

# 以音樂內容為基礎的情緒分析與辨識

陳若涵, 許肇凌, 張智星, 羅鳳珠

國立清華大學資訊系統與應用所

{annie, leon}@wayne.cs.nthu.edu.tw, jang@cs.nthu.edu.tw, gefjulo@saturn.yzu.edu.tw

## 摘要

音樂可以表達作曲者的意念、情感，而同樣聆聽者也會對音樂產生情緒，因此自動音樂情緒分類在音樂資訊檢索、人機互動或是音樂治療上都有重要的幫助。本文使用了階層式的分類法針對古典音樂以及流行音樂的音樂內容進行特徵萃取以及辨識，並加入歌詞情緒來比較其和音樂內容的辨識率，我們採用最近鄰居決定法則（K-nearest Neighbor Rule，簡稱 KNNR）、高斯混合模型（Gaussian Mixture Model，簡稱 GMM）和支向機(Support Vector Machine，簡稱 SVM)三種不同的分類器進行分類，實驗結果顯示 KNNR 效能最佳約 85%辨識率，且歌詞對於流行音樂的情緒辨識有很大的影響。雖然結合歌詞和音樂內容的情緒分類還有改進的空間，但已比單用音樂內容辨識的效果佳，所以仍可視為另一種可行的新方法。

## 1. 簡介與相關研究

現今音樂資料庫在家庭電腦或是網際網路的搜尋引擎都佔有重要的地位。在過去的搜尋方式，通常是針對音樂資料的名稱、演唱者、作曲者或是歌詞的關鍵字來搜尋。而近年來，在音樂旋律搜尋方面，可以針對旋律的音符或是節拍來搜尋；在音樂內容方面，可以針對音樂的類型以及情緒作為檢索。本篇論文主要是針對音樂內容的情緒進行分析以及辨識。

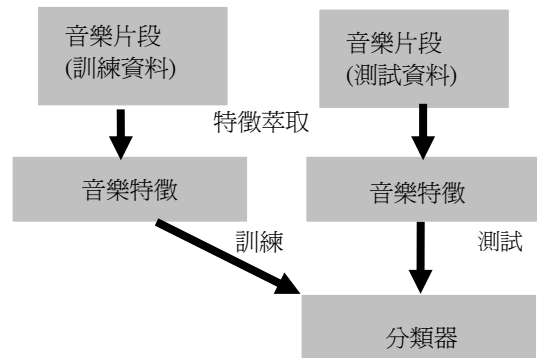
音樂心理學家表示當人們在聆聽音樂會對其產生情緒的反應，這也應用到近年來盛行的音樂治療方面，人們可以藉由聆聽音樂緩和情緒，也可以藉由聆聽音樂來提升士氣。Krumhansl[1]指出音樂可能含有情緒的意義，例如：有些音樂片段可以表現滿意的或是愉快的情緒；而有些音樂則表現焦慮的或是暴怒的情緒。除了音樂可以表現情緒之外，文字詩詞也都含有作者強烈的情感，因此在論文中我們希望除了透過音樂內容來偵測情緒之外，更能透過分析歌詞的情緒字彙來提高辨識結果。

在音樂情緒研究的領域，已有 Dan Liu 等[2]從單聲道的類比音樂中進行情緒辨識。在 MIDI(musical instrument digital interface)方面，Wang Muyuan 等人[3]提出利用人們對音樂的知覺特徵來進行音樂情緒辨

