J Korea Industr Inf Syst Res Volume 18 Number 4
http://dx.doi.org/10.9723/jksiis.2013.18.4.025

단위 신경망과 특징벡터 차원 축소 기반의 음악 분위기 자동판별†

(Music Mood Classification based on a New Feature Reduction Method and Modular Neural Network)

송 민 균*, 김 현 수**, 문 창 배**, 김 병 만***, 오 득 환****
(Min Kyun Song, HyunSoo Kim, Chang-Bae Moon, Byeong Man Kim, and Dukhwan Oh)

요 약 본 논문에서는 개인화된 분위기 분류 모델 대신에 대중의 분위기 분류 모델을 제안한다. 분위기 판별 성능을 개선하기 위해 두 가지 접근 방법을 선택하였는데, 그 첫 번째가 표준편차에 기초한 특징축소이다. 이는 음악의 특징을 추출하기 위해 사용하는 MIRtoolbox에서 추출되는 391개의 특징들을 모두 사용할 경우의 성능 저하 문제를 해결하기 위한 방법이다. 실험결과, 본 논문에서 제안한 특징축소 방법이 기존의 차원 축소 방법인 R-Square와 PCA보다 성능이 좋음을 확인할 수 있었다. 그리고 특징축소 방법만으로는 성능 개선에 한계가 있어 두 번째 개선 방법으로 단위 신경망을 사용하여 추가의 성능 개선을 시도하였다. 실험결과 이 역시 유효한 성능 개선이 이루어짐을 확인할 수 있었다.

핵심주제어 : 분위기 분류, 특징 차원 축소, 단위 신경망

Abstract This paper focuses on building a generalized mood classification model with many mood classes instead of a personalized one with few mood classes. Two methods are adopted to improve the performance of mood classification. The one of them is feature reduction based on standard deviation of feature values, which is designed to solve the problem of lowered performance when all 391 features provided by MIR toolbox used to extract features of music. The experiments show that the feature reduction methods suggested in this paper have better performance than that of the conventional dimension reduction methods, R-Square and PCA. As performance improvement by feature reduction only is subject to limit, modular neural network is used as another method to improve the performance. The experiments show that the method also improves performance effectively.

Key Words : Mood Classification, Feature Dimension Reduction, Modular Neural Network

† 본 연구는 금오공과대학교학술연구비에 의하여 연구된 논문
* 금오공과대학교 컴퓨터소프트웨어공학과, 제1저자
** 금오공과대학교 컴퓨터소프트웨어공학과
*** 금오공과대학교 컴퓨터소프트웨어공학과
**** 금오공과대학교 컴퓨터소프트웨어공학과, 교신저자
(dhoh@kumoh.ac.kr)

1. 서 론

많은 사람들이 MP3플레이어나 휴대폰, PC등을 통해 일상생활 속에서 음악을 수시로 접하고 있다. 최근 멀티미디어 데이터베이스에 저장되는 음악 정보는 양이 방대할 뿐만 아니라 시시각각 급속히 증가하고 있다. 또한, 인터넷의 발달로 인터넷을 이용한 디지털 음악도 많이 늘어 더 쉽고 빠르게 음악을 접할 수 있게 되었다. 이러한 음악 콘텐츠의 증가로 인해 사용자가 원하는 음악을 쉽게 찾고 관리하기 위해 음악의 속성에 대한 분류 시스템이 요구된다.

기존 음악을 분류하는 방법은 메타데이터 정보를 입력 받아 분류한 결과를 탐색하는 방식이었다. 하지만 텍스트 기반의 음악 검색은 메타데이터가 해당 음악의 정보를 충분히 기록하지 못할 경우 제대로 분류할 수 없고, 사용자가 검색할 때 잘못된 결과가 발생할 수 있다. 또한 수작업에 의하여 메타데이터를 기록하기 때문에 방대한 양의 음악에 적용되기는 지속적인 관리의 부담이 크게 된다. 이와 같은 여러 가지 이유로 음악으로부터 장르나 분위기 등을 음악의 내용 정보 즉 음향 정보에서 직접적으로 추출하거나 유추하는 기술이 필요하다.

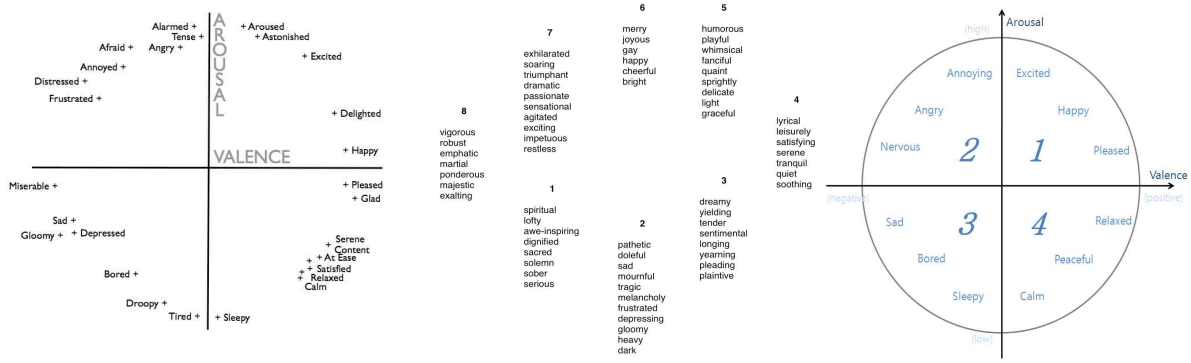
한 음악의 분위기는 음악 전체에 대하여 동일한 분위기가 아닌 시간 변화에 따라 지속적으로 변한다. 분위기뿐만 아니라 음률의 변화 또한 다양하다. 이렇게 음악의 분위기는 사람이 느끼는 감정만큼이나 다양하다. 하지만 [1, 2, 3, 4, 5]의 연구들을 보면 적은 수의 분위기 분류로 학습을 진행하거나, 많은 수의 분위기 분류로 학습을 진행할 때는 일반화 보다는 사용자 각각에 맞춘 개인화에 초점을 맞추는 방향으로 연구가 진행되어 왔다.

