

# 以音樂特徵為音樂資料之分類

羅有隆  
朝陽科技大學  
資訊管理系  
yllo@cyut.edu.tw

林奕昌  
朝陽科技大學  
資訊管理系  
s9714613@cyut.edu.tw

## 摘要

已被發展出的音樂分類技術可概分為兩個方向，其一為以學習機器為訓練基礎的分類方式，另一則為以音樂內容為基礎的分析分類方式，各有其優缺點。而其中，現有以音樂內容為基礎的分類方式，多以音樂之單特徵進行，在少數類型的音樂辨識可達70%左右的分類精確度，而多數類型的音樂，可正確被分類的精確度則低了許多。本研究報告在於分析與研究音樂資料內容，並利用音樂資料的多特徵性質，找出是否有每一類型音樂所特有的內容特徵，並以這些特徵設計出更高精確度的音樂資料分類技術。在我們的實驗中，我們發現所設計的技术利用音樂內容之多重特徵為分類基礎，可讓鄉村音樂、古典音樂、流行音樂與爵士音樂，分別達到85%、82%、80%與73%的分類精確度。

**關鍵字：**多媒體資料庫、音樂資料庫、數位音樂、音樂分類、內容擷取

## Abstract

The music classification techniques can be discriminated into two categories — based by music feature classification and training by learning machine classification. Both have their advantages and disadvantages. For music feature classifications, most of the approaches are based on single music feature, such as melody or chord, and the accuracy is about 70% in few genres of music. However, the accuracy for classification of most music genres is lower. In this research, we study the music contents and use the multi-features of music to design equation for more accuracy music classification. Our performance study shown that more than 85%, 82%, 80%, and 73% of folk, classic, pop, and jazz music can

be classified correctly, respectively, by using multi-feature of music content for classification.

**Keywords:** multimedia database, music database, digital music, music classification, content-based retrieval.

## 1. 導論

隨著行動式數位多媒體播放器(如：iPod、iPhone、iPad)的廣受歡迎，數位音樂的需求大增，對於數位音樂資料庫的應用也日趨熱門。音樂資料庫研究的相關議題有音樂的分類(classification)、特徵擷取(feature extraction)、多特徵索引(multi-feature indexing)、以及近似搜尋(approximate searching)等等，主要目的都是在幫助使用者能簡單而且快速的擷取到想要的樂曲[2][10][11][12][13][14]。

音樂可以陶冶性情，也可以安撫人心，使人心情獲得平靜。因此，有腦筋動得快的業者推出了所謂的音樂心情手機，例如：Sony Ericsson W508、W902與W910，就加入了心情點播器(SensMe)的功能，主要是將每一首歌依照歌曲的快節奏、慢節奏或是快樂、憂傷分為四個象限，聆聽音樂時，使用者就能以手機內心情點播器依照當時的心情，播放想要聆聽象限內的歌曲。但是，有關歌曲的分類，則還需要使用者自己建立音樂的分類清單，才能使用，使得人性化功能大打折扣。近年來有關音樂資料自動分類的研究，也因此越來越受到重視[3][4][6][7][10][15][16][19][20][22]。

由於音樂資料的內容(content)提供了多樣的特徵(features)可做為分析與查詢使用，例如：主旋律(key melody)、節拍(rhythm)與和弦(chord)等等，皆可呈現出每一首音樂獨特的曲風與特性。因此，以內容為主的音樂擷取(content-based music retrieval)，多年來是音樂資料庫研究的重要方向之一，而且也被當做音樂

分類的重要參考依據。相關研究例如：Lin 等學者的 music classification using significant repeating patterns [10]、Brecheisen 等學者的 hierarchical genre classification for large music collections [3]、以及 Cheng 等學者的 automatic chord recognition for music classification and retrieval [6]等，皆是探討以音樂內容來做音樂分類。不過，目前這些以內容為主的音樂分類方式，其分類之精確度，少部分音樂類型的分類精確度可達 70%左右，但是多數音樂類型的分類是低於此精確度，由此也顯示出現有的音樂分類技術尚有不少的進步空間。

本研究報告主要是利用音樂內容特徵 (music content feature) 之分析，設計公式，用於音樂之自動分類，以提升目前以音樂內容為主的分類精確度。我們期望所提出的方案，不僅對於音樂資料的正確分類，可以進一步的提昇，同時也希望這個研究的成果，在其他多媒體資料庫相關領域也可能有其它應用的價值。

## 2. 文獻探討

近年來音樂的自動分類方式，主要可概分為兩個研究方向：一個是以音樂內容為基礎 (music content-based) 的分析分類方式，是利用旋律 (melody)、節拍 (rhythm) 與和弦 (chord) 等等音樂特徵，作為分類的依據，如前所述 Lin 等 [10] 與 Brecheisen 等 [3]。另一個則是以 learning machine 為分類方式，利用貝式分析、線性分析和類神經網路等來建立音樂分類，如：Loh 等學者的 Extreme Learning Machine (ELM) [16] 與 Mandel 等學者的 Multiple-Instance Learning [17]。此外，Cheng 等學者的 automatic chord recognition [6] 則是以音樂特徵之一的和弦為分類依據，但也透過訓練等的學習機制。以下則簡要的介紹各音樂分類的方式：

### 2.1 SRP-Based Classification

Lin 等學者於 [10] 中利用 significant repeating patterns 來做音樂資料的分類，且以兩個於音樂內容擷取 (content-based music retrieval) 的特徵 (features)—旋律 (melody) 與節拍 (rhythm) 來代表音樂，以做為分類依據。重覆片段 (repeating patterns) 是指音樂中連續的片段，如：melody sequence 或 rhythm sequence，它重覆出現於音樂資料中至少 2 次或以上，這常常是自動判斷主旋律的依據。在很多音樂學及音樂心理學的研究中也認同重覆片段在音樂結構中為一普遍性的特徵 [9][18]，因為重覆片段的長度較實際音樂短，因此若以重覆片段

來表示實際音樂，則對以內容擷取式的音樂搜尋，將會使其效能大大提昇。作者將重覆片段做了一些的規範—maximum length、minimum length 與 minimum frequency 等，以產生 significant repeating patterns (SRP)，並提供方法來估算 SRP 於音樂分類的可用性 (usefulness)，以及與音樂類別的相關性。而人類的感知 (human perception) 與音樂理論 (musicology) 也被併入考量，以為做 SRP 比對時的相似計量。在經過 Usefulness of SRP for Classification、Similarity Measure for SRP Matching、以及 Class Determination 三個階段，透過公式計算，以最高分來決定一段音樂 (music piece) 是屬於那一類的音樂。

