未同時考量歌詞內容，而且經由社群標籤常會發生資料稀疏的現象。

#### 參、本研究提出之推薦機制

本研究建議的情感商品推薦機制如圖 1。圖 1 中兩大核心是建立情感分類器以及以實驗確定適合情境的推薦情感商品。分類器有兩方面，一個是「網路評論分類器」，期待有效了解目前網路評論意見傾向，當輸入某一篇評論的評論關鍵字矩陣輸入某個意見分類器後，即能指出該評論的特定情感傾向。另一個是「商品內容分類器」，期待瞭解該商品所傳遞的特定情感傾向。這不同於功能商品的規格，我們重視的是其試圖引起使用者的情緒。以本研究的試作對象音樂來說，這包含歌詞、音訊兩者；如以舞蹈表演來說，則包含視訊。然而監督式的學習法除非已有正負評價分數（如雅虎電影之星等評價），否則訓練資料需先由專家分類正負，而且更進一步，我們在意的是「蘊含的情感」，所以需要專家標記網路



評論及商品內容是否會激發各種情緒。此外，當然，不管是網路評論、歌詞、音樂資料檔案，均會需要如資料清洗、資料格式化、音訊轉檔等前處理工作。而且，網路評論是與日俱增的巨量，即使我們有此資訊處理能力，也需先檢視各評論的代表性，例如有 1000 人按讚的評論，其代表性應比只有 1 人按讚甚至 100 人表達不贊成有其代表性。最後，透過「網路評論分類器」與「商品內容分類器」，我們才可真正瞭解該情感商品所試圖傳遞的情感，此二分類器結果也有可能不見得一致，如一首歌本來作曲、填詞者試圖傳遞出快樂的情緒，但網路上卻不認為如此。確定某商品所試圖傳遞的情感後，何時該推薦此商品，則需一連串實驗確認。比如說，悲傷情境下，該聽何種歌？我們建議透過一連串實驗來確定其推薦規則。

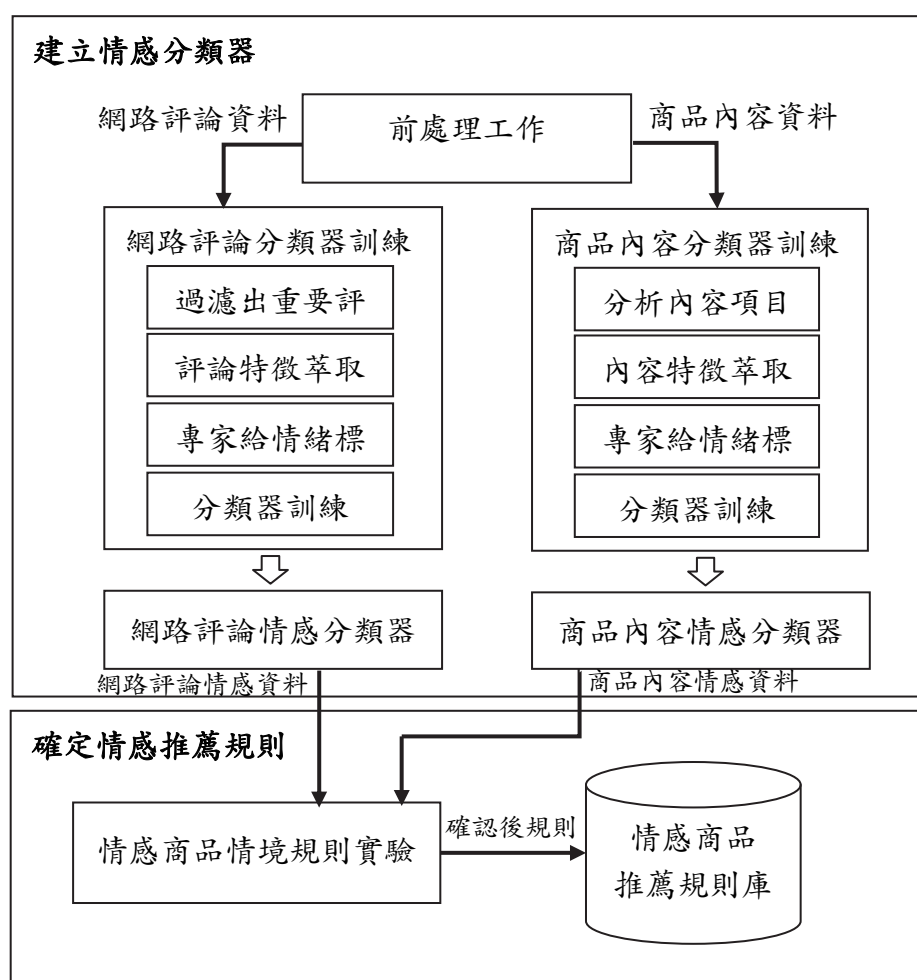


圖 1：建議的情感商品推薦機制

## 肆、華語流行音樂情感標籤分類器的建置

接下來，我們以華語流行音樂為例，來先說明上述分類器的建置方式。

### 一、網路評論及音樂內容資料的收集與前處理

前處理主要是要能完成自動化取得、清洗、格式化研究分析所需的相關資料。可分為音樂歌曲檔及音樂 meta data 主檔的取得、音樂評論的取得、動態歌詞的取得幾個項目。

音樂歌曲檔為著作版權物，惟為研究正當目的之必要，亦得引用已公開發表之著作。本研究從公開的 Youtube 網站上選取了 320 首華語歌曲，並轉換成 mp3 音樂檔，以供本研究的學術分析使用。除了取得音樂檔之外，為了選到音樂的相關的 meta data 資訊，本研究利用自製的爬蟲程式，從蝦米音樂網<sup>1</sup>抓取一些基本資料欄位如：歌名（title）、表演者（artists）、專輯名稱（album）、專輯封面圖檔（cover image）、作詞者（lyrics writer）、作曲者（composer）、編曲者（arranger）等。這些基本資料是為了以後建立推薦系統時，提供給使用者「自己找歌」的搜尋功能所需。在處理音樂主檔的時候，爬蟲系統也會將每首歌的評論首頁網址儲存下來。之後利用這個網址爬蟲系統可以從第一頁開始將所有的評論一一儲存下來。蝦米音樂網中的一個完整的評論資料包含了評論 ID（PK）、用戶 ID、用戶名（評論時的用戶名）、評論歌曲、評論時間、評論內容、按讚數（其它用戶認同的數量）、按弱數（其它用戶不認同的數量）及用戶使用的設備（從 iPhone、從 Android 來的用戶）等資料。針對此 320 首華文流行歌曲，共取得了 483,427 筆評論。

在蝦米音樂網站中，本來也有歌曲歌詞的資料，但並沒有動態歌詞資料。所以本研究的動態歌詞資料是由另一個繁體中文的魔鏡歌詞網 Mojim.com 中取得。所謂的動態歌詞是除了原有的歌詞之外還會增加標註每一行歌詞在歌曲中出現時間到百分之一秒（如[00:39.39]趕走了我的藍色星期一）。但是，魔鏡歌詞網並不是所有的歌曲都有動態歌詞，對於沒有動態歌詞的歌曲，則須抓取其一般歌詞，再利用本研究自行開發的動態歌詞編輯器以人工編輯為動態歌詞。在經過人工動態化及比對工作後，本研究共從目標的 320 首歌中，取得了 14,234 句的歌詞資料。

---

1 蝦米音樂網是中國音樂網站的前趨者，它擁有的音樂資料相當全面，除了華語、外文的流行音樂外，諸如藝術、經典音樂、遊戲配樂、民謠、自由音樂人創作等都有收錄。而因為發展早，所以累積了大量的音樂評論；其內容大多是對歌曲的簡潔評價，就算有長評論，通常也是與歌曲本身極其相關的，比如創作背景、背後奇聞逸事、評論人與此曲淵源等。

爬蟲系統取得評論的同時，本研究除了採用了中文繁簡轉換的動態連接函式庫 Microsoft Visual Studio International Pack 來進行由簡體轉繁體中文的轉換工作之外，我們也同時利用中央研究院的分詞斷句系統 SINICA CKIP 所提供的 API 介面在爬蟲程式內進行中文的分詞斷句，並將分詞斷句的結果一並儲存。