이에 본 논문에서는 일반화에 초점을 맞춘 대중화된 분류모델을 구축하기 위하여 많은 수의 피 실험자를 대상으로 데이터를 수집하고 이를 이용하여 대표 분위기를 정의한 후 신경망을 이용하여 대중의 분위기 모델을 학습하였다.

본 논문에서는 분위기 판별 성능을 개선하기 위해 두 가지 접근 방법을 선택하였는데, 그 첫 번째가 표준편차에 기초한 특징축소이다. 이는 음악의 특징을 추출하기 위해 사용하는 MIRtoolbox [6]에서 제공하는 391개의 특징들을 모두 사용할 경우의 성능 저하 문제를 해결하기 위한 방법이다. 실험결과, 본 논문에서 제안한 특징축소 방법이 기존의 차원 축소 방법인 R-Square[7]와 PCA[8]보다 성능이 좋음을 확인할 수 있었다. 그리고 특징축소 방법만으로는 성능 개선에 한계가 있어 두 번째 개선 방법으로 단위 신경망[9]을 사용하여 추가의 성능 개선을 시도하였다. 실험결과 이 역시 유효한 성능 개선이 이루어짐을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구, 3장에서는 본 논문에서 제안하는 알고리즘, 4장에서는 음악 분위기 수집 방법과 실험 결과, 마지막으로 5장에서는 결론을 짓도록 한다.

2. 관련연구



(a) Russell Model (b) Hevner Model (c) Expanded Thayer's 2-Dimensional Model

<Fig 1> Mood Collection Model

기존 음악 분위기 모델에는 <그림 1> (a)의 Russell 모델[10], <그림 1> (b)의 Hevner 모델[11] 그리고 Thayer 모델[12]이 있다. Russell 모델과 Hevner 모델은 형용사를 기반으로 한 모델로 의미가 중첩되거나 형용사적 표현상 모호한 단점을 가지고 있기 때문에 본 논문에서는 확장된 Thayer의 2차원 분위기 모델을 사용한다.

Thayer의 2차원 분위기 모델에서는 음악 분위기를 Arousal과 Valence로 이루어진 벡터 값으로 표현을 하는데 Arousal은 청취자가 음악에서 느끼는 자극의 강도를 나타내며 Valence는 음들의 안정감을 나타낸다. <그림 1> (c)는 확장된 Thayer의 2차원 분위기와 12개의 분위기/감정 형용사와의 관계를 나타낸 그림이다.

기존 음악 분위기 인식에 대한 연구들은 [13-23]등이 있다. Liu [13]는 음악 분위기 인식 시스템을 제안하였는데 이 시스템에서는 요한 슈트라우스의 왈츠를 다섯 가지로 분류하기 위해 퍼지 분류기를 사용하였으며 템포, 세기, 피치변화, note density, 음색(timbre) 등의 특징을 사용하였다.

Katayose [14]는 팝음악에 대해서 감정 (sentiment) 추출 시스템을 제안하였는데, 이 시스템에서는 단선율의 음향데이터가 음악 코드로 변환되고 이로부터 멜로디, 리듬, 하모니, 형식 (form)등이 추출된다.

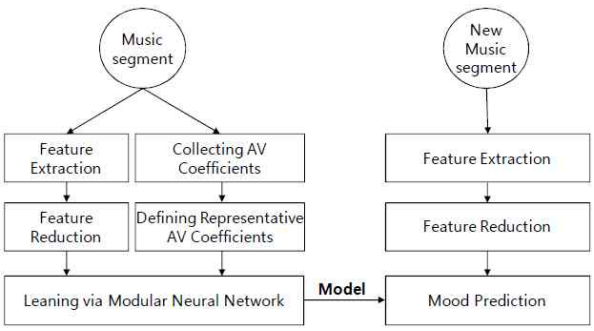
이러한 두 시스템은 나름대로 의미를 갖고 있으나 음향 데이터로부터 유용한 특징을 추출하기 어려운 관계로 MIDI 또는 기호적 표현을 사용하고 있다. 하지만, 많은 실세계의 음악이 기호적 표현으로 되어있지 않고 또한 음향 데이터를 기호적 표현으로 잘 번역할 수 있는 시스템도 존재하지 않는다[15]. 이로 인해 이전부터 음향 데이터로부터 직접적으로 분위기를 탐지할 수 있는 시스템에 대한 필요성이 제기되었다.

Feng [16]은 템포와 아티큘레이션 특징을 이용하여 분위기를 4개 - happiness, sadness, anger, fear -로 분류하는 방법을 제안하였으며 Li와 Ogihara [17]는 음색 (timbre texture), 리듬, 피치 특징을 이용하여 분위기를 탐지하는 방법을 제안하였다. 이 방법에서는 분위기의 분류법으로 Hevner의 검사목록(checklist) [18]을 Farnsworth [19]가 재구성한 13개의 형용사 그룹들을 사용하였다.