### 2.2 Hierarchical Genre Classification

Brecheisen 等學者於 [3] 中提出了階層式的曲風分類法 (hierarchical genre classification)，此方法可以處理音樂內容的多重特徵 (multiple features)，以達到較高的音樂分類準確度。此分類方法的基礎是使用到三種聽覺上 (acoustic realms) 的特徵—音色 (timbre)、節拍 (rhythm) 與音高 (pitch)。如此，一首音樂就可以此三特徵的多維向量 (multiple feature vectors) 來表示，而一段音樂就可以由一組特徵向量的集合來代表。接著將 feature vectors 做 hierarchical instance reduction，之後 reduced descriptions 再透過兩層的分類程式 (two layer classification process, 2LCP) 來處理階層分類問題，這過程還利用了 Support Vector Machines (SVM) [21] 做為分類者 (classifier)，經過訓練與歸類的程序。

### 2.3 Extreme Learning Machine, ELM

ELM 是用來解決傳統類神經網路 (traditional neural networks) 的問題，Loh 等學者於 [16] 提出了利用 ELM 來處理音樂資料的分類。ELM 技術處理出自於曲面的錯誤 (error surfaces) 的梯度下降 (gradient descent) 問題，而其倒傳遞學習演算法 (back-propagation learning algorithms) 已知是緩慢的，特別是當學習率參數小 (learning rate parameter small)，會增加收斂的時間。另一方面來說，較大的學習率，也可能形成無法收斂的網路。縱使全域的最低值 (global minimum) 比局部最低值 (local minimum) 都還低，Error Surface 演算法還是相當容易受到區域最小值 (local minima) 的影響，使得學習終止。ELM 也處理了過度學習 (over training) 的問題，增加了類神經網路的概括能力 (generalization capabilities)。

## 2.4 Multiple-Instance Learning

Mandel 等學者於[17]中分析了 multiple-instance learning 的音樂資料分類法。作者未提出新的技術，而是分析與比較 mi-SVM algorithm[1]與 Multiple-Instance Learning via Embedded Instance Selection (MILES)[5]，兩演算法應用於音樂分類之優劣。依據實驗的結果，其結論是 mi-SVM 優於 MILES，能較精確的做音樂資料的分類。

## 2.5 Chord Recognition

用來做音樂分類的 chord recognition system 是 Cheng 等學者於[6]所提出來的，它分為兩個階段。第一階段為 training phase，他們併用了 N-gram model 於 Hidden Markov Model(HMM)架構中[8]，做為和弦編曲(chord transcriptions)的訓練，以習得和弦發展(chord progression)的通則(common rules)。第二階段為 testing phase，利用前置訓練獲得的模式(pre-trained models)，chord sequence 可從 input 中被解碼(chord decoding)，得以做音樂分類。同時作者也提出了兩個新的和弦特徵—longest common chord subsequence 與 chord histogram，它們對音樂的分析、管理與擷取(retrieval)，都相當的有幫助。

上述的音樂資料分類技術，缺點在於精確度多為 70%左右或更低，或者需要先做費時的音樂資料訓練，以取得分類資訊。而本研究報告，即在於應用音樂資料的內容特徵，發展出更高精確度的音樂分類演算法。

## 3. 音樂分類公式設計

Lin 等學者於[10]中利用音樂資料的旋律(melody)與節拍(rhythm)等兩特徵，先找到 significant repeating patterns 來做音樂資料的分類，部分類型之分類精確度最高可達 70%左右，此支持了利用音樂資料的特徵，不需做資料訓練也可以做為音樂資料分類的依據，然而精確度還有待提升。因此，我們也繼續對音樂內容特徵的多樣性做分析，期望能有更精確的音樂分類方式。

我們使用的方法是找尋各別類型的音樂(如：classic、pop、folk 或 jazz)中可能有的音樂內容特徵性質，例如：那些音(notes)、那些節拍(rhythms)以及那幾種前後音差變化(pitch changes)發生頻率高與獨特性在某類的音樂中。我們的研究採取正向分析與逆向分析來進行，分別取樣(sampling)來計算發生頻率最高

(正向分析)與發生頻率最低(逆向分析)的  $n$  個特徵值  $x_1, x_2, \dots, x_n$ ，再以此  $n$  個特徵值當做多維空間中的中心座標  $(x_1, x_2, \dots, x_n)$ 。如此，一首音樂可透過這相對的  $n$  個特徵值  $y_1, y_2, \dots, y_n$ ，加上應有的權重  $(w_i)$ ，以計算此音樂距離中心座標遠或近，即可判斷其可能歸屬於那一類的音樂。而距離函式  $d(y_1, y_2, \dots, y_n)$  可設計如下：

$$d(y_1, y_2, \dots, y_n) = \sqrt{\sum_{i=1}^n w_i (x_i - y_i)^2} \quad \dots (1)$$

正向分析時為  $w_i=i$  正向分析的權重，是依每種類型音樂特徵的出現頻率的排名給予，例如古典音樂音符出現頻率的前三名分別為，Mi、Re 以及 So，我們給予第一名音符 Mi 的權重給予 1，第二名音符 Re 的權重給予 2 和第三名音符 So 的權重給予 3，而如果有排名第 4 名、第 5 名...等，則給予權重 4、5...等，以此類推，而在逆向時  $w_i=n-i+1$ ，逆向的權重，是依每種類型音樂特徵的出現頻率給予，例如古典音樂音符出現頻率的最後三名分別為低音 So、低音 Fa 以及低音 Mi，我們給予最後一名音符低音 So 的權重給予 3，倒數第二名音符低音 Fa 的權重給予 2 和倒數第三名音符低音 So 的權重給予 1，而如果有倒數排名第 4 名、第 5 名...等，則給予 4、3、2、1 的權重與 5、4、3、2、1 的權重，而不管是正向權重的給予還是逆向權重的給予，都是依每種類型的特徵出現頻率，出現頻率最高給予最小權重，出現頻率最低給予最大權重。給予這種權重方式的原因在於，如果再把出現頻率高的給予大的權重，而出現頻率低低的給予較小的權重，只會讓大的更大，小的雖然也有變大，但確不明顯，而出現頻率高的影響性就很大，而出現頻率低低的就會變的沒有什麼意義，所以才會給予，出現頻率較高的給予權重較小的，而出現頻率較低的給予權重較大的方式，其中權重都是為整數，且都是成一個等差，在設定權重時，我們有嘗試過許多的方式和數值，而本文採用的，為所測試過之最佳方式和數值。例如：統計某類音樂，其所屬音樂的旋律中，發生最高頻率的前三個音為 Mi(40%)、Re(30%) 以及 So(20%)，最低頻率的最後三個音為低音 Mi(5%)、低音 Fa(3%) 以及低音 So(2%)。我們可以這些音的發生頻率做為正向分析中心座標  $(x_1, x_2, x_3)=(0.4, 0.3, 0.2)$  與逆向分析中心座標  $(x_1, x_2, x_3)=(0.02, 0.03, 0.05)$ 。若有一首音樂之正向分析旋律(Mi, Re, So)的發生頻率為(0.25, 0.26, 0.24)與逆向分析旋律(低音 Mi, 低音 Fa, 低