## 二、音樂情緒標籤的萃取

所謂的情緒標籤指的是音樂使用後會引發收聽者感受的標籤，如喜悅、悲傷、興奮等。流行音樂會引發聽眾各種不同的獨立的、混合的情緒，本研究採用網路評論字詞分析來分析網路評論中常被使用的情緒關鍵字，以期找出可用來有效述敘流行音樂特徵的適當標籤。

我們取得的網路評論內容在所有完成分詞標註的 9,063,647 個詞當中，有 27,662 個不同的詞，其中出現超過 30 次的可能情緒字詞（名詞、動詞、形容詞）共有 2,004 個詞。再利用人工過濾的方式，過濾出可能會是情緒標籤的字詞，得到常用於評論歌曲的評論關鍵字詞共 508 個（表 1 列出了出現評論數前 20 名的字詞），我們也將這 508 個字當做候選的情緒標籤字詞。

表 1：在歌曲評論中常出現的情緒相關字詞前 20 名

| #  | 字詞 | 次數     | #  | 字詞 | 次數    |
|----|----|--------|----|----|-------|
| 1  | 喜歡 | 55,123 | 11 | 美  | 4,418 |
| 2  | 愛  | 32,794 | 12 | 笑  | 4,099 |
| 3  | 哭  | 10,575 | 13 | 開心 | 4,008 |
| 4  | 希望 | 10,391 | 14 | 謝謝 | 3,646 |
| 5  | 幸福 | 8,108  | 15 | 痛  | 3,332 |
| 6  | 快樂 | 6,482  | 16 | 懷念 | 3,292 |
| 7  | 美好 | 5,108  | 17 | 難過 | 3,213 |
| 8  | 感動 | 4,579  | 18 | 可惜 | 3,114 |
| 9  | 寂寞 | 4,541  | 19 | 溫柔 | 3,012 |
| 10 | 幸運 | 4,507  | 20 | 舒服 | 2,938 |

雖然我們已有在所有評論中出現的總次數做為關鍵字詞重要度判斷的依據，但經過了解評論的分佈狀況後，我們發現每一首歌曲擁有的評論數非常不平均。資料顯示擁有最多評論的一首歌有高達 9,795 篇評論，而最少的一首歌只有 38 篇評論，平均一首歌的評論數落在 1,510.7 篇，但標準差卻高達 2,150 篇；擁有最高

評論數的前 5% 的歌曲擁有了總評論數中 29.51% 的評論，擁有最高評論數的前 10% 的歌曲則更是擁有了總評論數中的 46.72% 評論。為了避免由超多評論數量的一群歌主導了整個情緒標籤的選擇，本研究用鎖定重要評論的方法來進行分析。在資料庫中的 320 首音樂及 483,427 篇評論裏，從每一首歌的評論中找出認同數（按讚數-按弱數）前 30 名的評論做為將要分析的評論。因為最少的一首歌也還有 38 篇評論，所以也能找出認同數前 30 名的評論。經過計算，雖然每一首歌只選擇了前 30 名較多人認同的評論來分析，但是選出的 9,600 篇評論所擁有的認同數已佔所有 48 萬多篇評論總認同數的 83.73%，我們認為這 9,600 篇評論是由用戶選出認為是較為言之有理的重要評論。

另外，為了避免分析出來的字詞會完全被熱門類型音樂（如悲傷的情歌）的評論支配，而忽略了其它較少出現但卻頗為重要的音樂類型評論所含的情緒關鍵字；所以我們採用了將音樂分群的方法來解決這個問題。

由於對歌曲可能會有的類別數量（分群數）並無法事先確定，所以本研究採用的並不是傳統的給定分群數的演算法，而是採用給定距離門檻類的 Jarvis-Patrick 分群演算法。我們用這 9,600 篇評論的情緒關鍵字將這 320 首音樂進行分群。首先計算每一首歌的 30 篇評論中候選情緒標籤的出現次數。完成這個工作後每首歌都可得一筆資料，共可得到 320 筆資料，每筆資料除了歌曲編號外都有 508 個數字欄位，分別記載著第 1~508 號候選情緒標籤字詞的出現次數，這些欄位的最大值為 30（前 30 名評論每篇都提到了這個情緒字詞），最小值為 0（前 30 名的評論都沒有提到這個情緒字詞）。完成之後，系統會兩兩計算歌曲候選字詞出現與否及出現次數的餘弦相似度（cosine similarity），每一組歌曲得到 CosSimA（只考慮是否出現）及 CosSimB（考慮出現次數）二個介於[0~1]之間的數。相似度為 0 者，代表這二首歌前 30 認同的評論出現的心情字詞沒有相同的部份；相似度為 1 者，則代表這二首歌前 30 認同的評論出現的心情字詞完全相同。另外，如果同時考量是否出現及出現次數，可以計算二個相似度的平方合的正根為 CosSimAB。本研究利用 CosSimA、CosSimB 及二個同時考慮的 CosSimAB 三組數據，使用 Jarvis-Patrick 分群方法與二組參數  $K=20, sk=12$ 、參數  $K=30, sk=17$  進行共六組不同相似度及參數的分群試驗。最後，若此 6 次試驗中有 5 次以上都被分成同組的二首歌，則被視為評論相似的同一組歌曲，而每次實驗都被分成同組的二首歌，我們也標記其為核心歌曲，最後得到了 11 組的歌曲，每一組歌曲的前 10 名心情字詞如表 2。

利用表 2，本研究從中選擇了較重要的 16 個情緒關鍵字：喜悅（開心）、悲傷（難過）、希望、失望、遺憾、孤獨、溫柔、悅耳、刺耳（吵）、興奮、平靜、

甜、苦、好笑（笑）、可愛、焦慮，做為主要流行音樂情緒標籤字。在經過與參與實驗喜好流行音樂的五位專家<sup>2</sup>討論之後，又追加了性感、憂鬱、嚴肅三個標籤字。為了更有效的區分情緒標籤字，我們參考了文獻上的分類方法。卓淑玲等（2013）將中文的情緒詞分成兩大類，一類是直接描述情緒感受的語詞（包含描述情緒的經驗、描述情緒所伴隨的動作、描述情緒所伴隨的認知狀態、描述情緒伴隨的生理感官、伴隨情緒的因應反應），如喜悅、悲傷、憤怒等，稱為情緒描述詞；另一類則是間接引發情緒感受的語詞，如體貼、優秀、失業等，稱為情緒誘發詞。所以，本研究將上述 19 個情緒標籤區分為二種，描述性情緒標籤有：喜悅、悲傷、興奮、平靜、憂鬱、焦慮、遺憾；誘導性情緒標籤有：希望、失望、孤獨、性感、好笑、嚴肅、甜、苦、溫柔、悅耳、刺耳、可愛。

表 2：歌曲各群的前 10 大情緒字詞

| 組號 | 歌數 | 核心歌數 | Top1 | Top2 | Top3 | Top4 | Top5 | Top6 | Top7 | Top8 | Top9 | Top10 |
|----|----|------|------|------|------|------|------|------|------|------|------|-------|
| 1  | 21 | 7    | 甜    | 喜歡   | 愛    | 溫馨   | 開心   | 笑    | 幸福   | 悅耳   | 無聊   | 悲傷    |
| 2  | 19 | 8    | 喜歡   | 愛    | 美    | 幸福   | 溫柔   | 希望   | 放棄   | 悲傷   | 疼    | 感謝    |
| 3  | 92 | 43   | 喜歡   | 愛    | 希望   | 哭    | 幸福   | 笑    | 孤獨   | 珍惜   | 開心   | 美好    |
| 4  | 72 | 23   | 喜歡   | 愛    | 遺憾   | 苦    | 可惜   | 難    | 孤獨   | 美好   | 悲傷   | 幸福    |
| 5  | 10 | 3    | 幸福   | 喜歡   | 愛    | 可愛   | 希望   | 關心   | 束縛   | 相戀   | 堅持   | 相信    |
| 6  | 11 | 4    | 喜歡   | 愛    | 可惜   | 哭    | 希望   | 相信   | 珍惜   | 感謝   | 平靜   | 悲傷    |
| 7  | 14 | 6    | 愛    | 喜歡   | 幸福   | 希望   | 受傷   | 孤獨   | 背叛   | 焦慮   | 心酸   | 後悔    |
| 8  | 16 | 7    | 寂寞   | 遺憾   | 愛    | 痛    | 孤獨   | 難過   | 失望   | 幸福   | 哭    | 開心    |
| 9  | 6  | 2    | 親愛   | 愛    | 幸福   | 喜歡   | 哭    | 珍惜   | 難過   | 堅定   | 希望   | 失望    |
| 10 | 5  | 2    | 喜歡   | 希望   | 悲傷   | 愛    | 感謝   | 寂寞   | 孤獨   | 心痛   | 哭    | 堅持    |
| 11 | 9  | 2    | 洗腦   | 希望   | 幸運   | 好笑   | 漂亮   | 驚艷   | 吵    | 興奮   | 嚇    | 喜歡    |

