

음악 추천을 위한 감정 전이 모델 기반의 음악 분류 기법

Emotion Transition Model based Music Classification Scheme for Music Recommendation

한 병준*, 황 인준***
Byeong-jun Han*, Een-jun Hwang***

Abstract

So far, many researches have been done to retrieve music information using static classification descriptors such as genre and mood. Since static classification descriptors are based on diverse content-based musical features, they are effective in retrieving similar music in terms of such features. However, human emotion or mood transition triggered by music enables more effective and sophisticated query in music retrieval. So far, few works have been done to evaluate the effect of human mood transition by music. Using formal representation of such mood transitions, we can provide personalized service more effectively in the new applications such as music recommendation. In this paper, we first propose our Emotion State Transition Model (ESTM) for describing human mood transition by music and then describe a music classification and recommendation scheme based on the ESTM. In the experiment, diverse content-based features were extracted from music clips, dimensionally reduced by NMF (Non-negative Matrix Factorization), and classified by SVM (Support Vector Machine). In the performance analysis, we achieved average accuracy 67.54% and maximum accuracy 87.78%.

요 약

최근까지 장르나 무드 등의 정적 분류 기술자를 이용한 음악 정보 검색에 관한 다양한 연구가 진행되어 왔다. 정적 분류 기술자는 주로 음악의 다양한 내용적 특징에 기반하기 때문에 그러한 특징에 유사한 음악을 검색하는 데 효과적이다. 하지만 음악을 들었을 때 느끼게 되는 감정 내지 기분 전이를 이용하면 정적 분류 기술자보다 더 효과적이고 정교한 검색이 가능하다. 사람이 음악을 들었을 때 발생하는 감정 전이의 효과에 관한 연구는 현재까지 미비한 실정이다. 감정 전이의 효과를 체계적으로 표현할 수 있다면 기존의 음악 분류에 의한 검색에 비해 음악 추천 등의 새로운 응용에서 더 효과적인 개인화 서비스를 제공할 수 있다. 본 논문에서는 음악에 의한 인간 감정 전이를 표현하기 위한 감정 상태 전이 모델을 제안하고 이를 기반으로 새로운 음악 분류 및 추천 기법을 제안한다. 제안하는 모델의 개발을 위하여 다양한 내용 기반의 특징을 추출하였으며, 고차원 특징 벡터의 차원 감소를 위하여 NMF (Non-negative Matrix Factorization)를 사용하였다. 성능 분석을 위한 실험에서 SVM (Support Vector Machine)을 분류기로 사용한 실험에서 평균 67.54%, 최대 87.78%의 분류 정확도를 달성하였다.

Key words : Music classification, emotion state transition model, musical feature extraction, SVM

* 고려대학교 전자전기공학과 박사과정

** 고려대학교 전기전자전파공학부 부교수

★ 교신저자 (Corresponding author)

※ 이 논문은 2007년도 정부재원(교육인적자원부 학술연구 조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-313-D00758).

I. 서론

최근 네트워크 및 IT 기술의 발전에 힘입어 다양한 장르의 음악 콘텐츠가 여러 경로를 통하여 생성되고 유통 및 소비되고 있다. 대용량 음악 데이터베이스가

보편화 되면서 효과적인 음악 정보 검색 (MIR; Music Information Retrieval)의 필요성이 크게 대두되고 있다. 현재 기존의 음악 정보 검색 분야가 통합된 MIREX[1]와 같은 음악 관련 연구자들의 콘테스트가 매년 진행되고 있다. MIREX 콘테스트에서 진행되는 연구 주제의 경우, 음악 아티스트 인식, 장르/무드 분류, 유사 검색, 노트 시작점(onset)/코드 검출, 실시간 악보 추적, 노래/허밍에 의한 질의 등 다양한 분야에서 음악 정보 검색을 다루고 있다.

하지만 음악에 의한 감정 전이 효과에 관한 연구는 아직도 미미한 실정이며, 음악 검색에 감정을 접목시킨 가장 근접한 연구로는 음악을 동적으로 분할하고, 분할된 구역마다 음악 감정을 추출하는 연구[2] 등을 들 수 있다. 한편, Xiang et al.[3] 등은 일반적인 사건에 대한 인간 심리의 전이를 설명하는 심리 상태 전이 네트워크 모델을 제안하였으나, 이에 대한 검증과 실제 어플리케이션에 대한 적용은 현재까지 보고된 바가 없다.

일반 사용자가 음악을 들을 때 발생하는 감정 전이 효과에 대한 연구는 다음과 같은 많은 분야에서 활용될 수 있다. 첫째, 사용자의 보편적인 음악 기호를 분석하고 이에 따라 음악을 분류하는 데에 활용될 수 있다. 둘째, 비슷한 성향을 가진 사용자끼리 군집 분류를 하여, 개인별 차별화 서비스를 제공할 수 있다. 셋째, 음악 치료나 유비쿼터스 상황 인지 등 감정의 변화가 수반되거나 예상되는 서비스 분야에 효과적으로 대처할 수 있다. 마지막으로, 음악 아티스트가 역으로 자신의 아트 성향을 분석하는 데에도 활용될 수 있다.