音 So)的發生頻率為(0.025,0.02,0.03)則可以分別帶入公式(1)求得距離如下：

正向分析：

$$d(0.25,0.26,0.24) = \sqrt{1(0.4-0.25)^2 + 2(0.3-0.26)^2 + 3(0.2-0.24)^2} = 0.17464249$$

逆向分析：

$$d(0.025,0.02,0.03) = \sqrt{3(0.02-0.03)^2 + 2(0.03-0.02)^2 + 1(0.05-0.025)^2} = 0.03354102$$

當此音樂的正向分析特徵與逆向分析特徵分別嘗試過各類型音樂之距離分析後，我們可將該音樂做最大可能性的歸類。

進一步，我們也可以考量同時以多個特徵來運算，加總於各特徵所計算的距離，以嘗試求得更精確的分類，假設有  $k$  個特徵，則多特徵的距離公式如(2)。其中  $md(k)$  為  $k$  個特徵的距離加總， $y_{1j}, y_{2j}, \dots, y_{nj}$  為第  $j$  特徵的  $1\sim n$  個特徵值之頻率。

$$md(k) = \sum_{j=1}^k d_j(y_{1j}, y_{2j}, \dots, y_{nj}) \quad \dots(2)$$

例如：統計某類音樂，其所屬音樂旋律中，發生最高頻率的前三個音符與前後音差，分別為 Mi(40%)、Re(30%)以及 So(20%)與升1度音之+1(40%)、降1度音之-1(20%)以及降2度音之-2(10%)，我們可以此各別三個音符與三個前後音差的發生頻率，做為音符正向分析中心座標  $(x_1, x_2, x_3)=(0.4, 0.3, 0.2)$  與前後音差正向分析中心座標  $(x_1, x_2, x_3)=(0.4, 0.2, 0.1)$ 。若有一首音樂之正向音符(Mi, Re, So)的發生頻率為(0.25, 0.26, 0.24)與正向前後音差(+1, -1, -2)的發生頻率為(0.34, 0.17, 0.16)，則可以利用公式(2)求得音符+前後音差之正向分析距離如下：

$$md(2) = \sqrt{1(0.4-0.25)^2 + 2(0.3-0.26)^2 + 3(0.2-0.24)^2} + \sqrt{1(0.4-0.34)^2 + 2(0.2-0.17)^2 + 3(0.1-0.16)^2} = 0.30192171$$

#### 4.精確度分析

於本節當中，我們嘗試先以公式(1)與公式(2)用來做古典音樂(classic)、流行音樂(pop)、鄉村音樂(folk)與爵士音樂(jazz)的分類實驗，並採用音樂旋律中的音符(note)、節拍(rhythm)以及前後音差變化(pitch change)來做為音樂內容的特徵，分析其所發生的頻率，以距離公式

來判別可能歸屬的音樂類別。而會採用前後音差變化，是因為它比較不受使用者的升音或降音(key)所影響，較適合音樂擷取的以樣本查詢(Query By Example)。

於實驗的音樂資料庫中，我們先對古典音樂、流行音樂、鄉村音樂與爵士音樂各建立 150 首真實音樂資料，使得資料庫中有完整 600 首真實音樂的音符、節拍與前後音差變化之資料。實驗以統計最高頻率之 3~7 個特徵值(n)以建立中心座標。公式(1)中的權重( $w_i$ )，仍設正向分析時為  $w_i=i$ ，而在逆向時  $w_i=n-i+1$ ，因為經我們實驗分析，這樣可以有較佳之實驗結果。實驗分析可分為兩個部份：

- (1)單特徵分類分析
- (2)多特徵分類分析

#### 4.1 單特徵分類分析

##### 4.1.1 單特徵正向分析對分類的影響

此實驗以音樂內容之正向單一特徵來做音樂分類，以探討其對分類之精確性的影響。首先我們分別對古典音樂、流行音樂、鄉村音樂與爵士音樂之音符資料取樣特徵值數(n)做分析，分別統計出最高頻率的 3~7 個音符，以做為中心座標，如圖 1 為古典音樂裡出現頻率最高的前 20 音符與出現頻率。而表 1 為頻率最高的七個音與出現頻率，分別為高音 Do、高音 Re、中音 Si、中音 La、高音 Mi、中音 So 與高音 Fa。因此，也可取平均之中心座標為(0.0847, 0.0821, 0.0776, 0.0759, 0.0746, 0.0729, 0.0677)。由於古典音樂、流行音樂、鄉村音樂與爵士音樂曲風上的差異，代表四類音樂之中心座標之音符與頻率，也就會有所不同。因此，我們可以透過距離公式(1)，來計算一首音樂與中心座標的距離，以分析它可能較接近那一類型的音樂。

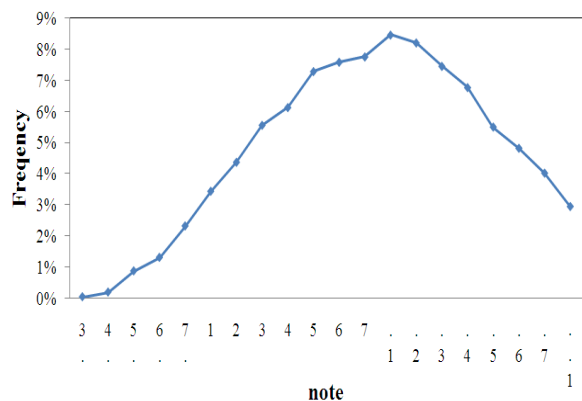


圖 1. 古典音樂音符分布圖

表 1. 古典音樂音符出現頻率前 7 名

排名	音符	出現頻率
1	高音 Do	0.0853
2	高音 Re	0.0814
3	中音 Si	0.0772
4	高音 Mi	0.0754
5	中音 La	0.0744
6	中音 So	0.0729
7	高音 Fa	0.0677