識。在本篇論文，我們主要針對分析古典音樂和流行音樂的情緒特徵並進行辨識，且使用 KNNR、GMM、SVM 三種不同的分類器進行分類，並比較其效能。

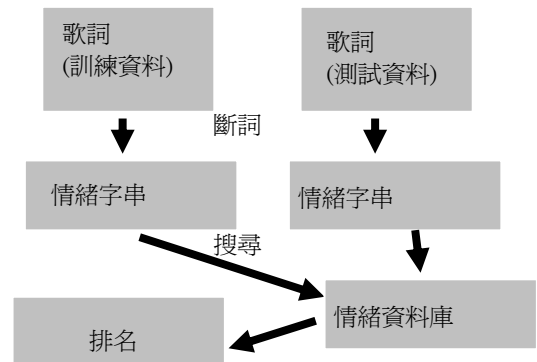
## 2. 系統架構

如圖一所示，在音樂部份由下列所架構成



圖一 音樂內容系統架構

如圖二所示，在歌詞部分由下列架構成



圖二 歌詞內容系統架構

## 2.1 歌曲資料庫

本文使用的音樂資料庫是以MIDI轉換出的資訊為主。由於我們只針對樂曲的旋律內容做分析，所以我們從所有的 MIDI 檔案中萃取和旋律相關（例如：音高、力度、音長等）的資訊。

## 2.2 歌詞資料庫

本文使用的歌詞資料庫是以KAR[4]檔案（內嵌歌詞的標準MIDI檔案）轉換出的資訊。我們從所有的KAR檔案中萃取音樂和歌詞兩部份，並將歌詞先經斷詞系統，所謂的斷詞系統[5]是指把輸入的字串分隔成詞串，詞串可以為單字詞或多字詞，例如：輸入詞串為『我到清華大學聽演講』，必須產生正確的字串為『[我][到][清華大學][聽][演講]』，而我們將斷詞後的結果，依照詞串的長度分為二字詞、三字詞、四字詞建立情緒資料表，並統計該情緒字串的詞頻且寫入資料表，再根據情緒字串以人工建立同義詞詞表。由於歌詞和音樂所要表達的情緒可能有所不同，因此我們分別針對音樂以及歌詞兩部分進行標準答案的人為標記。

## 2.3 分類器

本論文使用下列三種分類器進行分類，其簡述如下：

### 2.3.1 KNNR

最近鄰居法則(Nearest Neighbor Decision Rule)是指擁有相似特徵的資料，在以其特徵形成的空間中會聚集在一起。也就是說如果把要分類的特徵以高度空間來表示，則屬於同一類的點應該會距離比較近。因此對於一筆未知類別的資料，欲得知其所屬的類別，我們會先將特徵取出，再計算其和訓練資料特徵的距離，我們會判定資料的類別和最接近的點的類別是一樣的。當資料的雜訊較大時，只使用最接近的資料點來判斷可能會失之武斷，因此我們通常會使用 KNNR 進行分類。所謂的 KNNR (K Nearest Neighbor Rule) 是指先求取最接近的 K 個資料點，再根據對應的 M 個類別資訊來決定最後的類別，在此  $M \leq K$ 。

### 2.3.2 GMM

如果將每一筆資料視為在高維空間中的一點，而這些同一類別的資料點都是由一個高維高斯機率密度函數所產生，就可以用最佳參數估計法 (Maximum Likelihood Estimate) 來求出這個高斯密度函數的最佳參數值。但若資料的分布不是橢球狀，便無法用單一的高斯模型來模擬資料的分布，此時便要用數個高斯模型的加權平均來表示。這種方式就稱為高斯混合模型(Gaussian Mixture Model)。因此進行分類時，當高

斯模型的加權平均越高，則其屬於該類別的可能性越高。

### 2.3.3 SVM

支向機(Support Vector Machine)主要用來處理兩類別的問題。而支向機的目標是要在高維度的特徵空間，找出一個區分平面(Separating Hyperplans)來區分兩類資料。

## 2.3 歌詞計算方法

### 2.3.1 計算方法一

將測試資料經過斷詞處理拆解成二字詞、三字詞、四字詞，以這些詞組對資料庫做搜尋，會得到一個二到四字詞對快樂、焦慮類別的詞頻對應表，將其乘上一個權重，全部加總後分數最高者，則為該類。在本論文中，二字詞權重為 2;三字詞權重為 3;四字詞權重為 4。其計算公式如下：

- 1.計算對資料庫搜尋所得到的情緒詞分數

$$score = f_{d,t} * W_i$$

$f_{d,t}$  = 詞 t 在類別 d 出現的次數

$W_i$  = 詞 t 所對應的權重

- 2.將所計算出的情緒詞分數相加

$$score = \sum_{j=1}^j score_j$$

$j = 1 \sim N$  (N 為搜尋到的情緒詞總數)

