

폭소노미 분위기 태그를 이용한 음악의 분위기 유형 분석

Analysis of Music Mood Class using Folksonomy Tags

문창배* · 김현수* · 김병만**

Chang Bae Moon* · HyunSoo Kim* · Byeong Man Kim**

*금오공과대학교 소프트웨어공학과

*Dept. of Computer and Software Engineering, Kumoh National Institute of Technology

Abstract

When retrieving music with folksonomy tags, internal use of numeric tags (AV tags: tags consisting of Arousal and Valence values) instead of word tags can partially solve the problem posed by synonyms. However, the two predecessor tasks should be done correctly; the first task is to map word tags to their numeric tags; the second is to get numeric tags of the music pieces to be retrieved. The first task is verified through our prior study and thus, in this paper, its significance is seen for the second task. To this end, we propose the music mapping table defining the relation between AV values and music and ANOVA tests are performed for analysis. The result shows that the arousal values and valence values of music have different distributions for 12 mood tags with or without synonymy and that their type I error values are $P < 0.001$. Consequently, it is checked that the distribution of AV values is different according to music mood.

Key words : mood of music, folksonomy, ANOVA test, last.fm, AV tag

요약

폭소노미 (foxonomy) 분위기 태그를 이용한 음악 검색 시 내부적으로 단어 태그 대신에 수치 태그 (AV 태그 : Arousal과 Valence 값으로 이루어진 태그)를 이용하면 폭소노미의 문제점 중의 하나인 유사어 문제점을 일부 해결할 수 있다. 하지만 이를 위해서는 두 가지 선행 작업이 제대로 이루어져야 하는데, 그 첫 번째가 단어 태그를 수치 태그로 변환하는 작업이며 그 두 번째가 검색 대상인 음악을 수치 태그로 표현하는 작업이다. 첫 번째 작업에 대해서는 이전 연구를 통하여 그 유의성을 보였기 때문에 본 논문에서는 두 번째 작업에 대해서 그 유의성을 밝히고자 하였다. 이를 위하여 본 논문에서는 음악과 AV값 간의 관계를 정의하는 음악-분위기 매핑테이블을 제안하고, ANOVA 검증을 이용하여 분석 하였다. 실험 결과, 동의어 포함 유무에 무관하게 음악 구간의 A값과 V값 모두 12개 음악의 분위기에 대하여 분포차가 발생하고, 모두 제 1종 오류확률 $P < 0.001$ 를 만족하였다. 결론적으로 음악의 분위기에 따라 AV 값 분포가 다르다는 것을 확인할 수 있었다.

주제어 : 음악 분위기, 폭소노미, AVNOVA test, last.fm, AV 태그

+ 이 연구는 금오공과대학교학술연구비에 의하여 지원된 논문임

† 교신저자 : 김병만 (금오공과대학교 소프트웨어공학과 교수)

E-mail : bmkim@kumoh.ac.kr

TEL : 054-478-7544,

FAX : 054-478-7544

1. 서론

폭소노미는 사서나 운영자에 의해 관리되는 기존의 텍소노미 (taxonomy)와는 달리 자발적 참여자들에 의해 협력적으로 유지 관리되는 분류 체계이다. 텍소노미에서는 잘 구성된 고정된 분류체계를 이용하지만 폭소노미에서는 고정된 분류체계 없이 자유 분방한 형태로 분류 작업이 이루어진다. 일반적으로, 폭소노미에서는 참여자들이 콘텐츠 (예: 음악, 동영상)에 대해서 자신의 느낌이나 의견을 태그 형태로 부여하게 된다. 따라서 폭소노미 태그를 사용하여 콘텐츠를 검색할 경우 유사어에 대한 문제점이 발생하게 된다. 유사어에 대한 문제점은 평온한 음악과 잔잔한 음악과 같이 표현은 다르지만 의미적으로는 유사한 태그의 경우에 발생하는 문제점이다(문(2013)).

폭소노미 태그를 이용한 음악 검색 시에도 이러한 문제점이 발생하는데 이를 보완하기 위해 문(2013)의 연구에서는 단어 태그 대신에 내부적으로 수치 태그 (AV 태그 : Arousal과 Valence 값으로 이루어진 태그)를 사용하는 방법을 제안하였다. 이 방법이 제대로 동작하기 위해서는 두 가지 선행 작업이 이루어져야 하는데, 그 첫 번째가 단어 태그를 수치 태그로 변환하는 작업이며 그 두 번째가 검색 대상인 음악을 수치 태그로 표현하는 작업이다. 첫 번째 작업에 대해서는 이전 연구를 통하여 그 유의성을 보였기 때문에 본 논문에서는 두 번째 작업에 대해서 그 유의성을 밝히고자 하였다. 또한, 문(2013)에서는 음악의 대표 구간에 대한 분위기만을 고려하였는데 본 논문에서는 음악의 전체에 아우르는 분위기(전반, 대표, 후반 구간)를 고려하였다.

먼저, 본 논문에서는 음악-분위기 매핑 테이블을 구축하는 방법을 제시하였고, 음악과 분위기 (AV값으로 표현됨) 사이의 연관성을 파악하였다. 즉, 본 논문에서는 음악과 AV값과의 관계를 정의하는 매핑테이블을 제안하고, 실제 구축된 데이터를 이용하여 음악과 AV값의 연관 관계를 분석하였다. ANOVA Test를 실시한 결과, 제 1종 오류확률 $P < 0.001$ 로 음악의 분위기에 따라 각 구간의 분위기 분포가 다를 수 있었다.

또한, 음악-분위기 매핑테이블 분석 시 동의어에 대한 분석을 포함하기 위해 본 논문에서는 인터넷 동의어 사전을 이용하여 분위기의 동의어를 조사하였고, 음악의 전체적인 분위기를 고려하기 위해 음악의 전

반구간과 후반 구간을 추가하여 분석하였다. 분석 결과, 이 또한 제 1종 오류확률 $P < 0.001$ 로 음악의 분위기에 따라 각 구간의 분위기 분포가 다를 수 있었다.