有了中心座標後，接著我們將資料庫中的 600 首音樂，逐一以音符的發生頻率來分類，統計可成功分類到正確音樂類型的精確度 (Accuracy) 是多少。統計方式是將每首音樂，逐一以同古典音樂最高頻之 2~7 個特徵值(n)之發生頻率，以公式(1)計算與古典音樂中心座標的距離。同樣的，再將這首音樂以流行音樂、鄉村音樂與爵士音樂之最高頻之 3~7 個音符，也計算出流行音樂、鄉村音樂與爵士音樂中心座標之距離。如此，可將此首音樂歸類於較短距離之音樂類別。以音符特徵為分類之實驗結果如圖 2 所示。圖中，對古典音樂之音符分類精確度，約在 80% 左右，明顯優於對流行音樂、鄉村音樂與爵士音樂分類之精確度。而流行音樂、鄉村音樂與爵士音樂精確度之起伏較大，且流行音樂在特徵值數為 3、4 與 5 有著差不多的精確度，而鄉村音樂與爵士音樂在特徵值數為 5、6 與 7 時非常接近，三曲線相當接近，表示流行音樂、鄉村音樂與爵士音樂中之音符特徵出現頻率變化較大，不若古典音樂來得穩定，且精確度未能達到有效的結果。

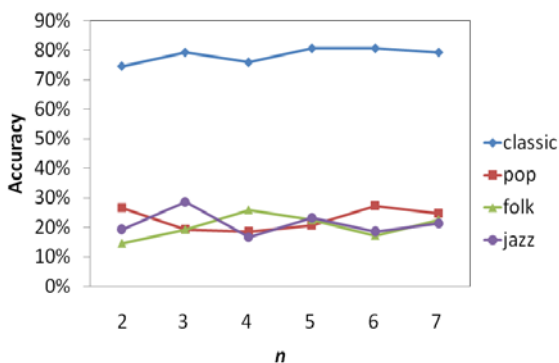


圖 2. 以音符特徵值正向分析分類之精確度

進一步利用單特徵-音符正向分析，以特徵值數目 n=2 與二維中心座標，取得各類音樂可涵蓋 90% 其所屬音樂之半徑如表 2 所示，並進一步標示於二維空間中，如圖 3。我們很容易的發現，古典音樂的半徑為最小。由此可以說明，以音符特徵做正向分析，古典音樂的分佈

較為集中於中心點的附近，也因此以音符來做音樂分類，古典音樂因與中心點的距離較短，可以有較高的準確度被分類。反觀流行音樂、鄉村音樂與爵士音樂，有著較大的半徑，意謂著所屬音樂分佈的較不集中，在每個圓多所重疊的情形下，容易被規類到距離較近的中心點，但卻是錯誤的分類。

表 2. 單特徵-音符正向分析半徑與圓心座標 (比率 90% 與 n=2)

音樂類型	半徑	二維圓心點座標
Classic	0.0079	(0.0011, 0.0014)
Pop	0.015	(0.0039, 0.0016)
Folk	0.035	(0.0094, 0.0042)
Jazz	0.032	(0.0074, 0.0043)

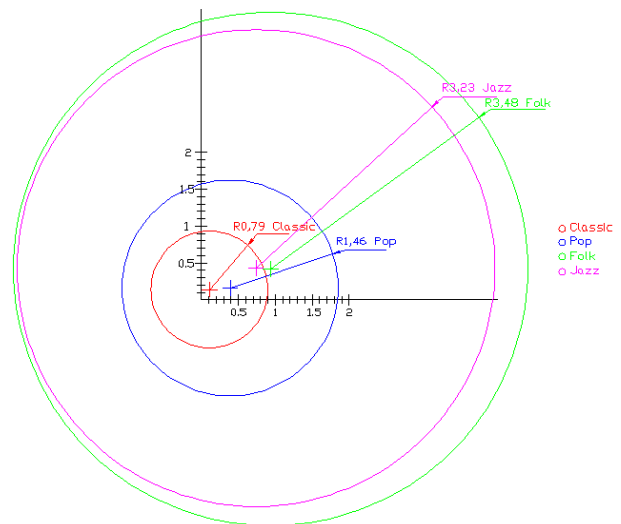


圖 3. 單特徵-音符正向分析(n=2)二維圓心圖

以上說明採各類可涵蓋 90% 其所屬音樂之半徑，主要目的在於排除各類音樂可能有部份音樂具較特殊或例外之特性。而我們也對於音符特徵值數目為 5 (n=5) 持續的實驗，並採可涵蓋同類音樂 100%、90%、80% 與 70% 之比率，分析其半徑與所涵概之音樂分配分率，結果如表 3 所示，在表 3 中，可得知半徑、分配比率與涵蓋比率的相關性，當涵蓋比率愈低，而四種音樂類型的半徑愈小，可排除的特殊音樂愈多，然而對於分配比率，古典音樂的分配比率會愈高，相對的，流行音樂、鄉村音樂與爵士音樂的分配比率卻愈低，從此就可以驗證，上述所說的，有較佳精確度的音樂類型，會分佈在圓心週圍。也由此可說明，單特徵音符的正向分析，可能比較適合於古典音樂的分類，而不適合於流行音樂、鄉村音樂與爵士樂的分類。



表 3.單特徵音符正向分析半徑與精確度(n=5)

	100%		90%		80%		70%	
	半徑	分配比率	半徑	分配比率	半徑	分配比率	半徑	分配比率
Classic	0.11	80.67	0.04	81.48	0.03	83.33	0.02	83.81
Pop	0.11	12.00	0.08	11.11	0.06	10.00	0.05	9.52
Folk	0.18	2.67	0.13	2.22	0.10	2.50	0.09	2.86
Jazz	0.13	4.67	0.09	5.19	0.07	4.17	0.06	3.81

接著我們以相同的實驗方式，繼續以節拍與前後音差變化之音樂內容單特徵正向分析來做音樂分類，實驗結果如圖 4、圖 5 與表 4、表 5。在圖 4 中，討論節拍正向分析的部份，對爵士音樂之節拍分類精確度，約在 70%~75% 之間，明顯優於古典音樂、流行音樂與爵士音樂。而古典音樂分類之精確度，則是明顯不如以音符來做為音樂分類，但爵士音樂明顯優於以音符來做音樂分類。而流行音樂與鄉村音樂以節拍之分類精確度，與以音符分類之精確度，表現得相當的相似。