이를 위하여 본 논문에서는 우선 음악에 의한 인간의 감정 전이를 기술하기 위한 감정 상태 전이 모델(ESTM)을 제안하고, 특정 음악이 감정의 전이에 미치는 긍정적/부정적 효과를 정량적으로 표현한다. 둘째, 음악의 특성을 표현하는 다양한 특징 추출 기법 및 이를 기반으로 하는 음악 분류 기법을 제안한다. 특징 추출을 위해 음악 문맥을 고려하는 프레임화(framing) 방법[4]을 사용하여 각 프레임으로부터 다양한 음악적, 통계적 특징들을 추출하고 NMF와 같은 차원 감쇄 방법을 사용하였다. 또한, Original SVM과 One class SVM를 사용하여 분류 성능을 측정하였다.

II. 관련 연구

1. 감정 모델

감정 모델은 많은 음악 분류 연구에서 직간접적으로 사용되는 소재이다. 심리학에서 감정은 다양한 방

법으로 표현하고 있으며 초기에는 쉽게 표현할 수 있는 감정 형용사를 사용하였다. 그러나 사람의 감정을 표현할 수 있는 형용사는 무수히 많을 뿐 아니라, 각각의 형용사가 지니는 의미가 서로 중첩되는 문제가 있다[5]. 이러한 문제를 해결하기 위한 다양한 시도의 하나로 Russel[6]의 circumflex 모델을 꼽을 수 있다. 이 모델에서는 대상 감정을 2 차원의 양방향 축에 표현하고 각 차원 내의 상반된 감정은 서로 멀리 표현하는 방식으로, 20가지 이상의 감정을 표현하였다. 한편, Thayer[7]는 Russel의 모델을 음악에 적용한 무드 모델(mood model)을 제안하였다. 그의 모델에서 모든 음악 감정은 각성(arousal)과 그에 대한 긍정/부정의 정도(valence)를 나타내는 2차원의 공간에 나타내어진다. 이러한 시도들은 현재 많은 음악 관련 연구[2,8]에서 활용되고 있다.

한편, Xiang et al.[3]은 기존의 형용사 표현을 보존하면서 감정과 감정 간의 전이를 확률적으로 표현한 심리 상태 전이 네트워크(Mental State Transition Network)를 제안하였다. 심리 상태 전이 네트워크는 각각의 감정을 개별적인 것으로 보고, 감정 간의 전이를 확률을 통해 직관적으로 나타낸다는 장점이 있지만, 어떤 시점에서의 감정과 감정 간의 상호 효과를 고려하지 않았으며, 어떤 사건으로 인해 현재 감정으로부터 다른 감정으로의 전이에 오히려 부정적 효과가 나타낼 수 있음을 고려하지 않았다. 본 논문에서는 이러한 심리 상태 전이 네트워크에서의 문제점을 다른 방식으로 해결하고자 한다.

2. 특징 추출

음악 정보 검색을 위한 특징 추출은 음악적 요소의 개입 여부에 따라 신호 분석 레벨에서의 특징 추출, 음악적 구조를 고려한 특징 추출, 그리고 이 두 가지를 혼합한 하이브리드 특징 추출 등으로 나눌 수 있다.

음성 신호 분석 레벨에서의 특징 추출은 다시 시간 도메인에서의 특징과 주파수 도메인에서의 특징으로 나눌 수 있다. 전자는 원 신호에 대한 간단한 연산으로 구현 가능하며 프레임 내 에너지량을 나타내는 RMS 에너지나 신호의 영교차율을 표현한 ZCR 등이 대표적인 예이다. 그러나 주파수 등 정보 추출에 한계가 있다. 후자는 음성 및 음악 신호 처리에 있어서 주파수 단위 별로 에너지 및 위상 분석이 가능하여 좀더 많은 정보를 추출할 수 있다. 대표적으로 예로는 통계적 분석을 도입한 Spectral Centroid(SC), Spectral Flatness(SF), Spectral Roll-off(SR), Spectral Flux(SFX) 등이 있다.

음악 분석(musical analysis)에 기반하는 특징 추출은 음악의 주요 요소인 리듬, 멜로디, 하모니 등을 분석하여 활용한다. 이러한 요소를 분석하기 위한 방법에는 여러 가지가 있는데, 가령 멜로디를 표현하는 특징의 일부인 조(key)나 화음(chord) 등을 추출하기 위해 크로마그램(chromagram)을 분석하여 단위 시퀀스 또는 프레임 내의 음의 주요 배치를 분석한다. Müller는 크로마그램을 변형한 CENS를 사용하여 기존 방식에 비해 음간 노이즈 등을 경감시키기도 하였다[9].