본 연구의 구성은 다음과 같다. 2장에서는 관련연구에 대하여 살펴보고, 3장에서는 음악의 전반, 대표, 후반 구간을 추출 방법과 추출한 각 구간의 AV값을 예측하는 방법 및 폭소노미 태그를 수집하는 방법에 대해서 설명한다. 4장에서는 폭소노미 분위기 태그의 동의어를 고려하여 음악의 분위기를 분류하는 방법을 설명한다. 5장에서는 4장에서는 폭소노미 기반에서 분류된 음악의 AV값을 분석한다. 마지막으로 6장에서는 결론 및 향후 연구에 대해 살펴본다.

2. 관련연구

음악의 감성적 반응을 개념화하는 연구의 대표적인 사례로서 Thayer(1989) 모델이 있고, 그림 1의 Thayer의 2차원 모델에서는 12개의 단어를 사용한다. 또한 본 논문에서도 분위기를 추출하기 위한 모델로 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.

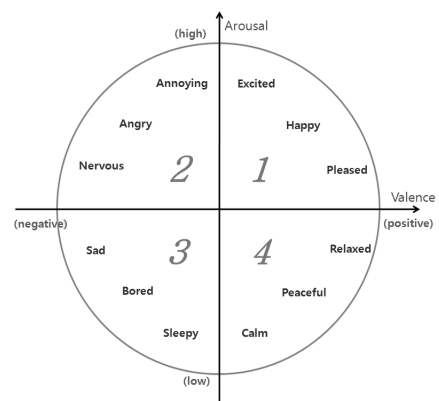


Figure 1. Music Mood Model(문(2003)의 문헌에서 발췌 p.61)

Liu(2003)는 음악 분위기 인식 시스템을 제안하였는데 이 시스템에서는 요한 스트라우스의 왈츠를 다섯 가지로 분류하기 위해 퍼지 분류기를 사용하였다. Katayose(1998)는 팝음악에 대해서 감정 (sentiment) 추

출 시스템을 제안하였다. 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 (Eric D. Scheirer, 2000).

Feng(2003)과 그 일행은 템포와 아티클레이션 특징을 이용하여 분위기를 4개의 분위기로 분류하는 방법을 제안하였으며 Li와 Ogihara(2003)는 음색 (timbre texture), 리듬, 피치 특징을 이용하여 분위기를 탐지하는 방법을 제안하였다. 이 방법에서는 분위기의 분류법으로 Hevner(1935)의 검사목록(checklist)을 Farnsworth(1958)가 재구성한 13개의 형용사 그룹들을 사용하였다. Yang(2006)에서는 퍼지 기반의 분류 방법을 사용하여 여러 분위기의 강도를 수치로 나타내는 연구를 하였다.

Yang(2006)에서는 분위기의 특성상 단일 분위기의 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 지적하였다(Yang, Y.H., Su, Y.F., Lin, Y.C. and Chen, H.H, 2007); Yang, Y.H., Liu, C.C., and Chen, H.H, 2008)). Yang(2007)의 연구에서는 다양한 사용자들로부터 얻은 AV계수와 비슷한 사용자 집단의 정보를 고려한 개인화에 맞춘 탐지 방법에 대해서도 연구하였다.

Lee(2009)의 연구에서는 Thayer의 2차원 분위기 모델과 구조 분석을 이용하여 음악의 분위기를 검출하는 방법에 대한 연구를 실시하였고, 문(2011)의 연구에서도 역시 Thayer의 2차원 분위기 모델을 사용하여 음악 장르에 따른 분위기 분포를 분석하였다. 본 논문에서는 문(2011)의 연구에서 수집하고 정의한 음악의 AV값과 폭소노미 사이트에서 수집한 음악의 태그를 이용하여 폭소노미 태그와 AV값과의 관계를 정의하였다.

음악의 폭소노미 태그를 이용한 연구에는 Steven(2009), Cyril(2009) 그리고 Kim(2011)의 연구가 존재하는데 Cyril(2009)과 Kim(2011)의 연구에서는 폭소노미 사이트(last.fm)에 존재하는 음악의 분위기 태그를 이용하여 음악의 분위기를 정의하고, 음악의 분위기를 예측하는 연구를 진행하였다. 즉, 이 연구들의 경우 하나의 분위기를 하나의 클래스로 취급하여 예측하는 방법이지만 본 논문의 경우는 모든 분위기를 아우르는 AV 모델을 구축하고, 폭소노미 분위기 태그 기반

에서 AV 값을 분석 하였다.

본 논문과 밀접하게 관련된 연구로 문(2013)의 연구가 있다. 이 연구는 태그-분위기 매핑테이블(태그와 AV값 간의 관계)을 제안하고, 태그-분위기 매핑테이블에 존재하는 폭소노미 태그와 이의 AV값과의 관계를 분석하였다. 본 연구에서는 태그와 AV값 간의 관계가 아닌 음악과 AV값 간의 관계를 분석 하였다. 분석 대상이 다르다는 점이 가장 큰 차이점이며 이외에 다음과 같은 두 가지 차이점이 존재한다. 그 첫 번째가 문(2013)의 연구에서는 음악의 대표구간에 대한 AV값을 정의하고, 정의한 대표구간의 AV값과 태그와의 관계를 분석 하였지만 본 논문에서는 음악의 전반, 대표, 후반구간에 대한 AV값을 정의하였고, 정의한 각 구간과 음악과의 관계를 분석 하였다. 즉, 문(2013)의 연구에서는 음악의 전체 분위기를 하나의 분위기(대표구간의 분위기)로 정의하여 분석하지만 본 논문에서는 음악 전체 분위기를 전반, 대표, 후반의 분위기로 정의하여 분석하였다. 두 번째는 문(2013)의 연구에서는 동의어를 고려하지 않고, 폭소노미 분위기 태그와 AV값과의 관계를 분석하였지만 본 논문에서는 동의어를 고려하여 음악과 AV값과의 관계를 분석하였다.