### 2.3.2 計算方法二

和計算方法一相似，但計算方法將每一個詞都視為同等重要，因此不乘上詞頻，將對資料庫搜尋之後的結果乘上一個權重。在本論文中，二字詞權重為 2;三字詞權重為 3;四字詞權重為 4。其計算公式如下：

- 1.計算對資料庫搜尋所得到的情緒詞分數

$$score = 1 * W_i$$

$W_i$  = 詞 t 所對應的權重

- 2.將所計算出的情緒詞分數相加

$$score = \sum_{j=1}^j score_j$$

$j = 1 \sim N$  (N 為搜尋到的情緒詞總數)

### 2.3.3 計算方法三

計算方法三是利用 Managing Gigabytes[6]一書中所提到的 TF\*IDF rule (Term Freq., Inverse Document Freq.)，其計算方法如下

1. 計算對資料庫搜尋所得到的情緒詞分數

$$score = f_{d,t} \cdot \log \frac{N}{f_t}$$

$f_{d,t}$  = 詞  $t$  在類別  $d$  中出現的次數

$N$  = 總類別數

$f_t$  = 出現過詞  $t$  的類別數

2. 將所計算出的情緒詞分數相加

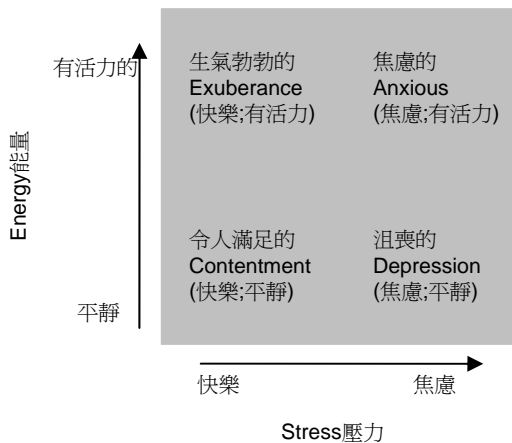
$$score = \sum_1^j score_j$$

$j = 1 \sim N$  ( $N$  為搜尋到的情緒詞總數)

這樣做法的意義在於，通常衡量重要性，皆是以該詞在文件內出現的次數做為決定性因子(TF 的含義)，但若該詞同時出現在多篇文件內，相對而言該詞比出現在少數文件的詞彙較不具價值(IDF 的含義)。

### 3. 情緒分類模型

要分類情緒之前需要先建立好情緒分類的模型。本篇論文是採用 1990 年 Thayer[7]所提出的二維情緒模型為基礎。如圖三所示



圖三：Thayer 情緒模型

Thayer 的情緒模型主要著重兩個特徵：能量 (Energy) 和壓力 (Stress)，因此在本篇論文的情緒模型中我們將能量用兩個情緒詞(「有活力(Energetic)」, 「平靜(Calm)」)來表示；壓力用兩個情緒詞(「快樂

(Happy)」, 「焦慮(Anxious)」)來表示。因此我們產生四種不同的情緒，其表示如下(表一)：

能量 Energy	壓力 Stress	情緒
快樂 Happy	有活力 Energetic	生氣勃勃的 Exuberance
焦慮 Anxious	有活力 Energetic	焦慮的 Anxious
快樂 Happy	平靜 Calm	令人滿足的 Contentment
焦慮 Anxious	平靜 Calm	沮喪的 Depression

表一：情緒模型

## 4. 特徵選取

### 4.1 古典音樂特徵

在古典音樂中由於時代以及當時環境的影響，作曲方式較嚴謹，很多作曲家都會承襲特定的風格或是曲式，我們要從古典音樂中找出情緒的特徵，會透過分析旋律所構成的特徵，例如：速度，調性等。而有些特徵(例如：音高、音長)可以利用統計的方式從 MIDI 檔案中取得，有些特徵(例如：調性)則需要透過演算法來計算。我們採用表二所列做為古典音樂的情緒特徵