마지막으로 두 가지를 혼합한 하이브리드 특징 추출의 대표적인 예로는 인간의 청각 특성을 고려한 멜스케일(Mel scale)에 따라 배치된 범위를 따라 추출하는 MFCC(Mel-Frequency Cepstral Coefficients)가 있다[10]. MFCC는 음악 분석뿐 아니라 신호 처리 분야 전반에서 널리 쓰이는 방법으로, 여러 인식 시스템에서 좋은 성능을 보이는 특징으로 알려져 있다[11].

3. 음악 추천

현재까지 효과적인 자동 음악 추천을 위한 다양한 연구가 진행되어 왔다. 특히 음악의 특징 추출을 통한 장르나 무드를 인식하고 이를 바탕으로 사용자에게 적합한 유형의 음악을 자동으로 추천하는 것에 대한 연구가 활발하였다.

Lu et al.[2]은 음악의 주요한 특징을 강도, 음색, 리듬으로 나누고, 이들을 이용하여 자동적으로 무드를 검출하고 추적하는 시스템을 제안하였다. 이들의 연구는 하나의 음악 내에서도 무드가 지속적으로 변화한다는 것을 보여준 것이며, 하나의 정적인 분류자(descriptor)로 음악을 제대로 표현하기 어렵다는 것을 보여주는 예이다. 하지만 여전히 무드를 기술하기 위한 형용사를 이용하여 음악의 구간을 표현하고 있다는 점에서 한계가 있다. Holzapfel et al.[24]은 차원 축소 방법의 하나인 NMF를 기반으로 분류 시스템을 제안하였다. 그들은 입력된 음악 신호로부터 분석된 주파수 도메인의 데이터를 구획화하고, 이들을 NMF를 사용하여 차원 축소하였다. 또한, GMM을 분류기로 사용하였고, EM 알고리즘을 훈련 알고리즘으로 사용하였다. 그러나 이들의 목적도 무드/장르와 같은 정적 분류자를 검출하고자 하는 것이다.

III. 감정 상태 전이 모델 (ESTM)

감정 상태 전이 모델은 사용자의 복합적인 현재 감정 상태가 음악에 의해 다른 감정 상태로 어떻게 전

이하는가를 정량적으로 표현하기 위한 방법이다. 이러한 감정 상태 전이 모델은 개인화 서비스나 소규모 음악 커뮤니티 등 다양한 범위의 사용자 그룹에 대하여 정의가 가능하며 동질의 사용자 그룹일수록 정확하고 효과적인 감정 상태 전이를 표현할 수 있다. 이를 위해 우선 사람의 복잡한 감정을 표현하는 방법을 정의하고 이를 바탕으로 감정의 전이 효과를 표현하는 방법을 정의한다.

1. 감정 전이의 개념

기존의 음악 정보 검색에서 분류의 대상은 장르나 무드와 같은 음악적 분류자를 기반으로 하고 있으나 다음과 같은 문제점으로 인해 한계를 드러내기도 한다.

첫째, 장르의 경우 퓨전 장르 혹은 장르 간 경계선에 있는 음악에 대해서는 분류하기가 어렵다. 둘째, Lu et al.[2]의 연구에서 볼 수 있듯이, 하나의 음악 안에서도 무드가 동적으로 변화한다. 그러나 MIREX[1] 등의 연구 콘테스트에서 의존하고 있는 allmusic.com 과 같은 무드 분류 레퍼런스는 음악에 대한 무드를 정적으로 정의하고 있다. 따라서 무드에 의한 분류 방법 또한 새로운 기준의 확립이 필요하다. 셋째, 장르나 무드 등은 인간의 음악 기호를 표현하기 위한 감각적인 분류 방법으로 보기 어려운 경우가 있다.

장르/무드 등의 정적인 기술자를 벗어나, 실제 사람이 음악을 들으면서 느끼게 되는 감정 변화의 정도를 분류의 기준으로 사용할 수 있다면 기존의 방법과는 다른 직관적이고 정서적인 음악적 분류가 가능할 것이다. 즉, 특정 음악이 사람에게 미치는 감정적 정서적 영향을 정량적으로 기술하고 표현할 수 있다면 사용자가 현재 원하는 유형의 노래를 좀 더 정확하게 표현할 수 있게 된다. 본 논문에서는 어떤 사건이나 현상 등으로 인하여 발생하는 감정의 변화를 ‘감정 전이(emotion transition)’라 정의하고, 음악에 의한 감정 전이의 정도를 정량적으로 표현하고자 한다.

2. 감정의 표현

어떤 시점에서의 사람의 감정을 표현하는 가장 쉬운 방법 중의 하나는 감정 형용사(adjective)를 사용하는 것이다. 이 때, 사용된 감정 형용사 간에는 중첩된 영역이 있을 수도 있고, 서로 상반된 영역이 있을 수도 있다. 가령, ‘슬픈(sad)’, ‘울적한(melancholy)’, ‘의기소침한(depressed)’ 등의 경우 AV[7] 관점에서 valance가 낮다는 공통점이 있다.