[4]에서는 퍼지 기반의 분류 방법을 사용하여 여러 분위기의 강도를 수치로 나타내는 연구를 하였다. 이 연구는 분위기의 특성상 단일 분위기로의 표현의 모호함을 해결하기 위해 퍼지 기반의 분위기 탐색 방법을 사용하였다. 하지만 개인화 서비스를 제공하는 시스템인 경우, 퍼지 방법을 사용하면 개인의 주관적 성향을 제대로 처리하지 못할 수 있음을 지적하고 [23, 24], 이를 해결하기 위해 분위기 클래스를 사용하는 것이 아닌 Thayer의 2차원 분위기 모델의 각 축의 값을 직접 -1~1사이의 실수로 두어 사용하였다. AV계수라 불리는 2차원 벡터로 이루어진 이 값은 각 값이 실수로 이루어지기 때문에 두 개의 회기 분석기를 통해 학습 및 추출이 가능하게 된다. AV계수를 얻기 위하여 피 실험자들이 각 음악마다 개개인이 생각하는 AV값을 직접적으로 입력하는 방식으로 데이터 수집을 하였다. [23]의 연구에서는 다양한 사용자들로부터 얻은 AV계수와 비슷한 사용자 집단(음악의 이해도 정도에 따라 전문가/비전문가로 구분)의 정보를 고려한 개인화에 맞춘 탐지 방법에 대해서도 연구하였다.

3. 단위 신경망과 차원 축소 기반의 음악 분위기 자동판별 알고리즘

본 논문의 음악 분위기 판별 구조는 <그림 2>와 같다. 최초 음악이 입력되면 음악의 구조를 파악하여 음악을 음원단위로 자르고, 대표 음원을 선택한다. 선택한 대표 음원의 특징과 대표 음원의 대표 분위기를 추출하여 신경망에 학습 및 판별하는 구조이다.

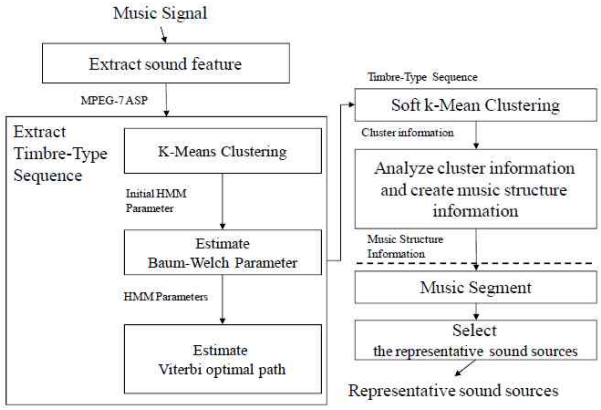


<Fig 2> Structure of mood classification

3.1 대표 음원 추출

대표 음원을 선정하기 위해 <그림 3>과 같이 음악의 구조 분석을 통한 세그먼트를 실행하였다. 대표 음

원 선정 방법은 음악 구조 정보를 추출하고, 분석된 구조정보를 이용하여 음악을 분리한 후 분리된 음원 중 에너지 값이 가장 큰 위치를 선택하여 대표 음원을 선정하였다. 음악의 구조 분석은 상태 열 기반 [25]의 유사 구간 클러스터링 방법을 사용하였다. 유사 구간 클러스터링 방법은 <그림 3>에서 점선부분까지로 음악 특징 벡터 추출, Timbre-Type 시퀀스 추출, Timbre-Type Soft k-Means 클러스터링 방법을 통하여 음악의 구조정보를 파악한다.



<Fig 3> Representative Sound Source Extraction

기존 상태 열 기반의 음악 구조 분석 연구[25]에서는 1차 음향 특징 추출을 위한 특징 추출 프레임의 길이 결정 방법으로 비트 탐색 알고리즘을 통하여 8개의 비트에 해당하는 길이를 프레임 윈도우의 홉사이즈로 사용하였다. 본 논문에서는 [1]의 연구와 마찬가지로 1.2s의 길이와 300ms의 홉사이즈를 가진 고정된 프레임을 사용하여 1차 음향 특징을 추출하였다.

본 논문에서는 유사구간을 획득한 후 시작부분부터 12초 단위로 음악을 분리시키고, 분리된 12초 단위의 음원들 중 음악의 도입부에서 1개와 종결부에서 1개를 선택한다. 그리고 음원들의 에너지를 계산하여 에너지가 가장 큰 샘플을 1개 선택하였다. 에너지는 식 (1)에 의하여 계산한다. 또한 음악당 최대 3개의 샘플을 선택하지만 도입부나 종결부에 에너지가 최대인 경우 음악당 2개의 샘플을 선택한다.

$$\epsilon_x = \sum_{-\infty}^{\infty} x(n)x(n)^* = \sum_{-\infty}^{\infty} |x(n)|^2 \qquad (1)$$

여기서, $x(n)$은 음원의 시퀀스, $x(n)^*$는 시퀀스의 켤레복소수를 의미한다.

3.2 대표 분위기 정의

한 음악에 대해 개개인이 느끼는 분위기는 다르기 때문에 본 논문에서는 피 실험자들로부터 음악의 분위기를 수집한 후 이를 기반으로 음악의 대표 분위기를 정의하였다. 대표 분위기를 정하는 방법은 우선 실험에 사용한 각 음원에 대해 피 실험자가 느낀 분위기를 분위기 별로 모든 피 실험자의 평가치를 식 (2)와 같이 더한다.

$$ed_i^s = \sum_{i=1}^{n} data_{ui}^s, \begin{cases} i = 1, 2, \ldots, 12 \\ n: \text{피실험자수} \end{cases} \qquad (2)$$

여기서, ed_i^s는 음원 S에 대한 i번째 분위기에 대한 피 실험자들의 평가치 합이고, $data_{ui}^s$는 음원 S에 대한 피 실험자 u의 i번째 분위기에 대한 평가치 이다. ed_i^s는 결국 음원 S를 사용자들이 얼마나 i번째 분위기로 인지하는 지를 나타내는 척도이다. 즉, 이 값이 크면 클수록 많은 사람들이 해당음악을 해당 분위기로 느낀다는 것이다. 향후 이 값을 음원 S의 i번째 분위기 강도라 칭한다.