3. 음악의 전반, 대표, 후반 AV값 정의 및 폭소노미 태그 수집

음악의 전반, 대표, 후반의 AV값을 정의하고, 폭소노미 태그를 수집하여야 하는데 음악들에 대하여 사용자로부터 개별적으로 AV값을 수집하여야 하고, 폭소노미 사이트에서 폭소노미 태그를 수집하여야 한다. 하지만 이러한 방법은 수집 비용이 너무 커 본 연구자들의 능력으로는 현실적이지 못하다. 따라서 본 논문에서는 음악과 AV값 사이의 연관관계를 분석하기 위해 문(2013)의 연구에서 사용한 데이터와 AV값 예측 모델을 재사용하였다. 하지만 문(2013)의 예측 모델은 음악의 대표구간에 대한 AV값만을 예측하였기 때문에 본 논문의 분석을 위해 음악의 전반과 후반 구간의 AV값을 추가로 예측할 필요가 있다. 본 논문에서는 이 부분에 중점을 두어 설명한다. 대표구간에 대한 내용은 문(2013)의 연구를 참조하기 바란다.

전반과 후반 구간의 AV값을 예측하는 방법은 그림 2와 같이 세 과정으로 구성된다. 그 첫 번째가 최초 음

악의 전반과 후반 구간을 추출하는 과정이고, 두 번째가 음악의 각 구간의 특징을 추출하는 과정이고, 마지막 단계가 문(2013)의 AV값 예측 모델을 이용하여 각 구간의 AV값을 예측하는 과정이다. 이 과정들을 실행함으로써 음악 각 구간의 AV값들이 예측되고, 예측한 AV값들을 이용하여 음악-분위기 매핑테이블을 생성한다.

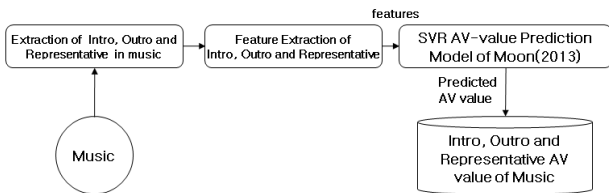


Figure 2. Prediction Process for Intro, Outro and Representative AV value

3.1. 음악의 전반 및 후반 구간 추출

음악의 전반 및 후반 구간을 추출하기 위해 본 논문에서는 문(2011)에서 사용한 음악의 구조 분석방법을 사용하였다. 전반 및 후반 구간 추출 방법은 음악 구조 정보를 추출하고, 분석된 구조정보를 이용하여 12s 단위로 음악을 분리한 후 분리된 음원 중 전반부 음원과 후반부 음원을 선택하였다.

음악의 구조 분석은 상태열 기반 Sandier(2006)의 유사 구간 클러스터링 방법을 사용하였다. Sandier(2006)의 유사 구간 클러스터링 방법은 음악 특징 벡터 추출, 팀버 타입(Timbre-Type) 시퀀스 추출, 팀버 타입(Timbre-Type) Soft k-Means 클러스터링 방법을 통하여 음악의 구조정보를 파악하는 방법이다.

3.2. 음악의 각 구간 AV값 예측

음악 전반과 후반의 AV값은 문(2013)의 연구에서 구축한 AV값 예측모델을 사용하여 예측하였다. 문(2013)의 연구에서 구축한 AV값 예측모델은 SVR(Support Vector Regression)과 문(2011)에서 수집한 데이터를 사용하여 구축한 모델이고, 음악 구간의 음향 특징은 Lartillot(2007)의 MIRToolbox를 사용하여 획득한다.

MIRToolbox를 사용하여 특징 추출하면 총 391개의 특징이 추출되는데 391개의 특징들 중 NaN이 추출된 특징들은 제거하였다. 또한 문(2013)의 연구에서 사용

한 SVR은 LIBSVM에서 제공하는 SVR 이다(Ryu(2008); Chang(2001); Lee(2009)).

음악 각 구간의 AV값을 예측하게 되면 그림 3과 같은 테이블을 구축할 수 있는데 이 테이블이 음악-분위기 매핑테이블 이다.

	Intro		Representative		Outro	
Music 1	V 1,1	A 1,1	V 1,2	A 1,2	V 1,3	A 1,3
Music 2	V 2,1	A 2,1	V 2,2	A 2,2	V 2,3	A 2,3
⋮						
Music n-1	V n-1,1	A n-1,1	V n-1,2	A n-1,2	V n-1,3	A n-1,3
Music n	V n,1	A n,1	V n,2	A n,2	V n,3	A n,3

Figure 3. Mapping Table of Music

4. 분석을 위한 매핑테이블 구축

폭소노미 기반에서 음악과 AV값 간의 연관 관계를 분석하기 위해 본 논문에서는 그림 4와 같은 순으로 분석용 매핑테이블을 구축 하였다.

본 논문에서는 분석용 매핑테이블을 두 가지 경우로 나누어 분석 하였는데 그 첫 번째가 음악의 폭소노미 분위기 태그를 이용하여 음악을 Thayer의 2차원 분위기 모델에 포함된 12개 분위기 기준으로 나누어 분석한 것이고(그림 4의 우측 진행), 두 번째가 12개 분위기의 동의어를 고려하여 분석한 것이다(그림 4의 좌측 진행). 본 논문에서는 12개 분위기 태그의 동의어를 고려하기 위해 인터넷 사전을 이용하여 이들의 동의어를 조사 하였다.

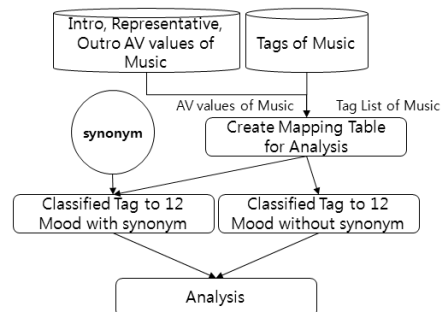


Figure 4. Defining mapping table for analysis

4.1. 분석용 매핑테이블 생성

음악-분위기 매핑테이블을 살펴보면 그림 3의 좌측

상단과 같이 음악 ID와 전반, 대표, 후반 구간의 AV 값만 존재하고, 음악의 분위기 태그는 존재하지 않는다. 음악의 분위기에 따른 AV값 분포를 분석하기 위해서는 음악의 AV값을 분위기 태그에 따라 분류할 필요가 있다. 이를 위해 본 논문에서는 그림 5의 하단과 같이 분석용 매핑테이블을 생성하였다. 분석용 매핑테이블은 그림 5 좌측상단 음악 각 구간의 AV값과 그림 5 우측상단 음악의 태그를 음악 ID를 기준으로 병합하는 방법으로 생성된다. 그림 5의 우측상단의 분위기 태그 데이터는 문(2013)의 데이터로 이 데이터는 널리 알려진 음악 폭소노미 사이트인 last.fm의 자료를 가공하여 만들어진 것이다.

