

## 음악의 분위기와 폭소노미 태그의 관계 분석

Analysis of Association between Mood of Music and Folksonomy Tag

문창배\* · 김현수\* · 장영완\* · 김병만\*\*

Chang Bae Moon\* · HyunSoo Kim\* · Young-Wan Jang\* · Byeong Man Kim\*\*

\*금오공과대학교 소프트웨어공학과

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

### Abstract

Folksonomies have potential problems caused by synonyms, tagging level, neologisms and so forth when retrieving music by tags. These problems can be tackled by introducing the mood intensity (Arousal and Valence value) of music as its internal tag. That is, if moods of music pieces and their mood tags are all represented internally by numeric values, A (Arousal) value and V (Valence) value, and they are retrieved by these values, then music pieces having similar mood with the mood tag of a query can be retrieved based on the similarity of their AV values though their tags are not exactly matched with the query. As a prerequisite study, in this paper, we propose the mapping table defining the relation between AV values and folksonomy tags. For analysis of the association between AV values and tags, ANOVA tests are performed on the test data collected from the well known music retrieval site last.fm. The results show that the P values for A values and V values are 0.0, which means the null hypotheses could be rejected and the alternative hypotheses could be adopted. Consequently, it is verified that the distribution of AV values depends on folksonomy tags.

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

### 요약

폭소노미는 폭소노미에 사용되는 태그에 대하여 유사어, 태깅 레벨, 신조어등의 문제점들이 있다. 본 연구자들은 이러한 문제점들을 해결하기 위해 음악의 분위기 강도(Arousal과 Valence의 강도)를 음악의 내부 태그로 활용하는 방법을 사용하고자 한다. 즉, A(Arousal)값과 V(Valence)값을 이용하여 음악의 분위기를 수치적으로 표현하고, 분위기 태그도 AV값으로 대응시켜 검색하게 되면 태그가 일치하지 않더라도 유사한 AV 값을 갖는 음악이 검색되어 결과적으로 분위기가 유사한 음악들을 검색할 수 있게 된다. 본 논문에서는 이의 선행연구로 AV값과 폭소노미 태그와의 관계를 정의하는 매핑테이블을 제안하고, 태그와 AV값의 연관 관계를 분석하기 위해 유명한 음악 검색 사이트인 last.fm에서 수집한 테스트 데이터에 대해 ANOVA 검증을 하였다. 검증결과, A값과 V값에 모두에 대하여 제 1종 오류확률 P가 0.0으로 귀무가설을 기각하고 대립가설을 채택할 수 있었다. 결론적으로 폭소노미 태그에 따라 AV 값 분포가 다르다는 것을 검증 할 수 있었다.

**주제어** : 음악 분위기, 폭소노미, 태그, last.fm, AV 태그

\* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2010-0021097)

\*\* 교신저자 : 김병만(금오공과대학교 소프트웨어공학과 교수)

E-mail : bmkim@kumoh.ac.kr

TEL : 054-478-7544,

FAX : 054-478-7544

## 1. 서론

음악 분류 방법에는 텍소노미와 폭소노미가 존재한다. 텍소노미는 계층적인 구조로 사서나 운영자에 의하여 분리 체계를 전문적으로 관리하는 방식이지만 확장성이 결여된다는 문제점이 있다. 이를 보완하기 위해 폭소노미가 대두되고 있는데 폭소노미는 평형적인 구조로 참여자의 참여로 분리체계를 나누는 방법이라 할 수 있다.

폭소노미의 장점은 텍소노미의 문제점인 확장성을 보완하고, 배타적인 문제점들을 보완할 수 있다. 하지만 폭소노미는 사용되는 태그에 대한 세 가지 문제점이 있는데, 그 첫 번째는 유사어에 대한 문제점이고, 두 번째는 태깅레벨에 대한 문제점이며 세 번째는 신조어에 대한 문제점이다. 유사어에 대한 문제점은 평온한 음악과 잔잔한 음악과 같이 표현은 다르지만 의미적으로는 유사한 태그의 경우에 발생하는 문제점이다. 즉, 태그 “평온한”을 사용하여 음악을 검색하는 경우 태그 “잔잔한”은 검색하지 못한다. 태깅레벨에 대한 문제점은 평온한 음악과 참 평온한 음악과 같이 어원의 의미는 같지만 표현 방식이 다른 경우에 발생하는 문제점이다. 즉, 유사어에 대한 문점과 동일하게 태그 “평온한”을 사용하여 음악을 검색하는 경우 태그 “참 평온한”은 검색하지 못한다. 마지막으로 신조어에 대한 문제점은 “옛지있다”의 의미는 모서리, 날카로움, 뾰족함 등으로 해석할 수 있는데 이를 의역하면 개성있다는 의미로 문화의 변화에 의해 발생하는 문제점이다.

본 연구자들은 폭소노미의 문제점을 보완하기 위한 방법으로 음악의 분위기를 표현하는 Arousal과 Valence의 값을 음악의 내부 태그로 활용하는 방법을 사용하고자 한다. A(Arousal)값과 V(Valence)값을 이용하여 음악의 분위기를 수치적으로 표현하고, 분위기 태그도 AV값으로 대응시켜 검색하게 되면 검색 태그의 AV값과 음악의 AV값을 이용하여 분위기가 유사한 음악들을 검색할 수 있게 된다. 앞에서 기술한 폭소노미의 세 가지 문제점 모두 음악 분위기는 비슷하지만 표현이 달라 검색되지 않는다는 점에서 일치한다. 따라서 AV값을 내부 태그로 사용하게 되면 이러한 문제점들이 자연스럽게 해결된다. 이러한 방법을 적용하기 위해서는 먼저 태그와 AV값 사이의 연관성 파악이 선행되어야 한다. 이를 위해 본 논문에서는 AV값과 분위기 폭소노미 태그와의 관계를 정의하는

매핑테이블을 제안하고, 매핑테이블의 태그와 AV값의 연관 관계를 분석하였다.