古典音樂特徵	
速度	調(24 大小調)
音高平均	調性(大調小調)
力度平均	拍號

表二：古典音樂情緒特徵

#### 4.1.1 調性

一首古典音樂作品的調性對於情緒有很大的影響力，例如大調大多有活潑快樂的感覺；小調多有平靜哀傷的感覺，因此調性成為辨識古典音樂作品情緒一個很重要的特徵。本篇論文我們先採用 Krumhansl 等 [8]在 1982 年所提出計算調的演算法，其對 24 個調的權重定義如表三，再由所計算出的調來定義其屬於大調或是小調。其計算方法如下：

$$T = \arg \max \sum_{i=1}^N \frac{S^t_i}{N}$$

$S^t_i$  表音符對應表三所得到的權重(假設該片段有 I、II 兩音，則我們可得  $S^t_i = 6.35 + 2.23$ )

N 表輸入計算音樂片段的音符數目

調	I	II	III	IV	V	VI
大調	6.35	2.23	3.48	2.33	4.38	4.09
小調	6.33	2.68	3.52	5.38	2.6	3.53
調	VII	VIII	IX	X	XI	XII
大調	2.52	5.19	2.39	3.66	2.29	2.88
小調	2.54	4.75	3.98	2.69	3.34	3.17

表三：Krumhansl 24 個調對應權重表

計算時需要針對不同的調做移調的步驟，假設今天我們要計算該樂曲屬於 D 大調的分數，則我們需要把 D 視為 I、E 視為 II，依此類推，再對照表三去計算，最後找出 24 個調分數最高者，則判定其屬於該調。

## 4.2 流行音樂特徵

流行歌曲也受到時代以及文化的影響，現今的作曲方式比較不會遵守過去嚴謹的作曲形式，作曲的人有些甚至沒有受過正統的音樂訓練，因此我們較難從旋律所構成的特徵來做為辨識情緒的依據。但在編曲形式方面，卻可以從目前台灣編曲的風格找出特定的規則，因此在流行音樂的部份，除了和古典音樂相同的特徵（速度、音高）之外，我們也加入了可以從編曲形式中統計出來的特徵（鼓組出現的時間）。我們採用表四所列做為流行音樂的情緒特徵

流行音樂特徵	
速度	鼓組出現的時間
音高平均	鼓組密度

表四 流行音樂的情緒特徵

### 4.2.1 鼓組編曲形式

現今流行音樂可以分為前奏(Intro)、主歌 (Verse)、副歌 (Chorus)、音樂過門及結尾 (Instrumental and Ending) 四個部份。一般的歌曲大多作 AA'BA'的曲式。A 代表主歌，而 B 段是副歌。也就是一首歌的構造是由前奏，兩段主歌 AA'，一段副歌 B，過門音樂，再重覆到主歌 A'副歌 B，以及結尾音樂順序地連接而成的。在編曲方面，通常較活潑輕快的歌曲會在 A 就編制鼓組且鼓點較密集；而較平靜哀傷的歌曲會在 B 或是第二次的 A'才編制鼓組且鼓點較鬆散。因此我們統計第一拍鼓組在歌曲中所出現的時間做為特徵。

### 4.2.2 歌詞特徵

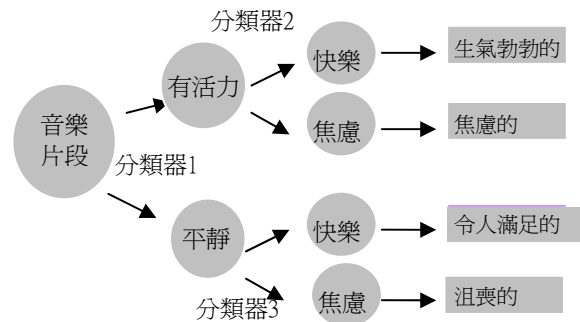
流行音樂中歌詞可以代表整首歌要給聆聽者的感覺，因此從歌詞中可以找到代表歌曲情緒的特徵。我們先從已分類好的訓練資料中透過斷詞，統計出每個具有情緒的字彙出現的頻率，並透過人工建立同義詞詞表，最後完成情緒詞彙資料庫。

