

가사의 감정 분석과 구조 분석을 이용한 노래 간 유사도 측정

(Similarity Evaluation of Popular Music based on Emotion and Structure of Lyrics)

이 재 환 [†] 임 혜 원 [†] 김 형 주 ^{**}
(Jaehwan Lee) (Hyewon Lim) (Hyoung-Joo Kim)

요약 음악 스트리밍 서비스의 대중화로 음악의 소유 없이 언제든지 원하는 듣고 싶은 노래를 들을 수 있게 되었다. 넓어진 선택권은 역설적으로 노래를 선택하기 어렵게 만들었다. 이러한 선택의 어려움을 극복하기 위해 음악 추천에 대한 관심이 높아졌고 판도라와 Last.fm과 같은 상용 서비스뿐 아니라 음악 정보 검색 분야의 연구자들도 다양한 추천 시스템을 제안하였다. 내용 기반 필터링과 협업 필터링 방식이 주류인 기존의 추천 시스템은 음악 감상의 주요 요인인 맥락을 고려하지 않았다는 한계점을 지니고 있다. 본 논문에서는 음악을 선택하는 맥락 중 주요한 요인인 감정을 이용한 노래간 유사도 측정 방법을 제안하여 새로운 추천 시스템에 대한 가능성을 탐색한다. 노래의 감정 추출에 가사를 이용하였고 가사에서 노래의 구조도 추출해 노래의 의미적 분석을 시도하였다. 실험을 통해 제안한 모델이 기존의 추천 시스템에 비해 작은 계산 복잡성으로 기존 모델과 유사한 성능을 보일 수 있음을 보였다.

키워드: 음악 추천, 추천 시스템, 감정 분석, 가사 분석

Abstract People can listen to almost every type of music by music streaming services without possessing music. Ironically it is difficult to choose what to listen to. A music recommendation system helps people in making a choice. However, existing recommendation systems have high computation complexity and do not consider context information. Emotion is one of the most important context information of music. Lyrics can be easily computed with various language processing techniques and can even be used to extract emotion of music from itself. We suggest a music-level similarity evaluation method using emotion and structure. Our result shows that it is important to consider semantic information when we evaluate similarity of music.

Keywords: music recommendation, recommendation system, emotion analysis, lyrics analysis

· 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No. R0113-15-0005, 대규모 트랜잭션 처리와 실시간 복합 분석을 통합한 일체형 데이터 엔지니어링 기술 개발)

[†] 비 회 원 : 서울대학교 컴퓨터공학부(Seoul National Univ.)
jlee@idb.snu.ac.kr
hwlim@idb.snu.ac.kr
(Corresponding author)

^{**} 중 신 회 원 : 서울대학교 컴퓨터공학부 교수
hjk@snu.ac.kr

논문접수 : 2016년 1월 28일
(Received 28 January 2016)
논문수정 : 2016년 6월 29일
(Revised 29 June 2016)
심사완료 : 2016년 7월 1일
(Accepted 1 July 2016)

Copyright©2016 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다. 정보과학회 컴퓨팅의 실제 논문지 제22권 제10호(2016. 10)

1. 서 론

음악 스트리밍 서비스는 네트워크가 연결되어 있는 기기를 이용해 음악을 저장하지 않아도 음악을 감상할 수 있는 서비스이다. 스트리밍 서비스를 이용하면 거의 모든 노래에 언제든지 접근할 수 있지만 듣기 위한 노래를 선택하는데 어려움을 겪게 된다.

음악 추천 시스템은 사용자들의 청취 히스토리를 이용해 사용자들에게 음악을 추천하여 이런 선택의 어려움을 해소한다. 음악 추천 시스템에서는 음악을 하나의 정보로 여기고 정보 검색 기법(Information Retrieval)을 유사하게 적용한다. 음악 정보 검색(Music Information Retrieval)이라고 불리는 이 연구 분야는 음악학, 심리학, 기계 학습 등 여러 학문이 결합되어있고 실생활에 다양하게 적용이 가능하다는 장점이 있다.

세계 음악 시장은 빠르게 변화하고 있다. 국제 음반산업협회(IFPI)에서는 ‘디지털 음악 보고서(Digital Music Report) 2011’에서 음악을 이용하는 방식이 다운로드와 같은 소유에서 스트리밍과 같은 접근으로 패러다임 변화를 이루고 있다고 지적하기도 하였다. 전 세계 디지털 음악 시장 규모 추이와 스트리밍 음악 시장 규모의 추이를 나타내는 그림 1¹⁾을 보면, 디지털 음악 시장의 성장률은 2011년 10.9%에서 2012년 9.8%, 2013년 5.4%로 성장세가 둔화되는 양상을 보인 반면 스트리밍 음악 시장은 2011년 39.0%, 2012년 63.1%, 2013년 51.4%로 빠른 성장세를 보이고 있다[1].