또한 사람의 감정은 하나의 감정 형용사만으로도 순수하게 표현될 수 있지만, 대개 여러 감정 형용사의 조합으로 표현되는 것이 일반적이며 또한 현재의 감정 상태를 표현하는 감정 형용사 중에 영향을 미치는 정도가 서로 다를 수 있다.

마지막으로, 동일한 사건에 대한 감정 전이의 결과는 현재의 감정 상태에 따라 달라질 수 있다는 것이다. 따라서 이전 상태의 단일 감정 혹은 복수 감정의 조합과 사건에 의한 다음 상태의 감정 조합의 구성을 정량적으로 표현할 수 있다면 사건에 의한 감정 전이 결과를 계산하는 것이 가능하다.

3. 감정 상태 전이 모델

감정 상태의 전이를 수학적으로 정의하면, 우선 n 개의 감정 형용사를 사용해서 사람의 감정을 표현한다고 가정하고, 현재의 감정 상태는 구성 감정 형용사의 정도를 나타내는 행렬 벡터 수식을 사용하여 다음과 같이 표현할 수 있다.

$$\mathbf{ES}(t) = [es_1(t) \ es_2(t) \ \dots \ es_n(t)] \quad (1)$$

감정 상태 벡터를 나타내는 $\mathbf{ES}(t)$ 는 시간에 대한 함수 형태를 취하며 벡터를 구성하는 $es(t)$ 는 감정 형용사의 추상적 정도를 나타내며, 0 이상 1 이하의 범위를 가지는 실수로 표현된다.

어떤 두 시점 t_1, t_2 의 감정 상태 벡터를 $\mathbf{ES}(t_1)$ 및 $\mathbf{ES}(t_2)$ 라 할 때, $\mathbf{ES}(t_2) - \mathbf{ES}(t_1)$ 은 상태 감정의 전이된 양을 나타낸다. 그러나 이러한 차이는 동일한 감정 간의 전이된 양을 나타낼 뿐이므로, 처음 상태(t_1)로부터 전이된 감정의 정도를 나타내기 위해서는 다음과 같은 감정 상태 전이 행렬을 정의한다.

$$\begin{aligned} \mathbf{ETM}(t_i, t_f) &= \mathbf{ES}(t_f)^T \cdot (\mathbf{ES}(t_f) - \mathbf{ES}(t_i)) \\ &= \{p_{n,m} | p_{n,m} = es_n(t_f)(es_m(t_f) - es_m(t_i))\} \\ &= \begin{bmatrix} p_{1,1} & p_{1,2} & \dots & p_{1,m} & \dots & p_{1,N} \\ p_{2,1} & \ddots & & & & \vdots \\ \vdots & & \ddots & & & \vdots \\ p_{n,1} & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ p_{N,1} & \dots & \dots & \dots & \dots & p_{N,N} \end{bmatrix} \end{aligned} \quad (2)$$

이 때 t_i 와 t_f 는 각각 처음 상태와 나중 상태를 나타내며, $p_{n,m}$ 은 감정 상태 전이 행렬의 (n,m) 번째 요소를 가리킨다.

예를 들어, 한 사람의 감정 상태를 5차원 벡터 $\mathbf{ES}(t)$ 로 표현하고, 각각의 성분이 순서대로 즐거움, 슬픔, 분노, 침울, 기타에 대응한다고 가정하자. 사람

이 음악을 감상하기 전($t=t_1$)과 감상한 후($t=t_2$)의 감정 상태가 $\mathbf{ES}(t_1)=[0.5 \ 0.1 \ 0.1 \ 0.2 \ 0.1]$, $\mathbf{ES}(t_2)=[0.1 \ 0.5 \ 0.2 \ 0.1 \ 0.1]$ 라고 하면, 음악이라는 이벤트로 인해 변화하는 전후 감정의 $\mathbf{ETM}(t_1, t_2)$ 는 다음과 같다.

$$\mathbf{ETM}(t_1, t_2) = \begin{bmatrix} -0.20 & 0.20 & 0.05 & -0.05 & 0 \\ -0.04 & 0.04 & 0.10 & -0.10 & 0 \\ -0.04 & 0.04 & 0.10 & -0.10 & 0 \\ -0.08 & 0.08 & 0.02 & -0.02 & 0 \\ -0.04 & 0.04 & 0.10 & -0.10 & 0 \end{bmatrix} \quad (3)$$

한편, 제안하는 감정 전이 상태 모델과 심리 상태 전이 네트워크[3]는 구현에 있어서 다른 점을 가지고 있다. 심리 상태 전이 네트워크의 경우 이전 감정 상태로부터 다음 감정 상태로의 전이를 확률적으로 표현한다. 그러나 심리 상태 전이 네트워크는 이전 감정 상태로부터 다음 감정 상태로 전이 시 역효과나 사람의 복합적인 감정 조합을 고려하지 않고 있다.

IV. 특징 추출

본 장에서는 앞에서 제안한 감정 상태 전이 모델을 구축하기 위한 기초 데이터로 사용되는 다양한 음악의 저수준 특징들과 그들을 추출하는 방법을 제시한다.