我們所建立的情緒資料庫包含快樂 182 個情緒詞彙(例如:感動、戀愛、快樂等)以及焦慮 217 個情緒詞彙(例如:眼淚、傷心、分離等)。

## 5. 情緒辨識

### 5.1 階層式情緒分類

由於我們採用 Thayer[7]所提出的情緒模型，其使用二維的情緒詞彙組合(「有活力(Energetic)」、「平靜(Calm)」);(「快樂(Happy)」、「焦慮(Anxious)」)，分出四類的情緒，因此在本篇論文中，我們使用階層式的分類方法來進行分類。其表示如圖四：



圖四 階層式情緒分類

### 5.2 歌詞輔助情緒分類

由於歌詞可以表達出作詞者想要闡述的意境，因此一首流行歌曲的曲和詞可謂密不可分。我們希望可以結合分析歌詞和分析歌曲所得到的特徵，但因為歌詞部分在「能量」類別區分不明顯，在此我們只將歌詞以「壓力」類別區分，也就是用{‘快樂’}以及{‘焦慮’}兩個詞彙來描述。當有某一類別的權重比例較高時，代表其重要性較高，而實驗也證明當某一類別有較高的權重比例時，其正確性較高，因此我們設計了下列的計算流程：

1. 計算由歌詞內容所分析出的兩類權重比例相減之值，得到一門檻值。
2. 若符合門檻值者則保留以歌詞分析計算出的答案。
3. 未符合門檻值者，我們音樂內容的答案做為辨識結果。
4. 在此門檻值我們透過實驗，使用暴力法找出最佳值。

其強度運算流程如下

計算透過歌詞內容分類方法所得的百分比：

$$lyric\_score(j^i) = 100 * \frac{score_j}{\sum_1 score_j}$$

$j = 1, 2$  ( $j$  表快樂、焦慮兩類別)

$i = 1, 2, \dots, 50$  ( $i$  表測試資料數目)

## 6. 實驗

### 6.1 實驗環境

#### 6.1.1 古典音樂

(1) 測試平台：Windows XP (Pentium4 2.4G/512MB)

(2) 訓練資料：

- 來源：由網路上蒐集的 MIDI 格式歌曲再經程式處理轉為便於比對的格式後存於資料庫中
- 數目：200 首(92 首快樂、108 首焦慮)
- 內容：包含巴洛克樂派、古典樂派、浪漫樂派時期的作品
- 屬性：MIDI 包含單軌多軌
- 答案：由一位有音樂背景者根據聆聽結果標記四類情緒

(3) 測試資料：

- 來源：由網路上蒐集的 MIDI 格式歌曲再經程式處理轉為便於比對的格式後存於資料庫中
- 數目：50 首(22 首快樂、28 首焦慮)
- 內容：包含巴洛克樂派、古典樂派、浪漫樂派時期的作品
- 屬性：MIDI (包含單軌與多軌)

#### 6.1.2 流行歌曲

(1) 測試平台：Windows XP (Pentium4 2.4G/512MB)

(2) 訓練資料：

- 來源：由網路上蒐集的 KAR 格式歌曲再經程式處理轉為便於比對的格式後存於資料庫中。
- 數目：200 首(98 首快樂、102 首焦慮)
- 內容：包含近年發行過的流行專輯歌曲
- 屬性：多軌
- 答案：由一位有音樂背景者根據聆聽結果標記四類情緒

(3) 測試資料：

- 來源：由網路上蒐集的 KAR 格式歌曲再經程式處理轉為便於比對的格式後存於資料庫中
- 數目：50 首(18 首快樂、32 首焦慮)
- 內容：包含近年發行過的流行專輯歌曲

- 屬性：多軌

#### 6.1.3 歌詞

(1) 測試平台：Windows XP (Pentium4 2.4G/512MB)