음원 S의 i번째 분위기 강도가 주어지면, 이로부터 i번째 분위기에 의한 Valence(V_i^s)와 Arousal(A_i^s) 값을 아래 식 (3)을 이용하여 계산할 수 있다

$$\begin{aligned} V_i^s &= \sin(f\theta_i) ed_i^s,\ i = 1, 2, 3, \ldots, 11, 12 \\ A_i^s &= \cos(f\theta_i) ed_i^s,\ i = 1, 2, 3, \ldots, 11, 12 \\ f\theta_i &= f\theta_{i-1} + 30,\ 2 \le i \le 12, f\theta_1 = 15 \end{aligned} \qquad (3)$$

여기서, $f\theta_i$는 i번째 분위기의 중심축 각도를 의미한다. 예를 들어, <그림 1>에서 첫 번째 분위기 "pleased"는 0°~30° 범위에 해당하여 따라서 $f\theta_1$는 15°이다.

이렇게 구한 음원의 분위기별 V_i^s와 A_i^s의 값을 이용하여 식 (4)와 같이 12 분위기 전체의 평균을 구해서 최종적인 V_{total}^s와 A_{total}^s의 값을 구한다. 이렇게

구한 V_{total}^{s}와 A_{total}^{s}가 그 음원의 최종 Valence와 Arousal 값이다.

$$V_{total}^{s} = \frac{1}{12}\sum_{i=1}^{12} V_i^s,\ A_{total}^{s} = \frac{1}{12}\sum_{i=1}^{12} A_i^s \qquad (4)$$

이 Valence와 Arousal 값을 이용해서 이 값에 해당하는 분위기를 구하게 된다. 먼저, 확장된 Thayer의 2차원 분위기 모델의 경우 12개의 분위기가 있고 한 분위기당 30°의 영역을 가지기 때문에 음원의 Valence와 Arousal 값을 이용해서 이 값에 해당하는 각도를 식 (5)를 사용하여 구한다.

$$\theta = atan2(total_V, total_A) \times \frac{180}{\pi} \qquad (5)$$

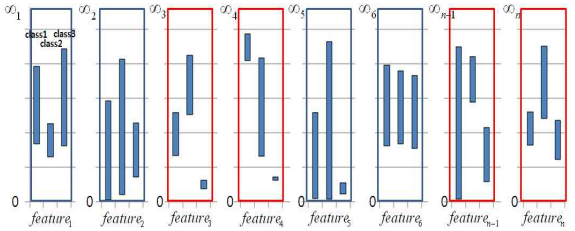
각도를 구한 후에는 그 각도에 해당하는 분위기를 음원의 대표 분위기로 정의한다.

3.3 표준편차를 이용한 특징 축소

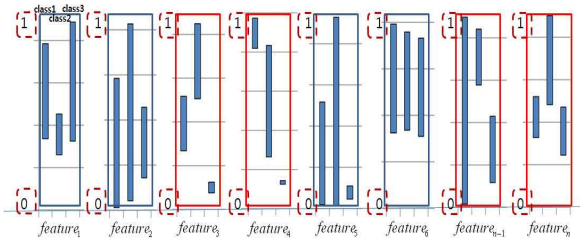
본 논문에서는 MIRtoolbox [6]를 이용하여 음원의 특징을 추출하였고, 이 특징들의 대분류는 Dynamics, Fluctuation, Rhythm, Spectral, Timbre, Tonal 이고, 중분류는 MFCC, Tempo, Chromagram, Rolloff등을 포함한 28개이다. 이 28개의 중분류들 각각에 대해 Mean, Std, Slope, PeriodFreq, PeriodAmp, PeriodEntropy등의 통계 값들을 추출하여 이들을 특징값으로 사용하였는데 최종 특징벡터의 크기는 391차원이 된다. 하지만 MIRtoolbox를 이용하여 특징을 추출하는 경우 “NaN”의 값이 발생하는데 “NaN”은 수로 표현할 수 없는 경우로 본 논문에서는 “NaN”을 포함하는 특징을 제거하여 최종적으로 347차원을 사용하였다.

사전 실험결과 MIRtoolbox [6]를 이용하여 획득한 347개의 특징들을 모두 사용할 경우 잡음 특징들 때문에 오히려 역효과가 발생 하였다. 따라서 본 논문에서는 표준편차를 이용한 특징 축소 방법을 사용하여 특징을 선별하여 사용하였다. 표준편차를 이용하여 특징을 축소하는 방법은 두 가지로 나뉘어 볼 수 있는데 첫 번째 방법은 기본 표준편차를 이용한 특징축소 방법이고, 두 번째 방법은 확장된 표준편차를 이용한 특징축소 방법이다.

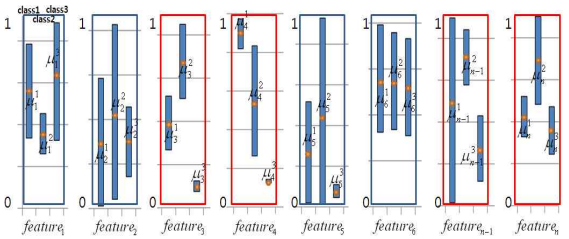
첫 번째 기본 표준편차를 이용한 특징 축소방법은 정규화, 기준정의, 특징선택의 과정으로 구성되며 정규화 과정은 <그림 4> (a)에서 보는 것과 같이 각 특징들은 서로 다른 최댓값을 가지기 때문에 <그림 4> (b)의 점선 사각형과 같이 특징들을 0~1의 사이 값으로 변환하는 과정이다. 정규화 과정은 최초 각 특징들