1. 음악 신호 분석과 프레임화

음악 분류 및 특징 검출을 위해 제안된 방법들의 공통점 중의 하나는 음악 신호를 분석하기에 앞서 신호를 프레임화(framing)하는 것이다. 프레임화는 신호를 더 작은 동일한 길이의 프레임(frame)으로 자르는 과정이다. 이 때, 프레임의 길이가 작아지면 시간 해상력은 증가하나 주파수 해상력이 떨어지며, 프레임 길이가 길어지면 주파수 해상력이 좋아지는 반면 시간 해상력이 떨어진다. 즉, 프레임의 길이에는 시간 해상력과 주파수 해상력의 trade-off가 있음이 알려져 있다[12].

그러나 이러한 trade-off는 음악 신호 분석에 있어서 큰 약점으로 작용하지 않는다. 대부분의 음악은 90~240 BPM의 템포(tempo)를 가지며, 코드(chord)는 일반적으로 한 마디 이상을 거친 후 변화한다. 시작점 검출(onset detection) 시 240 BPM의 템포를 가지는 음악의 16분 음표의 경우 62.5ms의 지속 시간을 가지며, 이는 일반적인 오디오 신호 처리에서 사용되는 프레임의 일반적인 최소 길이로 알려진 10~15ms를 훨씬 상회하는 길이이다.

한편, 이와 같은 기존 프레임화는 자료의 특성이

알려지지 않은 신호 데이터에 대해 통계적 분석을 할 때에는 유용하나, 본 연구와 같이 음악적 리듬 특성이 알려진 경우 비효율적일 수 있다. 즉, 음악의 내용과 관계없이 동일한 길이의 프레임으로 분할한다는 단점이 발생한다. 따라서 본 연구에서 음악 비트에 기반하는 프레임화 방법[4]을 사용하였다.

2. 음색 특징

음색(timbre)은 사람의 인지에서 파생된 주관적 특성으로, 물리적 특징과 일대일로 연결하기 어려운 특성이다. 본 연구에서는 음색의 측정을 위해 주로 사용되는 MFCC[11]를 추출하였다. MFCC는 인간의 청각 구조에 맞게 기존 주파수 도메인을 멜-스케일 도메인으로 변환하여 특정 계수를 추출한다.

3. 음악적 변화 특징

음악을 구분하기 위해 가장 많이 사용되는 요소로 리듬, 가락, 화성 등이 있으며, 이러한 특성들은 음악으로 인한 사람의 감정 상태 전이와 연결하여 다양하게 설명될 수 있다. 리듬은 보통 에너지의 변화로 설명된다. 음의 전체적인 강도(intensity)와 그의 변화가 클 경우 사람은 리듬감을 쉽게 느낀다.

박자의 흐름이 변화하는 것도 리듬감을 인지하는데에 큰 영향을 미친다. 특히, 재즈의 경우 연주자의 성향에 따라 박자의 변화가 다양하다. 이러한 박자의 변화는 물리적으로 각 비트 간 길이의 차이로 표현할 수 있는 데 이를 수식으로 표현하면 다음과 같다.

$$VBL(n) = \frac{60}{Tempo} (BL(n+1) - BL(n)) \quad (4)$$

여기서, VBL(Variance-of-Beat Length)는 각 비트 간 길이의 차이이며, BL은 한 비트의 길이이다.

서양 음악의 화성은 주로 한 마디의 음조 구성을 결정하는 코드에 의해 결정된다. 코드 추출을 위해서 우선 서양 음조의 대표 화음인 C~Bb의 강도를 표현하는 크로마그램을 추출한다. 이에 Key Profile Matrix[14]를 적용하여 해당 프레임의 주된 코드를 추출한다.

4. 통계적 특징

몇몇 잘 알려진 통계적 특징들은 효과적으로 프레임의 특징을 나타낸다고 알려져 있다[10,12]. 본 연구에서는 이들 중에서 Spectral Centroid(SC), Spectral Flatness(SF), Spectral Spread(SS), Spectral

Flux(SFX) 등을 활용하여 특징을 추출하였다.

V. 특징 추출

본 장에서는 추출된 다양한 특징들과 분류 데이터(ground-truth)를 연결짓기 위한 분류(classification)에 대해 설명한다. 우선, 분류기에 넣을 특징 데이터의 차원 축소를 위한 NMF를 설명한 다음, 분류기로 사용한 SVM과 이의 다양한 커널 변종을 설명한다.

1. 특징 차원 감쇄

특징 추출 단계에서 각각의 프레임에서 다양한 음악 특징들이 추출되기 때문에, 훈련을 위한 전체 특징 데이터의 차원은 커질 수밖에 없다. 또한, 각각의 음악은 서로 다른 재생 시간을 가지기 때문에, 프레임화를 통한 프레임의 개수 또한 서로 다르다. 따라서 추출된 데이터의 차원을 동일한 차원으로 줄이면서, 특징을 유지하는 방법이 필요하다.