註：有 45 首歌曲無法分組，不在列表中

### 三、網路評論意見傾向分類器的訓練

由於目前各音樂網站並無給評論順便評分的做法，更無由評論者自己表達是否已被激發各種情緒，故在要進行分類器訓練前，得先取得網路評論的情緒標記作為目標資料。本研究設計了一個網路評論標記系統來取得人們認為評論作者的意見傾向。可是因為網路評論的資料量通常都很大（本研究是 483,427 筆評論），

2 這五位專家年齡在 27-42 歲間，聽熱門流行音樂至少已有 15 年。

我們採用的又是監督式學習法，每一篇收集到的網路評論都進行受測者標記在資源上是不現實且不必要的。一方面，本研究只是示範性的雛形試作，我們無法對近五十萬的評論去一一標記，可是，試作的每首歌均應對其重要的評論有所標註。所以，我們必須同時刪減試作的歌曲數目，以及採用上節的認同數概念來過濾出重要的評論。分析結果顯示在全部的 320 首歌裏面，擁有最多認同數音樂的所有評論共有 143,606 人次的認同數；而擁有最低認同數音樂的所有評論則是只有 101 人次的認同數。我們設定下列條件來進行評論的過濾：(1)入選的歌曲所有評論的總認同數需至少有 10,000 人次的認同，經過此條件的過濾，歌曲數從 320 首歌減少到 71 首歌曲。但這較熱門的 71 首歌也還有 7,029 篇評論。(2)對入選的每首歌至少選 3 篇、最多選 7 篇，在 3~7 篇間，如果選的評論的認同數已經超過總認同數的 70%的話，就停止不再選取。我們將入選的每一首歌曲的評論按認同數多寡順序列排，前 3 篇一定要被選，從第 3 篇開始，就選到累積超過同首歌所有評論的總認同數的 70%或到最多的 7 篇為止。按這個條件，每一首歌被選到的評論數不一定相同，以編號 972 號的音樂為例，讀到前 7 名認同數的評論時，累計只佔了 62.09%的認同數，所以會共選 7 篇。而編號 1105 號的音樂所有的網路評論認同數為 20,653，它的前幾名認同數的網路評論認同數分別是為 7,907、4,582、2,446 及 900。20,653 的 70%為 14,457 個認同數，所以系統只需讀最少的 3 篇便已達可以停止的總認同數 70%門檻。以這種方式，從 71 首歌曲中可以選出 341 篇較多人認同的熱門音樂評論，而選這 341 篇評論佔全部 7,029 篇評論的 72.38%的總認同數。

再者，由於時間及資源限制，本研究的試作也無法去收集上述所有 19 個情緒標籤的傾向標記，其實各情緒標籤資料收集的訓練方式是類似的。不過，我們也不希望只憑研究者自己的主觀來挑要試作的情緒，所以我們先公開在網路上邀請有習慣聽流行音樂者來填簡單問卷，請他們每個人從 19 個情緒標籤中選 5 個聽音樂時會特別在意的情緒標籤。所謂的「在意」並非只是指「喜歡」這類音樂，也包含「想避免」這類音樂。最後得到 31 個有效的回應資料。本研究依得票數從中優先選擇了「喜悅」、「悲傷」、「溫柔」、「希望」、「興奮」與「平靜」這 6 個較多人有興趣的為本研究接下來的試作情緒。

接下來，本研究設計了一個網站版的意見傾向評定系統，邀請平日會閱讀網路評論的大專學生針對一段網路評論的文字內容來評定：(1)網路評論作者的正負傾向（1 至 5 等），(2)該作者於評論中所欲表達的情緒感受，請其對六種情緒標籤各選擇「未觸發(0)」或「觸發(1)」二種狀態。當受測者進行評定時，系統讓他看到的只有該篇網路評論的文字內容及需要他標註的意見傾向問題而已，其它諸如網路評論的發言人、發言時間、認同數、評論的歌曲、歌詞、音樂等均不揭露，以免其受到影響。

經過利用此網站共邀請具名的大專學生受測者共有 171 人，共完成了 9,782 筆網路評論標註資料。在經過清洗後<sup>3</sup>，共捨棄了無效問卷的 237 筆資料、12 個易極端的受測者的 516 筆答案及 88 筆太少停留時間的答案，得到了 8,941 筆的有效回答資料，其中被最多人答題的評論共有 34 筆答案；最少答題的評論則有 23 筆答案。這些資料被視為是網路評論的專家意見傾向基準值，也將做為分類器訓練的目標值。而對於各情緒標籤是否為「觸發」是以「多數決」，也就是平均所有的受測者觸發值是否大於 0.5 來認定。

在這 341 篇評論當中，不同的情感受測者認為有觸發的比例非常的不同，如表 3，溫柔及興奮這二種情感的觸發比例不到 20%，這種資料不平衡 (unbalanced data) 會影響分類器正確率。

表 3：主要評論中專家意見傾向觸發佔比

|     | 喜悅    | 悲傷    | 溫柔    | 希望    | 興奮    | 平靜    |
|-----|-------|-------|-------|-------|-------|-------|
| 觸發  | 86    | 151   | 60    | 80    | 47    | 95    |
| 未觸發 | 255   | 190   | 281   | 261   | 294   | 246   |
| 觸發比 | 25.2% | 44.3% | 17.6% | 23.5% | 13.8% | 27.9% |

如前一節所述，從文本我們已找到 508 個候選的情緒標籤字詞，將這些關鍵字詞所對映的英文單字，利用 ConceptNet 中找尋二英文字關聯數值的功能，針對這 508 個關鍵字，分別計算每個關鍵字與 6 個主要情緒標籤字之間的關係數值即可得如表 4。然後，訓練系統開始掃描所有 341 目標評論，對每篇評論中出現的 508 個關鍵字進行次數加權平均，計算出每篇評論的主要情緒特徵矩陣後，配合受測者收集到的意見傾向平均值當作訓練目標值，就完成了訓練資料的準備工作，便可以開始進行網路評論分類器的訓練。

本研究採用 SVM 進行網路評論分類器的訓練。我們利用 5 折交叉驗證的方法來進行正確率的測試。首先將所有 341 篇網路評論特徵資料平分成 5 個資料集，每次選 1 個資料集出來當作測試資料，其它 4 個資料集則當成訓練資料，來進行每一個網路評論分類器的訓練與測試。我們依訓練資料狀況計算後，採用

3 本研究使用下列二個策略來做資料的清洗：以 5 等正負傾向來計算平均值與變異數，平均值加減 3 倍變異數為正常區間。當答題者的回答如果在正常區內，那就是代表這個答題是正常的，否則就是極端值。當一份答卷裏有超過五個答案是極端值時，就認為這份答卷是無效的答卷，將予刪除；當同一個受測者回答的問題有超過 1/5 是極端值的話，則是認為這位受測者是易極端的受測者，那這位受測者的所有答題都將捨棄不用。第二個清洗策略是參考回答的停留時間；當受測者在回答評論問題時，他每個答題所花的停留時間系統會一並與答案記錄起來。某評論回答一個問題停留時間少於 3 秒時，本研究認為該題的答案是沒有經過正常理性判斷，也會予以捨棄。

cost=10 及 gamma=0.5 的參數值進行訓練。

表 4：利用 ConceptNet 找到 508 個關鍵字與試作的 6 個情緒間關係表

| 情緒關鍵字    | Joy<br>喜悅 | Sadness<br>悲傷 | Tender<br>溫柔 | Hope<br>希望 | Excitement<br>興奮 | Claim<br>平靜 |
|----------|-----------|---------------|--------------|------------|------------------|-------------|
| abuse    | -0.064    | -0.01         | 0.035        | 0.003      | -0.078           | -0.001      |
| beloved  | 0.09      | 0.118         | 0.098        | 0.046      | -0.03            | -0.021      |
| confused | -0.018    | 0.022         | -0.03        | -0.033     | 0.147            | -0.05       |
| ...      | ...       | ...           | ...          | ...        | ...              | ...         |