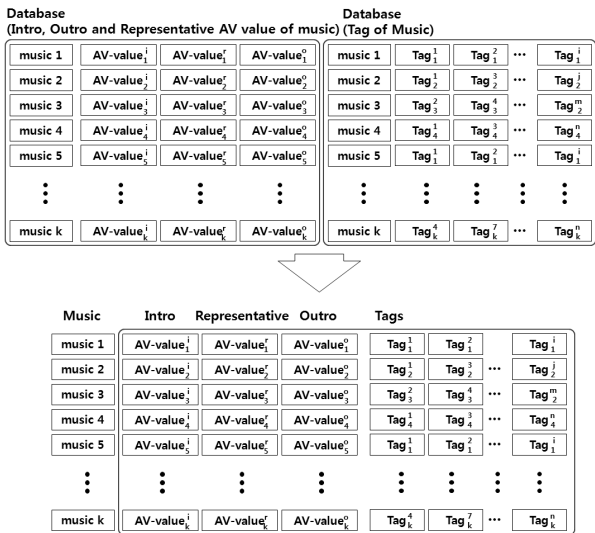


Figure 5. mapping table reconstitution for analysis

4.2. 분위기 태그의 동의어

분석용 매핑테이블에 포함된 태그의 동의어를 조사하기 위해 인터넷 사전(www.synonym.com)을 이용하였고, 모든 분위기 태그의 동의어를 고려하여 음악의 분위기에 따른 AV값 분포를 분석하는 것은 불가능하기 때문에 본 논문에서는 그림 1의 (c)에 포함된 12개의 형용사 분위기 단어만을 고려하여 이들의 동의어를 조사하였다.

본 논문에서 수집한 동의어는 annoying 동사 14개와 형용사 13개, angry 형용사 31개, nervous 형용사 13개, sad 형용사 19개, bored 동사 4개와 형용사 4개, sleepy 형용사 3개, calm 동사 23개와 형용사 11개, peaceful 형용사 9개, relaxed 동사 27개의 동사와 형용사 4개, pleased 동사 7개와 형용사 11개, happy 형용사

13개, excited 동사 37개와 형용사 27개 이다.

4.3. 분위기 태그를 이용한 음악 분류

본 논문에서 구축한 분석용 매핑테이블의 경우 12 분위기 태그와 12 분위기의 동의어 태그 그리고 기타 태그들로 구성된다. 본 논문에서는 12 분위기를 기준으로 분석하기 위해 기타 태그들을 제거하였고, 4.2에서 조사한 동의어와 12 분위기를 이용하여 12 분위기 그룹으로 분류하였다.

본 논문의 분류 방법은 동의어 포함 유무에 따라 두 가지가 존재하는데 그 첫 번째가 동의어를 고려한 12 분위기 그룹으로 분류하는 방법이고, 두 번째가 동의어를 포함하지 않은 12 분위기 그룹으로 분류하는 방법이다.

첫 번째 방법의 경우 그림 6의 좌측 상단과 같이 태그들 중에서 12 분위기 단어와 이의 동의어를 파악하고, 우측 상단과 같이 12 분위기 태그와 이들의 동의어 태그만을 선별한다. 선별한 분위기를 그림 6의 하단과 같이 음악을 12개의 분위기 그룹으로 분류한다. 이의 예로, 그림 6에서 음악 1의 “angered”는 “angry”의 유사어로 이 음악은 angry 그룹으로 분류하고, 음악 3의 “asleep”는 “sleepy”의 동의어로 이 음악은 sleepy 그룹으로 분류하여 12 분위기로 모든 음악을 분류한다.

두 번째 방법의 경우 그림 7의 좌측 상단과 같이 태그들 중에서 12 분위기 단어가 포함된 분위기 태그만을 선별한 후, 우측 상단과 같이 음악을 기준으로 분위기 태그들을 정리하고 이를 이용하여 그림 7의 하단과 같이 음악을 12개의 분위기로 분류하는 방법이다. 참고로, 음악 1의 태그가 “angered”로 angry의 동의어이기 때문에 그림 6에서는 angry 그룹으로 속하지만 그림 7에서는 angry 그룹으로 속하지 못한다. 반면에 음악 4의 경우 태그가 “Happy song”으로 “Happy” 분위기 단어를 포함하기 때문에 그림 7에서도 happy 그룹으로 분류된다.

본 논문에서 분석에 사용한 음악의 수는 총 1,243개이고, 이 음악들은 적어도 하나의 분위기 태그들을 갖고 있으며 여러 개의 분위기가 존재하는 경우도 있다. 또한 동의어를 포함한 경우 음악의 수는 calm 450개, Pleased 202개, sad 492개, Excited 268개, Nervous 134개, peaceful 68개, Relaxe 469개, happy 269개, Bored 161개, sleepy 74개, Angry 196개 그리고 Annoying 234

개 이고, 동의어를 포함하지 않은 경우 음악의 수는 calm 151개, Pleased 3개, sad 480개, Excited 15개, Nervous 20개, peaceful 44개, Relaxe 34개, happy 256개, Bored 26개, sleepy 39개, Angry 90개 그리고 Annoying 30개 이다.