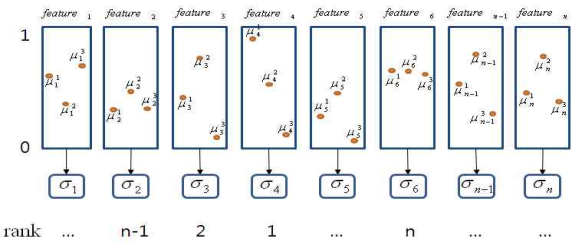
(a) Examples of features



(b) Examples of normalization



(c) Examples of calculating mean point of each class for each feature



(d) Example of selection of features

<Fig 4> Examples for feature reduction processes using standard deviation

에 대하여 최댓값을 구하고, 각 특징에 대하여 최댓값으로 나누는 방법으로 식 (6)을 사용하여 정규화 한다.

$$\widetilde{d_{k,n}^c} = (d_{k,n}^c - \widehat{\infty_n^c})/(\infty_n^c - \widehat{\infty_n^c}),$$
$$\infty_n^c = \max(d_{1,n}^c, d_{2,n}^c, \ldots, d_{k-1,n}^c, d_{k,n}^c), \quad (6)$$
$$\widehat{\infty_n^c} = \min(d_{1,n}^c, d_{2,n}^c, \ldots, d_{k-1,n}^c, d_{k,n}^c)$$

여기서, n은 특징수이고, k는 데이터 인덱스이다. $d_{k,n}^c$는 클래스 c에 속한 k번째 데이터의 n번째 특징 값을 의미하고 $\widetilde{d_{k,n}^c}$는 $d_{k,n}^c$를 정규화한 값, ∞_n^c은 클래스 c에 속한 n번째 특징의 특징 값들 중 가장 큰 값, $\widehat{\infty_n^c}$은 클래스 c에 속한 n번째 특징의 특징 값들 중 가장 작은 값을 의미한다.

정규화 후에는 특징별로 각 클래스의 기준점을 구해야 하는데, 이 기준점은 식 (7)과 같이 클래스에 속한 데이터의 해당 특징 값들의 평균으로 구한다. <그림 4> (c)에 특징별 클래스의 기준점을 예시하였는데 막대 위에 표기된 점이 기준점들이다.

$$\mu_n^c = \frac{1}{k}\sum_{i=1}^{k}\widetilde{d_{i,n}^c} \quad (7)$$

여기서, k는 c클래스에 속한 데이터의 개수이며 μ_n^c는 n번째 특징에 대한 c번째 클래스의 평균을 의미한다.

마지막으로 특징선택은 특징별로 기준점들의 표준편차를 계산하고, 표준편차를 이용하여 분별력이 좋은 특징을 선택한다. 즉, 클래스의 개수를 m개라 하면 모든 클래스에 대하여 j번째 특징의 기준점 $\mu_j^1, \mu_j^2, \ldots, \mu_j^m$을 구한 후 이들의 표준편차 σ_j를 식 (8)을 이용하여 구한다. 모든 특징에 대해 동일하게 표준편차를 구하였으면 표준편차에 대한 순위를 구한 후 순위가 일정 이상인 특징(표준편차가 큰 값)을 선택한다.

$$\sigma_j = \sqrt{\frac{1}{m}\sum_{c=1}^{m}(\mu_j^c - \overline{\mu_j})^2} \quad (8)$$

여기서, $\overline{\mu_j}$은 $\mu_j^1, \mu_j^2, \ldots, \mu_j^m$의 평균이다.

두 번째로 확장된 표준편차를 이용한 특징 축소방법은 표준편차를 이용한 특징축소 방법을 적용하기 이전에 식 (9)와 같이 각 클래스의 특징에 대하여 표준편차를 적용 후 표준편차가 큰 특징을 제거하는 방법이다. 즉, 클래스 내 분산은 최소화하면서 클래스 간 분산은 최대화하는 특징을 선택하는 방법이다.

$$\sigma_j^c = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(d_{i,j}^c - \overline{d_j^c})^2} \quad (9)$$

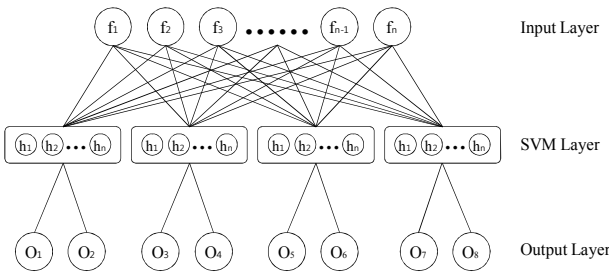
여기서, σ_j^c는 클래스 c에 속하는 데이터의 j번째 특징의 표준편차이고, $d_{i,j}^c$는 i번째 데이터의 j번째 특징값, $\overline{d_j^c}$는 $d_{1j}^c, d_{2j}^c, \ldots, d_{kj}^c$의 평균이다. 식 (8)의 표준편차는 클래스 간 표준편차이며 식 (9)의 표준편차는 클래스 내 표준편차 이다.

모든 클래스와 특징에 대하여 클래스 내 표준편차를 구하고, 그 표준편차에 대하여 순위를 구한 후 순위가 일정 이상인 값을 가지는 특징을 제거한 후, 즉, 클래스 내 분산이 큰 특징을 제거하고 남은 특징들을 사용하여 기본 표준편차를 이용한 특징축소 알고리즘에 적용하면 된다.

3.4 단위 신경망을 이용한 분위기 학습 및 평가

본 논문에서 구축한 신경망의 구조는 <그림 5>와 같다. 입력층은 3.3에서 사용한 확장된 표준편차를 사용하여 획득한 특징들을 사용하고, 모듈층은 은닉층으로만 구성되며 은닉노드의 개수를 조절할 수 있도록 하였다. 출력층은 2개로 구성된다. 본 논문에서는 학습 및 판별에 사용할 특징 수를 50개로 설정하였다.