본 연구의 구성은 다음과 같다. 2장에서는 관련연구에 대하여 살펴보고, 3장에서는 음악의 대표 구간을 추출한 후 추출한 음악의 대표 구간을 SVR(Support Vector Regression)로 학습하여 새로운 음악의 AV값을 예측하는 방법과 폭소노미 태그를 수집하는 방법에 대해서 설명한다. 4장에서는 본 논문에서 제안한 음악의 분위기와 폭소노미 태그의 관계를 정의하는 매핑 테이블 생성방법과 이를 분석하기 위한 태그 분류 방법을 설명한다. 5장에서는 3장에서 사용한 SVR의 AV값 예측성능을 분석하고, 본 논문에서 제안한 매핑 테이블을 분석 한다. 마지막으로 6장에서는 결론 및 향후 연구에 대해 살펴본다.

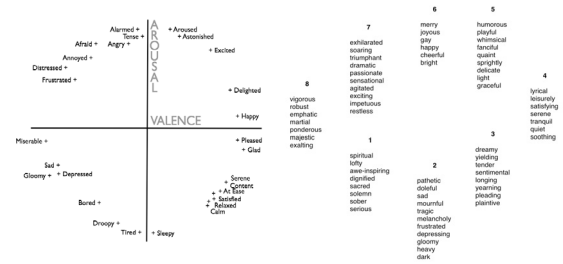
## 2. 관련연구

현재까지 제안된 음악 검색 형태를 보면 5가지 (텍스트 질의(query-by-text), 허밍 질의(query-by-humming), 부분 질의(query-by-part), 예시 질의(query-by-example), 클래스 질의(query-by-class)) 정도로 요약할 수 있을 것이다. 텍스트 질의 검색은 음악정보 데이터베이스에 저장한 서지 정보 (작가, 곡명, 장르 등)를 기반으로 기존의 정보검색 시스템의 질의 (query) 처리 방식대로 처리한다. 허밍 질의 방식은 사용자가 허밍(humming)을 입력하면 이것을 질의로서 인식하고 이와 유사한 멜로디를 갖는 곡들을 찾아주는 방식이고 부분 질의 방식은 사용자가 레스토랑에서 나오는 음악을 듣다가 곡이 좋아 이 곡이 현재 본인 단말기에 저장되어 있는지 알고 싶지만 곡명이나 멜로디를 모를 경우, 흘러나오는 음악을 입력으로 유사한 곡들을 찾아주는 방식이다. 예시 질의 방식은 사용자가 특정 곡을 선택하면 유사한 곡들을 찾아주는 방식으로 부분 질의 방식과 비슷하나 예시 질의에서는 곡 전체를 입력으로 하지만 부분 질의는 곡의 일부분만 입력으로 이용한다. 또한, 예시 질의에서는 실제 음악 대신에 곡명을 입력으로 하지만 부분 질의는 실제 음악을 입력으로 한다. 클래스 질의 방식은 사전에 음악을 장르나 분위기에 따라 분류를 해 놓고 텍소노미(taxonomy)에 따라 선택해 나가는 방식이다.

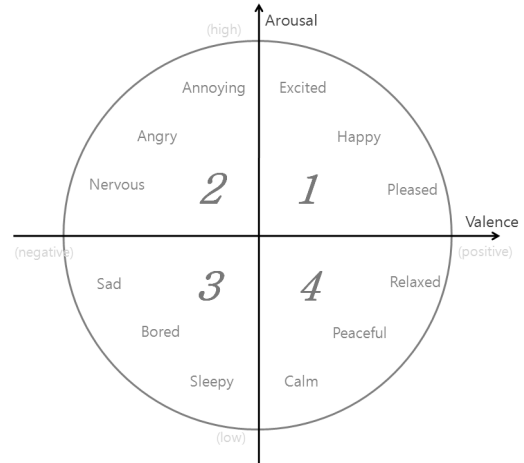
5가지 검색 방법 중 허밍 질의, 부분 질의, 예시 질의는 일반적인 검색 방법이 아니고 특수한 상황에서 사용 가능한 방법이다. 일반적인 방법은 텍스트 질의

나 클래스 질의 형태일 것이다. 하지만 이 두 방법 모두 전문가나 운영자의 개입을 요구한다. 즉, 새로운 음악이 나왔을 경우, 필요한 서지정보를 입력하거나 택소노미에 따라 어떤 분류에 속하는지를 결정해야 한다. 요즘처럼 새로운 음악이 계속 쏟아지는 상황에서는 이러한 방식은 더욱더 문제가 된다. 가능한 해결책 중의 하나는 택소노미에 따라 자동으로 태깅하는 방법을 사용하는 것이다. 자동으로 분류하여 그 분류에 해당하는 서지정보를 자동 입력하거나 분류코드를 할당하는 것이다. 하지만 택소노미에 따른 분류는 사서나 운영자등과 같이 사이트를 관리하는 일부 특정 계층이 직접 분류하는 방법이고, 특정 체계의 지식이 필요함으로 새로운 아이템이 추가될 경우 확장이 결여될 수 있다. 또한 정해진 하나의 카테고리에 하나의 의미를 부여하기 때문에 하나 이상의 주제를 가지고 있는 아이템에 대한 해결책이 필요하다. 이를 보완하기 위해 폭소노미가 대두되고 있는데 폭소노미는 전통적인 분류방식인 택소노미의 계층적인 방식과는 달리 평형적인 구조를 가지고, 사서나 운영자에 의하여 분리 체계를 전문적으로 관리하는 방식이 아닌 참여자의 참여로 분리체계를 나누는 방법이라 할 수 있다.

음악의 감성적 반응을 개념화하는 연구의 대표적인 사례로서 그림 1 (a)의 Russell(1980) 모델, 그림 1 (b)의 Hevner(1936) 모델 그리고 Thayer(1989) 모델이 있다. Russell 모델과 Hevner 모델은 형용사를 기반으로 한 모델로 의미가 중첩되거나 형용사적 표현상 모호한 단점이 있다. Russell 모델과 Hevner 모델의 모호한 단점을 보완한 것이 확장된 Thayer의 2차원 모델이고, 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. 그림 1 (c)는 확장된 Thayer의 2차원 분위기와 12개의 분위기/감정 형용사와의 관계를 나타낸 그림이다.