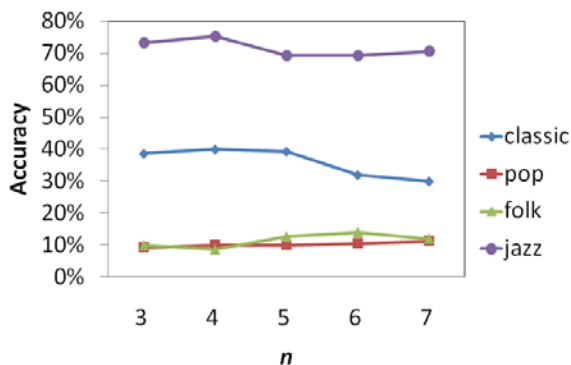


圖 4. 以節拍正向分析分類之精確度

表 4 中，討論節拍正向分析，涵蓋比率、半徑跟分配比率的關係，四種類型的音樂，隨著涵蓋比率的降低，半徑跟隨著縮小，對於分配比率，流行音樂與爵士音樂隨著涵蓋比率的下降，分配比率逐漸的上升，但對於爵士音樂表現的較佳，而對於古典音樂與鄉村音樂，則是隨著涵蓋比率的下降，分配比率也隨下降。

表 4.單特徵節拍正向分析半徑與精確度(n=5)

	100%		90%		80%		70%	
	半徑	分配比率	半徑	分配比率	半徑	分配比率	半徑	分配比率
Classic	1.93	6.00	0.56	4.44	0.41	3.33	0.34	2.86
Pop	1.57	14.00	0.57	15.56	0.36	17.50	0.25	17.14
Folk	1.34	10.67	0.49	8.89	0.32	6.67	0.23	5.71
Jazz	1.38	69.33	0.45	71.11	0.28	72.50	0.19	74.29

於圖 5，以前後音差變化正向分析來做音樂分類之精確度比較，古典音樂與爵士音樂之分類精確度，都有明顯的起伏，沒有鄉村音樂與流行音樂來的穩定。古典音樂、流行音樂、鄉村音樂與爵士音樂在特徵值增加，對於精確度來說，並未漸增，反而有遞減的狀況，而古典音樂表現最佳，鄉村音樂稍優，而流行音樂與爵士音樂，表現的相當相似。整體而言，以單特徵正向分析來做音樂分類，於古典音樂效果較佳，且以音符之分類方式最佳，而對於爵士音樂，以節拍之分類方式最佳，表現略遜於古典音樂。

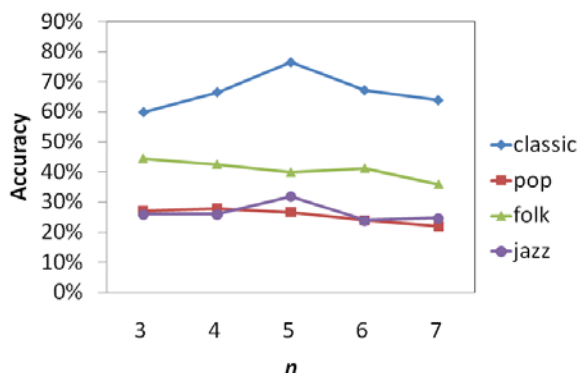


圖 5. 以前後音差正向分析分類之精確度

表 5 中，討論前後音差正向分析，比率、半徑跟分配比率的關係，對於四種類型的音樂，隨著比率的降低，而半徑跟隨著縮小，對於分配比率，古典音樂隨著比率的下降，分配比率逐漸的上升，但不若特徵音符來的好，對於流行音樂，隨著比率的下降，有著差不多的分配比率，而對於鄉村音樂與爵士音樂，隨著比率的下降，分配比率也跟著下降。

表 5.單特徵音差正向分析半徑與精確度(n=5)

	100%		90%		80%		70%	
	半徑	分配比率	半徑	分配比率	半徑	分配比率	半徑	分配比率
Classic	0.19	76.67	0.06	77.04	0.04	79.17	0.03	81.90
Pop	0.19	10.67	0.08	11.11	0.05	10.83	0.04	11.43
Folk	0.22	8.00	0.08	7.41	0.06	6.67	0.05	3.81
Jazz	0.21	4.67	0.08	4.44	0.06	3.33	0.05	2.86

#### 4.1.2 單特徵逆向分析對分類的影響

此實驗以音樂內容之單一特徵逆向分析來做音樂分類，也就是從較稀少出現之特徵值來做分類，以探討其對分類之精確性的影響。我們以流行音樂為例，做如圖 1 的音符分佈圖，結果如圖 6 顯示出現頻率最高的前 21 音

符與出現頻率，而其它的音符低於低音 Mi 之下與高於高十六度音 Do 之上的音符，皆歸類於 other，而我們在實驗時也發現，如果把 other 忽略不用，對於精確度是有不利影響的。而表 6 為頻率最低的七個音與出現頻率，分別為低音 Mi、other、低音 Fa、高十六度音 Do、高音 Si、高音 La 與低音 So，而取頻率為中心座標為 (0.0024, 0.0025, 0.0044, 0.0051, 0.0053, 0.0159, 0.0166)。

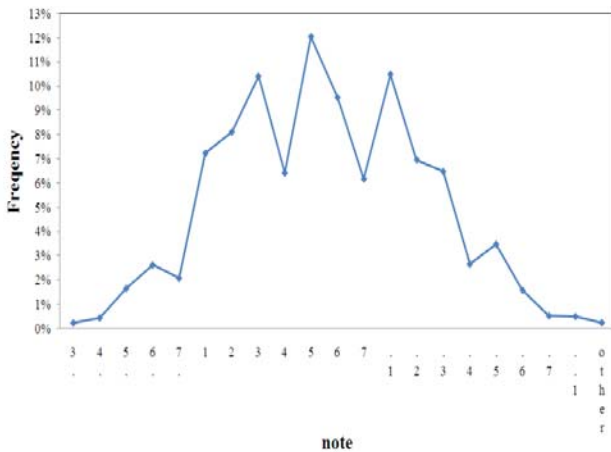


圖 6. 流行音樂音符分布圖

表 6. 流行音樂音符出現頻率最後 7 名

排名	音符	出現頻率
1	低音 Mi	0.0024
2	other	0.0025
3	低音 Fa	0.0044
4	高八度 Do	0.0051
5	高音 Si	0.0053
6	高音 La	0.0159
7	低音 So	0.166

我們分別對古典音樂、流行音樂、鄉村音樂與爵士音樂之音符資料取樣特徵值數(n)，分別統計出最低頻率的 3~7 個音符，以做為中心座標。接著透過距離公式(1)，來計算一首音樂與中心座標的距離，以分析它可能較接近那一類型的音樂。有了中心座標後，我們將資料庫中的 600 首音樂，逐一以音符的發生頻率來分類，統計可成功分類到正確音樂類型的精確度是多少。統計方式是將每首音樂，以古典音樂最低頻率之 3~7 個特徵值(n)，以公式(1)計算與古典音樂中心座標的距離。同樣的，再將這首音樂以流行音樂、鄉村音樂與爵士音樂之最低頻之 3~7 個音符，也計算出流行音樂、鄉村音樂與爵士音樂中心座標之距離。如此，可將此首音樂歸類於較短距離之音樂類別。以音符特徵為分類之實驗結果如圖 7 所示。對鄉村音樂