본 논문에서 구성된 모듈층은 12개의 분위기로 총 6개로 구성해야 하지만 아주 적은 수의 데이터를 가



<Fig 5> Modular neural network structure

지는 4개 분위기를 제거하고 8개의 분위기에 대해 거리가 먼 분위기를 한 쌍으로 하여 총 4개의 모듈층을 구성하였다. 12개의 분위기로 6개의 모듈층을 구성하는 방법은 <그림 1> (c)에서 극과 극의 분위기 (“Excited”와 “Sleep”같은 분위기)를 하나의 모듈에서 학습하는 방법으로 진행되지만 본 논문에서는 데이터 수의 문제로 Angry, Nervous, Pleased, Sad를 제외한 8개의 분위기를 사용하기 때문에 분위기의 거리가 가장 먼 분위기(본 논문에서 사용한 모듈 구성은 Annoying와 Peaceful, Bored와 Relaxed, Calm와 Excited, Happy와 Sleepy를 각 하나의 모듈로 구성함)를 한 쌍으로 모듈에 학습한 후 새로운 값이 입력되면 $\arg max(o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8)$로 판별하는 방법을 사용하였다. 예를 들어, o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8에서 o_1이 가장 클 경우 분위기는 Annoying이고, o_2가 가장 클 경우 분위기는 Peaceful이다.

4. 실험 및 성능 평가

3장에서 기술한바와 같이 본 연구는 구조분석을 통하여 음악을 각 구간으로 나눈 뒤 대표 구간과 실험에 사용할 구간을 지정하여, 각 구간에 대해서 피 실험자들로부터 분위기 값을 평가 받아 각 구간에 대한 대표 분위기를 지정하였다. 이렇게 구축된 각 음원에 대한 대표 분위기와 음향 특징을 사용하여 분위기 판별 성능을 살펴보았다.

실험에 참여한 약 200명의 피 실험자는 확장된 Thayer의 2차원 분위기 모델의 기본이 되는 AV 모델과 분위기의 관계에 대한 교육을 사전에 받은 뒤, 본인의 느낌을 바탕으로 평가하도록 하였다. 실험 데이터의 분위기 평가는 총 3일에 걸쳐서 받았는데, 약 200명의 피 실험자는 280개의 음원 중 랜덤으로 47개의 음원에 대한 분위기를 평가를 받아서 실험에 사용하였다.

실험에 사용한 음악 데이터의 포맷은 범용적으로 사용되는 음악 포맷인 44100Hz 샘플링 레이트의 스테레오 채널 MP3파일로부터 구조 분석 방법을 사용하여 추출한 구간을 동일한 포맷의 Wav 파일로 각각 저장하여 사용하였다.

각 음원의 구간에 대해서 대표 분위기로 지정된 분위기별 음원의 수는 아래 <표 1>과 같다.

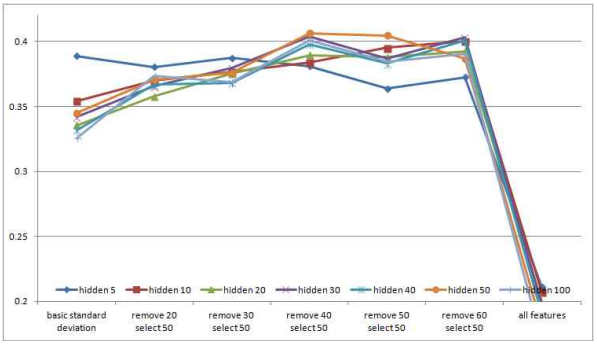
<Table 1> Number of music segments for each representative mood

mood	number of data	mood	number of data
Angry	10	Nervous	7
Annoying	48	Peaceful	38
Bored	12	Pleased	5
Calm	51	Relaxed	18
Excited	30	Sad	10
Happy	32	Sleepy	20

본 논문에서는 데이터 수가 10개 이하인 분위기 Angry, Nervous, Pleased, Sad를 제외한 나머지에 대하여 실험을 실시하였고, 은닉층의 개수는 5, 10, 20, 30, 40, 50, 그리고 100개로 변경하여 실험하였다. 본 논문에서 사용한 검증 방법은 Leave-one-out Cross-validation을 사용하였고, 10,000번 학습하였다. 특징축소 알고리즘의 성능을 비교하기 위해 PCA와 R-Square[7]를 사용한 특징축소 방법과 비교하였고, 단위 신경망의 성능을 비교하기 위해 일반 신경망과 비교하였다.

4.1 표준편차를 이용한 특징축소 성능 평가

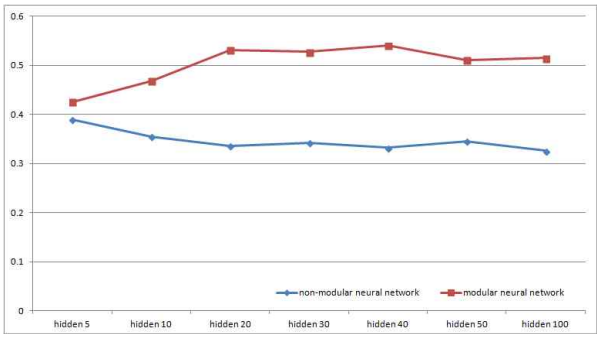
표준편차를 이용한 특징축소 실험은 일반 신경망을 대상으로 하였으며 그 결과는 <그림 6>과 같다. 실험 결과, 특징 축소를 하지 않고 NaN값을 가지는 일부 특징만 제거한 경우 21%의 성능을 보이는 반면, 기본 표준편차를 적용한 경우 은닉층 5에서 39%로 가장 좋은 성능을 보이고, 확장된 표준편차 방법에 의해 20개의 특징을 제거한 후 나머지 특징에서 상위 50개를 선택한 경우(은닉층 노드의 수는 5개) 38%, 특징 30개 제거한 경우 39%, 특징 40개 제거한 경우 41%, 특징 50개 제거한 경우 40%, 특징 60개 제거한 경우 40%의 성능을 보인다. 은닉층의 개수에 따른 성능의 차이는 미미하였다. 또한 전체 특징에 비하여 기본 표준편차의 성능이 18% 향상되었고, 기본 표준편차에 비하여 확장된 표준편차가 2%의 성능이 향상되었다. 결론적으로, 전체 특징을 사용하는 것 보다 표준편차를 이용한 특징축소 방법을 사용하는 경우가 더 좋은 성능을 보였다.