(a) Russell Model (b) Hevner Model



(c) two-Dimensional Model of Thayer

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 사용하였으며 템포, 세기, 피치변화, note density, 음색 (timbre) 등의 특징을 사용하였다. Katayose(1998)는 팝음악에 대해서 감정 (sentiment) 추출 시스템을 제안하였는데, 이 시스템에서는 단선율의 음향데이터가 먼저 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(Eric D. Scheirer, 2000). 이로 인해 이전부터 음향 데이터로부터 직접적으로 분위기를 탐지할 수 있는 시스템에 대한 필요성이 제기되었다.

Feng(2003)과 그 일행은 템포와 아티큘레이션 특징을 이용하여 분위기를 4개 - happiness, sadness, anger,

fear -의 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. 하지만 퍼지기반 시스템은 단선율의 음향에 대해서 좋은 성능을 보이지만 일반적인 복합음이 혼재한 음악의 경우 일반화된 성능을 나타내지 못하기 때문에 번역상의 오류가 분위기 탐지 성능에 개입될 수 있다.

Yang(2006)에서는 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 (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)), 이를 해결하기 위해 분위기 클래스를 사용하는 것이 아닌 Thayer의 2차원 분위기 모델의 각 축의 값을 직접 -1~1사이의 실수로 두어 사용하였다. AV 계수라 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. Yang(2007)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.

Lee(2009)의 연구에서는 Thayer의 2차원 분위기 모델과 구조 분석을 이용하여 음악의 분위기를 검출하는 방법에 대한 연구를 실시하였고, 문창배외 3(2011)의 연구에서도 역시 Thayer의 2차원 분위기 모델을 사용하여 음악 장르에 따른 분위기 분포를 분석하였다. 본 논문에서는 문창배외 3(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 존재하

는데 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, 이를 이용하여 새로운 음악의 AV 값을 예측한다.

Steven(2009)의 3의 연구에서는 기존 태그들을 이용하여 새로운 음악이 입력되었을 때 시스템에서 자동으로 태그를 부여하는 방법을 제안하였다. 이 방법은 음악을 세부 단위로 나누고, 나누어진 음악의 특징을 추출한다. 그리고 이 음악 특징을 SVM(Support Vector Machine)을 이용하여 학습시킨다. 새로운 음악이 입력되면 예측값을 이용하여 새로운 음악에 태그를 부여하는 방식이다. Steven(2009)의 연구에서는 작사, 작곡, 분위기 등의 태그를 사용하였지만 본 논문에서는 음악의 AV 값을 사용 하였다.

### 3. 음악의 대표 AV값 정의 및 폭소노미 태그 수집

본 논문에서는 음악분위기와 폭소노미 태그 사이의 연관관계를 분석하기 위한 데이터를 구축하기 위해 대표적인 음악 폭소노미 사이트인 last.fm을 활용하였다. 먼저 last.fm에서 제공하는 API를 이용하여 음악 분위기 태그를 갖는 음악 1,000여곡을 선별하였고, 이 음악들에 대해 AV값과 태그들을 수집하였다. 분위기 태그들은 last.fm 자체에서 사용자들에 의해 제공되지만 분위기를 나타내는 AV값은 last.fm에서 제공하지 않는다. 따라서 본 연구자들이 개별적으로 1,000여곡에 해당하는 음악들에 대하여 사용자로부터 AV값을 수집하여야 한다. 하지만 이러한 방법은 수집 비용이 너무 커 본 연구자들의 능력으로는 현실적이지 못하다. 그래서 본 논문에서는 차선택으로 문(2011)의 연구에서 구축한 AV 데이터를 이용하여 AV 값을 예측할 수 있는 모델을 구축하고, 이 모델을 이용하여 last.fm에서 선별한 1,000여곡의 AV값을 예측하는 방법을 사용하였다.

음악의 대표 AV값을 예측하는 과정은 학습 과정과 판별과정으로 나뉜다. 학습 과정에서는 학습 데이터를 이용하여 SVR(Support Vector Regression)을 이용하

여 AV 예측 모델을 학습하는 과정이고, 판별 과정은 SVR 예측 모델을 이용하여 새로 추가한 음악의 AV 값을 예측하는 과정이다 (그림 2 참고). 본 논문에서는 AVR 예측 모델을 구축하기 위한 학습 데이터로 기존 문(2011)의 데이터를 재사용 하였다.

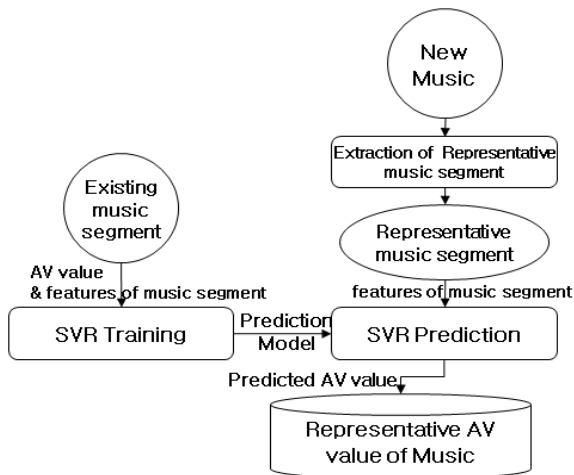


Figure 2. Process for AV value Prediction

### 3.1. 음악의 대표 구간 추출

분석대상 음악의 대표 구간을 추출하기 위해 본 논문에서는 문(2011)에서 사용한 음악의 구조 분석방법을 사용하였다. 대표구간 추출 방법은 음악 구조 정보를 추출하고, 분석된 구조정보를 이용하여 음악을 분리한 후 분리된 음원 중 음악의 에너지 값이 가장 큰 위치를 선택하여 대표구간을 추출하였다. 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 구조 정보를 파악한다.