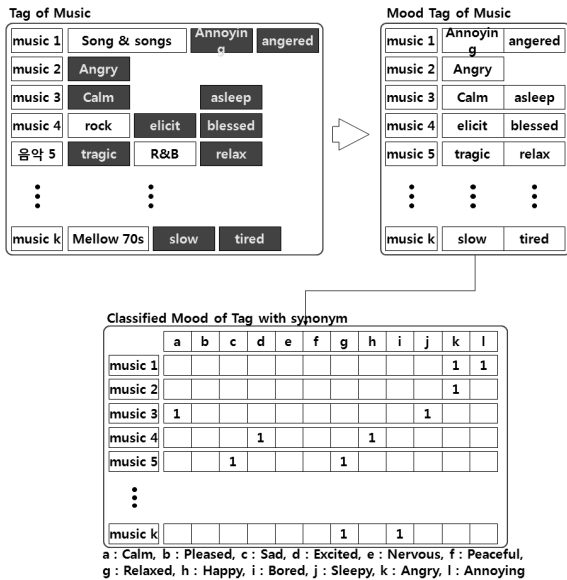


Figure 6. Classified Music with synonym

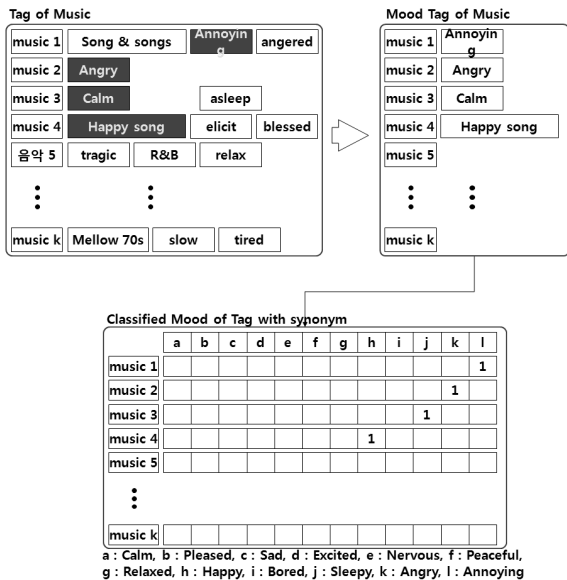


Figure 7. Classified Music without synonym

5. 분석

본 논문에서는 4장에서 분류한 것과 같이 두 가지 분류 방법에 대하여 음악의 AV값이 태그의 분위기를

기준으로 서로 다른 분포를 가지는지를 파악하기 위해 ANOVA Test를 실시하였다.

5.1. 동의어를 포함하지 않은 음악의 AV 분포 분석

동의어를 포함하지 않은 음악의 AV 분포를 분석하기 위하여 본 논문에서는 AV값을 각도로 변경하였고, 이 각도를 Thayer의 모델에 맞게 다시 12개의 분면으로 나누어 분석 하였다.

Table 1. AV Distributions without Synonym for Intro (a : pleased, b : happy, c : excited, d : annoying, e : angry, f : nervous, g : sad, h : bored, i : sleepy, j : calm, k : peaceful, l : relaxed, A : angle, B : mood)

B	A	0-30	30-60	60-90	90-120	120-150	150-180	180-210	210-240	240-270	270-300	300-330	330-360
a		0	33	0	33	0	0	0	0	33	0	0	
b		7	20	13	7	0	7	0	0	0	13	27	7
c		4	8	12	27	0	0	0	8	0	8	15	19
d		0	15	20	20	0	0	0	0	10	20	10	5
e		6	7	6	5	2	1	1	5	10	27	18	11
f		0	5	0	7	2	5	0	0	11	32	27	11
g		19	22	15	7	0	0	0	1	5	11	9	11
h		3	20	16	27	1	0	0	0	6	18	7	3
i		13	3	23	13	3	0	3	7	3	13	10	7
j		9	7	4	3	3	3	1	4	11	30	19	7
k		5	0	3	10	0	3	0	8	15	28	26	3
l		9	6	3	3	0	0	3	3	6	35	21	12

전반구간에서 pleased의 경우 30~60도, 90~120도, 270~330도에서 각 33%의 분포를 보이고, happy의 경우 300~330도에서 33%, excited의 경우 90~120도에서 27%, annoying의 경우 60~90도, 90~120도, 270~300도에서 20%, angry의 경우 270~300도에서 27%, nervous의 경우 270~300도에서 32%, sad의 경우 30~60도에서 22%, bored의 경우 90~120도에서 27%, sleepy의 경우 60~90도에서 27%, calm의 경우 270~330도에서 30%, peaceful의 경우 270~330도에서 28%, relaxed의 경우 270~330도에서 35%, 로 가장 많은 분포를 보인다.

Table 2. AV Distributions without Synonym for Representative (a : pleased, b : happy, c : excited, d : annoying, e : angry, f : nervous, g : sad, h : bored, i : sleepy, j : calm, k : peaceful, l : relaxed, A : angle, B : mood)

A \ B	0-30	30-60	60-90	90-120	120-150	150-180	180-210	210-240	240-270	270-300	300-330	330-360
a	0	33	33	0	0	0	0	0	0	33	0	0
b	0	20	27	40	0	0	0	0	7	7	0	0
c	8	4	19	46	4	0	4	0	0	12	4	0
d	0	20	20	45	10	0	5	0	0	0	0	0
e	10	9	12	14	6	3	1	4	11	13	12	5
f	5	5	7	7	7	0	0	5	11	25	23	7
g	6	24	32	12	2	0	0	3	3	4	8	7
h	7	13	37	40	0	0	0	0	0	1	2	0
i	17	20	23	17	3	7	3	0	3	0	0	7
j	8	9	4	9	2	3	0	9	9	20	23	5
k	13	8	8	5	8	5	0	5	13	18	13	5
l	12	21	12	0	0	0	0	0	6	18	15	18

Table 3. AV Distributions without Synonym for Outro (a : pleased, b : happy, c : excited, d : annoying, e : angry, f : nervous, g : sad, h : bored, i : sleepy, j : calm, k : peaceful, l : relaxed, A : angle, B : mood)

