

Multiple Regression-Based Music Emotion Classification Technique

Dong-Hyun Lee[†] · Jung-Wook Park[†] · Yeong-Seok Seo^{**}

ABSTRACT

Many new technologies are studied with the arrival of the 4th industrial revolution. In particular, emotional intelligence is one of the popular issues. Researchers are focused on emotional analysis studies for music services, based on artificial intelligence and pattern recognition. However, they do not consider how we recommend proper music according to the specific emotion of the user. This is the practical issue for music-related IoT applications. Thus, in this paper, we propose a probability-based music emotion classification technique that makes it possible to classify music with high precision based on the range of emotion, when developing music related services. For user emotion recognition, one of the popular emotional model, Russell model, is referenced. For the features of music, the average amplitude, peak-average, the number of wavelength, average wavelength, and beats per minute were extracted. Multiple regressions were derived using regression analysis based on the collected data, and probability-based emotion classification was carried out. In our 2 different experiments, the emotion matching rate shows 70.94% and 86.21% by the proposed technique, and 66.83% and 76.85% by the survey participants. From the experiment, the proposed technique generates improved results for music classification.

Keywords : Emotion, Music Analysis, Music Classification, IoT, Emotional Intelligence

다중 회귀 기반의 음악 감성 분류 기법

이 동 현[†] · 박 정 욱[†] · 서 영 석^{**}

요 약

4차 산업혁명 시대가 도래하면서 기존 IoT에 감성지능이 포함된 신기술들이 연구되고 있다. 그 중 현재까지 다양하게 진행된 음악 서비스 제공을 위한 감성 분석 연구에서는 인공지능, 패턴인식 등을 활용한 사용자의 감성 인식 및 분류 등에만 초점을 맞추고 있는 상황이나, 사용자의 특정 감성에 해당하는 음악들을 어떻게 자동적으로 분류할지에 대한 감성별 음악 분류기법들에 대한 연구는 매우 부족한 상황이다. 본 연구에서는 최근 각광을 받고 있는 사람들의 감성과 관련된 음악관련 서비스를 개발할 시, 음악을 감성 범위에 따라 높은 정확도로 분류할 수 있도록 하는 감성 기반 자동 음악 분류기법을 제안한다. 데이터수집 시 Russell 모델을 바탕으로 설문조사를 하였으며, 음악의 특성으로 평균파장크기(Average amplitude), peak-평균(Peak-average), 파장 수(The number of wavelength), 평균파장 길이(Average wavelength), BPM(Beats per minute)을 추출하였다. 해당 데이터들을 바탕으로 회귀 분석을 이용하여 다중회귀식을 도출하였으며, 각 감성에 대한 표준 수치들을 도출하여 새로운 음악 데이터와 해당 각 감성에 대한 표준 수치들과의 거리 비교를 통해 음악의 감성을 분류시키는 작업을 실시하였다. 이를 통해 나온 결과에 회귀분석을 통하여 나온 데이터를 대입하여 해당 데이터와 각 감성들의 비율을 통해 최종적으로 판단된 감성을 추출하였다. 본 연구에서 실험한 감성 일치율의 2가지 방식에 대해서 제안한 기법의 경우 70.94%, 86.21%의 일치율이 나왔고, 설문참가자들의 경우 66.83%, 76.85%의 일치율이 나오므로써, 연구 기법을 통한 감성의 판단이 설문참가자들의 평균적인 판단보다 4.11%, 9.36%의 향상된 수치를 제공함을 알 수 있었다.

키워드 : 감성, 음악 분석, 음악 분류, IoT, 감성 지능

1. 서 론

과거 SF(Science Fiction) 영화에서나 등장했던 인공지능(AI)이 우리 현실 깊숙이 파고들고 있다. 인공지능은 다양한

정보통신기술(ICT)과 결합돼 전에 없던 새로운 제품과 서비스를 만들어내는 4차 산업혁명의 총아로 부상했다. 인공지능은 개인의 생활패턴이나 습관을 이해하고 각종 다양한 생활 및 환경으로부터 최적의 해법을 제시하는 등 인간의 편의성을 최대한 향상시키는 핵심 기술로서 인식되고 있기 때문에 전세계 곳곳에서 인공지능이 가지고 올 사회에 많은 관심을 가지고 있다.

인공지능 기술 연구 개발 필요성이 확산되고 심도있는 연구가 진행됨에 따라 영화에 나오는 로봇처럼 사람의 마음과 감성을 파악하는 ‘감성 인공지능’ 개발에 산·학·연 각 분야에서 많은 연구자들이 연구역량을 쏟아 붓고 있다. 감성 분석이

※ 이 성과는 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2017R1C1B5018295).

† 비 회 원 : 영남대학교 컴퓨터공학과 학사

** 종신회원 : 영남대학교 컴퓨터공학과 교수

Manuscript Received : January 12, 2018

Accepted : February 5, 2018

* Corresponding Author : Yeong-Seok Seo(ysseo@yu.ac.kr)

란 인간의 감성이 갖는 법칙을 파악하고자 하는 본원적인 호기심과 생체 데이터를 해석한다는 난제가 공존하는 연구 이슈이다. 실제로 감성 표현을 구현하기 위해 IT산업을 이끌어가는 주요 기업인 페이스북, 구글, 마이크로소프트, 애플 등도 감성을 추출하거나 학습하는 기술을 개발하거나 관련 기업을 인수하고 있는 실정이다[1, 2].

이러한 감성 분석을 위한 세부적인 연구 분야 중 하나가 바로 “음악” 분야이다. 방대한 양의 멀티미디어 콘텐츠가 보편화되면서 음악은 우리 일상에서 쉽게 접하고 들을 수 있게 되었지만, 개인별로 상황에 따른 감성이나 기분에 따라 선택하는 음악 유형은 매우 크게 달라질 수 있다. 따라서 이러한 측면에 대한 편의성을 도모하기 위해 인간의 감성을 분석한 후 적절한 분위기의 음악을 추천하는 기술들이 개발되고 있다. 최근 아마존 알렉사(Alexa), 애플 시리(Siri), 구글 어시스턴트(Google Assistant)가 탑재된 인공지능 스피커(AI Speaker)들이 공개 및 활용됨에 따라 사용자의 기분에 따른 적절한 음악을 선정하여 들려준다[3, 4]. 이렇게 사람의 감성을 분석하여 상황에 맞는 음악 추천 등의 음악 관련 서비스는 더욱더 확대될 전망이고, 마케팅, 교육, 엔터테인먼트, 게임, 헬스케어 등 다양한 산업군에서 고객들을 대상으로 한 감성 분석 비즈니스에도 활용할 수 있다[5].