依網路意見調查本研究共訓練了六個情緒標籤的網路評論分類器，5 折平均正確率資料報導如表 5 所列。其中興奮及溫柔二個評論分類器，因為資料不平均的現象太嚴重所以 F1 低於 65%之外，其它的分類器 F1 正確率都在 73%~77% 間。為了進一步分析資料極不平均的兩類資料，我們調整未觸發的興奮評論資料由 294 篇隨機移除至 153 篇，與 47 篇已觸發的評論資料，總資料雖降為 200 篇，但觸發率則上昇到 23.5%；類似的也用相同的方式將溫柔類資料總資料降為 255 篇，觸發率也上昇為 23.5%。在平衡資料進行 5 折訓練之後，得到調整平衡度後的評論興奮在訓練階段與測試階段均有提升。

表 5：情緒標籤網路評論分類器的正確率

|       | 訓練階段      |        |        | 測試階段      |        |        |
|-------|-----------|--------|--------|-----------|--------|--------|
|       | Precision | Recall | F1     | Precision | Recall | F1     |
| 評論喜悅  | 67.41%    | 86.89% | 75.92% | 68.57%    | 79.08% | 73.45% |
| 評論悲傷  | 77.34%    | 81.28% | 79.26% | 76.75%    | 77.48% | 77.12% |
| 評論平靜  | 68.80%    | 87.37% | 76.98% | 68.63%    | 85.26% | 76.05% |
| 評論興奮  | 48.55%    | 84.01% | 61.54% | 44.76%    | 80.89% | 57.63% |
| 評論溫柔  | 46.66%    | 78.75% | 58.60% | 48.13%    | 80.00% | 60.10% |
| 評論希望  | 65.62%    | 86.25% | 74.53% | 65.13%    | 88.75% | 75.13% |
| 調整後興奮 | 55.34%    | 84.64% | 66.92% | 56.00%    | 80.89% | 66.18% |
| 調整後溫柔 | 60.36%    | 84.58% | 70.45% | 60.92%    | 85.00% | 70.97% |

得到這些網路評論分類器後，如有新的評論出現時，這些分類器應能正確的將新評論進行意見傾向分類。進一步，對同一首歌，利用每個單篇評論的認同度比例做為權重，可以針對目前完成的六個情緒網路評論分類器進行整首歌的意



見傾向彙總，如表 6。舉例說明其算法，如編號 3 號的音樂，共有 3 篇的網路評論，評論 1 有 10 個認同數，六個分類器分類的結果依序為{1、0、1、0、1、0}；評論 2 有 5 個認同數，六分類器分類結果依序為{1、1、1、0、0、0}；評論 3 有 3 個認同數，六分類器分類結果依序為{1、1、0、0、1、1}。編號 3 號音樂的網路評論第一個分類器（喜悅）的百分數值= $(10*1+5*1+3*1)/(10+5+3)=100\%$ ；第二個分類器（悲傷）的百分數值= $(10*0+5*1+3*1)/(10+5+3)=44.4\%$ ，這對聽眾及音樂製作人等來說應是容易使用及了解的

表 6：網路評論意見傾向彙總表

| 歌曲 ID | 喜悅     | 悲傷     | 溫柔     | 希望     | 興奮     | 平靜     |
|-------|--------|--------|--------|--------|--------|--------|
| 21    | 72.97% | 0.61%  | 71.06% | 2.56%  | 5.22%  | 69.85% |
| 102   | 4.57%  | 90.17% | 6.15%  | 1.07%  | 58.95% | 23.16% |
| 332   | 5.64%  | 18.17% | 81.27% | 16.17% | 11.93% | 20.22% |
| ...   | ...    | ...    | ...    | ...    | ...    | ...    |

#### 四、音樂內容情感傾向分類器的訓練

##### （一）音樂內容專家標註資料的收集

本研究設計了視窗版的專家標籤評定程式，交由前述五名專家進行情緒標籤標註。這裡我們是針對全部 320 首歌，而非僅是納入分類實作的 71 首歌，乃是希望有不受侷限的判定。在專家們的前導會議中，先確定作業進行方式、各標籤的標註標準。我們以黑白素色做為設計主風格。畫面上除了歌曲的基本資訊之外，在播放歌曲的同時，隨著進行的時間，會出現演唱者目前正在唱的歌詞，但無人物風景等其他視訊。專家在邊聽音樂、邊看歌詞的狀況下，可以利用介面，進行標籤初稿的標註。在歌曲結束之後，專家要對標籤的標註狀況進行確認後儲存。每一個標籤都有三種狀態，未觸發（數值為 0）、部份觸發（數值為 1）及完全觸發（數值為 2）。為了怕單日一次聆聽過多首歌曲造成標註的品質下降，一位專家原則上一天聆聽的上限為 10 首歌，系統在列表區也有支援標示當天可以聽的部份。

經過三個月的標註後，經由標註系統共得到 1,600 人次的音樂標籤標註資料。對於不同專家間標註情形相差過大的狀況，本研究採取的清洗策略如下：如果 5 位專家的平均分數加減一倍標準差的區間範圍中沒有 3 位以上的專家標註資料，該標籤就得列入委員會進一步當面討論出一致的結果。在這個規則之下，大部份的歌曲標籤是不需要進一步討論的；在 320 首歌中共有 23 首歌的標籤資料

經過委員會討論之後，得到一致的結果。一首歌某個音樂標籤的 5 筆標註資料平均值，將作為音樂標籤分類器的訓練目標值，於音樂內容分類器訓練模組中被使用。

## （二）音樂內容分析-歌詞分析

歌詞資料在完成分詞斷句後，也可針對每一個字詞進行分析。在所有 55,453 筆的字詞中，共出現 5,366 種不同的字詞。同樣針對可能影響心情的關鍵字詞進行過濾，得到共 40 個常使用的歌詞情緒關鍵字。表 7 列出了前 20 名的常用歌詞心情關鍵字。利用這 40 個歌詞心情關鍵字，可以掃描出每首歌的歌詞中出現的關鍵字詞次數。另外因為這 40 個歌詞候選心情字詞都包含在 508 個網路評論關鍵字中，所以可以直接利用已儲存的表 2，配合歌詞的出現關鍵字詞次數加權計算出每一首歌詞的主要情緒特徵矩陣。

## （三）音樂內容—音訊特徵的萃取

本研究選定的音樂特徵參數如表 8，選擇這些參數乃是參考過去學者在音樂類別分類文獻曾經使用的音樂特徵參數（彙整於表 9）。

表 7：在歌詞中常出現的情緒相關字詞前 20 名

| #  | 字詞 | 次數  | #  | 字詞 | 次數 |
|----|----|-----|----|----|----|
| 1  | 愛  | 785 | 11 | 相信 | 72 |
| 2  | 好  | 229 | 12 | 陪  | 72 |
| 3  | 世界 | 123 | 13 | 飛  | 71 |
| 4  | 夢  | 105 | 14 | 痛  | 68 |
| 5  | 幸福 | 98  | 15 | 離開 | 64 |
| 6  | 快樂 | 92  | 16 | 笑  | 61 |
| 7  | 眼淚 | 86  | 17 | 溫柔 | 59 |
| 8  | 時間 | 79  | 18 | 懂  | 59 |
| 9  | 哭  | 75  | 19 | 風  | 56 |
| 10 | 寂寞 | 74  | 20 | 回憶 | 53 |