A \ B	0-30	30-60	60-90	90-120	120-150	150-180	180-210	210-240	240-270	270-300	300-330	330-360
a	33	0	33	0	0	0	0	0	0	0	33	0
b	7	13	20	40	0	0	0	7	0	13	0	0
c	8	8	19	23	12	4	0	0	4	12	4	8
d	5	5	35	25	5	0	0	0	0	5	15	5
e	6	5	10	9	5	1	3	2	13	26	11	9
f	0	7	5	2	7	0	2	7	16	32	14	9
g	13	15	26	9	3	1	1	1	6	6	10	8
h	1	6	46	29	4	1	1	1	7	1	2	1
i	13	20	23	10	13	7	0	0	0	0	10	3
j	3	9	5	5	5	1	1	2	15	28	17	10
k	10	5	5	5	5	0	3	0	13	28	23	3
l	3	9	3	9	3	0	0	0	15	18	38	3

대표구간에서 pleased의 경우 30~60도, 60~90도, 270~300도에서 각 33%의 분포를 보이고, happy의 경우 90~120도에서 40%, excited의 경우 90~120도에서 46%, annoying의 경우 90~120도에서 45%, angry의 경우 90~120도에서 14%, nervous의 경우 270~300도에서 25%, sad의 경우 60~90도에서 32%, bored의 경우 90~120도에서 40%, sleepy의 경우 60~90도에서 23%, calm의 경우 300~330도에서 23%, peaceful의 경우 270~300도에서 18%, relaxed의 경우 330~360도에서 18%로 가장 많은 분포를 보인다.

후반구간에서 pleased의 경우 0~30도, 60~90도,

300~330도에서 각 33%의 분포를 보이고, happy의 경우 90~120도에서 40%, excited의 경우 90~120도에서 23%, annoying의 경우 60~90도에서 35%, angry의 경우 270~300도에서 26%, nervous의 경우 270~300도에서 32%, sad의 경우 60~90도에서 26%, bored의 경우 60~90도에서 46%, sleepy의 경우 60~90도에서 23%, calm의 경우 270~300도에서 28%, peaceful의 경우 270~300도에서 28%, relaxed의 경우 300~330도에서 38%로 가장 많은 분포를 보인다.

표 1~3에서 보는 바와 같이 구간에 따라 AV값의 분포가 달라짐을 알 수 있었다. 이는 전반구간, 대표구간, 후반구간의 분위기가 다르기 때문에 나타나는 현상으로 보인다.

5.2. 동의어를 포함한 음악의 AV 분포 분석

동의어를 포함한 음악의 AV 분포 분석 역시 5.1과 동일하게 12개의 분면으로 나누어 분석 하였고, 그 결과는 표 4~6과 같다.

Table 4. AV Distributions with Synonym for Intro (a : pleased, b : happy, c : excited, d : annoying, e : angry, f : nervous, g : sad, h : bored, i : sleepy, j : calm, k : peaceful, l : relaxed, A : angle, B : mood)

A \ B	0-30	30-60	60-90	90-120	120-150	150-180	180-210	210-240	240-270	270-300	300-330	330-360
a	6	16	20	10	1	1	0	2	5	15	14	9
b	13	22	14	8	1	0	0	1	3	16	11	10
c	9	11	15	12	3	2	1	2	5	17	15	8
d	9	16	20	12	4	2	1	1	6	14	11	4
e	7	13	17	21	4	1	2	1	5	14	10	6
f	8	10	13	10	3	1	1	0	7	26	13	7
g	8	12	10	8	3	2	1	2	7	23	15	9
h	11	15	10	8	2	1	1	4	4	17	19	9
i	5	1	7	11	0	1	1	7	15	20	22	9
j	9	11	9	6	3	1	1	3	8	24	18	9
k	3	4	4	19	3	3	0	0	7	28	18	10
l	8	11	8	6	3	1	1	3	8	25	18	9

전반구간에서 calm, nervous, relaxed, sad, excited, peaceful의 경우 270~300도에서 각 24%, 26%, 25%, 23%, 17%, 28%로 가장 많은 분포를 보이고, bored와 sleepy의 경우 300~330도에서 각 19%, 22%, pleased와 annoying의 경우 60~90도에서 20%, angry의 경우 90~120도에서 21%, happy의 경우 30~60도에서 22%로 가장 많은 분포를 보인다.

Table 5. AV Distributions with Synonym for Representative (a : pleased, b : happy, c : excited, d : annoying, e : angry, f : nervous, g : sad, h : bored, i : sleepy, j : calm, k : peaceful, l : relaxed, A : angle, B : mood)

A \ B	0-30	30-60	60-90	90-120	120-150	150-180	180-210	210-240	240-270	270-300	300-330	330-360
a	7	24	22	19	2	0	0	2	4	8	5	6
b	7	20	29	16	3	0	0	4	4	5	6	6
c	6	15	22	24	5	1	2	2	4	9	7	3
d	8	14	28	25	5	1	1	1	2	5	4	5
e	5	11	29	34	5	1	2	1	2	3	4	4
f	3	13	21	29	6	2	1	3	1	10	6	4
g	8	11	19	19	4	1	1	3	7	11	10	7
h	7	18	19	17	5	1	2	1	6	13	6	6
i	11	4	4	12	3	4	0	7	15	16	16	8
j	9	14	16	14	4	2	0	3	6	13	14	7
k	3	7	12	19	6	0	1	3	7	16	16	9
l	10	13	15	10	3	2	1	3	7	13	14	10

대표구간에서 calm, angry, happy, annoying, relaxed, bored의 경우 60~90도에서 각 16%, 29%, 29%, 28%, 15%, 19%로 가장 많은 분포를 보이고, nervous, excited, peaceful의 경우 90~120도에서 각 29%, 24%, 19%, pleased의 경우 30~60도에서 24%, sad의 경우 60~90도, 90~120도에서 19%, sleepy의 경우 270~300도, 300~330도에서 16%로 가장 많은 분포를 보인다.

후반구간에서 pleased, happy, annoying, nervous, excited의 경우 60~90도에서 22%, 25%, 23%, 26%, 21%로 가장 많은 분포를 보이고, calm, relaxed, sad, peaceful의 경우 270~300도에서 19%, 19%, 17%, 26%, angry의 경우 90~120도에서 28%, bored의 경우 30~60도, 60~90도, 16%, sleepy의 경우 270~300도, 300~330도, 23%로 가장 많은 분포를 보인다.