현재까지 음악 서비스 제공을 위한 감성 분석 연구에서는 인공지능, 패턴인식 등을 활용한 사용자의 감성 인식 및 분류 등에만 초점을 맞추고 있는 상황이다. 그러나 사용자의 특정 감성에 해당하는 음악들을 어떻게 분류할지에 대한 감성별 음악 분류기법들에 대한 연구는 매우 부족한 상황이다. 따라서 본 논문에서는, 최근 각광을 받고 있는, 사람들의 감성과 관련된 음악관련 서비스를 개발할 시 감성에 따른 음악 분류기법을 제안하고자 한다. 제안한 기법을 통해 음악을 감성 범위에 따라 높은 정확도로 분류할 수 있도록 하는데 목적을 두고 있다.

제안한 기법은 크게 데이터수집, 통계적 분석, 감성 분류의 3단계로 이루어져 있다. 다양한 장르의 기존 음악들에 대해서 사람들에게 설문을 실시하여 사람들이 음악에서 느끼는 감성의 요소들(Emotional Features)과 음악의 특성들(Music Feature)을 체계적으로 수집 분석하였다. 수집한 각 요소들을 바탕으로 다중 회귀 분석을 통해 요소별 통계적 관계를 도출하였고, 이러한 데이터를 기반으로 감성 표준 수치를 도출하여 그 거리를 이용하여 음악을 분류할 수 있도록 하였다.

논문의 구성은 다음과 같다. 2장에서는 관련 연구를 소개한다. 3장에서는 본 논문에서 제안한 감성 기반 자동 음악 분류 기법에 대해 설명한다. 4장에서는 제안한 기법에 대한 실험 설계 및 실험 결과에 대해 분석하고 마지막 5장에서는 결론 및 향후과제를 기술한다.

2. 관련 연구

2.1 감성 모델

감성 분석을 위해 현재까지 감성을 수치화할 수 있는 방법에 대한 연구가 활발히 진행되어 왔다. 그러한 연구 결과 여

러 가지 감성 모델들이 제안되었다.

첫 번째, 인지과학이나 심리학분야에서 유명하고 자주 언급되는 모델인 Russell의 Circumflex 모델이 있다. 해당 모델은 2차원 좌표 평면 상에서 2가지 축을 긍정·부정(Valence)과 각성도(Arousal)라는 2가지 척도를 사용해 감성을 나타내게 된다(여기서 Valence는 감성의 긍정과 부정의 정도를 나타내고, Arousal은 감성의 강도를 나타낸다).

각성의 정도와 그 각성의 여부가 긍정인지 부정인지에 따라 감성의 축을 그린다. 감성을 각 축의 끝 부분에 원형으로 배열하거나 6가지 주요 감성을 뽑아 감성 지도를 그리기도 한다. Fig. 1은 Russell 모델을 나타낸다[6].

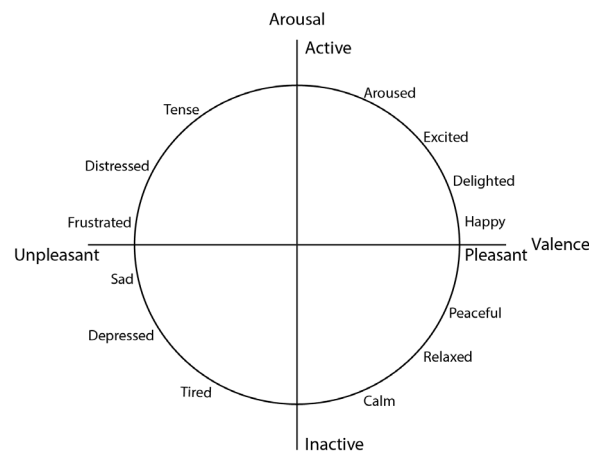


Fig. 1. Russell's Circumflex Model

두 번째, Thayer의 모델이 있다. 해당 모델은 Russell 모델을 음악에 적용시킨 모델로써 2차원 좌표 평면 상에서 Energy와 Stress 라는 두 가지 척도를 사용한다. Energy의 경우 음악에서 소리의 크기(Volume)나 정도(Intensity)를 의미하고, Stress의 경우 음악의 조성(Tonality)과 템포(Tempo)를 의미한다. Fig. 2는 Thayer 모델을 나타낸다[7, 8].

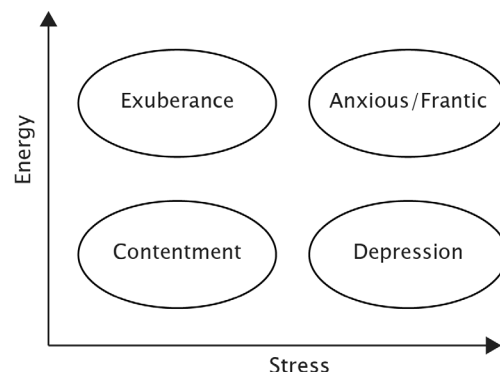


Fig. 2. Thayer model

세 번째, Tellegen-Watson Clark 모델이 있다. 해당 모델에선 긍정과 부정(positive/negative affect), 즐거움과 불쾌함(pleasantness/unpleasantness), 시끄러움과 조용함(engagement/

disengagement)을 이용해 이전 모델들보다 더 다양한 감성을 분류하고 있다. Fig. 3은 Tellegen-Watson Clark 모델을 나타낸다[9].

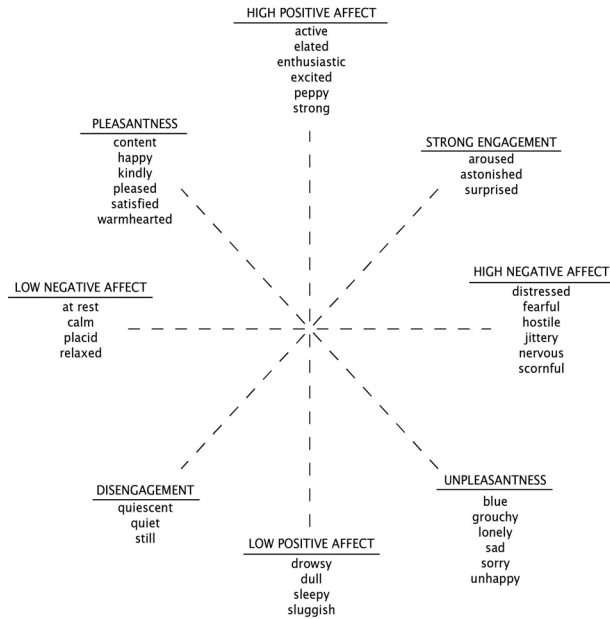


Fig. 3. Tellegen-Watson Clark Model

2.2 음악 분류

De La Salle 대학에서 실시한 음악 분류 연구[10]에서 가사를 바탕으로 연구가 진행되었다. 해당 연구에서 200곡의 필리핀 음악을 가지고 50명의 사람들에게 각각 20곡씩, 각 곡에 대해 5번씩 조사를 실시하였다. 설문조사(Manual annotation)와 자동적 기법(Automated way)을 사용하여 감성을 2가지, ‘행복한’과 ‘슬픈’으로 나눠 실험을 실시하였다. 먼저 설문조사를 실시할 때 참여자에게 음악적 성향, 빈도(Exposure)를 조사하였으며 준비된 각 곡에 대해 감성을 5단계로, ‘가장 슬픈(Saddest)’, ‘슬픈(Sad)’, ‘중립(Neutral)’, ‘행복한(Happy)’, ‘가장 행복한(Happiest)’으로 나눈 수치와 각 곡에 대한 친밀도를 조사하였다.