유사구간을 획득한 후에는 유사구간의 시작부 부터 12초 단위로 음악을 분리시킨 후 에너지를 계산하여 에너지가 가장 큰 샘플을 대표 구간으로 설정 하였다. 에너지는 식 1에 의하여 계산된다.

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

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

켈레복소수를 의미한다.

### 3.2. SVR 학습 및 판별

SVR에 학습 및 판별하기 위해서는 음원의 특징을 추출하여야 하는데 본 논문에서는 Lartillot(2007)의 MIRToolbox를 사용하였다. MIRToolbox의 구조는 그림 3과 같은 구조로 최초 오디오 시그널 파형(Audio signal waveform)이 입력되면 영교차율(Zero-crossing rate), RMS 에너지(energy), 엔빌로프(Envelope), 스펙트럼(Spectrum) 그리고 필터 बैं크(Filter bank)들이 계산되고 최종적으로 각각의 mean, std, slope, period freq, period amp 그리고 period entropy 등과 같은 특징들 391개가 추출 된다.

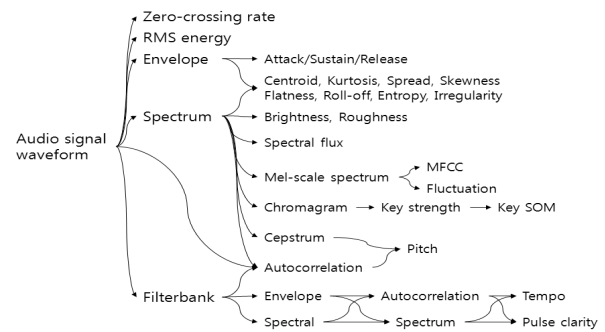


Figure 3. Extracted Musical Features by MIRToolbox(Lartillot(2007)의 문헌에서 발췌 p.238)

MIRtoolbox를 사용하여 특징을 추출하는 경우 NaN이 추출되기도 하는데 NaN은 수로 표현할 수 없는 값으로 본 논문에서는 NaN이 발생한 특징은 제거하였고, 또한 모든 음에 대한 특징 값이 동일한 경우에도 특징을 제거 하였다.

학습 및 판별에 사용한 SVR은 LIBSVM에서 제공하는 SVR을 사용하였다. LIBSVM에서 제공하는 SVR은 입실론(epsilon)-SVR과 nu-SVR을 이고, 커널 함수는 표 1과 같이 선형(Linear), 비선형(Polynomial), RBF(Radial basis function), 시그모이드(Sigmoid)의 네 가지이다(Ryu(2008); Chang(2001); Lee(2009)). 여기서  $\gamma$ ,  $r$ ,  $d$ 는 각 커널에서의 파라미터를 의미한다. 본 논문에서는 모든 커널을 사용하였고, 커널의 파라미터는 LIBSVM[11]에서 초기 값으로 제공하는 옵션을 사용 하였다.

Table 1. SVR Kernel Function

| Kernel Function       | Formula   |
|-----------------------|---|
| Linear                | $K(x_i, x_j) = x_i^T x_j$   |
| Polynomial            | $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$              |
| Radial basis function | $K(x, x_i) = \exp\left(-\frac{1}{2\sigma^2} \ x - x_i\ ^2\right)$ |
| Sigmoid               | $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$                       |

본 논문에서 사용한 SVR과 학습 및 판별에 사용한 파라미터는 실험에 의해 가장 좋은 성능을 가진 파라미터를 선택하였고, 파라미터 선택 관련 실험은 5장에서 자세히 다룬다.

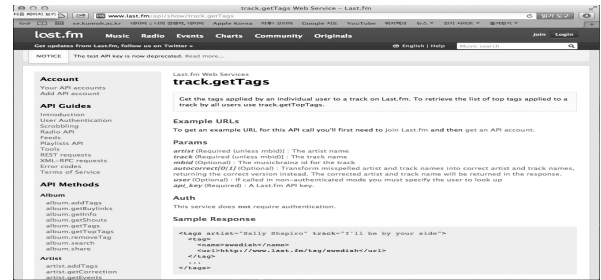
### 3.3. 폭소노미 태그 수집

본 논문에서 사용한 폭소노미 태그는 last.fm에서 제공하는 태그를 사용하였다. last.fm은 음악을 제공하는 대표적인 폭소노미 사이트로서 음악 청취자들이 음악을 청취 후 태그를 입력 하면 그림 4와 같이 해당 음악의 태그들을 확인할 수 있고, 태그를 선택하면 해당 태그를 포함하는 음악을 검색 할 수 있다.



Figure 4. Tag List of Music

last.fm에서 음악의 태그를 자동으로 수집하는 방법은 그림 5 (a)의 내용과 같이 artist와 track를 입력하면 (b)와 같이 해당 음악을 포함한 태그들을 리턴 받는다. 그림 5는 입력 값 artist="Sally Shapiro"과 track="I'll be by your side"의 결과이고, 해당 음악의 태그는 (b)의 .. 위치에 표기된다.



(a) Call of API

```
<tags artist="Sally Shapiro" track="I'll be by your side">
  <tag>
    <name>swedish</name>
    <url>http://www.last.fm/tag/swedish</url>
  </tag>
  ...
</tags>
```

(b) Result of API call

Figure 5. API for tag extraction from last.fm

본 논문에서 수집한 태그별 음악의 개수는 표 1과 같다. 표 2에서 모든 태그에 대해 음악의 수를 합하면 분석 음악의 수보다 많은데, 이는 한 음악에 여러 분위 태그가 달려있기 때문이다. 표 2는 이의 일부분을 요약한 것이다.