Table 6. AV Distributions with Synonym for Outro (a : pleased, b : happy, c : excited, d : annoying, e : angry, f : nervous, g : sad, h : bored, i : sleepy, j : calm, k : peaceful, l : relaxed, A : angle, B : mood)

A \ B	0-30	30-60	60-90	90-120	120-150	150-180	180-210	210-240	240-270	270-300	300-330	330-360
a	8	14	22	16	5	2	2	2	4	12	7	7
b	9	16	25	12	3	1	1	2	6	9	9	8
c	7	10	21	18	5	0	3	1	8	14	5	7
d	7	12	23	20	5	2	3	0	3	7	9	9
e	7	6	27	28	6	1	3	1	6	4	7	7
f	6	13	26	19	4	1	2	1	7	12	7	2
g	6	10	15	14	4	1	2	2	10	17	11	8
h	7	16	16	12	4	1	2	1	11	14	9	5
i	4	5	4	11	3	0	3	1	16	23	23	7
j	7	11	13	10	4	1	1	2	11	19	13	8
k	1	9	7	13	4	3	1	4	10	26	12	7
l	6	10	13	7	4	1	3	1	11	19	16	8

표 4~6에서도 구간에 따라 분포가 달라지기는 하지만 표 1~3에 비해 그 정도가 적다. 이렇게 동의어

포함유무에 따라 분포의 차이가 발생하는 이유는 한 분위기 단어가 여러 기본 분위기 단어 (Thayer의 12개 분위기 단어)와 동의어 관계에 있기 때문으로 보인다. 예를 들어, mad는 12 분위기단어 중 angry와 excited의 동의어로 mad가 포함된 음악은 angry의 분포와 excited의 분포에 모두 영향을 줄 수 있다.

5.3. ANOVA Test

본 논문에서는 5.1과 5.2의 분류에 의해 구축한 음악의 12 분위기 그룹에 대하여 서로 독립적인 분포를 가지는지를 확인하기 위해 ANOVA Test를 실시하였고, 각 구간의 A값과 V값에 대하여 실험 하였다. 본 논문의 독립변수는 폭소노미 태그를 기준으로 분류한 음악의 12개 분위기(happy, sad, annoying, pleased, excited, nervous, bored, sleep, calm, peaceful, relaxed, angry) 그룹으로 설정하였고, 종속변수는 각 음악 구간(전반, 대표, 후반)의 A값과 V값으로 설정하였다. 본 논문의 실험 결과는 표 7과 같고, 모두 p-value가 0.000으로 귀무가설 H_0 을 기각할 수 있으므로 대립가설 H_1 을 채택할 수 있다. 즉, 동의어 포함 유무에 무관하게 음악의 AV값을 12 분위기 그룹으로 분류할 경우 AV값 분포의 차이가 발생한다고 결론을 내릴 수 있다.

6. 논의 및 결론

폭소노미는 사용되는 태그에 대한 세 가지 문제점이 있는데, 그 첫 번째는 유사어에 대한 문제점이고, 두 번째는 태깅레벨에 대한 문제점이며 세 번째는 신조어에 대한 문제점 이다. 이러한 문제점 중 유사어에 대한 문제점은 단어 태그 대신에 수치 태그를 사용하면 일부 해결할 수 있다. 단어 태그 대신에 수치 태그를 이용하여 음악을 검색하기 위해서는 두 개의 매핑 테이블이 필요한데 그 첫 번째가 음악과 AV값과의 관계를 정의하는 음악-분위기 매핑테이블이고, 두 번째가 태그와 AV 값과의 관계를 정의하는 태그-분위기 매핑테이블이다. 태그-분위기 매핑테이블 구축방법과 이의 분석에 대해서는 기존 연구인 문(2013)에서 선행 연구가 되어졌고, 본 논문에서는 음악-분위기 매핑 테이블을 구축하는 방법을 제시하고, 음악의 분위기와 음악의 AV값 사이의 연관성을 파악하였다. 또한, 음악-분위기 매핑테이블 분석 시 분위기 태그의 동의어를 고려하기 위해 인터넷 동의어 사전을 이용하였다.

Table 7. Result of ANOVA Test

		Source	DF	SeqSS	Adj SS	Adj MS	F	P
without Synonym	Intro V	mood	11	1.647	1.647	0.150	3.980	0.000
		Error	1225	46.035	46.035	0.038		
		Total	1236	47.682				
	Intro A	mood	11	25.031	25.031	2.276	20.060	0.000
		Error	1225	138.933	138.933	0.113		
		Total	1236	163.964				
	Representative V	mood	11	2.655	2.655	0.241	6.820	0.000
		Error	1225	43.376	43.376	0.035		
		Total	1236	46.032				
	Representative A	mood	11	37.040	37.040	3.367	34.330	0.000
		Error	1225	120.162	120.162	0.098		
		Total	1236	157.201				
	Outro V	mood	11	1.735	1.735	0.158	4.430	0.000
		Error	1225	43.576	43.576	0.036		
		Total	1236	45.311				
	Outro A	mood	11	38.148	38.148	3.468	28.650	0.000
		Error	1225	148.294	148.294	0.121		
		Total	1236	186.442				
with Synonym	Intro V	mood	11	2.190	2.190	0.199	5.210	0.000
		Error	3030	115.673	115.673	0.038		
		Total	3041	117.863				
	Intro A	mood	11	25.875	25.875	2.352	17.210	0.000
		Error	3030	414.156	414.156	0.137		
		Total	3041	440.031				
	Representative V	mood	11	2.724	2.724	0.248	6.680	0.000
		Error	3030	112.401	112.401	0.037		
		Total	3041	115.125				
	Representative A	mood	11	37.608	37.608	3.419	29.080	0.000
		Error	3030	356.204	356.204	0.118		
		Total	3041	393.812				
	Outro V	mood	11	1.751	1.751	0.159	4.360	0.000
		Error	3030	110.572	110.572	0.036		
		Total	3041	112.323				
	Outro A	mood	11	36.228	36.228	3.293	22.600	0.000
		Error	3030	441.633	441.633	0.146		
		Total	3041	477.860				

음악의 각 구간에 대한 AV값을 정의하기 위해 문(2013)에서 구축한 AV 예측 모델을 이용하여 음악의 전반과 후반의 AV값을 예측하였고, 예측한 음악 각 구간의 AV값 분포를 파악하기 위해 두 가지 방법으로 실험 대상을 분류 하였다. 그 첫 번째가 태그의 동의어를 고려하여 음악의 AV값을 12 분위기 그룹으로 분류하는 것이고, 두 번째가 동의어를 고려하지 않고 음악의 AV값을 12 분위기 그룹으로 분류하는 것이다.