자동적 기법에선 단어의 목록에서 각 단어에 Arousal와 Valence 수치를 추가하였고, 이를 통해 두 감성, ‘행복한’과 ‘슬픈’에 대한 레이블(label)을 2가지 방법을 이용해 분류하였다. 첫 번째 방법은 클러스터를 이용하는 방법이고, 두 번째 방법은 거리 공식을 사용해 분류하였다. 그렇지만 이러한 방법을 통해 나타난 결과에서 한 가지 감성으로 편향된 결과가 나오게 되어 올바른 라벨을 표시하는데 충분하지 못하다는 결론이 나오게 되었다.

각 단어에 대해 TF-IDF(Term Frequency-Inverse Document Frequency)와 keyGraph, 그리고 그 둘과 ANEW를 합친 특징(Combination)을 통해 2가지 실험을 실시하였다. 여기서 TF-IDF란, 여러 문서로 이뤄진 문서군에서 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수

치를 뜻하며[11], keyGraph의 경우 주제와 관련된 여러 가지 키워드 중 서로 관련성이 있는 핵심 키워드를 연결지어 각각의 그래프 그룹을 형성시키는 기법이다[12]. 첫 번째 실험에선 단어들을 제거하지 않고 모든 단어에 대해 실시하였고, 두 번째 실험에선 단어를 두 목록으로 분류하고, 두 목록에서 충돌(conflict)되는 단어를 제거한 후 실시하였다. 첫 번째 실험의 결과, TF-IDF의 경우, 임계값으로 0.001에서 0.7까지 설정할 경우 64.7%의 정확도가 나오게 되었으며, keyGraph의 경우 75%의 정확도가 나오게 되었다. 두 번째 실험의 결과, TF-IDF의 경우 90.5%의 정확도가 나오게 되었지만 임계값의 범위를 줄여 0.1에서 0.7로 설정하게 되면 정확도가 67.1%로 낮아지게 되었다. keyGraph의 경우, 80%~89%의 정확도가 나오게 되어 TF-IDF의 경우보다 더 나은 정확도가 나온다는 사실을 통해, keyGraph 방법이 TF-IDF 방법이 직면하는 문제를 겪지 않는다는 걸 알 수 있다.

3. 접근법

본 논문에서는 음악에 대한 감성 분류를 수행하기 위해 음악들에 대한 감성 식별 설문조사와 해당 음악들로부터 감성 분류를 위한 데이터 추출을 실시하였다[13]. 이를 통해 수집된 데이터들을 기반으로 회귀분석을 진행하였고, 각 감성에 대한 표준 수치와 새로운 음악 데이터의 수치를 거리 비교하여 해당 음악이 가지는 각 감성의 비율을 도출하였다. Fig. 4는 본 논문에서 제안한 감성 기반 자동 음악 분류 기법의 전체 접근법을 표현한 것이다.

3.1 Step 1: 데이터 수집(Data Collection)

가장 첫 번째 단계로 사람들로부터 특정 음악들에 대한 감성 데이터를 측정하였고, 이와 함께 해당 음악들로부터 진폭, 진동수 등 특성 데이터들을 수집하였다.

음악의 경우 한 곡 내에서도 여러 가지 분위기가 포함되어 있는 경우가 있다. 또한, 데이터 추출 시, 음악의 전 구간을 기준으로 할 경우, 편차가 커지게 되므로 보다 의미 있는 데이터로 판단하기 위해 음악의 각 구간을 20초 단위로 나누어 데이터를 측정하였다.

1) Step 1-1: 사용자 감성 조사(Survey Emotions from Participants)

사람들은 음악을 들으며 기쁨, 신남, 화남, 슬픔 등 셀 수 없이 많은 감성들을 느끼고 있으며, 이러한 감성들은 같은 음악에 대해서도 서로 다르게 느낄 수 있다. 그러한 음악에 대한 감성들은 객관적이지 못하고 매우 주관적이기에 이를 객관적으로 측정하기란 매우 어려운 일이다. 따라서 본 논문에서는 음악에 대한 감성 설문조사를 실시할 때 기존 Russell의 감성 모델[6]을 기반으로 조사 대상자의 감성을 긍정-부정(Valence)과 각성도(Arousal)라는 2가지 척도로 조사하였다.

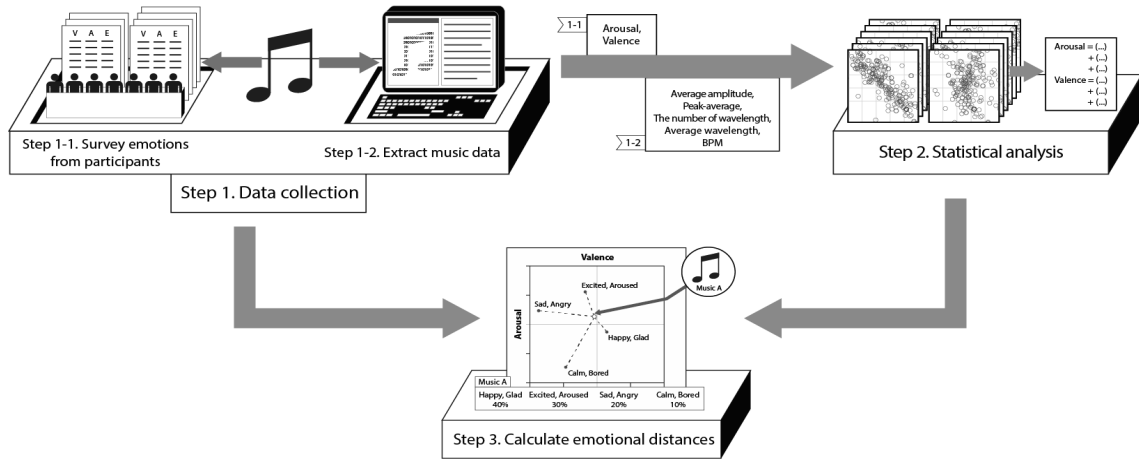


Fig. 4. Overall Approach

2) Step 1-2: 음악 데이터 추출(Extract Music Data)

소리는 크기, 높이, 음색 등으로 구성되어 있으며 이는 물리적으로 분석해 보았을 때 진폭, 진동수, 파형 등과의 관계로 볼 수가 있다. 따라서 본 논문에서는 이와 같은 특성 데이터들을 음악 데이터로서 수집하였다.