表 8：本研究選擇的音樂特徵

| #  | 中文名   | 英文名             | 分類 | 說明           |
|----|-------|-----------------|----|--------------|
| #1 | 平均響度  | Avg. Energy     | 響度 | 區間中的總響度除於區間數 |
| #2 | 方均根響度 | RMS. Energy     | 響度 | 區間中的響度的方均根值  |
| #3 | 低響度率  | Low Energy Rate | 響度 | 低於平均響度的區間比率  |

|     |         |                |    |                             |
|-----|---------|----------------|----|-----------------------------|
| #4  | 音調值     | Pitch          | 音調 | 區間中的音調值                     |
| #5  | 倒頻譜     | Cepstrum       | 音調 | 將原來信號的頻譜先轉成類似分貝的單位，再作逆傅里葉變換 |
| #6  | 節奏值     | Tempo          | 韻律 | 區間中的節奏值                     |
| #7  | 脈衝清晰度   | Pluse Clarity  | 韻律 | 區間中脈衝的清晰數值                  |
| #8  | 節拍數     | Beat Sum       | 韻律 | 區間中的節拍數                     |
| #9  | 最強節拍強度  | Strongest Beat | 韻律 | 區間中最強的節拍強度                  |
| #10 | 梅爾倒頻譜係數 | MFCC           | 音色 | 區間內對不同頻率的感受程度               |
| #11 | 亮度      | Brightness     | 音色 | 區間內聲音的清亮程度                  |
| #12 | 粗糙度     | Roughness      | 音色 | 區間內聲音的粗細程度                  |
| #13 | 過零率     | ZCR            | 其它 | 區間內頻度由正到負/負到正的比率            |
| #14 | 中心性     | Centroid       | 其它 | 區間內頻譜的質量中心                  |
| #15 | 滾降值     | Roll-off       | 其它 | 區間內頻率的傳輸函數的斜率               |

表 9：文獻中使用音樂特徵的對應表

| 特徵                             | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | #12 | #13 | #14 | #15 |
|--------------------------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| Bergstra et al.2006            |    |    |    |    | ✓  |    |    |    |    | ✓   |     |     | ✓   | ✓   | ✓   |
| Buger et al. 2013              |    |    |    | ✓  |    | ✓  | ✓  | ✓  |    |     | ✓   | ✓   |     |     |     |
| Kaur & Kumar.2017              |    |    |    |    |    |    |    |    |    | ✓   |     |     | ✓   |     | ✓   |
| Lee et al. 2009                |    |    |    |    |    |    |    |    |    | ✓   |     |     |     |     |     |
| Jothilakshmi & Kathiresan 2012 |    | ✓  |    |    |    | ✓  |    |    |    | ✓   |     |     | ✓   | ✓   | ✓   |
| Patel & Trivedi 2017           | ✓  | ✓  | ✓  |    |    |    |    | ✓  | ✓  | ✓   |     |     | ✓   | ✓   | ✓   |
| Tzannetakis et al. 2002        |    | ✓  | ✓  | ✓  |    | ✓  |    | ✓  | ✓  | ✓   |     |     |     | ✓   | ✓   |
| Xu et al. 2003                 | ✓  | ✓  |    |    |    |    |    | ✓  | ✓  | ✓   |     |     | ✓   |     |     |
| Xu et al. 2005                 | ✓  |    |    |    | ✓  |    |    |    | ✓  | ✓   |     |     | ✓   | ✓   |     |
| Zhang et al. 2014              |    |    |    |    |    |    |    |    |    | ✓   |     |     |     |     |     |
| 陳威佑 2012                       | ✓  | ✓  |    |    |    | ✓  | ✓  |    |    |     | ✓   | ✓   | ✓   | ✓   | ✓   |

我們使用 Matlab 的 MIRtoolbox 工具套件做為音樂特徵的採樣計算器，利用自行撰寫的批次計算檔去讀取每個音樂檔，並計算表 8 每個選定的參數。每個參數除了總數值之外，依不同的分隔區（frames）可以計算出各區的數值。輸出的音樂特徵數據將轉存為文字檔再導入資料庫。如此一來所有歌曲都可以取得這些數值。之後，我們對每個計算出的原始數值做 Z-Score 的轉換並儲存在音樂特徵

資料庫中。

#### (四) 考慮商品內容為分類依據的分類器訓練

到這個階段為止，已取得二類要當作分類器輸入的資料，一是利用 MIRtoolboxs 計算出來的音訊特徵資料、二是每首歌的歌詞情緒特徵矩陣；也已從專家處取得了做為分類器目標值的標籤資料。五位專家針對每一首歌的情緒標籤資料都有設定標籤{無觸發(0)、有觸發(1)、非常觸發(2)}的三種觸發值；分類器的訓練需要一個結果標籤，但計算五位專家的平均觸發值是 0 到 2 之間的連續數值，無法直接當做分類器訓練的目標值。本研究分配當平均觸發值  $x$  在  $0.0 < x \leq 0.6$  之間時認為目標值是{無觸發(0)}、當平均觸發值  $x$  在  $0.6 < x \leq 1.4$  之間時認為目標值是{有觸發(1)}、當平均觸發值  $x$  在  $1.4 < x \leq 2.0$  時認為目標值是{非常觸發(2)}。

接下來就是採用 SVM，進行音樂內容標籤分類器的訓練。在音樂內容標籤分類器的訓練時本研究採用 3 折交叉驗證的方法來進行正確率的測試。首先將所有 320 首音樂資料依各類比例亂數平分分成 3 個資料集，每次選 1 個資料集出來當作測試分類器的資料，其它 2 個資料集則當成訓練資料，來進行每一個 SVM 標籤分類器的訓練與測試。

在實際操作時 SVM 訓練參數依訓練資料狀況計算適當的 cost 值及 gamma 參數值再進行訓練，採用  $\text{cost}=10$  及  $\text{gamma}=0.5$ 。由於每個標籤都是屬於{無觸發(0)、有觸發(1)、非常觸發(2)}的三類狀況的多分類問題，在分類績效報導上則是參考 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，如表 10。

表 10：本研究三分類音樂內容標籤分類器的正確率

|      |    | 訓練階段      |        |        | 測試階段      |        |        |
|------|----|-----------|--------|--------|-----------|--------|--------|
|      |    | Precision | Recall | F1     | Precision | Recall | F1     |
| 喜悅標籤 | 巨觀 | 71.99%    | 84.78% | 77.87% | 72.03%    | 88.88% | 79.57% |
|      | 微觀 | 86.72%    | 86.72% | 86.72% | 83.43%    | 83.43% | 83.43% |
| 悲傷標籤 | 巨觀 | 68.57%    | 70.24% | 69.39% | 72.19%    | 75.38% | 73.75% |
|      | 微觀 | 72.49%    | 72.49% | 72.49% | 73.74%    | 73.74% | 73.74% |
| 平靜標籤 | 巨觀 | 73.37%    | 86.16% | 79.25% | 64.52%    | 78.06% | 70.65% |
|      | 微觀 | 87.03%    | 87.03% | 87.03% | 82.19%    | 82.19% | 82.19% |
| 興奮標籤 | 巨觀 | 68.42%    | 85.32% | 75.94% | 64.42%    | 78.80% | 70.89% |
|      | 微觀 | 83.13%    | 83.13% | 83.13% | 80.90%    | 80.90% | 80.90% |

|      |    |        |        |        |        |        |        |
|------|----|--------|--------|--------|--------|--------|--------|
| 溫柔標籤 | 巨觀 | 72.07% | 73.63% | 72.84% | 68.21% | 70.48% | 69.32% |
|      | 微觀 | 79.69% | 79.69% | 79.69% | 74.99% | 74.99% | 74.99% |
| 希望標籤 | 巨觀 | 66.66% | 83.95% | 74.31% | 63.91% | 81.00% | 71.45% |
|      | 微觀 | 80.63% | 80.63% | 80.63% | 78.11% | 78.11% | 78.11% |

如果合併三類狀況中的{有觸發(1)}及{非常觸發(2)}為{觸發(1)}，在情緒標籤{無觸發(0)、觸發(1)}的二類狀況之下，除了興奮標籤的分類器 F1 是 68.41%之外，其它的分類器 F1 正確率都超過 70%，詳細的分類器正確率資料如表 11。