NMF[15,16]는 기존의 차원 축소 방법에 비해 데이터의 특징을 좀 더 잘 보존한다는 장점이 있다. 하지만 데이터의 모든 요소가 양수여야 한다는 제약 조건을 가진다. 따라서 본 연구에서는 특징 데이터를 모두 양수로 변환하여 NMF를 적용하였다.

2. 분류 기법

특징 분류를 위해서는 SVM을 사용하였다. 그러나 SVM은 한 번에 한 가지 클래스만 분류할 수 있으므로, 본 논문의 감정 상태 전이 행렬을 분류하기 위해서는 감정 상태 전이 행렬 요소 별로 SVM을 훈련하여 대응시켰다[17].

VI. 실험 결과

본 장에서는 제안하는 감정 상태 전이 모델 및 음악 특징 추출을 기반으로 하여 수행한 음악의 분류 및 추천 관련 실험 결과를 소개한다.

1. 음악 분류

본 논문에서 음악 분류는 다음과 같은 단계로 진행된다. 우선 음악 데이터로부터 비트 검출 기반의 프레임화를 통한 특징을 추출한 다음, NMF를 이용한 차원 축소 및 SVM를 이용한 분류 과정을 거친다. 마지막으로 사용자가 표현할 수 있는 감정 형용사를 기반으로 감정 표현을 수집하여, 분류기의 분류 데이

터로 활용한다. 그림 1은 본 논문에서 제안하는 음악 분류 과정을 보여주고 있다.

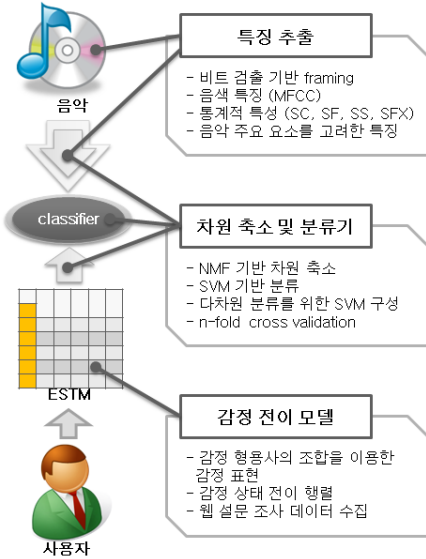


Fig. 1. Steps for music classification
그림 1. 음악 분류 절차

2. 데이터 수집

감정의 표현은 매우 주관적이기에 사용자로부터 하나의 음악에 대한 다양한 감정 데이터를 수집하는 것이 중요하다. 실험을 위해 20대에서 50대에 이르는 다양한 연령층의 자발적 참가자 30명을 선정하였다. 선정된 참가자들에게는 웹 기반의 감정 수집 프로그램을 통해 준비된 120곡의 다양한 음악을 듣고 체험한 감정 전이 정도를 기입하게 하였다. 음악은 팝송의 경우, allmusic.com[18]에 제시된 무드 분류에 균일하도록 선정되었으며, 한국/일본 음악의 경우 멜론[19]을 활용하였다.

자신의 감정을 기술하기 위한 감정 형용사로는 분노(angry), 흥분(excited), 행복(happy), 기쁨(pleased), 이완(relaxed), 평화로움(peaceful), 졸림(sleepy), 슬픔(sad), 지루함(bored), 짜증남(nervous), 그리고 고요함(calm)의 11가지 형용사를 사용하였다.

많은 수의 감정 형용사가 존재하기에, 참가자들은 적은 수의 감정 형용사를 사용하여 자신의 감정을 기술하였다. 본 실험에서는 하나의 노래에 대해 적어도 121개의 감정 상태 전이에 대한 데이터가 수집되어야 하므로, 참가자가 다양한 감정 상태에서부터 음악을 들

고 전이되는 효과를 측정하였다. 최종적으로 모든 참가자들은 32,573 건의 실험 데이터를 입력하였으며, 이들을 이용하여 분류 데이터로 활용하였다.

3. 실험 결과

시스템 구현을 위해 MATLAB과 Matweb의 연동 인터페이스를 사용하였다. SVM 구동을 위해서는 LIBSVM[20]을 활용하였는데 LIBSVM은 다양한 종류의 인터페이스와 커널 함수들을 제공하고 있다. 분류를 위해서 Original SVM[21,22]과 One-Class SVM[23]을 대상으로 그들의 분류 성능을 비교하였다.

성능 평가를 위한 평가지표로는 전체 분류 정확도, 정확도 표준편차, 그리고 최대 정확도 등을 사용하였다. 전체 분류 정확도는 모든 클래스에 대한 분류기의 분류 정확도를 나타내는 지표이며, 다음과 같이 정의된다.

$$Overall Accuracy = \frac{\sum_{s=1}^N \sum_{e=1}^N CD(s,e)}{M \cdot N^2} \quad (5)$$

여기서 N은 감정 형용사의 종류를 나타내며, M은 실험을 위해 사용된 노래의 수를 나타낸다. 그리고 CD(s,e)는 감정 s로부터 감정 e로 전이 시 올바르게 분류된 노래의 수이다.