음악의 특성 분석 시 음악을 20초 간격으로 분석을 수행하였고, 전체 음악에 대한 평균 파장 크기(Average amplitude), peak 평균(Peak-average), 파장 수(The number of wavelength), 평균 파장 길이(Average wavelength), BPM(Beats Per Minute)을 추출하였다. Fig. 5에서 도식화하여 보여주고 있듯이, 평균 파장크기는 파장 크기의 평균을 의미하고, peak평균은 파장에서 양의 값을 갖는 파장 중 자신을 기점으로 전과 후의 값들과 비교해 봤을 때 자신의 값이 더 큰 값(Peak)들의 평균을 의미한다. 그리고 파장 수의 경우 파장의 수치가 0에서 양의 값이 된 순간부터 다시 0이 되는 구간의 수를 의미하며, 파장의 길이는 그러한 구간의 폭을 의미한다.

3.2 Step 2: 통계적 분석(Perform Regression Analysis)

음악은 리듬, 음색, 빠르기 등 여러 가지 요소로 구성되어 있다. 사람의 경우 해당 요소들을 식별하여 음악의 분위기에 대해 판단하는 것은 어렵지 않은 일이다. 그러나 평균파장 크기, 파장 수, 평균파장 길이 등과 같은 음악의 특성들만 살펴보고 해당 음악의 감성을 알아내기란 상당히 어려운 작업이다.

Step 2에서는 음악의 특성들만으로도 해당 음악에 대해 사

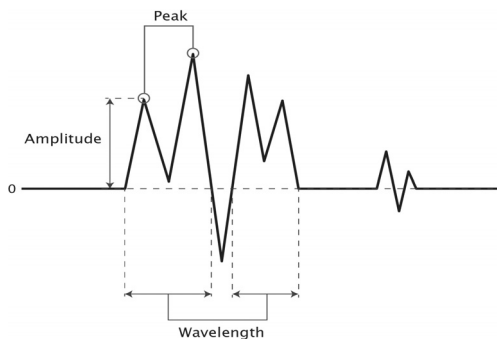


Fig. 5. Music Feature Explanation

람이 느끼는 감성이 어떤 것일지 유추해 낼 수 있도록 하기 위해 통계적 기법 중 하나인 회귀분석을 이용하였다. 즉, Step 1에서 얻은 Valence 및 Arousal 수치들과 음악 분석을 통해 수집한 데이터들 간의 연관관계를 분석하여 해당 음악에 대한 음악 데이터 값에 대해 사람이 느끼는 감성이 어떤 것인지 판단할 수 있도록 한다.

회귀 분석 시, 독립 변수로는 앞서 언급한 평균 파장 크기, peak평균, 파장 수, 평균 파장 길이, BPM을 사용하였다. 그리고 종속 변수로는 사람으로부터 수집한 Valence와 Arousal 데이터를 사용하였다. 여러 독립변수들 중 종속 변수에 적합한 독립 변수를 찾기 위해 단순 회귀 분석을 통한 각 요소간의 결정계수(Coefficient of determination)를 분석하여 상관관계가 적은 요소들은 배제하고 다중 회귀식을 도출하였다.

3.3 Step 3: 감성 식별 분류(Calculate Emotional Distances)

Step 3에서는 Step 2의 회귀분석을 통해 확보한 음악데이터별 Valence 및 Arousal 수치들이 실제 어떠한 감성과 연관되는지 분류하는 작업을 수행한다.

감성의 분류를 위해서는 기존 Russell의 감성 모델[6]에 기반하여 각 감성별로 얻어진 Valence, Arousal의 수치들의 평균을 도출한다. 각 감성의 평균 수치는 해당 감성의 Valence, Arousal 수치의 표준이 된다. Fig. 6은 음악의 Valence 및 Arousal 수치에 따라 ‘기쁜·달달한’(Happy, Glad), ‘신난·흥분된’(Excited, Aroused), ‘슬픈·화난’(Sad, Angry), ‘조용한·잔잔한’(Calm, Bored)의 4가지 감성에 대한 표준 수치를 나타낸 예시이다.



Fig. 6. The Classified Emotional Categories

이러한 연구 결과를 바탕으로 추후 발매된 신곡을 분류할 시, 우선 신곡 음악 데이터를 추출한 이후 회귀분석을 이용하여 도출된 감성 수치(Valence 및 Arousal)를 각 감성들의 표준 수치와 비교하여, 2차원 평면상에서의 각 감성에 대한 거리를 도출한다. 각 감성에 대해 도출된 거리가 짧을수록 해당 감성과 유사한 것으로 판단한다. 도출된 결과를 바탕으로 해당 수치들의 역을 이용해 각 감성에 대한 확률을 나타내어 각 감성의 정도를 분석한다. 이를 통해 음악에 대한 감성을 자동적으로 분류할 수 있도록 한다.

4. 실험 및 실험 결과 분석

4.1 실험 설계

감성 조사는 음악에 대한 사람들이 느끼는 감성을 알아보기 위해 100여곡의 음악에 대해 20대의 7명의 인원에게 주변 소음이 없는 동일한 공간에서 Table 1의 양식을 통해 설문조사를 실시하였으며, 이때 감성 조사에서 사용된 음악은 전 세계적으로 인기가 많은 Korean pop song을 활용하였다. 장르의 경우 댄스, 발라드, R&B, 힙합, 락 등 다양하게 선택되었다. 음질은 320kbps인 mp3 파일을 사용하였다.

해당 설문에서 사용된 설문지에는 Music Title, Time, Arousal, Valence, Emotion 항목을 두었다. Music Title은 곡명을 나타내고, Time은 곡을 20초로 나눈 구간을 나타낸다. Arousal와 Valence는 사람들이 음악에서 느끼는 Arousal와 Valence 수치를 기입하는 항목이다. 해당 수치의 범위는 최대 100에서 최소 -100까지로 하였다. Emotion은 각 구간에서 느끼는 감성을 선택하는 항목이다. Emotion 항목에는 4가지 항목, 'Happy, Glad', 'Excited, Aroused', 'Sad, Angry', 'Calm, Bored'를 선택할 수 있도록 구성하였다.