表 11：本研究二分類音樂內容標籤分類器的正確率

|      | 訓練階段      |        |        | 測試階段      |        |        |
|------|-----------|--------|--------|-----------|--------|--------|
|      | Precision | Recall | F1     | Precision | Recall | F1     |
| 喜悅標籤 | 68.52%    | 91.03% | 78.18% | 61.64%    | 87.18% | 72.22% |
| 悲傷標籤 | 78.72%    | 83.67% | 81.12% | 77.05%    | 84.87% | 80.77% |
| 平靜標籤 | 80.15%    | 89.29% | 84.47% | 72.99%    | 89.68% | 80.48% |
| 興奮標籤 | 59.39%    | 88.62% | 71.12% | 57.14%    | 85.21% | 68.41% |
| 溫柔標籤 | 87.86%    | 86.17% | 87.01% | 85.57%    | 80.42% | 82.91% |
| 希望標籤 | 62.82%    | 90.61% | 74.20% | 60.14%    | 86.32% | 70.89% |

## 伍、舒緩情緒之音樂實驗：以悲傷為例

文獻上，Moore (2013) 提出在當人們當聽到如喜歡的、熟悉的音樂或是自己演唱時，會激發與情緒調節 (affect/emotion regulation) 有關的正向神經活動；當人們聽到複雜、不協調或非期待中的音樂時，則會激發情緒調節不想要的神經活動。Van Goethem 與 Sloboda (2011) 對受測者進行了為期一到數週的日誌研究也發現音樂有助於情感調節，尤其對於在創造快樂和放鬆心情上有很重要的作用。例如音樂可以幫助收聽者更加正向積極，或是可以幫助收聽者抽離某個情緒。

透過上述的「網路評論分類器」與「商品內容分類器」建立，我們已可瞭解某歌曲所試圖傳遞的情感後，然而，何時該推薦此該歌曲，我們建議透過一連串實驗來確定其推薦規則。在此，我們以悲傷情境為例<sup>4</sup>，試圖回答：「悲傷時，該

4 單一種主要情緒比較容易操作。如果比較二種最常見的情緒：悲傷與快樂；快樂是正面的情緒，當快樂的時候，可能只有很少的情況下人們會想要主動的利用情感商品來離開快樂的情緒。而相對的因為悲傷是負面情緒，人們會更想要經由情感商品來緩和或轉換悲傷情緒。所以本研究選擇以悲傷為例來進行舒

聽何種歌？」

### （一）悲傷情緒的量測與誘發

然而實務上，我們很難剛好找到一群悲傷的人來進行實驗。在心理學領域中，研究者如果希望觀察人類在某個情緒下的行為，會使用誘發媒體對受測者進行情緒的操弄。本研究參考 Nadler 等（2010）、盧毓文（2011）操弄受測者的方法與量表，來進行實驗所需要的悲傷情緒的誘發與量測。我們的量表乃由受測者以 1-7 尺度自我評估其悲傷程度的強烈（2 題，信度 Cronbach's Alpha 為 0.928）與對悲傷的可控制度（2 題，信度 Cronbach's Alpha 為 0.872）。而在情緒誘發媒體方面，我們先由網路上資訊，對「戰爭難民」及「311 日本地震」的記錄片各剪接成 16 分鐘的兩個短片，先對 10 名大專學生進行觀看媒體誘發情緒的前測，透過前測結果量表與質性訪談，難民短片較能誘發悲傷效果，是以被選為本實驗的誘發媒體。

另外實驗中要撥放的音樂因為需要與誘發媒體配合的關係，在實驗中所播放的歌曲是由音樂內容標籤分類器所分出的{喜悅、平靜、悲傷}三種類情緒標籤的非情歌歌曲：每一類本研究選的歌都希望它的該類的標籤是單獨的高度被激發的，而其它二類的標籤則是無激發，例如喜悅類的歌，選擇的條件是它只有喜悅是高度激發的，而悲傷和平靜的標籤則是無激發。利用這樣的條件，本研究利用「網路評論分類器」與「商品內容分類器」綜合判斷選擇了二首喜悅的歌，分別是《慶祝》<sup>5</sup>與《稻香》<sup>6</sup>；平靜的歌選擇《火花》<sup>7</sup>；悲傷的歌則是選擇《翅膀之歌》<sup>8</sup>。為了再次確認音樂內容標籤分類器的正確性，本研究也針對這四首歌進行標籤測試，共邀請了 54 位大專學生在聽過此四首歌曲後，對每一首歌都進行三類標籤的標註（可多選）：測試資料顯示，對上述四首歌，均有超過九成答案認可我們分類器的標籤判定。

### （二）實驗設計與前測

在決定好誘發媒體後，我們徵求 10 名未參與誘發媒體前測的大專學生進行實驗前測，前測資料顯示，音樂效果是符合預期，此外，前測參與者也提供實驗系統介面的一些修改建議。前測後，我們邀請另外的 96 位受測者參與正式實驗，其中男性為 46 名、女性為 50 名；年紀則皆在 18~24 歲間。整個實驗時間約 45 分鐘，在電腦教室中利用實驗程式進行。受測者被要求自己攜帶耳機，在就位

---

緩情緒的音樂實驗。

5 慶祝 <https://www.youtube.com/watch?v=FDAut-oMSbU>

6 稻香 <https://www.youtube.com/watch?v=9InC9n5obus>

7 火花 <https://www.youtube.com/watch?v=IH0RDxdIYey>

8 翅膀之歌 <https://www.youtube.com/watch?v=-QOavxQ8Dwc>

後，會先聽一段約五分鐘的實驗注意事項，隨後以我們設計的實驗系統先作初始情緒的線上量測。在確定所有的受測者量測繳交後，會統一播放戰爭難民短片的誘發媒體。看完短片之後，做第二次情緒的線上量測。第一次量測是瞭解進入實驗室者今天的情緒，第二次量測則為本實驗接下來分組的依據。本實驗將受測者分成三組：A 組及 B 組都是看完影片後悲傷程度大於等於中間值 4 的受測者，系統會計算比較其初始狀況及誘發狀況儘量將有類似數據的受測者平均分到二組；而 C 組則是誘發後悲傷度小於 4（也就是情緒操弄對其沒有成功）的受測者組成。96 人中有 5 位受測者因為資料不正確或無耐心中斷實驗的原因而無法採用，剩下的 91 人中誘發成功者 70 人平均分配在 A、B 兩組、每組 35 人；而未誘發成功的 21 人則分配在 C 組。

每組都是聆聽三首歌，受測者均會配戴自用耳機，聆聽時系統會與歌曲同步展示動態字幕，但無背景影像也不告知其它資訊（如演唱者姓名）。A 組的聆聽順序是平靜的歌《花火》→喜悅的歌《慶祝》→喜悅的歌《稻香》；B 組聆聽順序是喜悅的歌《慶祝》→平靜的歌《花火》→喜悅的歌《稻香》。C 組的受測者則是聆聽悲傷的歌《翅膀之歌》→平靜的歌《花火》→喜悅的歌《稻香》。在聽完每一首歌之後，受測者都會進行當時的悲傷度、控制度的量測，並對剛聽完的那首歌給予滿意度評價。此處有特別提醒，關於滿意度的評分請受測者以其當下的情緒來考量是否適合這首歌來表達滿意度，而不要考量對歌手、歌曲風格之個人偏好的影響。

### （三）實驗的結果分析

表 12 是各組在悲傷、控制、滿意度的平均值的表現，以下幾點的進一步觀察與分析：

各組在實驗初始時，情緒均未悲傷。

看過誘發媒體後，A、B 組被成功誘發悲傷情緒，亦無法控制此情緒，而 C 組悲傷度雖略提升，但並未至可稱為悲傷（1-7 點量表，顯著低於中間值 4）。故我們實驗主要針對 A、B 組人，C 組其實只是給實驗安慰劑做對照組。

粗略而言，A、B 組在聽過實驗的歌曲後，其悲傷度均有下降，控制度均會提升。

表 12：各組初始及誘發後悲傷、控制、滿意度的平均值

| 變數  | 悲傷度   |       |      |       |       | 控制度   |       |       |       |       | 滿意度  |       |       |
|-----|-------|-------|------|-------|-------|-------|-------|-------|-------|-------|------|-------|-------|
|     | 初始    | 誘發    | 第一首歌 | 第二首歌  | 第三首歌  | 初始    | 誘發    | 第一首歌  | 第二首歌  | 第三首歌  | 第一首歌 | 第二首歌  | 第三首歌  |
| A 組 | 2.77* | 5.20* | 3.56 | 2.46* | 2.34* | 5.66* | 4.71* | 5.54* | 6.06* | 6.06* | 4.17 | 4.54* | 5.71* |

|     |       |       |       |       |       |       |       |       |       |       |      |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|-------|
| B 組 | 2.54* | 5.44* | 2.94* | 3.07* | 2.00* | 5.51* | 4.41  | 5.87* | 5.99* | 6.27* | 3.60 | 4.17  | 5.60* |
| C 組 | 1.52* | 2.76* | 2.45* | 2.55* | 1.71* | 6.24* | 5.77* | 6.31* | 6.26* | 6.62* | 4.19 | 4.86* | 5.86* |