정확도 표준편차는 측정된 정확도의 표준편차를 뜻하며, 최대 정확도는 분류기가 제시할 수 있는 최고의 정확도를 뜻한다. 최대 정확도가 높다는 것은 특정한 음악 데이터나 특정한 감정 상태 전이에 대해 좋은 결과를 보임을 뜻한다.

그림 2는 Original SVM과 One-Class SVM의 정확도 표준편차와 최대 정확도를 비교하고 있다. 그림에서 정확도 표준편차와 최대 정확도가 서로 비례 관계에 근접함을 보이고 있다. 일반적으로 정확도 표준편차가 높은 경우, 최대 정확도 또한 높음을 알 수 있었다.

그림 3은 Original SVM과 One-Class SVM의 전체 분류 정확도와 최대 정확도를 비교하고 있다. 그림에서 점선은 Original SVM과 One-Class SVM의 우세를 가늠하기 위한 경계선이다. 결과적으로 전체 분류 정확도에서 Original SVM이 그리고 최대 정확도에서 One-Class SVM이 더 나은 성능을 보이고 있다. 다시 말해서, Original SVM은 평균적으로 좋은

분류 정확도를 보여주지만, 특정한 음악에서는 정확도가 떨어지는 반면, One-Class SVM은 특정한 음악에서 좋은 결과를 낼 수 있다.

따라서, 감정 상태 전이 기반의 음악 추천에 있어, 단일 청취자의 음악 취향이 어느 특정 영역에 편중되어 있는 경우, One-Class SVM에 기반한 분류가 좀더 정확한 결과를 보여주며, 일반적인 대중 취향의 평균적 분류를 위해서는 Original SVM이 더 정확한 결과를 보여줄 수 있다.

VII. 결론

본 논문에서는 음악의 분류나 질의 및 추천을 위한 방법으로 기존의 무드나 장르와 같은 정적 분류자 대신 음악을 통해 사람이 느끼게 되는 감정의 전이 효과를 이용하고자 하였다. 이를 위해 음악에 의한 인간 감정 전이의 정도를 정량적으로 표현하기 위한 감정 상태 전이 모델 (ESTM)을 제안하고 이를 기반으로 새로운 음악 분류 및 추천 기법을 제안하였다. 제안하는 모델의 유용성을 보이기 위하여 다양한 음악 클립에서 다수의 특징을 추출하였으며, 고차원 특징 벡터의 차원 감소를 위하여 NMF를 사용하였다. 성능 분석을 위한 실험에서 두 가지 유형의 SVM을 분류기로 사용하여 다양한 실험을 수행하였으며 결과적으로 평균 67.54%, 최대 87.78%의 분류 정확도를 달성하였으며 이를 바탕으로 특정 응용에 적합한 분류기를 선정하는 가이드라인을 제시하였다.

참고문헌

[1] MIREX 2008, <http://www.music-ir.org/mirex/2008/index.php>
 [2] L. Lu, et al., "Automatic Mood Detection and Tracking of Music Audio Signals," IEEE Transactions on Audio, Speech, and Language Processing, vol.14, no.1, pp.5-18, 2006.
 [3] H. Xiang, et al., "An Experimentation on Creating a Mental State Transition Network," Proceedings of IEEE International Conference on Information Acquisition, pp.432-436, 2005.
 [4] 한병준, 노승민, 황인준, "음악 무드 분류에서 음향 특성과 차원 감쇄 기법의 효과 연구," 한국컴퓨터종합학술대회, vol.35, no.1(C), pp.406-411, 2008년.
 [5] 박창호 등, "인지공학심리학: 인간-시스템 상호작용의 이해," 시그마프레스, 2007.
 [6] J.A. Russel, "A circumplex model of affect," Journal of Personality Social Psychology, vol.39,

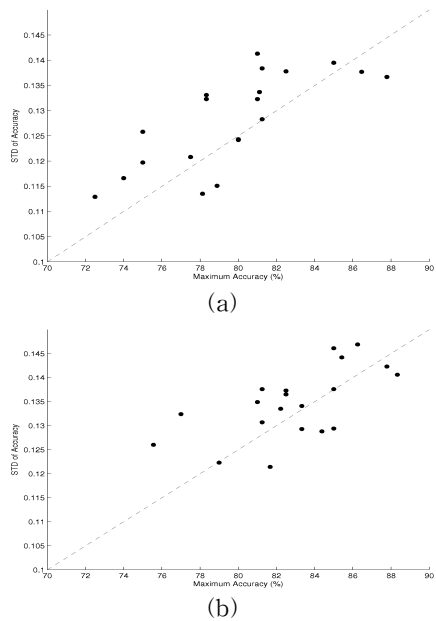


Fig. 2. Accuracy standard deviation and maximum accuracy (a) Original SVM (b) One-Class SVM
 그림 2. 정확도 표준편차와 최대 정확도 (a) Original SVM (b) One-Class SVM