(2) 訓練資料：

- 來源：由上述 200 首 KAR 檔案抽取出的歌詞檔案，經處理為便於比對的格式存於資料庫中
- 數目：200 首(98 首快樂、102 首焦慮)
- 內容：包含近年發行過的流行專輯歌曲
- 屬性：TXT 文字檔
- 答案：由一位有音樂背景者根據聆聽結果標記兩類情緒，由於歌詞可能會和歌曲產生不同情緒，在此我們會另外針對歌詞標記答案。

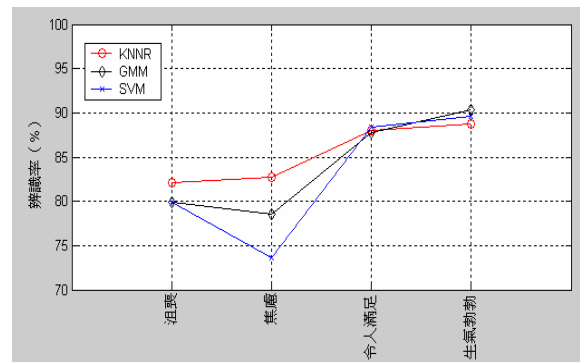
(3) 測試資料：

- 來源：由上述測試資料 50 首 KAR 檔案抽取出的歌詞檔案，經處理為便於比對的格式存於資料庫中
- 數目：50 首(18 首快樂、32 首焦慮)
- 內容：包含近年發行過的流行專輯歌曲
- 屬性：TXT 文字檔

## 6.2 實驗結果

### ● 古典音樂

使用三種方法 KNNR、GMM、SVM 表示如圖五；三種分類器辨識結果表示如表五：



圖五：三種分類方法對古典音樂分類結果

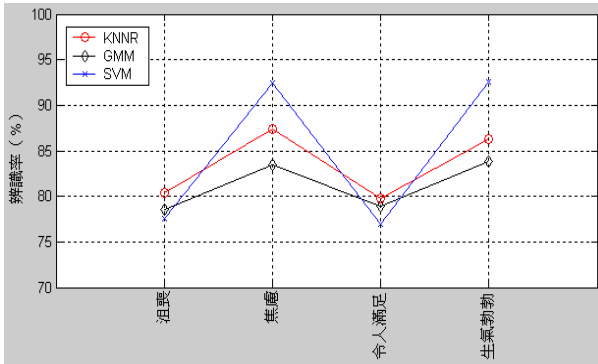
<b>KNNR</b>	85.56%
<b>GMM</b>	84.34%
<b>SVM</b>	83.16%

表五：三種分類器分類結果

### ● 流行歌曲

# 2006 International Workshop on Computer Music and Audio Technology

使用三種方法 KNNR、GMM、SVM 表示如圖六；三種分類器辨識結果表示如表六。



圖六：三種分類法對流行歌曲的分類結果

<b>KNNR</b>	82.44%
<b>GMM</b>	80.47%
<b>SVM</b>	79.11%

表六：三種分類器分類結果

## ● 歌詞

由於實驗結果得知使用 KNNR 以及計算方法一辨識結果最佳，因此我們利用 KNNR 分類器再結合計算方法一的歌詞分類做為測試。表七表只聽音樂的標記答案；表八表只看歌詞標記答案；表九表同時聽音樂加歌詞所標記的答案；圖七為三種計算方法對歌詞分類結果；圖八為歌詞權重比例相減門檻值的辨識率；圖九為使用歌詞計算結果所保留歌曲佔全部歌曲的比例；圖十為 KNNR 結合歌詞分類結果。

從表七我們得知若只聽音樂為標記答案，以歌詞所計算出的結果較不具效益；但以只看歌詞做為標記答案時，音樂和歌詞可以得到較平均的辨識率，由此可推論出當聆聽者只聽音樂時，所感受到的情緒會和作詞者所要表達的情緒差異較大，因此當聆聽者在判別一首歌曲的情緒時，可能會受到歌詞影響比較大；此外，從圖八可以看出當兩類權重比例相減所得的值越大，則正確率越高，由此可推論出當某一類別有較強權重比例時，則其計算出的答案較可靠，但由圖九可知，當權重比例相減差異越大時，符合相減門檻值的歌曲數目會遞減。圖十我們使用暴力法找到最佳的門檻值，其值為 93，保留歌曲比例從圖九可看出約為 30%，辨識率可由使用音樂內容計算出的 86% 提升到 88%。