감성이란 것은 매우 주관적이고 Russell모델은 실험 참가자들에게 익숙하지 않기 때문에 실제 설문 전, 설문에 대한 상세

설명을 제공하여 설문참가자들이 보다 정확한 지식을 가지고 객관적인 실험을 수행할 수 있도록 하였다.

음악의 특성(Music Feature)에 대한 추출 및 분석 시스템은 Java 언어를 사용해서 직접 구현하였고, BPM의 경우 오픈소스인 Easy BPM Calculation Java[14]를 사용하였다. 이러한 자동화 도구들을 구현하고 활용하여 다음 Table 2와 같은 형식으로 음악 데이터들을 추출하였다.

감성 조사와 데이터 추출을 통해 얻어진 데이터들을 바탕으로 회귀 분석을 수행하기 전에 극단적인 값을 가지는 데이터(Outlier)를 제외하여 분석의 신뢰성을 높일 수 있도록 하였다. 그 후 통계 프로그램인 R[15]을 이용하여, 종속변수로 Arousal와 Valence를, 독립변수로 평균과장크기, peak평균, 과장 수, 평균과장길이, BPM을 사용하여 회귀 분석을 수행하였다. 그 중 가장 대표적으로 Fig. 7은 Valence-BPM의 단순 회귀분석 결과를 보여주고, Fig. 8는 Arousal-The number of wavelength의 단순회귀분석 결과를 보여준다.

위 데이터를 바탕으로 다중회귀분석 시 correlation분석을 통해 Valence와 Arousal의 유효한 독립변수를 확인하였다. 이를 통해 아래와 같이 Arousal와 Valence의 수식이 도출되었다. Equation (1)과 같이, Arousal에서는 독립변수로 평균과장크기, peak평균, 평균과장 길이, 과장 수가 유효한 변수로 확인되어 수식에 적용하였으며, 이 중 과장수의 회귀계수가 0.235205로 가장 크게 나타났다. 그리고 다음 Equation (2)와 같이, Valence에서는 독립변수로 BPM, 평균과장크기, peak 평균이 유효한 변수로 확인되어 수식에 적용하였으며, 이 중 BPM의 회귀계수가 2.673으로 가장 크게 나타났다.

$$\begin{aligned}
 Arousal = & (-66.413186) \\
 & +(0.001763*Average\ amplitude) \\
 & +(-0.008526*Peak\ -average) \\
 & +(-0.020090*Average\ wavelength) \\
 & +(0.235205*The\ number\ of\ wavelength)
 \end{aligned}
 \tag{1}$$

Table 1. Survey Example

Music Title	Time(s)	Arousal (-100~100)	Valence (-100~100)	Emotion			
				Happy, Glad	Excited, Aroused	Sad, Angry	Calm, Bored
A	~20	-37	-56				✓
	~40	-30	-53				✓
	~60	-10	-46				✓
	~80	-39	3				✓
	~100	-33	33			✓	
	~120	-30	52			✓	
B	~20	15	31	✓			
	~40	23	36	✓			
	~60	9	32	✓			
	~80	36	42		✓		
	~100	57	53		✓		
	~120	74	61		✓		

Table 2. Extracted Music Data Set

Music Title	Time(s)	Average amplitude	Peak-average	The number of wavelength	Average wavelength	BPM
A	~20	34245	4318	111	1260	83
	~40	42570	5078	125	1081	83
	~60	37599	4456	111	1237	83
	~80	29604	3464	105	1354	83
	~100	38488	4929	198	695	83
	~120	36033	4565	164	839	83
B	~20	83611	10735	277	444	107
	~40	81514	10510	317	392	107
	~60	71338	8894	193	654	107
	~80	68925	8453	182	694	107
	~100	84434	10804	306	402	107
	~120	88222	11446	349	351	107

$$\begin{aligned}
 \text{Valence} = & (-240.8) \\
 & + (2.673 * \text{BPM}) \\
 & + (0.0009863 * \text{Average amplitude}) \\
 & + (-0.002892 * \text{Peak-average})
 \end{aligned} \tag{2}$$

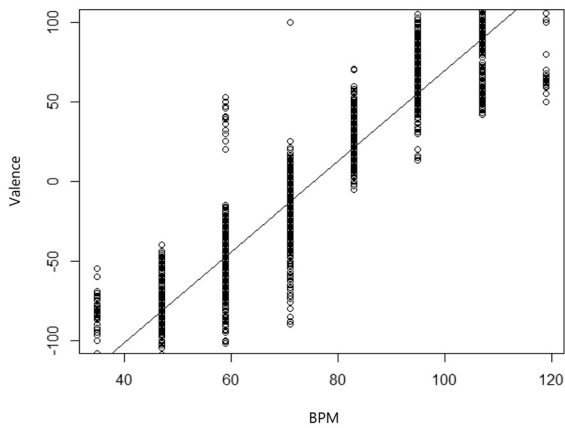


Fig. 7. Valence Simple Regression Analysis Graph

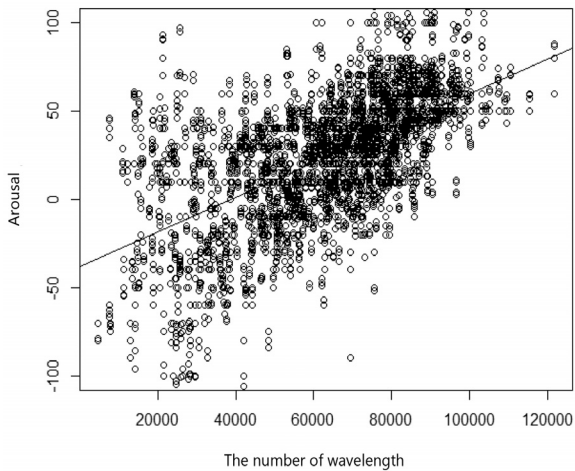


Fig. 8. Arousal Simple Regression Analysis Graph

사람들로부터 설문 조사를 통한 감성 데이터들을 통해 각 감성에 대한 표준 수치들을 도출한 결과 Table 3과 같은 결과가 도출되었다.

새롭게 발매되는 신곡들을 분류할 시 다중회귀분석을 통해 얻어낸 결과를 표준 수치들과 거리 비교하여 각 감성의 비율을 도출한 후 자동적으로 감성을 분류하도록 한다.

본 실험에 대한 유효성을 검증하기 위하여 테스트 셋으로서 7명의 실험참가자에게 28곡의 새로운 음악에 대해 설문조사를 실시한 후, 28곡의 새로운 음악에 대한 본 기법의 감성 분석 결과와 비교 분석하였다.