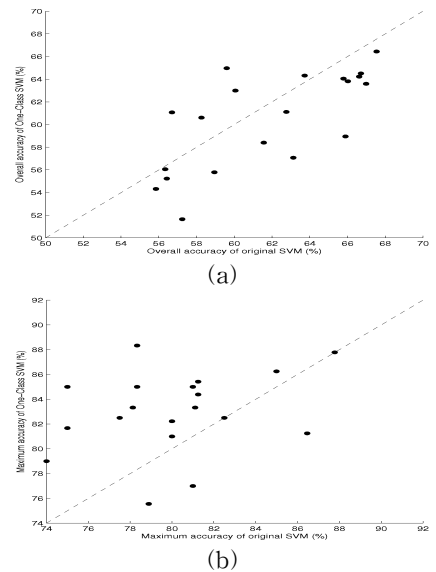


Fig. 3. Comparison of Original SVM and One-Class SVM (a) Overall classification accuracy (b) Maximum Accuracy
 그림 3. Original SVM과 One-Class SVM의 비교 (a) 전체 분류 정확도 (b) 최대 정확도

pp.1161-1178, 1980.

- [7] R.E. Thayer, "The Biopsychology of Mood and Arousal," Oxford Univ. Press, 1989.
- [8] S. Jun, S. Rho, B. Han, E. Hwang, "A Fuzzy Inference-based Music Emotion Recognition System," IEE International Conferences on Visual Information Engineering, 2008.
- [9] Meinard Müller, "Information Retrieval for Music and Motion," Springer, 2007.
- [10] L. R. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition," Prentice Hall, 1993.
- [11] M. Cord, P. Cunningham, "Machine Learning Techniques for Multimedia," Springer, 2008.
- [12] A. Klapuri and M. Davy, "Signal Processing Methods for Music Transcription," Springer Science+Business Media LLC, 2006.
- [13] D. Ellis and G. Poliner, "Identifying 'Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking," Proceedings of IEEE ICASSP 2007, vol.4, pp.1429-1432, 2007.
- [14] Krumhansl C (1990) Cognitive Foundations of Musical Pitch. Oxford University Press.
- [15] D.D.Lee and S.Seung, "Learning the parts of objects by non-negative matrix factorization," Nature vol.401, no.6755 pp.788-791, 1999.
- [16] D.D.Lee and S.Seung, "Algorithms for Non-negative Matrix Factorization," Advances in Neural Information Processing System 13: Proc. of the 2000 Conference, pp.556-562. MIT Press, 2001.
- [17] E. Allwein, et al., "Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers," Journal of Machine Learning Research, vol.1, pp.113-141, 2000.
- [18] All Music Guide, <http://www.allmusic.com/>
- [19] 뮤직인터페이스 멜론(melon), <http://melon.com/>
- [20] C.-C. Chang and C.-J. Lin, "LIBSVM : a library for support vector machines," Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm> of subordinate document. Access 30 Jun 2008.
- [21] B. E. Boser, et al., "A training algorithm for optimal margin classifiers," In proc. of the Fifth Annual Workshop on Computational Learning Theory, pp.144-152, ACM Press, 1992.
- [22] C. Cortes and V. Vapnik, "Support-vector network. Machine Learning," vol.20, 1995.
- [23] B. Schölkopf, J.C. Platt et al., "Estimating the support of a high-dimensional distribution,"

Microsoft research corporation technical report MSR-TR-99-87, 1999.

- [24] A. Holzapfel and Y. Stylianou, "Musical Genre Classification Using Nonnegative Matrix Factorization-Based Features," IEEE Trans. on ASLP, vol.16, no.2, pp.424-434, 2008.

저 자 소 개

한 병 준 (학생회원)



2005년 : 고려대학교 전자과 졸업 (공학사)
2008년 : 고려대학교 대학원 전자과 (공학석사)
2008년~현재 : 고려대학교 대학원 전자과 (공학박사 과정)

<주관심분야> 멀티미디어 특성 추출, 음향/음악 분석, 소리/영상 검색 시스템

황 인 준 (정회원)



1988년 : 서울대학교 컴퓨터공학과 졸업 (공학사)
1990년 : 서울대학교 대학원 컴퓨터공학과 졸업 (공학석사)
1998년 : Univ. of Maryland at College Park 전산학과 졸업 (공학박사)

1998~1999 : Bowie State Univ. 조교수
1999~1999 : Hughes Research Lab. 연구교수
1999~2004 : 아주대학교 정보통신전문대학원 조교수/부교수
2004~현재 : 고려대학교 전기전자전공공학부 조교수/부교수

<주관심분야> 데이터베이스, 멀티미디어 검색, 스크린 데이터 처리, 웹 응용, 영상처리, 유비쿼터스 컴퓨팅

ehwang04@korea.ac.kr