Table 2. Number of Music depending on Tag (NoM : Number of Music)

| tag        | NoM | tag       | NoM |
|------------|-----|-----------|-----|
| aggravated | 1   | calm down | 8   |
| agitated   | 2   | care      | 80  |
| agog       | 2   | chill out | 61  |
| amused     | 4   | comfort   | 30  |
| angered    | 1   | composed  | 19  |
| Angry      | 98  | console   | 1   |
| Annoying   | 37  | cool it   | 6   |

|          |     |             |     |
|----------|-----|-------------|-----|
| mellow   | 440 | sad         | 564 |
| mournful | 3   | Happy       | 337 |
| Nervous  | 33  | satisfy     | 8   |
| Peaceful | 67  | settle down | 6   |
| pensive  | 9   | simmer down | 3   |
| pitiful  | 16  | sleepy      | 43  |

### 4. 음악의 분위기와 폭소노미 태그의 관계정의

음악의 분위기와 폭소노미 태그의 관계를 정의하는 과정은 그림 6과 같이 3장에서 정의한 음악의 AV값과 last.fm에서 수집한 음악의 태그를 사용하여 맵핑 테이블을 생성한다. 생성한 맵핑 테이블을 분석하기 위해 본 논문에서는 확장된 Thayer의 2차원 분위기모델에 포함된 분위기 단어로 클러스터링 하고, 클러스터링 한 그룹을 기준으로 태그와 AV 관계를 분석 하였다. 확장된 Thayer의 2차원 분위기모델에 포함된 분위기 단어로 클러스터링하는 내용은 4.2에서 상세히 다룬다.

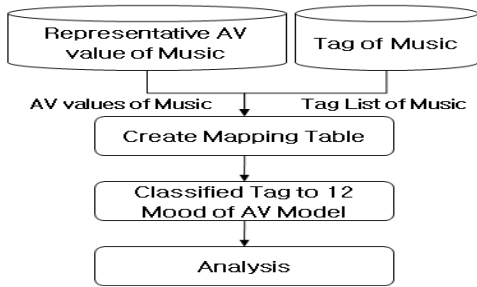


Figure 6. Defining Relation between Music Moods and Folksonomy Tags

맵핑 테이블 생성과정은 두 개의 과정으로 구성되는데, 첫 번째가 태그들의 중복 제거이고, 두 번째가 태그의 AV값을 정의하는 것이다.

#### 4.1. 맵핑 테이블 생성

음악의 AV값과 태그 리스트의 관계를 정리하면 그림 7의 좌측과 같이 각 음악에 대해 하나의 AV값이 존재하고, 하나의 음악에 k개의 태그가 존재한다. 태그의 맵핑 테이블을 생성하기 위해 전체 음악에 속한 모든 태그들의 중복을 제거하여 그림 7의 우측과 같이 중복이 없는 n개의 태그만을 유지한다.

태그의 중복을 제거한 후, 태그의 대표 AV값을 계산하기 위해 해당 태그를 포함하는 음악들의 AV값들을 구한 후 이들의 평균을 태그의 대표 AV값으로 설정하였다. 즉, 그림 7에서 태그1을 가진 음악은 1번 음악과 2번 음악이고, AV 값 역시 AV 값1, AV 값2이다. 또한 태그 2를 가진 음악은 1번 음악이고, AV

값 역시 AV 값1로 이를 태그를 중심으로 AV값을 재정리하면 그림 8의 좌측과 같이 나타낼 수 있다. 이 태그들에 포함된 A값의 평균과 V값의 평균을 구함으로써 그림 8의 우측과 같이 하나의 대표 AV 값을 구할 수 있다. 즉, 그림 8의 우측이 맵핑 테이블 이다.

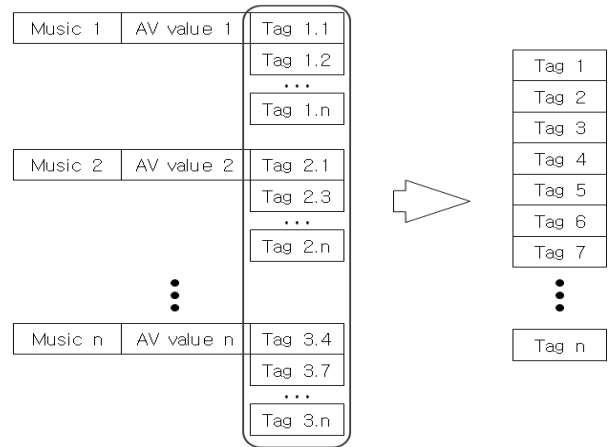


Figure 7. De-Duplication of Tags

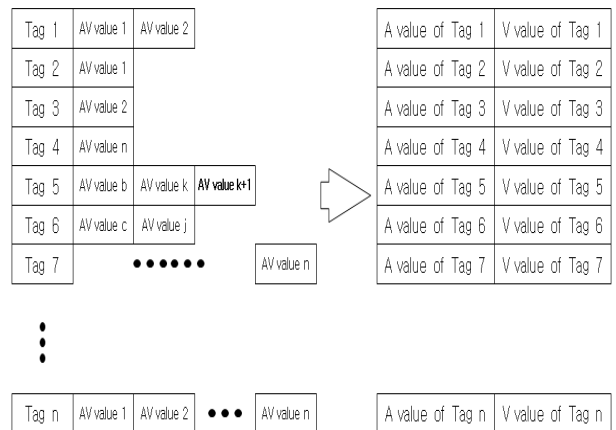


Figure 8. Defining the Representative AV value of a Tag

실제 음악의 분위기 대한 매핑 테이블의 결과는 표 3과 같다. 표에서 보는 바와 같이 본 논문에서는 424개의 분위기 태그에 대하여 AV 값을 예측하여(그림 1(c) 참고) 사용하였다.