註：1-7 點量表，中間值為 4，打\*號者，為與中間值比較在信心水準 0.05 下顯著

進一步統計檢定，我們發現 A 組在前兩階段「聽第一首歌後」相對「誘發後」、「聽第二首歌後」相對「聽第一首歌後」的悲傷度、控制度差異均達統計顯著（0.05 水準）。而第三階段「聽第三首歌後」相對「聽第二首歌後」的差異就不顯著。然而，滿意度反之，在第三階段提升最顯著。

而 B 組則在第一階段「聽第一首歌後」相對「誘發後」、第三階段「聽第三首歌後」相對「聽第二首歌後」悲傷度、控制度差異均達統計顯著（0.05 水準），而「聽第二首歌後」相對「聽第一首歌後」的差異就不顯著。不過與 A 組相同地，滿意度在第三階段提升最顯著。

我們進而將 A、B 組悲傷、控制度的變化以圖顯示，如圖 2、3，並進而以統計去比較其變化量差異，我們得知：(1)相對實驗初始，A 組誘發後悲傷度上升（2.43）、控制度下降（-0.95）與 B 組誘發後悲傷度上升（2.90）、控制度下降（-1.10）並無統計顯著差異，證明了一開始的即時分組功能是有順利的將受測者們依誘發程度平均的分配到兩組。(2)第一單曲效果：相對誘發後，A 組聽完第一首歌後悲傷度下降（-1.64）、控制度上升（0.83）與 B 組聽完第一首歌後悲傷度下降（-2.50）、控制度上升（1.46）達統計顯著差異（0.05 水準）。(3)第二單曲效果：相對第一首歌後，A 組聽完第二首歌後悲傷度下降（-1.10）、控制度上升（0.51）與 B 組聽完第二首歌後悲傷度微微的上升（0.13）、控制度上升（0.12）達統計顯著差異（0.05 水準）。(4)第三單曲效果：由於 B 組在第二首歌後悲傷度有微幅的增加，相對第二首歌後，故 B 組聽完第三首歌後悲傷度下降（-1.07）與 A 組聽完第三首歌後悲傷度下降（-0.12）達統計顯著差異（0.05 水準），而其兩組的控制度並無統計顯著差異（A 組前後無變化、B 組上升 0.28）。(5)前兩首歌曲的累積效果、全部三首歌曲的累積效果均呈現 A 組、B 組的全部悲傷度下降、控制度上升並無顯著差異。這說明 A、B 兩種不同的播放順序所整體造成的效果並沒有太顯著的不同。



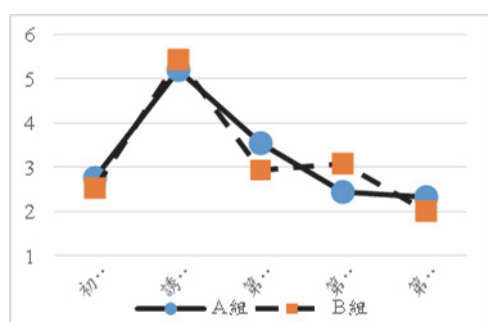


圖 2：AB 兩組悲傷度的變化

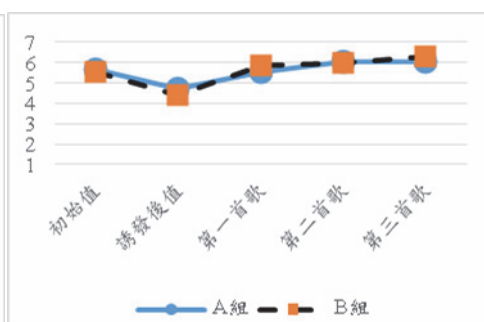


圖 3：AB 兩組控制度的變化

我們的安排第一首歌，A 組是聽平靜的歌《花火》，B 組是聽喜悅的歌《慶祝》，第二首歌，A 組是聽喜悅的歌《慶祝》，B 組是聽平靜的歌《花火》。由上述得知，如果只聽一首歌的話，推薦喜悅的歌曲較能有效的讓悲傷降低與控制度上昇。但是我們也觀察到，《慶祝》這首歌在 B 組的滿意度（3.60）不論跟 A 組第一首歌《花火》（4.17）或是 A 組第二首歌聽同一首《慶祝》的滿意度（4.54）相比都較低。我們推測可能這是因為 B 組剛看完讓人傷心的誘發媒體馬上就聽喜悅的歌所造成的反差感所造成的。雖然單比悲傷度的下降效用的話，馬上聽喜悅的歌會比較有效，但是聽眾對這首歌的滿意度不會太好。甚至，B 組在聽第二首平靜的歌，經反思自省後，其悲傷度還會微幅的增加。

是以經由實驗，我們得到的推薦規則是：「如果在悲傷時，要推薦兩首歌的話，雖先播放平靜或喜悅的先後順序對悲傷度與控制度的累積改變是沒有統計明顯不同的，但因為滿意度的關係，建議先聽平靜的歌再聽喜悅的歌。」

另外，從 C 組數據指出，在聽完第一首的悲傷歌曲或第二首平靜歌曲後，悲傷度與控制度所產生的效果與聽該歌曲前相比並沒有明顯的差異；直到聽了第三首的喜悅的歌時，悲傷度下降與控制度上升才會顯著。這似乎顯示，對於不易悲傷或較冷漠的人，悲傷或平靜歌曲也不會改變其悲傷度，倒是喜悅的歌還有一定效果。

## 陸、結論

本研究建議一個同時參考網路評價及商品內容的情感商品推薦架構，並以音樂為例來說明這架構，我們同時使用了網路評論及音樂內容（包含歌詞及音訊特徵）做為商品推薦的資料依據來源。從分類器的訓練及測試的結果資料看來，不論是網路評論分類器或是音樂內容分類器也都有近 70%以上的一定程度穩定正確率。進而，本研究利用情緒實驗以了解在用戶悲傷情緒下應推薦的音樂，發現若只聽一首歌的話，推薦喜悅的歌曲較能有效的讓悲傷降低與控制度上昇；若可要

推薦兩首歌的話，建議先聽平靜的歌再聽喜悅的歌。

本研究的貢獻不在情感商品所蘊含的情感辨識技術方法的改進，而是提出一個完整的推薦機制，強調情感商品的情感辨識應同時考量到商品內容的資訊（以我們的例子而言，是音樂音訊、歌詞）以及網路評價，以及其推薦程序應先經過一連串的情緒實驗，方能有合適的推薦。國內外以往文獻並無類似的情感商品推薦機制，是以本研究有其創新性。不過，相對音樂只有音訊，其他的情感商品如舞蹈表演加上了視訊，其所激發的情緒分析會更複雜。後續研究，可實驗其他情緒情境下應推薦的音樂，或考量自動感知情境（如睡眠、運動時）的音樂推薦；或進一步嘗試舞蹈、戲劇等包含影像之情感商品的推薦問題。

### 誌謝

作者感謝匿名評審及主編寶貴意見。同時，本文受科技部補助研究計畫（NSC 107-2410-H-004-097-MY3）出版補助，特此致謝。

### 參考文獻

- 卓淑玲、陳學志、鄭昭明（2013），『台灣地區華人情緒與相關心理生理資料庫－中文情緒詞常模研究』，*中華心理學刊*，第五十五卷，第四期，頁 493-523。
- 陳威佑（2012），『基於前進選擇之特徵選取之流行音樂曲風辨識與分析』，未出版碩士論文，國立中山大學電機工程學系研究所，高雄。
- 楊亨利、林青峰（2017），『微網誌短句的情感指數分析－以新浪微博為例』，*中華民國資訊管理學報*，第二十四卷，第一期，頁 1-28。
- 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 335-361。
- 盧毓文（2011），『情緒對類別學習之影響』，未出版碩士論文，國立政治大學心理學研究所，台北。
- Bergstra, J., Casagrande, N., Erhan, D., Eck, D. and Kegl, B. (2006), 'Aggregate features and ADABOOST for music classification', *Machine Learning*, Vol. 65, No. 2-3, pp. 473-484.
- Bu, J., Tan, S., Chen, C., Wang, C., Wu, H., Zhang, L. and He, X. (2010), 'Music recommendation by unified hypergraph: combining social media information and music content', *Proceedings of the 18th ACM international conference on Multimedia*, Firenze, Italy, October 25-29, pp. 391-400.
- Burger, B., Thompson, M.R., Luck, G., Saarikallio, S. and Toivainen, P. (2013), 'Influences of rhythm-and timbre-related musical features on characteristics of