음악의 각 구간(전반, 대표, 후반)에 대한 AV값의 분포를 분석한 결과, 음악의 분위기에 따라 각 구간의 분위기 분포가 다르고, 특히 동의어를 포함하지 않은 경우 AV모델에서 분위기 분포 차가 두드러졌다. 또한 ANOVA Test를 실시한 결과, 동의어 포함 유무에 무

관하게 음악 각 구간의 A값과 V값 모두 분포차가 발생하고, 모두 제 1종 오류확률 P<0.001를 만족하였다.

향후, 본 논문에서 분석한 데이터를 폭소노미 분위기 태그를 이용한 검색 기초자료로 사용할 계획이다. 하지만 실제 시스템 환경에 적용하기 위해서는 보다 다양한 분석이 필요하다. 또한 본 논문에서 제안한 방법을 실제 환경에 적용하여 실제 사용자가 느끼는 반응에 대해서도 분석할 필요가 있다.

REFERENCES

Chang, C.-C. and Lin, C.-J. (2001). LIBSVM: a library

- for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- Farnsworth, P. R. (1958). The Social Psychology of Music. *The Dryden Press*.
- Feng, Yazhong. Zhuang, Yueting. Pan, Yunhe. (2003). Popular Music Retrieval by Detecting Mood. *Proc. of ACM SIGIR 2003*.
- Hevner, K. (1935). Expression in music: a discussion of experimental studies and theories. *Psychological Review*, 42, 186-204.
- Hevner, K. (1936). Experimental studies of the elements of expression in music. *The American Journal of Psychology*, Vol. 48, No. 2, 246 - 68.
- Katavose, H., Ihal, H., and Inokuchi, S. (1988). Sentiment extraction in music. *Proc. of Int. Conf. Pattern Recognition*, 2, 1083-1087.
- Kim, J. H., Lee, S., Kim, S. M., Yoo, W. Y. (2011). Music mood classification model based on Arousal-Valence values. In: *Proc. ICACT*. pp. 292 - 295.
- Lartillot, O. and Toivainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, pp. 237-244, Bordeaux, France, September 10-15.
- Laurier, C., Meyers, O., Serra, J., Blech, M., Herrera, P. (2009). Music Mood Representation from Social Tags. *Proceedings of the 10th International Society for Music Information Conference*, Kobe, Japan.
- Lee, J. I., Yeo, D.-G., Kim, B. M., Lee, H.-Y. (2009). Automatic Music Mood Detection through Musical Structure Analysis. *International Conference on Computer Science and its Application CSA 2009*, 510-515.
- Levy, M. Sandler, M. and Casey, M. (2006). Extraction of High-Level Musical Structure From Audio Data and Its Application to Thumbnail Generation. *Proc. of ICASSP'06, Vol. 5, Toulouse, France*, 13-16.
- Li, T., Ogihara, M. (2003). Detecting emotion in Music. *Proc. of ISMIR 2003*.
- Liu, D. Zhang, N. Y. & Zhu, H. C. (2003). Form and mood recognition of Johann Strauss's waltz centos. *Chinese Journal of Electronics*, 3.
- Moon, C. B., Kim, H. S., Song, M. K. and Kim, B. M. (2011). Mood and Color Distribution Characteristics of Music Genres(음악 장르에 따른 분위기와 색상 분포의 특성), *Proc. of the Science of Emotion & Sensibility*, Vol.14, No.1, pp. 59-72.
- Moon, C. B., Kim, H. S., Jang, Y.-W. and Kim, B. M. (2013). Analysis of Association between Mood of Music and Folksonomy Tag(음악의 분위기와 폭소노미 태그의 관계 분석), *Proc. of the Science of Emotion & Sensibility*, Vol.16, No.1, pp. 53-64.
- Ness, S. R. Theocharis, A., Tzanetakis, G. and Martins, L. G. (2009). Improving Automatic Music Tag Annotation Using Stacked Generalization Of Probabilistic SVM Outputs. *Proc. of ACM MM'09*, pp.705-708.
- Russell, J. A. (1980). circumplex model of affect. *Journal of Personality and Social Psychology*, No. 39, 1161.
- Ryu, S.-J., Lee, H.-Y., Cho, I.-W. and Lee, H.-K. (2008). Document Forgery Detection with SVM Classifier and Image Quality Measure. *Lecture Notes in Computer Science*, vol.5353, pp.486-495.
- Scheirer, E. D. (2000). Music-listening Systems. *Ph. D. Thesis, MIT Media Lab*.
- Thayer, R. E. (1989). The Biopsychology of Mood and Arousal, New York. *Oxford University Press*.
- Yang, Y. H., Su, Y. F. Lin, Y. C. and Chen, H. H. (2007). Music emotion recognition: the role of individuality. *Proc. of ACM SIGMM Int. Workshop on Human-centered Multimedia 2007*, pp. 13-21, Augsburg, Germany.
- Yang, Y.-H. Liu, C.-C and Chen, H.-H. (2006). Music emotion classification: a fuzzy approach. *Proc. of ACM MM'06, pp. 81-84*, Santa Barbara, CA, USA.
- Yang, Y. H., Liu, C. C. and Chen, H. H. (2008). A regression approach to music emotion recognition. *Audio, Speech, and Language Processing, IEEE Transactions on, Vol. 16*, 448-457.

원고접수: 2013.07.19

수정접수: 2013.09.13

게재확정: 2013.09.13