<Fig 6> Performance of suggested feature reduction methods

4.2 단위 신경망 성능 평가

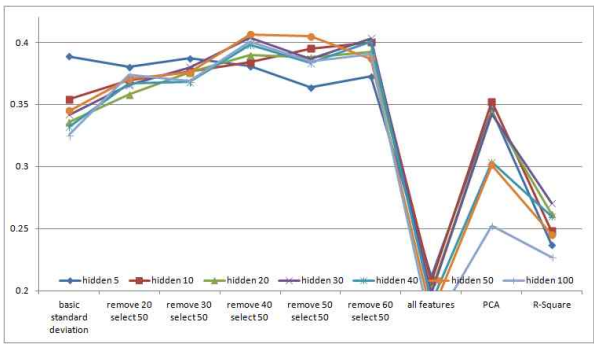
기본 표준편차 특징 축소방법을 적용한 경우 일반 신경망과 단위 신경망의 성능은 <그림 7>과 같다. 그림에서 보는바와 같이, 각 은닉층별 성능적 차이는 은닉층 5인 경우 4%, 은닉층 10인 경우 11%, 은닉층 20인 경우 20%, 은닉층 30인 경우 19%, 은닉층 40인 경우 21%, 은닉층 50인 경우 17%, 은닉층 100인 경우 19%로 단위 신경망의 성능이 향상되어 평균 16%의 성능이 향상 되었다. 즉, 본 논문에서 사용한 단위 신경망이 일반 신경망에 비해 좋은 성능을 보임을 알 수 있다.



<Fig 7> Comparison of performance between general neural network and MNN (using basic standard deviation)

4.3 종합 성능 평가

표준편차를 이용한 특징축소 방법과 단위 신경망을 결합한 경우의 성능은 <그림 8>과 같다. 그림에서 보는바와 같이 기존의 대표적인 특징 축소 방법인 R-Square를 이용한 방법과 PCA를 이용한 방법과의 비교 실험도 하였다. 실험결과, 기본 표준편차에서는 은닉층 40개에서 54%, 확장된 표준편차의 경우 특징제거 20개 은닉층 30개에서 60%, 특징제거 30개 은닉층 50개에서 60%, 특징제거 40개 은닉층 50개에서 58%, 특징제거 50개 은닉층 50개에서 61%, 특징제거 60개 은닉층 40개에서 57%, 전체특징을 사용한 경우 23%, PCA를 이용한 경우 53%, R-Square를 이용한 경우 43%의 성능을 보였다. 즉, 확장된 표준편차를 이용하여 50개의 특징을 제거한 후 상위 50개의 특징과 은닉층 50개를 사용하여 학습한 경우가 가장 좋은 성능을 보였다.



<FIg 8> Performance when feature reduction method and MNN are combined

5. 결 론

본 연구에서는 음악의 분위기 분류를 위하여 기존 적은 수의 분위기나 개인화에 초점을 맞춘 연구가 아닌 일반화에 초점을 맞춘 연구를 수행하였으며 판별 성능 향상을 위해 특징 축소 방법으로 기본 표준편차를 이용한 특징 추출 방법, 확장된 표준편차를 이용한 특징추출방법을 제안하였고, 학습 방법에서는 단위 신경망을 사용하였다.

실험 결과, 특징 축소 방법의 경우 PCA, R-Square, 기본 표준편차를 이용하여 특징을 축소시켜 학습 및 판별하는 방법보다 확장된 표준편차를 이용하여 특징을 축소하여 학습 및 판별하는 방법이 더 좋은 성능을 보이고, 일반 신경망의 성능보다는 본 논문에서 단위 신경망이 더 좋은 성능을 보임을 알 수 있었다.

향후, PCA 차원 축소방법 역시 본 논문에서 제안한

방법과 같은 다양한 차원으로의 축소를 적용하여 성능을 비교할 필요가 있다. 또한 판별 성능을 높이기 위하여 대표구간 선택방법, 대표 분위기 선택방법 그리고 특징축소 방법 및 학습 방법에 대한 보다 세밀한 연구가 필요하다. 성능을 좀 더 개선하기 위해 본 제안 방법을 퍼지화하는 방법도 차후 연구할 필요가 있다.

References

[1] Jong In Lee, Dong-Gyu Yeo, Byeong Man Kim, and Hae-Yeoun Lee, "Automatic Music Mood Detection through Musical Structure Analysis," International Conference on Computer Science and its Application CSA 2009, pp. 510-515, 2009

[2] Hyun Soo Kim, Dong Won Lee, Chang Bae Moon, Byeong Man Kim, and Jong-Yeol Yi, "Mood Lighting System Representing Music Mood," Korea Information Processing Society, 2011 Conference, Vol 18, No. 1, 2011.

[3] Lu, Lie, Dan Liu, and Hong-Jiang Zhang, "Automatic mood detection and tracking of music audio signals," IEEE Trans. Audio, Speech, and Language Processing, Vol. 14, No. 1, pp. 5-18, 2006.

[4] Y. H. Yang, C. C. Liu and H. H. Chen, "Music Emotion Classification: A Fuzzy Approach," Proc. of ACM Multimedia 2006 (ACM MM'06), pp. 81-84, 2006.

[5] P. Singh, A. Kapoor, V. Kaushik, and H. B. Maringanti, "Architecture for Automated Tagging and Clustering of Song Files According to Mood," International Journal of Computer Science Issues, Vol. 7, Issue 4, No 2, pp. 11-17, 2012.