4.2 기법 실행 과정의 예

본 논문의 기법을 통하여 A곡을 감성 분류하면 다음의 절차로 수행된다.

먼저 3.1에서 제시된 방법에 따라 A라는 곡의 음악 데이터 추출을 한다. 그 결과 Average amplitude은 109361, Peak-average는 14101, The number of wavelength는 286, Average wavelength는 419, BPM은 95가 추출된다.

두 번째로, 3.2에서 제시된 방법에 따라 음악 데이터 추출을 통하여 얻어진 데이터들을 다중회귀 분석으로 도출된 수식(4.1에서 제시된 Arousal, Valence 수식)에 대입하면 Arousal는 65.01605, Valence는 80.21766으로 도출된다.

마지막으로, A곡의 감성을 식별하기 위해서, 3.3에서 제시된 방법에 따라 두 번째 과정에서 얻어진 A곡의 Arousal, Valence 수치와 4가지 감성의 표준 수치들과의 거리를 구한다. 그 결과 ‘기쁜, 달달한’과의 거리는 약 69.00, ‘신난, 흥분된’과의 거리는 약 35.00, ‘슬픈, 화난’과의 거리는 약 138.53, ‘조용한, 잔잔한’과의 거리는 약 149.70이다. 도출된 거리들에 역수를 취한 수치들을 기준으로 각 감성의 비율을 구할 경우, ‘기쁜, 달달한’은 약 25.44%, ‘신난, 흥분된’은 약 50.16%, ‘슬픈, 화난’은 약 12.67%, ‘조용한, 잔잔한’은 약 11.73%로 도출된다. 결과적으로, A곡은 약 50.16%의 가장 높은 비율을 차지한 ‘신난, 흥분된’ 감성으로 분류된다.

Table 3. Emotion Standard Value

	Happy, Glad	Excited, Aroused	Sad, Angry	Calm, Bored
Arousal	20.17013	54.65951	28.45534	34.8582
Valence	27.77813	46.78834	53.3976	31.2927

Table 4. The Emotion Match Rate (%)

	TPM	A	B	C	D	E	F	G
MAX	79.31	72.41	79.31	75.86	82.76	65.52	75.86	82.76
MIN	55.17	58.62	55.17	48.28	58.62	48.28	65.52	65.52
AVG	70.94	65.02	69.95	67.49	69.46	58.62	69.95	71.43

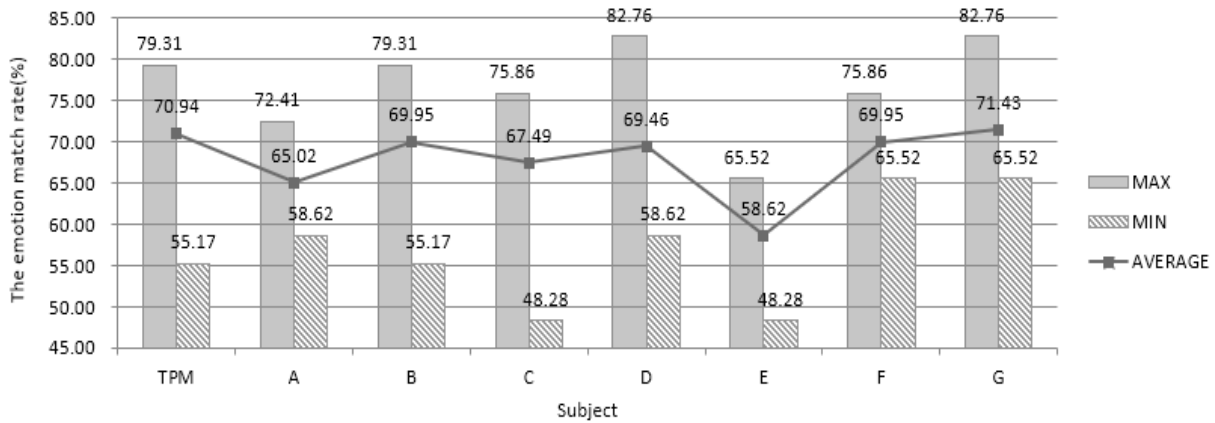


Fig. 9. The Emotion Match Rate

4.3 실험 결과 분석

4.1을 통해 얻어진 “감성 표준 수치”에 대해 새로운 28곡을 활용하여 유효성 검증을 수행하였다. 우선 7명의 실험참가자에게 28곡에 대한 감성 분류 조사 결과를 확보하고 본 기법을 통해 도출된 감성 범위와 비교 분석하였다. Table 4는 본 기법과 각 설문참가자의 데이터를 각각 비교하여 같은 감성 일 확률(%)의 최대(MAX), 최소(MIN), 평균(AVG) 수치를 나타낸다. TPM(The Proposed Method)은 본 연구에서 제안한 기법이며 A부터 G까지는 설문참가자이다. 각 칸의 수치는 본 연구에서 제안한 기법과 설문참가자가 판단한 감성에 대한 일치율을 나타낸 것이다. 기법과 설문 간의 감성의 일치율은 평균 70.94%, 최대 79.31%의 결과를 보여주었다. Fig. 9는 이를 나타낸 그래프이다.

또한, 설문참가자로부터 한 음악에 대해서 4가지 감성(‘기쁜·달달한’, ‘신난·흥분된’, ‘슬픈·화난’, ‘조용한·잔잔한’)의 비율을 도출하였을 때, 가장 많은 비율로 도출된 감성이 얼마나 많은 비율을 차지하는지에 따라 기법이 얼마나 정확하게 해당 감성을 도출하는지 분석하였다. Fig. 10은 이를 나타낸 그래프이다. 해당 그래프에서 가로축은 설문 참가자들의 감성 결과들 중 가장 많이 도출된 감성이 차지하는 비율을 나타내

며, 세로축은 감성이 해당 비율로 도출되었을 경우에 대한 기법의 정확도를 의미한다. 가령 설문참가자들에게서 각 감성의 비율이 도출되었을 때, 그 중 가장 많은 비율을 차지하는 감성의 비율이 60~70%인 경우에 대해서 기법은 약 80%의 정확도를 보여준다는 것을 의미한다.