Table 3. Result of Mapping Table

| NO  | tag                   | Representative V | Representative A |
|-----|-----------------------|------------------|------------------|
| ~   |                       |                  |                  |
| 10  | sad                   | 0.07             | -0.03            |
| 11  | Smooth Soul           | 0.20             | 0.04             |
| 12  | sadness               | -0.30            | -0.31            |
| ~   |                       |                  |                  |
| 45  | happy                 | 0.70             | -0.07            |
| 46  | get happy             | 0.33             | 0.32             |
| 47  | always makes me happy | 0.37             | 0.01             |
| ~   |                       |                  |                  |
| 419 | angry                 | 0.54             | -0.10            |
| 420 | angry hate music      | 0.58             | 0.15             |
| 421 | angry grrrr           | 0.20             | 0.23             |
| 424 | angry songs           | 0.78             | -0.01            |

#### 4.2. 12개의 분위기로 태그 분류

맵핑 테이블에서 정의한 태그와 대표 AV값의 관계를 분석하기 위해 본 논문에서는 확장된 Thayer의 2차원 분위기 모델의 12개 단어를 기준으로 분석하였고, 이를 위해 12개 단어를 기준으로 맵핑 테이블에 있는 424개의 태그를 분류(클러스터링)하였다. 즉, 표 2에서 “always makes me happy”와 같은 태그는 happy로 분류하고, “elton john - sad songs”와 같은 태그는 sad로 분류하는 방식으로 12개의 분위기에 대한 AV값을 맵핑 테이블에서 발췌하였다. 또한 일부 태그들의 경우 “happy sad”와 같이 두 개의 분위기를 포함하는 태그도 존재하는데 이러한 태그의 경우 happy와 sad 두 곳 모두 포함시켰다.

### 5. 분석

본 논문에서는 총 3가지를 분석하였는데 첫 번째가 SVR과 관련된 실험, 즉 AV값 예측 결과를 분석하였고, 두 번째가 12 분위기 태그의 AV값 분포를 분석하였고, 마지막으로 12 분위기 태그의 AV 분포에 대하여 ANOVA Test한 결과를 분석하였다.

#### 5.1. SVR 성능 분석

본 논문에서 학습 파라미터를 선택하기 위한 실험 방법은 10-fold 교차검증으로 280개의 음원을 10개의 그룹으로 나누고, 10그룹 중 9개의 그룹은 학습용으로 1개의 그룹은 판별용으로 사용하였다. 파라미터 설정의 경우 C를 1~100까지 5단위로 증가시키고, 입실론(epsilon)을 0.1~1.0까지 0.1단위로 증가시키면서 실험결과가 가장 좋은 성능을 가진 파라미터를 선택 하였다.

실험결과는 표 4와 같고, A값의 경우 입실론(epsilon)-SVR에서 RBF커널을 사용하였을 때 epsilon 0.3, C 11에서 결정계수(Squared Correlation Coefficient)의 값이 0.80으로 가장 좋은 성능을 보였고, V값의 경우 nu-SVR에서 비선형(polynomial)커널을 사용하였을 때 epsilon 0.2, C 96에서 결정계수의 값이 0.35로 가장 좋은 성능을 보였다. 표 4의 결과는 LIBSVM에서 제공하는 결과를 기반으로 작성 하였다.

Table 4. SVR performance (SCC : Squared Correlation Coefficient, E : epsilon, L : linear, P : polynomial, RBF : radial basis function)

| AV      | kernel     | epsilon-SVR |           |             | nu-SVR     |           |             |
|---------|------------|-------------|-----------|-------------|------------|-----------|-------------|
|         |            | E           | c         | SSC         | E          | c         | SCC         |
| A value | L          | 0.3         | 1         | 0.77        | 1          | 1         | 0.77        |
|         | P          | 0.1         | 56        | 0.79        | 0.1        | 66        | 0.80        |
|         | <b>RBF</b> | <b>0.3</b>  | <b>11</b> | <b>0.80</b> | 0.4        | 1         | 0.80        |
| V value | L          | 0.3         | 1         | 0.30        | 0.4        | 1         | 0.29        |
|         | <b>P</b>   | 0.2         | 41        | 0.34        | <b>0.2</b> | <b>96</b> | <b>0.35</b> |
|         | RBF        | 0.4         | 11        | 0.31        | 0.2        | 11        | 0.32        |

본 논문에서는 표 4에 의해 구한 파라미터, 커널 그리고 SVR을 사용하여 분위기 예측 모델을 구축하였고, 구축한 예측 모델을 이용하여 새로 추가한 음악의 AV값을 예측하여 해당 음악의 AV값으로 설정 하였다.

#### 5.2. 12 분위기 태그의 AV 분포 분석

본 논문에서는 음악 분위기와 태그와의 관계를 분석하기 위해 AV 모델에 포함된 12개의 단어가 포함된 태그들을 분석 하였다. 분석에 사용된 태그들은 표 5와 같다. 4.2에서 설명한 바와 같이 해당 태그에 AV 모델의 단어가 포함되면 그 태그를 AV모델의 단어로



포함 시켰다. 즉, 표 5에서 태그 “another sad love song”의 경우 태그에 “sad”가 포함되기 때문에 해당태그는 “sad”에 포함시키고, 태그 “excite me”의 경우 태그에 “excite”를 포함하기 때문에 해당 태그는 “excite”에 포함시켰다.

Table 5. Tag Groups for Analysis

| happy   | sad   | annoying   |
|---|---|--|
| always makes me happy, be happy dammit, becks happy songs, bright and happy, catchy happy 외 117 | a sad weightlessness, alkiviadis sadness, Angst and Sadness, <u>another sad love song</u> 외 186 | all things annoying in the world put together into one, annoying, not annoying 외 1 |
| pleased   | excite  | nervous  |
| pleased   | Shivering Excitement, excite, <u>excite me</u> , the word excite                                | nervous, nervousbreakdown  |
| bored   | sleep   | calm   |
| bored, bored to death, boredom, get bored 외 6   | makes you sleepy without being boring, sleepy, sleepy baby, sleepy music, sleepy song 외 8       | Beth Orton verde si calm, calm, calm and nice, calm country, calm down 외 27        |
| Peaceful  | relaxed   | angry  |
| Peaceful Sounds, Peacefulspot, calm-peaceful, peaceful  | relaxed, relaxed not bored, relaxed piano, relaxed rock, warm soft relaxed                      | 8 angry, a bit angry, aaaangry, angry, angry afternoon 외 32                        |