之音符逆向分析分類精確度，約在 80% 左右，明顯優於對古典音樂、流行音樂與爵士音樂分類之精確度。其中流行音樂與爵士音樂兩類的精確度，相當接近。流行音樂的起伏較大，並沒有古典音樂、鄉村音樂與爵士樂來的穩定，但在精度上，特徵數 4~7 有漸漸提升的現象。

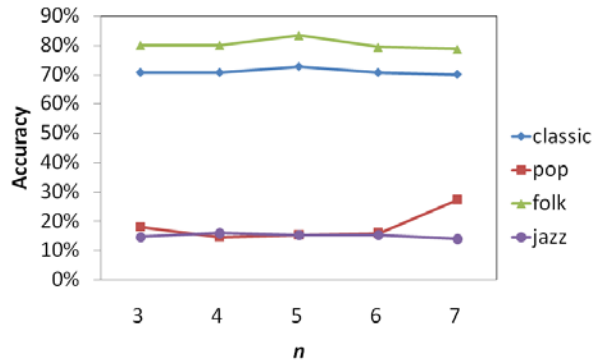


圖 7. 以音符特徵值逆向分析分類之精確度

接著我們以相同的實驗方式，繼續以節拍與前後音差變化之音樂內容特徵逆向分析來做音樂分類，實驗結果如圖 8、圖 9。

在圖 8 中，討論節拍逆向分析的部份，其流行音樂分類精確度特徵值數 4~7，隨著特徵值數增加而提升。而古典音樂與流行音樂起伏較大，鄉村音樂與爵士音樂雖然平穩，但精確度非常的不理想，未能達到一個有效的精確度。

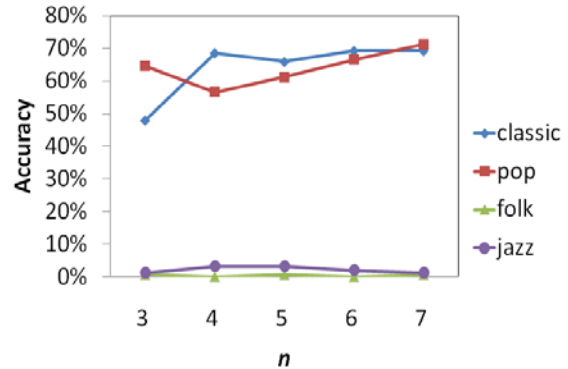


圖 8. 以節拍特徵值逆向分析分類之精確度

於圖 9，以前後音差變化逆向分析來做音樂分類之精確度比較，鄉村音樂之分類精確度，平均約在 85% 左右，明顯優於流行音樂、鄉村音樂與爵士音樂分類精確度，但隨著特徵值數目的增加，對於精度有遞減的狀況，而在前後音差變化逆向分析精確度上，古典音樂精確度起伏較穩定，而流行音樂與爵士音樂精確度起伏較大，但相對於以音符逆向分析的分類方式，音符較穩定。而其他結果，古典音樂則略遜，流行音樂與爵士樂相當相似。整體而言，以單特徵逆向分析來做音樂分類，於鄉村音樂效果較佳，且以音符

之逆向分析分類方式最佳且穩定，而對於流行音樂，以節拍之逆向分析分類方式較佳。

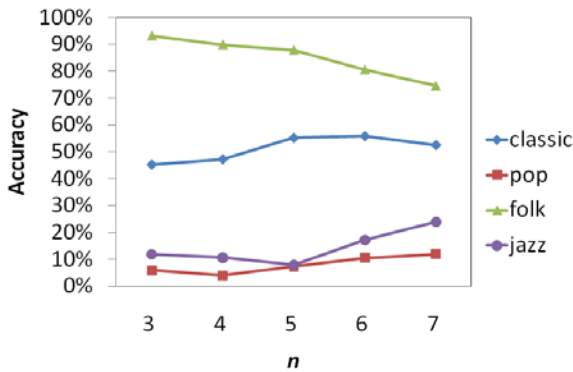


圖 9. 以音差特徵值逆向分析分類之精確度

## 4.2 多特徵對分類的分析

### 4.2.1 多特徵正向分析對分類的影響

此實驗當中，探討了結合 2 個或 3 個音樂內容特徵(k)，是否可以提升音樂分類的精確度。實驗程序與參數設定，仍然與前一節(4.1.1 節)相同，但多用了公式(2)，以加總多特徵各別與中心座標的距離，以判斷可能的音樂類型。我們總共做了(音符+節拍)、(音符+前後音差變化)、(節拍+前後音差變化)，以及(音符+節拍+前後音差變化)等四種特徵組合之正向分析。實驗結果分別呈現於圖 10~圖 13。

圖 11 中以音符+前後音差變化來做古典音樂之正向分析分類，有較好之精確度，從單以音符特徵正向分析分類約 80% 的精確度，提升到約 83%。於圖 12 中以節拍+正向前後音差變化正向分析來做爵士音樂之分類，有較好之精確度，從單以節拍特徵正向分析分類，從約 70% 的精確度，提升到約 73%。反觀流行音樂與鄉村音樂，採多特徵正向分析來做分類，精確度並未見提升，但在多特徵正向分析中，隨著特徵值數目增加，精確度也會有正向的增加。