[6] O. Lartillot, and P. Toiviainen, "A Matlab toolbox for musical feature extraction from audio," International Conference on Digital Audio Effects, pp. 237-244, 2007.

[7] H. Simões, G. Pires, U. Nunes, and V. Silva, "Feature Extraction and Selection for Automatic Sleep Staging using EEG," 7th International Conference On Informatics in Control, Automation and Robotics - ICINCO2010, 15-18 June 2010

[8] C. M. Bishop, "Pattern Recognition and Maching Learning," Springer, 2006.

[9] B. L. Happel, and J. M. Murre, "Design and evolution of modular neural network architectures," Neural networks, Vol. 7, No. 6, pp. 985-1004, 1994.

[10] J. A. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, Vol. 39, No. 6, pp. 1161-1178, 1980.

[11] K. Hevner, "Experimental studies of the elements of expression in music," The American Journal of Psychology, Vol. 48, No. 2, pp. 246 - 268, 1936.

[12] R. E. Thayer, "The Biopsychology of Mood and Arousal," New York, Oxford University Press, 1989.

[13] D. Liu, N. Y. Zhang, and H. C. Zhu, "Form and Mood Recognition of Johann Strauss's Waltz Centos," Chinese Journal of Electronics, vol. 12, No. 4, pp. 587-593, 2003.

[14] H. Katayose, M. Imal, and S. Inokuchi, "Sentiment Extraction in Music," Proc. of Int. Conf. on Pattern Recognition, Vol. 2, pp. 1083-1087, 1988.

[15] Eric D. Scheirer, "Music-listening Systems," PhD Thesis, Massachusetts Institute of Technology, 2000.

[16] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 375 - 376, 2003.

[17] T. Li, and M. Ogihara, "Detecting emotion in Music," Proc. of ISMIR 2003, 2003.

[18] K. Hevner, "Expression in music: a discussion of experimental studies and theories," Psychological Review, Vol. 42, No. 2, pp. 186-204, 1935.

[19] P. R. Farnsworth, "The social psychology of music," the Dryden Press, 1958.

[20] Young In Kim, “Content-based Music Retrieval by TIP-indexing Techniques and Features of Audio files”, Korea Industrial Information Systems Society, Vol. 11, No, 3, pp. 10-14, 2006.
[21] Young In Kim, and Seon-Jong Kim, “Performance Analysis of the Time-series Pattern Index File for Content-based Music Genre Retrieval”, Korea Industrial Information Systems Society, Vol. 11, No. 5, pp. 18-27, 2006.
[22] Jun-il Choi, Soon-Cheol Kim, and Joong-Hyuk Chang, “A Multi-functional Memorandum System for Managing Information Intelligently”, Korea Industrial Information Systems Society, Vol. 15, No. 5, pp. 89-95, 2010.
[23] Y. H. Yang, Y. F. Su, Y. C. Lin, and H. H. Chen, “Music emotion recognition: the role of individuality,” Proc. of ACM SIGMM Int. Workshop on Human-centered Multimedia 2007, pp. 13-21, 2007.
[24] Y. H. Yang, C. C. Liu, and H. H. Chen, “A regression approach to music emotion recognition,” Audio, Speech, and Language Processing, IEEE Transactions on, Vol. 16, pp. 448-457, 2008.
[25] M. Levy, M. Sandier, and M. Casey, “Extraction of High-Level Musical Structure From Audio Data and Its Application to Thumbnail Generation,” Proc. of IEEE Int. Conf. Acoustics, Speech, Signal Processing 2006(ICASSP'06), Vol. 5, pp. 13-16, May 2006.

송 민 균 (Min Kyun Song)

- 2009년: 대구대학교 제어계측공학과 공학사
- 20012년: 금오공과대학교 소프트웨어공학과 공학석사

- 관심분야 : 인공지능, 감성공학, 정보검색

김 현 수 (HyunSoo Kim)

- 2008년: 금오공과대학교 컴퓨터공학과 공학사
- 2010년: 금오공과대학교 소프트웨어공학과 공학석사
- 2010년 ~ 현재: 금오공과대학교 소프트웨어공학과 박사과정

- 관심분야 : 인공지능, 소프트웨어공학, 디자인패턴

문 창 배 (Chang-Bae Moon)

- 2007년: 금오공과대학교 컴퓨터공학과 공학사
- 2010년: 금오공과대학교 소프트웨어공학과 공학석사
- 2013년: 금오공과대학교 소프트웨어공학과 공학박사

- 관심분야 : 패턴인식, 멀티미디어처리, 감성공학, 인공지능, 지식검색

김 병 만 (Byeong Man Kim)

- 정회원
- 1987년: 서울대학교 컴퓨터공학과 공학사
- 1989년: 한국과학기술원 전산학과 공학석사
- 1992년: 한국과학기술원 전산학과 공학박사
- 1992년 ~ 현재: 금오공과대학교 교수
- 1998년 ~ 1999년: 미국 UC, Irvine 대학 방문교수
- 2005년 ~ 2006년: 미국 콜로라도 주립대학 연구교수
- 관심분야 : 인공지능, 정보검색, 정보보안

오 득 환 (Dukhwan Oh)

- 1982년: 경북대학교 공과대학 전자공학과 공학사
- 1985년: 한국과학기술원 전산학과 공학석사
- 1994년: 한국과학기술원 전산학과 공학박사
- 1986년 ~ 현재: 금오공과대학교 교수
- 관심분야 : 인공 신경망, 임베디드 시스템, 컴퓨터 네트워크

논 문 접 수 일 : 2013년 05월 06일
1차수정완료일 : 2013년 06월 07일
2차수정완료일 : 2013년 07월 03일
게 재 확 정 일 : 2013년 07월 25일