그림 9는 표 5에 있는 태그들의 AV값을 좌표로 표기한 그림이다. 좌측 그림의 x축은 A값을 의미하고, y축은 V값을 의미한다. 우측 그림은 AV값을 각도로 변환하여 그래프에 표기한 그림으로 x축은 태그를 의미하고, y축은 각도를 의미한다.

annoying의 경우 그림 9 (a)에서 보는 것과 같이 AV값은 우 상단에 주로 분포하고, 0~50도의 범위에 분포한다. angry의 경우 (b)에서 보는 것과 같이 AV값은 우 상단부터 우 하단에 주로 분포하고, -30~30도의 범위에 분포한다. nervous의 경우 (c)에서 보는 것과 같이 AV값은 우 하단 원점에 근접한 지점에 주로 분포하고, 300~360도의 범위에 분포한다. sad의 경우 (d)에서 보는 것과 같이 AV값은 원점을 기점으로 좌

상단과 우상단에 주로 분포하고, 각도의 범위는 50~200도의 범위에 분포한다. bored의 경우 (e)에서 보는 것과 같이 AV값은 우상단에 주로 분포하고, 각도의 범위는 -30~30도의 범위에 분포한다. sleepy의 경우 (f)에서 보는 것과 같이 AV값은 y축에 근접하여 좌상단과 우상단에 주로 분포하고, 각도의 범위는 100~200도의 범위에 분포한다. calm의 경우 (g)에서 보는 것과 같이 AV값은 원점을 기점으로 좌상단에 주로 분포하고, 각도의 범위는 150도에 근접하여 분포한다. peaceful의 경우 (h)에서 보는 것과 같이 AV값은 y축을 기준으로 좌상단과 우상단에 주로 분포하고, 각도의 범위는 100~150도의 범위에 분포한다. relaxed의 경우 peaceful과 유사한 분포를 가지지만 각도에 있어 peaceful은 150에 근접하지만 relaxed는 100도에 근접한다. pleased의 경우 (i)에서 보는 것과 같이 AV값은 우상단에 분포하고, 각도는 55도 이다. happy의 경우 (k)에서 보는 것과 같이 AV값은 우상단에 주로 분포하고, 각도의 범위는 0~50도의 범위에 분포한다. 마지막으로 excite는 (l)에서 보는 것과 같이 AV값은 x축을 기준으로 우상단과 우하단에 주로 분포하고, 각도의 범위는 -30~30도의 범위에 분포한다.

### 5.3. ANOVA Test

본 논문에서는 5.2의 분류에 의해 구축되어진 12 분위기 태그 분포에 대하여 서로 독립적인 분포를 가지는가를 확인하기 위해 ANOVA Test를 실시하였고, 각 A값과 V값에 대하여 실험 하였다. ANOVA Test의 독립변수는 12개 분위기 태그(happy, sad, annoying, pleased, excite, nervous, bored, sleep, calm, peaceful, relaxed, angry)로 설정하였고, 종속변수는 A값과 V값으로 설정하였다. 실험 결과는 표 6과 같고 모두 p-value가 0.000으로 귀무가설  $H_0$ 을 기각할 수 있고, 대립가설  $H_1$ 을 채택할 수 있기 때문에 A값과 V값 모두 12 분위기 태그에 대하여 분포차가 발생한다고 결론을 내릴 수 있다.