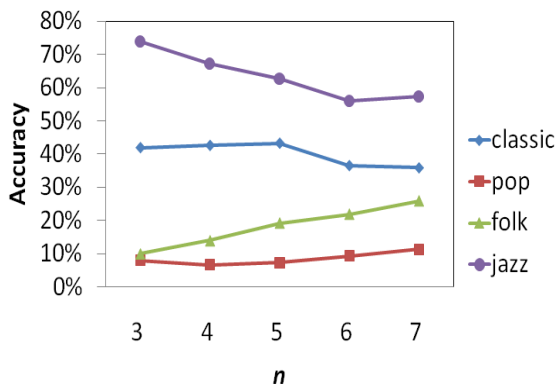


圖 10. 音符+節拍特徵值數目正向分析分類之精確度

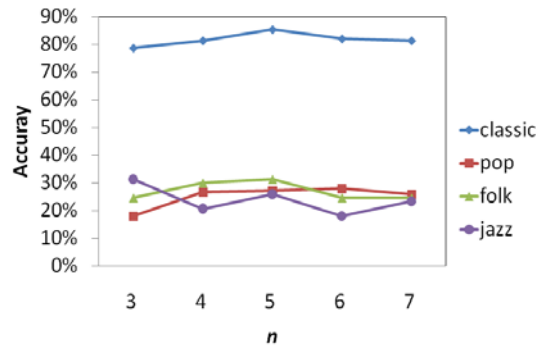


圖 11. 音符+前後音差特徵值數目正向分析分類之精確度

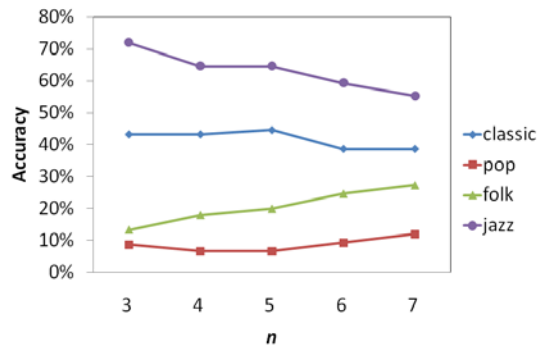


圖 12. 節拍+前後音差特徵值數目正向分析分類之精確度

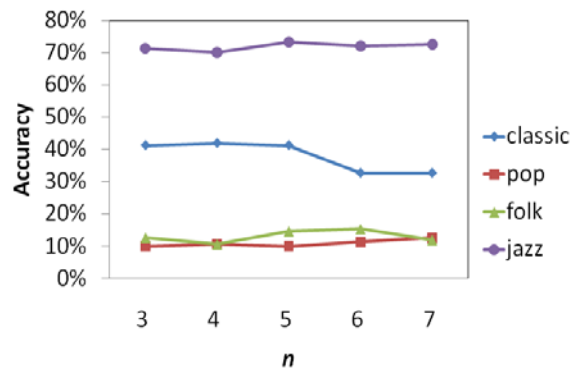


圖 13. 音符+節拍+前後音差特徵值數目正向分析分類之精確度

### 4.2.2 多特徵逆向分析對分類的影響

在此實驗當中，我們探討了結合 2 個或 3 個音樂內容特徵(k)之逆向分析，以了解是否可以提升音樂分類的精確度。實驗程序與參數設定，仍然與前一節(4.1.2 節)相同，但多用了公式(2)，以加總多特徵各別與中心座標的距離，以判斷可能的音樂類型。我們總共做了(音符+節拍) 逆向分析、(音符+前後音差變化) 逆向分析、(節拍+前後音差變化) 逆向分析以及(音符+節拍+前後音差變化) 逆向分析等四種特徵組合，實驗結果的精確度，則是分別呈現於圖



14~圖 17。分析圖 14~圖 17，於圖 14 中以音符+前後音差變化逆向分析來做流行樂之分類，有較好之精確度，從原本單特徵節拍逆向分析之精確度 61%左右，提升到 77%左右，且隨著特徵數目增加，精確度也有穩定的增加。於圖 15 中以節拍+前後音差變化逆向分析來做鄉村音樂之分類，有較好之精確度，但與單特徵節拍逆向分析分類有著差不多的精確度，不過不若單特徵節拍逆向分析來的穩定。反觀古典音樂爵士音樂，採多特徵逆向分析來做分類，精確度並未見提升。

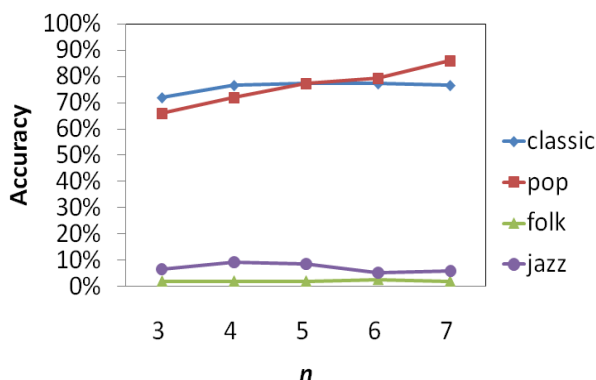


圖 14. 音符+節拍特徵值數目逆向分析分類之精確度

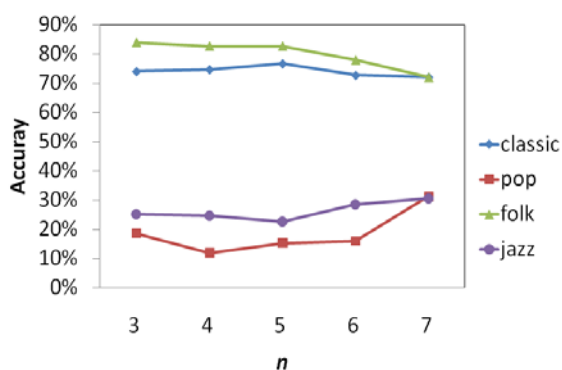


圖 15. 音符+前後音差特徵值數目逆向分析分類之精確度

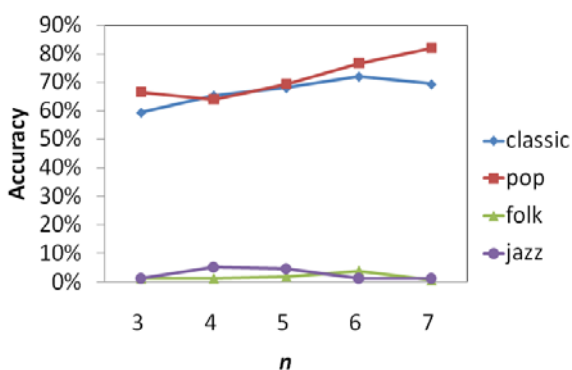


圖 16. 節拍+前後音差特徵值數目逆向分析分類之精確度

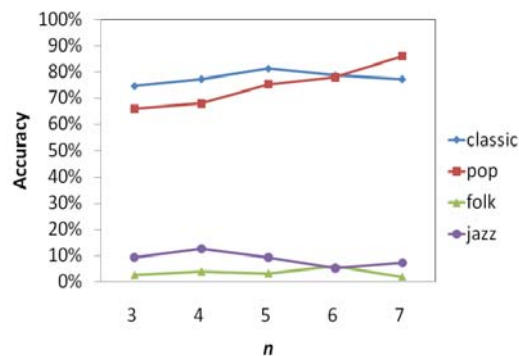


圖 17. 音符+節拍+前後音差特徵值數目逆向分析分類之精確度

此外，我們的實驗組合中，並無考慮混合正向分析+逆向分析，因為逆向分析與中心點的距離相當的短，與正向分析的距離不成比例，兩者加在一起，無法對正向分析產生明顯的影響，也無法凸顯逆向分析稀有性特徵值的意義。