	快樂	焦慮	平均
--	----	----	----

音樂內容	92.52	82.35	88.02
歌詞內容	84.00	60.00	72.00

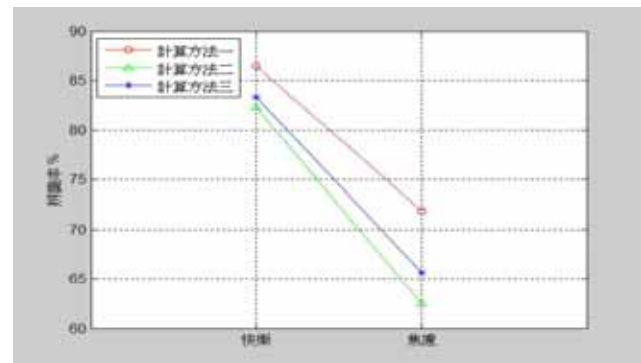
表七：只聽音樂的標記答案

	快樂	焦慮	平均
音樂內容	94.12	72.73	80.00
歌詞內容	88.24	87.88	88.00

表八：表只看歌詞標記答案

	快樂	焦慮	平均
音樂內容	90.00	80.32	86.00
歌詞內容	84.38	77.79	82.00

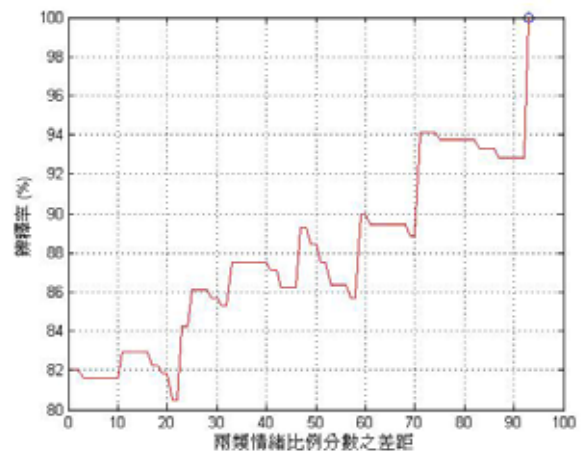
表九：以歌詞加音樂為標記答案



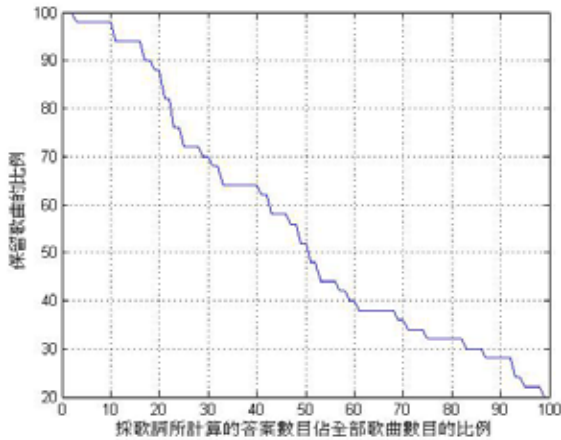
圖七：三種計算方法對歌詞分類結果

計算方法一	88.00
計算方法二	78.00
計算方法三	82.00

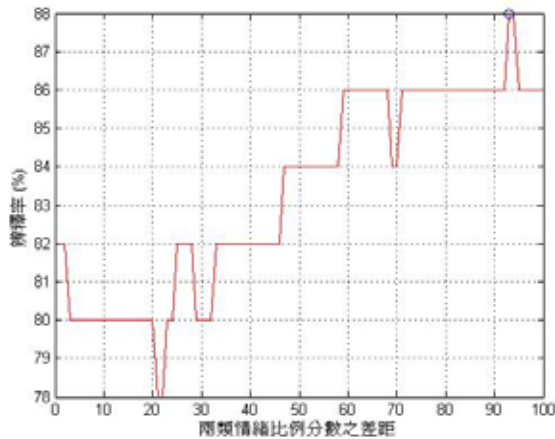
表十：三種計算方法對歌詞分類結果



圖八：歌詞權重比例相減門檻值的辨識率



圖九:使用歌詞計算結果所保留歌曲佔全部歌曲的比例



圖十：KNNR 結合歌詞分類結果

## 6.3 錯誤分析

### 6.3.1 音樂內容

1. MIDI 輸入單位有誤：從網路上收集到的 MIDI 檔案大多不是以採譜為目的而製，很多 MIDI 檔案是以聆聽為考量，因此在節拍單位常出現和原曲單位不一致的情況。例如：節拍為 100 的 4 分音符可被記錄為 50 的 8 分音符，所以當我們在萃取節拍特徵時，便會產生錯誤。
2. 人對情緒的感覺不同：由於心情會影響一個人聆聽一首歌的感覺，因此同一首歌對於不同的人，可能會產生不同的答案。

### 6.3.2 歌詞內容

1. 有隱藏涵義：許多歌詞都有隱藏的涵義，例如「還記得不停流著眼淚，對天空喊著，我恨你的那年冬季，雖然當時心情還清晰，關於恨早隨他遠去」，其作詞人是要描述揮別過去的戀情，重新開始期待下一次的戀愛，應該是正向的情緒，但

是就文句來看，負向的情緒字彙卻比正向強烈，因此會發生判斷錯誤。

2. 歌曲和歌詞相反：約佔 20%。例如五月天的志明與春嬌，歌曲部分表現的是生氣勃勃情緒，但是歌詞描述是分手的情緒，兩者產生矛盾。
3. 情緒詞彙分佈情況與假設不同：例如梁靜茹的分手快樂，歌詞中根據斷詞會大量出現「分手」、「快樂」造成正面和負面的情緒分數相近，而發生分類錯誤的情況。

## 7. 結論

本篇論文我們使用了階層式的分類法針對不同時代的音樂作品的音樂內容進行情緒分析以及辨識，並加入歌詞來提高單從音樂內容分析情緒的辨識率。此外，我們也比較了 KNNR、GMM、SVM 三種分類器的效能。