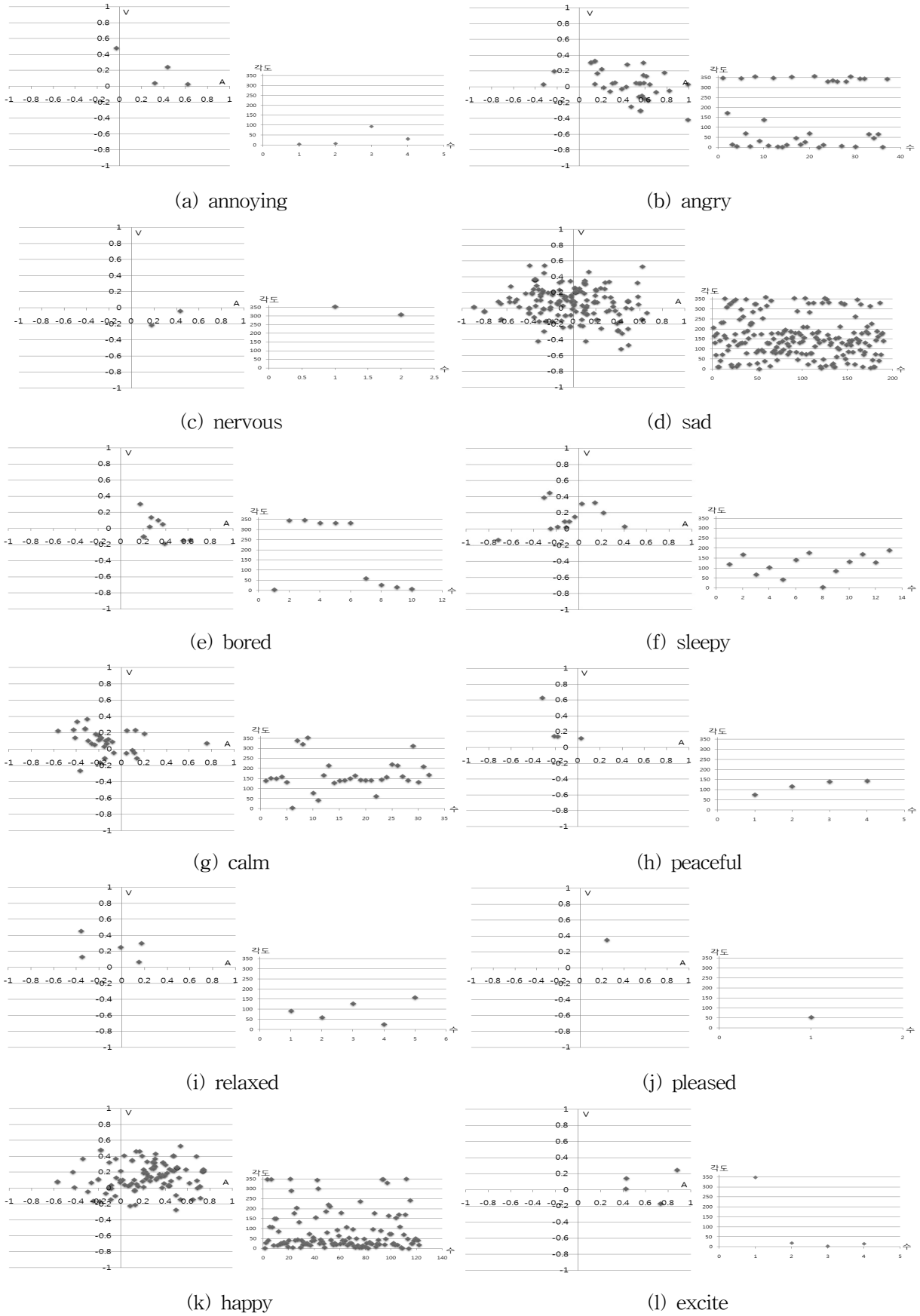


Figure 9. Distributions of AV values of Tags for 12 Mood Groups

Table 6. Result of ANOVA Test (A value and V value)

| AV     | A-value |       |       | V-value |       |       |
|--------|---------|-------|-------|---------|-------|-------|
|        | Source  | tag   | Error | Total   | tag   | Error |
| DF     | 11      | 412   | 423   | 11      | 412   | 423   |
| Seq SS | 16.11   | 38.67 | 54.78 | 1.21    | 13.92 | 15.13 |
| Adj SS | 16.11   | 38.67 |       | 1.21    | 13.92 |       |
| Adj MS | 1.47    | 0.09  |       | 0.11    | 0.03  |       |
| F      | 15.60   |       |       | 3.24    |       |       |
| P      | 0.00    |       |       | 0.00    |       |       |

## 6. 논의 및 결론

텍소노미에 따른 분류는 사서나 운영자등과 같이 사이트를 관리하는 일부 특정계층이 직접 분류하는 방법이다. 하지만 텍소노미의 경우 확장이 결여될 수 있다는 문제점이 발생한다. 이러한 문제점을 해결할 수 있는 방법이 폭소노미 이다. 폭소노미는 평형적인 구조를 가지고, 참여자의 참여로 분리체계를 나누는 방법이다.

폭소노미의 장점은 텍소노미의 문제점인 확장성을 보완하고, 배타적인 문제점들을 보완할 수 있지만 폭소노미는 사용되는 태그에 대한 세 가지 문제점이 있는데, 그 첫 번째는 유사어에 대한 문제점이고, 두 번째는 태그레벨에 대한 문제점이며 세 번째는 신조어에 대한 문제점 이다.

본 연구자들은 폭소노미의 문제점을 보완하기 위한 방법으로 음악의 분위기를 표현하는 Arousal과 Valence의 값을 음악의 내부 태그로 활용하는 방법을 사용하고자 한다. 이 방법을 적용하기 위해서는 먼저 태그와 AV값 사이의 연관성 파악이 선행되어야 한다. 이를 위해 본 논문에서는 AV값과 분위기 폭소노미 태그와의 관계를 정의하는 매핑테이블을 제안하고, 매핑 테이블의 태그와 AV값의 연관 관계를 분석하였다.

본 논문에서는 음악의 분위기와 폭소노미 태그와의 관계를 분석하기 위해 last.fm에서 분위기 태그가 포함된 음악과 태그를 수집 하였고, 수집한 음악의 AV값을 정의하기 위해 기존 문(2011)에서 수집한 분위기 데이터(280개의 음원과 AV값)를 SVR에 학습하여 판별모델을 구축하였다. 구축한 판별모델에 분위기 태그가 포함된 음악의 AV값을 예측하였고, 예측한 AV

값을 이 음악의 대표 AV값으로 정의하였다.

이렇게 구축된 데이터를 이용하여 본 논문에서는 폭소노미 태그와 음악의 AV값의 관계를 정의하는 맵핑 테이블을 구축하였고, 구축한 맵핑 테이블을 분석하기 위해, 본 논문에서는 발췌한 태그들을 12개의 분위기로 나누어 AV값의 분포를 분석하였다.

음악의 분위기와 태그와의 관계를 분석한 결과, 폭소노미 태그의 분위기에 따라 분위기 값이 다르게 분포되고, 특히 AV모델의 영역을 기준으로 분석한 경우 분위기 분포의 차이가 두드러졌다. 또한 폭소노미 태그의 분위기에 따라 AV 값 분포의 차이가 발생하는지를 확인하기 위해 ANOVA Test를 실시한 결과, A값과 V값의 분포차가 발생하고, 모두 제 1종 오류확률  $P < 0.001$ 를 만족하였다.

향후, 본 논문에서 분석한 데이터를 폭소노미 분위기 태그를 이용한 검색의 기초자료로 사용할 계획이다. 하지만 실제 시스템 환경에 적용하기 위해서는 유사어등을 포함한 보다 다양한 분석이 필요하다. 또한 각 태그간의 유의성을 검증할 필요가 있고, 본 논문에서 제안한 방법을 실제 환경에 적용하여 실제 청취자가 느끼는 반응에 대해서도 분석할 필요가 있다. 현재는 SVR 판별 성능 실험 시 280개의 음원을 이용하였기 때문에 음원 수의 부족으로 통계적 유의성은 검증하지 못하였다. 앞으로 이 부분에 대한 보완도 필요하다.

## 참고문헌

- 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). xperimental studies of the elements of expression im 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 Toiviainen, 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.
- 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: 2012.12.13

수정접수: 2013.01.21

게재확정: 2013.02.07