보다 심도있는 분석을 위해 2가지 관점으로 분류하여 감성의 일치율을 비교해보았다. 첫 번째, 각 곡에 대해 “TPM에서 도출한 감성 결과”와 “설문참가자들로부터 조사한 감성 결과”가 모두 하나의 감성으로 일치하는 경우(ALL)를 Table 5와 같이 비교 분석해보았다. Table 5의 첫 번째 row는 TPM과 각 설문참가자들을 나타내고 두 번째 row의 각각의 수치는 각 곡에 대해 TPM 혹은 각 설문참가자로부터 수집한 감성 결과와 다른 설문참가자들로부터 수집한 감성 결과가 모두 하나의 감성으로 일치하는 일치율을 나타낸다. 즉, TPM이 70.94%는 TPM에서 도출한 감성 결과와 설문참가자들로부터 수집한 감성 결과가 하나의 감성으로 일치하는 비율이 70.94%라는 것을 의미한다. Table 5의 결과 본 연구에서 제안한 TPM의 수치가 70.94%로 가장 우수한 결과를 보여주었다. 설문참가자들의 수치들도 59.20%부터 70.69%의 분포를 가지고 있음을 확인할 수 있었다.

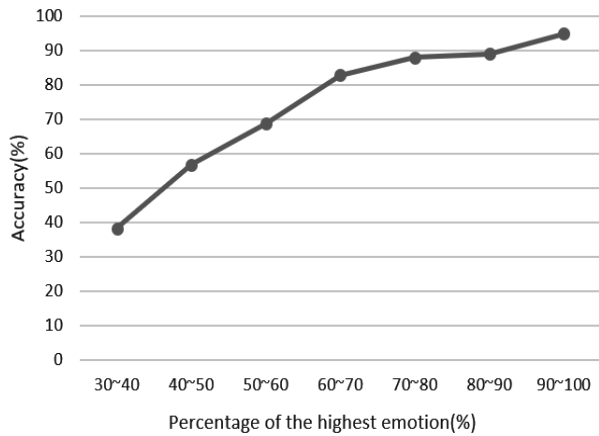


Fig. 10. The Accuracy of TPM for Emotion that Occupies the Most

Table 5. The Emotion Match Rate for the "ALL" Case

(%)	TPM	A	B	C	D	E	F	G
ALL	70.94	64.94	68.39	66.09	68.39	59.20	70.11	70.69

두 번째, 각 곡에 대해 “TPM에서 도출한 감성 결과”와 “설문참가자들로부터 조사한 감성 결과들 중 가장 많이 도출된(dominant) 감성 결과”가 서로 일치하는 경우(MOST)를 Table 6과 같이 비교 분석해보았다. Table 6의 내용은 Table 5와 같은 양식으로 구성되어 있다. Table 6에서 알 수 있듯이, TPM의 결과는 86.21%를 보여주었는데, 이는 “TPM에서 도출한 감성 결과”와 “설문참가자들로부터 수집한 감성 결과들 중 가장 많은 감성 결과”가 서로 일치하는 비율이 86.21%라는 것을 나타낸다. Table 6의 결과 본 연구에서 제안한 TPM의 수치가 86.21%로 가장 우수한 결과를 보여주었다. 이는 대부분의 설문참가자가 느끼는 감성을 TPM에서 판별해 줄 수 있다는 것을 의미한다.

Table 6. The Emotion Match Rate for the 'MOST' Case

(%)	TPM	A	B	C	D	E	F	G
MOST	86.21	75.86	86.21	79.31	79.31	58.62	79.31	79.31

Fig. 11은 Table 5, Table 6에서의 2가지 관점에 따른 실험 결과를 통합적으로 비교 분석해보았다.

Fig. 11은 2가지 관점별 전체 결과를 보여준다. Fig. 11에서 Survey participants는 설문참가자들 감성 일치율들의 평균을 나타내고, TPM은 본 연구에서 제안한 기법을 통해 도출된 감성 일치율을 나타낸다, 2가지 관점별로 도출된 결과를 각각 나타내었다.

2가지 관점에 대한 TPM의 감성 일치율이 86.21%, 70.94%로 도출되었고, 2가지 방식에 대한 설문참가자들의 평균적인 감성 일치율이 76.85%, 66.83%로 도출되었다. 따라서 TPM과

설문참가자들의 결과를 비교해 보았을 때, TPM의 결과가 2가지 관점 모두 설문참가자들의 평균보다 우수함을 알 수 있었다. 그 중에서 MOST 관점에서의 일치율이 86.21%로서 설문참가자 평균인 76.85%에 비해 9.36% 이상 높았다. 이를 통해 본 연구에서 제안한 기법이 특정 음악의 가장 대중적인 감성을 분류하는 기준으로써 충분히 활용될 수 있음을 확인할 수 있었다.

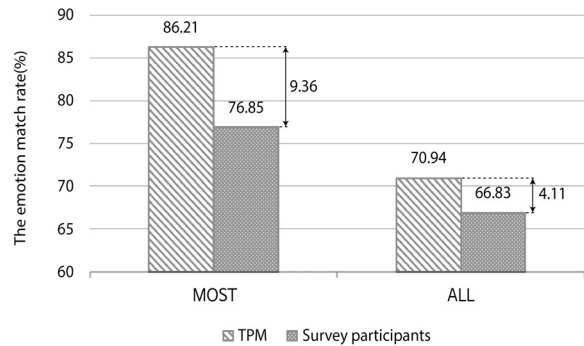


Fig. 11. The Comparison of the Emotion Match Rates

본 실험에서 한 가지 유의해야 하는 사항은 음악의 음질에 따라 감성이 달라질 수 있다는 것이다. 이를 위해 Cool Edit Pro[16]를 사용하여 음질에 따른 파형을 분석하였다. Fig. 12는 특정 음악(M)이 고음질일 경우의 파형이고 Fig. 13은 동일한 음악이 저음질일 경우의 파형을 나타낸다. 본 연구에서 제시한 기법을 적용해 본 결과 M의 고음질 버전인 경우 ‘신난, 흥분’ 결과를 보여주었고 저음질 버전인 경우 감성 분류 결과 ‘슬픈, 화난’이 도출되었다. 실험참가자의 경우 대부분 ‘신난, 흥분’으로 판단하였기 때문에 저음질 버전의 경우에는 유효한 결과가 나오지 않을 수 있다고 판단할 수 있다. 본 연구에서는 기본적으로 320kbps인 고음질 버전의 음악 파일들을 이용하여 실험을 수행하였다. 저음질 버전의 음악 파일들의 경우 고음질의 파형과 비교했을 때 상당히 다른 파형을 보여줄 수 있기 때문에 실제 실험 시 유의해야 할 이슈이다.