綜合以上實驗分析，以(音符+前後音差變化)兩特徵組合正向分析來對古典音樂(classic)做分類，效果最好，準確度可達 85%；對鄉村音樂的分別是(音符)單特徵逆向分析分類，效果最佳，準確度可達 83%；對流行音樂(pop)的分類，則是(音符+節拍)兩特徵組合的逆向分析分類，效果較佳，準確度可達 77%；而爵士音樂對於(節拍與前後音差)兩特徵組合正向分析分類，準確度可達 73%，雖然較遜於前三類，但仍然優於現有的音樂分類技術。

## 5. 結論

本研究報告提出了以音樂內容特徵(feature of music content)做為音樂分類之方式，以音樂特徵出現較高與較低的頻率值，做為音樂分類的中心座標。對於一首音樂可能屬於那一類的音樂，則是計算其與中心座標的距離來決定。我們分別對古典音樂(classic)、流行音樂(pop)、鄉村音樂(folk)與爵士音樂(jazz)四類音樂，搭配音符(notes)、節拍(rhythms)與前後音差變化(pitch changes)等三個音樂內容特徵，做音樂分類的實驗分析。而其中，以(音符+前後音差變化)兩特徵組合來對古典音樂做正向分析分類，準確度可達 85%；對鄉村音樂則是(音符)單特徵逆向分析分類，準確度可達 83%；對流行音樂的分類，則是(音符+節拍)兩特徵組合的逆向分析分類，準確度可達 77%；而爵士音樂對於(節拍+前後音差)兩特徵組合正向分析分類，準確度可達 73%。整理如表 7，我們的分析準確度優於現有的技術。

表 7. 各音樂類型分類特徵與精確度

類型	分析	特徵	精確度
古典音樂	正向分析	音符+前後音差	85%
流行音樂	反向分析	音符+節拍	77%
鄉村音樂	反向分析	音符	83%
爵士音樂	正向分析	節拍+前後音差	73%

## 參考文獻

- [1] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," In S. Thrun and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 561-568, MIT Press, Cambridge, MA, 2003.
- [2] S. G. Blackburn and D. C. DeRoure, "A Tool for Content-based Navigation of Music," *Proc. of the ACM Multimedia*, pp.361-368, 1998.
- [3] S. Brecheisen, H.-P. Kriegel, P. Kunath, and A. Pryakhin, "Hierarchical Genre Classification for Large Music Collections" *IEEE 7th int'l conf. on multimedia and Expo*, pp. 1385-1388, 2006.
- [4] W. Chai and B. Vercoe, "Folk Music Classification Using Hidden Markov Models," *Proc. of Int'l Conf. on Artificial Intelligence*, 2001.
- [5] Y. Chen, J. Bi, and J. Z. Wang, "MILES: Multiple-instance learning via embedded instance selection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12):1931-1947, 2006.
- [6] H. T. Cheng, Y. H. Yang, Y. C. Lin, and H. H. Chen, "Automatic Chord Recognition for Music Classification and Retrieval," *IEEE Int'l Conf. on Multimedia and Expo*, pp. 1505-1508, 2008.
- [7] R. B. Dannenberg, B. Thom, and D. Watson, "A Machine Learning Approach to Musical Style Recognition," *Proc. of Int'l Computer Music Conf.*, 1997.
- [8] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*, Prentice Hall, 2001.
- [9] C. L. Krumhansl, *Cognitive Foundations of Musical Pitch*, Oxford University Press, New York, 1990.
- [10] C-R. Lin, N-H. Liu, Y-H. Wu, and A. L. P. Chen, "Music Classification Using Significant Repeating Patterns," *Lecture Notes in Computer Science*, Springer-Verlag, Vol. 2973, pp. 506-518, 2004.
- [11] C. C. Liu, J. L. Hsu, and Arbee L. P. Chen, "Efficient Theme and Non-Trivial Repeating Pattern Discovering in Music Databases," *Proc. of the IEEE Data Engineering*, pp.14-21,1999.
- [12] Y. L. Lo, C. H. Lee, and C. H. Wang, "Scalable Multi-feature Index Structure for Music Databases," *Information Sciences*, Elsevier, Vol.179, Issue 15, pp. 2662-2675, July 2009.
- [13] Y. L. Lo, W. L. Lee, and L. H. Chang, "True Suffix Tree Approach for Discovering Non-trivial Repeating Patterns in a Music Object," *Journal of Multimedia Tools and Applications*, Springer, Vol. 37, No. 2, pp. 169-187, April 2008.
- [14] Y. L. Lo and L. Y. Tsai, "Approximate Searching for Music Data in Real-Valued Feature Indexing," *Journal of Convergence Information Technology*, Vol. 4, No. 4, pp. 87-95, Dec. 2009.
- [15] Yu-lung Lo and Yi-Chang Lin, "Content-Based Music Classification," *the 3rd IEEE Int'l Conf. on Computer Science and Information Technology (IEEE ICCSIT 2010)*, pp. 112-116, July 9-11, 2010.
- [16] Q. J. B. Loh, and S. Emmanuel, "ELM the Classification of Music Genres," *9th Int'l conf. on Control, Automation, Robotics and Vision (ICARCV'06)*, pp. 1-6, 2006.
- [17] I. Mandel, and P. W. Ellis, "Multiple-Instance Learning for Music Information Retrieval," *9th Int'l Conf. on Music Information Retrieval*, pp.577-582, 2008.
- [18] E. Narmour, *The Analysis and Cognition of Basic Melodic Structures*, The University of Chicago Press, Chicago, 1990.
- [19] M.-K. Shan, F.-F. Kuo, and M.-F. Chen, "Music Sytle Mining and Classification by Melody," *Proc. IEEE Int'l Conf. on Multimedia and Expo (ICME'02)*, pp. 97-100, 2002.
- [20] G. Tzanetakis, G. Essl, and P. Cook, "Automatic Musical Genre Classification of Audio Signals," *Proc. of Int'l Symposium on Music Information Retrieval*, Vol. 10, No. 5, pp. 293-302, 2001.
- [21] V. N. Vapnik, *The Nature of Statistic Learning Theory*, New York, Springer-Verlag, 1995.
- [22] Y. Zhang, and J. Zhou, "A Study on Content-based Music Classification," *IEEE 7th Int'l Symposium on Signal Processing and Its Applications*, Vol. 2, pp.113-116, 2003.