- music-induced movement, *Frontiers in Psychology*, Vol. 4, Article 183, pp. 1-10.
- Celma, O. (2006), 'FOAFing the music: Bridging the semantic gap in music recommendation', *Proceedings of the International Semantic Web Conference*, Springer, Berlin, Heidelberg, November 5-9, pp. 927-934.
- Chaudhuri, A. and Holbrook, M.B. (2001), 'The chain of effects from brand trust and brand affect to brand performance: the role of brand loyalty', *Journal of Marketing*, Vol. 65, No. 2, pp. 81-93.
- Chen, H.C. and Chen, A.L. (2001), 'A music recommendation system based on music data grouping and user interests', *Proceedings of the 10th International Conference on Information and Knowledge Management*, Atlanta, U.S.A., November 5-10, pp. 231-238.
- Corona, H. and O'Mahony, M.P. (2015), 'An exploration of mood classification in the million songs dataset', *Proceedings of the 12th Sound and Music Computing Conference*, Maynooth, Ireland, July 25-August 1, 2015.
- Esuli, A. and Sebastiani, F. (2006), 'Sentiwordnet: A publicly available lexical resource for opinion mining', In *LREC*, Vol. 6, pp. 417-422.
- Gómez, L.M., and Cáceres, M.N. (2017), 'Applying data mining for sentiment analysis in music', *Proceedings of the 15th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS 2017)*, Porto, Portugal, June 21-23, pp. 198-205.
- Hu, X., Downie, J.S., West, K. and Ehmann, A.F. (2005), 'Mining music reviews: Promising preliminary results', *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, United Kingdom, September 11-15, pp. 536-539.
- Hu, X., and Downie, J.S. (2010), 'Improving mood classification in music digital libraries by combining lyrics and audio', *Proceedings of the 10th annual joint conference on Digital libraries*, Surfer's Paradise, Australia, June 21-25, pp. 159-168.
- Hu, X., and Yang, Y.H. (2017), 'The mood of Chinese pop music: Representation and recognition', *Journal of the Association for Information Science and Technology*, Vol. 68, No. 8, pp. 1899-1910.
- Jothilakshmi, S. and Kathiresan, N. (2012), 'Automatic music genre classification for indian music', *Proceedings of the International Conference on Software and Computer Applications (ICSCA 2012)*, Singapore, June 9-10.
- Kaur, C. and Kumar, R. (2017), 'Study and analysis of feature based automatic music

- genre classification using Gaussian mixture model', *Proceedings of the 2017 International Conference on Inventive Computing and Informatics (ICICI 2017)*, Coimbatore, India, November 23-24, pp. 465-468.
- 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.
- 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.
- Kumar, V. and Minz, S. (2013), 'Mood classification of lyrics using SentiWordNet', *Proceedings of the 2013 International Conference on Computer Communication and Informatics (ICCCI 2013)*, Coimbatore, India, January 9-11, pp. 1-5.
- Lee, C.H., Shih, J.L., Yu, K.M. and Lin, H.S. (2009), 'Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features', *IEEE Transactions on Multimedia*, Vol. 11, No. 4, pp.670-682.
- Lee, S.K., Cho, Y.H. and Kim, S.H. (2010), 'Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations', *Information Sciences*, Vol. 180, No. 11, pp. 2142-2155.
- Li, H. and Lu, W. (2017), 'Learning latent sentiment scopes for entity-level sentiment analysis', *Proceedings of the 31st AAAI conference (AAAI-17)*, San Francisco, U.S.A., February 4-9, 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.
- Liu, H. and Singh, P. (2004), 'Focusing on ConceptNet's natural language knowledge representation', *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES 2004)*, Wellington, New Zealand, September 20-25.
- Moore, K.S. (2013), 'A systematic review on the neural effects of music on emotion regulation: implications for music therapy practice', *Journal of Music Therapy*, Vol. 50, No. 3, pp. 198-242.
- 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.
- Nadler, R.T., Rabi, R. and Minda, J.P. (2010), 'Better mood and better performance:

- Learning rule-described categories is enhanced by positive mood', *Psychological Science*, Vol. 21, No. 12, pp. 1770-1776.
- Nanopoulos, A., Rafailidis, D., Symeonidis, P. and Manolopoulos, Y. (2010), 'Musicbox: Personalized music recommendation based on cubic analysis of social tags', *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 2, pp. 407-412.
- Oramas, S., Espinosa-Anke, L. and Lawlor, A. (2016), 'Exploring customer reviews for music genre classification and evolutionary studies', *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR 2016)*, New York, U.S.A., August 7-11.
- 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.
- Patel, D. and Trivedi, K. (2017), 'Research of music classification based on mood recognition', *International Education and Research Journal*, Vol. 3, No. 5, pp. 544-555.
- Rho, S., Han, B.J., and Hwang, E. (2009), 'SVR-based music mood classification and context-based music recommendation', *Proceedings of the 17th ACM International Conference on Multimedia*, Beijing, China, October 19-22, pp. 713-716.
- Schein, A.I., Popescul, A., Ungar, L.H. and Pennock, D.M. (2002), 'Methods and metrics for cold-start recommendations', *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Tampere, Finland, August 11-15, pp. 253-260.
- Sharma, V., Agarwal, A., Dhir, R. and Sikka, G. (2016), 'Sentiments mining and classification of music lyrics using SentiWordNet', *Proceedings of the 2016 Symposium on Colossal Data Analysis and Networking (CDAN 2016)*, Indore, India, March 18-19, pp. 1-6.
- 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.
- Tzanetakis, G. and Cook, P. (2002), 'Musical genre classification of audio signals', *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 5, pp. 293-302.
- Oord, A., Dieleman, S. and Schrauwen, B. (2013), 'Deep content-based music

- recommendation', *Advances in Neural Information Processing Systems*, pp. 2643-2651.
- Van Goethem, A. and Sloboda, J. (2011), 'The functions of music for affect regulation', *Musicae Scientiae*, Vol. 15, No. 2, pp. 208-228.
- Wang, X. and Wang, Y. (2014), 'Improving content-based and hybrid music recommendation using deep learning', *Proceedings of the 22nd ACM International Conference on Multimedia*, Orlando, U.S.A., November 3-7, pp. 627-636.
- Wang, J. and Zhao, X. (2019), 'Deep learning based mood tagging for Chinese song lyrics', arXiv preprint, *arXiv:1906.02135*.
- 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.
- Xu, C., Maddage, N.C. and Shao, X. (2005), 'Automatic music classification and summarization', *IEEE Transactions on Speech and Audio Processing*, Vol. 13, No. 3, pp. 441-450.
- Xu, C., Maddage, N.C., Shao, X., Cao, F. and Tian, Q. (2003), 'Musical genre classification using support vector machines', *Proceedings of the International Conference of Acoustics, Speech, and Signal (ICASSP'03)*, Hong Kong, China, April 6-10, Vol. 5, pp. 429-432.
- 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.
- Ye, Q., Shi, W. and Li, Y. (2006), 'Sentiment classification for movie reviews in Chinese by improved semantic oriented approach', *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, Koloa, U.S.A., January 5-8.
- Zhang, C., Evangelopoulos, G., Voinea, S., Rosasco, L. and Poggio, T. (2014), 'A deep representation for invariance and music classification', *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, pp. 6984-6988.
- Zhuang, L., Jing, F. and Zhu, X.Y. (2006), 'Movie review mining and summarization', *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, Arlington, U.S.A., November 6-11, pp. 43-50.