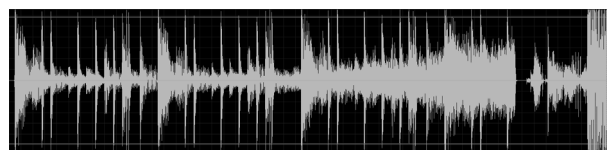


Fig. 12. The Waveform of the High Quality Version of the M Music

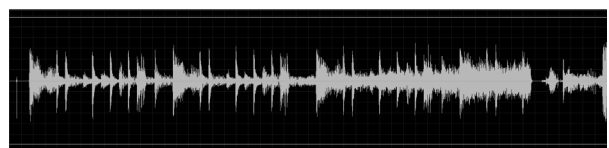


Fig. 13. The Waveform of the Low Quality Version of the M Music

5. 결 론

본 논문에서는 다양한 종류의 음악관련 감성지능 IoT 서비스를 개발할 시, 음악을 감성 범위에 따라 높은 정확도로 분류할 수 있도록 하는 감성 기반 자동 음악 분류기법을 제안하였다. 감성을 수치화하기 위해 잘 알려진 Russell모형을 활용하였고, 다양한 장르의 음악에 대해 실험참가자들의 설문문을 통해 음악별 감성 결과를 확보하였으며, Java기반의 도구를 개발하여 음악 데이터를 추출하였다. 추출한 많은 데이터들을 기반으로 다중회귀분석을 수행하고 최종적으로 각 감성에 대한 표준 수치들과의 거리 비교를 통해 해당 음악의 감성 비율을 도출하였다. 이를 통해 다양한 장르의 음악들을 자동적으로 분류할 수 있도록 하였다.

본 연구에서 제안한 기법을 통해 도출한 감성 분석 결과를 설문참가자들을 통해 얻어진 결과와 2가지 방식으로 비교분석해 보았을 때, 본 연구에서 제안한 기법의 경우 70.94%, 86.21%의 감성 일치율을 나타내었고, 설문참가자들의 경우 66.83%, 76.85%의 감성 일치율을 보여줌으로써, 본 연구에서 제안한 기법을 통한 감성의 판단이 설문참가자들의 평균적인 판단보다 우수한 결과를 보여주었다.

본 연구를 통해 상당히 고무적인 연구결과가 있었지만 향후 다음과 같은 몇 가지 이슈에 대해 추가적인 연구를 진행하고자 한다. 본 연구에서는 음악에 대한 감성의 영역으로 ‘기쁜·달달한’, ‘신난·흥분된’, ‘슬픈·화난’, ‘조용한·잔잔한’의 4가지 감성의 영역을 정의하였지만 보다 더 세밀한 영역으로 분류해서 연구를 진행할 예정이다. 그리고 음악의 음질에 따라 감성 분석이 달라질 수 있기 때문에 음악의 과장 외에도 가사 등을 활용하는 연구들을 수행해 나가고자 한다.

References

[1] Rolfe Winkler, Biography, Daisuke Wakabayashi, and Elizabeth Dwoskin, Apple Buys Artificial-Intelligence Startup Emotient [Internet], <https://www.wsj.com/articles/apple-buys-artificial-intelligence-startup-emotient-1452188715>.

[2] Hannah Jane Parkinson, Happy? Sad? Forget age, Microsoft can now guess your emotions [Internet], <https://www.theguardian.com/technology/2015/nov/11/microsoft-guess-your-emotions-facial-recognition-software>.

[3] Matt Clinch, Streaming has now taken over the music industry and A.I. will push it into the stratosphere [Internet], <https://www.cnn.com/2017/04/04/google-home-streaming-alexa-apple-music-spotify.html>.

[4] Yu Jeong Song, Su Yeon Kang, Sun Young Ihm, and Young Ho Park, "Multimedia Processing: An Implementation of a Classification and Recommendation Method for a Music Player Using Customized Emotion," *KIPS Transactions on Software and Data Engineering*, Vol.4, No.4, pp.195-200, 2015.

[5] ReportLinker, How Music Streaming Services Are Using AI to Revolutionize the Listening Experience [Internet], <https://www.reportlinker.com/insight/music-streaming-services-using-ai-revolutionize-listening-experience.html>.

[6] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, Vol.39, No.6, pp.1161-1178, 1980.

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

[8] Donglim Kim, Lim Bin, and Younghwan Lim, "Music Emotion Control Algorithm based on Sound Emotion Tree," *Journal of the Korea Contents Association*, Vol.15, No.3, pp.21-31, 2015.

[9] Auke Tellegen, David Watson, and Lee Anna Clark, "On the dimensional and hierarchical structure of affect," *Psychological Science*, Vol.10, No.4, pp.297-303, 1999.

[10] Emil Ian V. Ascalon and Rafael Cabredo, "Lyric-Based Music Mood Recognition," in *Proceedings of the DLSU Research Congress*, 2015, pp.1-8.

[11] Wikipedia, tf-idf [Internet], <https://en.wikipedia.org/wiki/Tf%E2%80%93idf>.

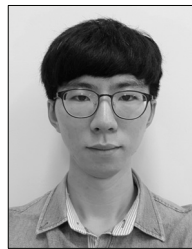
[12] H. Sayyadi and L. Raschid, "A Graph Analytical Approach for Topic Detection," *ACM Transactions on Internet Technology*, Vol.13, No.2, pp.1-23, 2013.

[13] D. Patel and K. Trivedi, "Research of Music classification based on mood recognition," *International Education and Research Journal*, Vol.3, No.5, pp.544-555, 2017.

[14] Peter Backx, Easy BPM Calculation in Java [Internet], <http://www.streamhead.com/easy-bpm-calculation-in-java/>.

[15] The R Project for Statistical Computing [Internet], <https://www.r-project.org/>.

[16] Cool Edit Pro [Internet], <https://www.adobe.com/special/products/audition/syntrillium.html>.



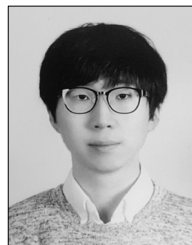
이 동 현

<https://orcid.org/0000-0001-6091-6254>

e-mail : sksaaa00@daum.net

2011년~2018년 영남대학교 컴퓨터공학과
학사

관심분야 : Data mining, Artificial
Intelligence, Software
Engineering



박 정 욱

<https://orcid.org/0000-0001-9070-7197>

e-mail : pjwook9112@naver.com

2011년~2018년 영남대학교 컴퓨터공학과
학사

관심분야 : Program Analysis, Software
Engineering



서 영 석

<https://orcid.org/0000-0002-5319-7674>

e-mail : ysseo@yu.ac.kr

2006년 숭실대학교 컴퓨터학부(학사)

2008년 KAIST 전산학과(석사)

2012년 KAIST 전산학과(박사)

2014년~2016년 한국산업기술시험원(KTL)

선임연구원

2016년~현재 영남대학교 컴퓨터공학과 교수

관심분야 : Data Mining, Software Modularization, Software
Cost Estimation, Software Measurement and
Analysis, Mining Software Repositories, and
Software Process Improvement