而未來我們希望可以從音樂內容中找出其他可以代表情緒的特徵，例如音色、教會調式、和弦進行走向；在歌詞方面，能加入程度詞的判斷，以提高辨識率；而在結合歌詞來輔助音樂內容的分類實驗中，我們要改進計算歌詞權重的方法，找到更適合結合其和音樂的方法。

## 8. 參考

- [1] Krumhansl, C.L. (2002). Music: a link between cognition and emotion. *Current Directions in Psychological Science*, 11(2), 45-50.
- [2] Dan Liu, Lie Lu, Hong-Jiang Zhang, "Automatic Music Mood Detection from Acoustic Music Data", *Proc. of International Symposium on Music Information Retrieval 2003 (ISMIR03)*, pp. 81-87, Baltimore, MD, Oct. 26-30, 2003
- [3] Wang Muyuan; Zhang Naiyao; Zhu Hancheng;"User-adaptive music emotion recognition", *Signal Processing*, 2004. *Proceedings. ICSP '04. 2004 7th International Conference on Volume 2*, 31 Aug.-4 Sept. 2004 Page(s):1352 - 1355 vol.2
- [4] <http://www.notation.com/midifiles.htm>
- [5] 朱怡霖 (2002) 中文斷詞與專有名詞辨識之研究, 碩士論文, 國立台灣大學資訊工程所, 台北, 2002.
- [6] Ian H. Witten, Alistair Moffat, Timothy C. Bell, "Managing Gigabytes: Compressing and Indexing, Documents and Images", Van Nostrand Reinhold, New York, 1994

## 2006 International Workshop on Computer Music and Audio Technology

- [7] Thayer, R. E. (1989). *The biopsychology of mood and arousal*. Oxford University Press.
- [8] Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, 334-368