주요 글로벌 음악 스트리밍 서비스로는 스웨덴에서 설립되어 세계 28개국에서 서비스하고 있는 스포티파이(Spotify)와 세계 최대의 인터넷 라디오 스트리밍 서비스로 꼽히는 판도라(Pandora), 프랑스에서 설립되어 전 세계 182개국에서 서비스를 제공하고 있는 디저(Deezer)가 있다. 스포티파이는 2,000만곡 이상, 판도라는 100만곡 이상, 디저는 3,069만 곡 이상을 보유하고 있다. 세 글로벌 스트리밍 서비스의 특징은 표 1에 정리되어 있다.

주요 스트리밍 서비스는 음악 추천 서비스를 제공한다. 특히 미국의 판도라는 차별화된 추천 서비스로 다른 서비스에 비해 적은 양의 음악 보유량에도 불구하고 많은 사용자를 확보하고 있다. 판도라의 추천 시스템은 음악 교육을 받은 음악 전문가들이 각 노래를 최대 450개에 해당하는 특성 값으로 분류한다. 분류된 450개의 특성을 450차원의 좌표로 치환하여 좌표간 거리를 노래간의 유사도로 사용하는데 사용자의 취향을 반영하여 특성 간 다른 가중치를 둔 유클리드 거리를 사용한다[2].

2002년에 서비스를 시작한 영국의 Last.fm²⁾ 도 음악

표 1 글로벌 스트리밍 서비스 특징

Table 1 Characteristics of the global streaming services

	Spotify	Pandora	Deezer
Price per Month	\$4.99	\$9.99	\$4.99
Music	20M	1M	30M
Active User	24M	73M	12M
Offline Service	Unlimited mobile download	-	Unlimited download

추천 서비스를 제공하고 있다. 노래에 초점을 맞춘 판도라의 추천 시스템과 달리 Last.fm은 노래를 듣는 사용자에 초점을 맞추는데 Audioscrobbler이라고 불리는 플러그인을 이용해 데이터를 수집한다. 각 사용자의 청취 기록은 곧 사용자의 프로필이 된다. 이렇게 Audioscrobbler를 통해 수집된 사용자 프로필을 이용해 유사한 사용자를 찾고 유사 사용자의 청취 기록을 탐색하여 새로운 곡을 추천한다.

두 추천 시스템을 비롯한 기존 음악 추천시스템은 맥락 정보를 전혀 고려하지 않는다. 음악의 추천에 있어 청취자의 감정이나 행위, 계절과 같은 맥락 정보는 매우 중요한 역할을 한다[3]. 가사는 노래를 나타내는 메타데이터이므로 가사를 이용해 노래의 감정을 추출해 맥락 정보로 사용할 수 있다.

본 연구에서는 기존 연구들에서 제시된 음악 추천 시스템의 한계라고 할 수 있는 연산의 복잡성과 맥락에 대한 무시를 극복하고자 새로운 음악 추천 시스템을 제안한다. 특히 대중가요(Popular Music)에 집중하여 노래라는 용어를 대중가요로 한정하여 사용한다. 본 연구에서 제안하는 추천 시스템에서는 노래의 빠르기와 높

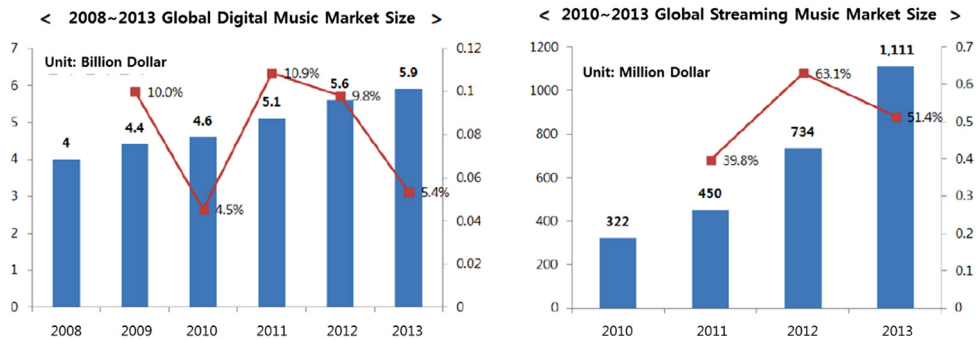


그림 1 전 세계 디지털 음악 시장과 스트리밍 음악 시장 규모 추이[1]

Fig. 1 The trends of the global digital music market and the streaming music market[1]

1) 그림 1은 [1]에서 인용한 그림으로, 한글로 작성된 그림을 한국정보과학회 논문 투고 규정에 따라 영어로 번역한 것임
 2) <http://www.last.fm>

낮이, 음계, 가사에 드러난 감정, 음악적 구조를 주요 요소로 선정하여 노래의 분위기를 추출하고 분위기를 비교하여 노래간 유사도를 계산한다. 유사도 계산의 복잡성을 줄이기 위해 감정과 음악적 구조의 추출에 가사를 사용하였다.

본 논문의 구성은 다음과 같다. 1장에서는 최근의 음악 시장 동향을 살펴보고 음악 추천 시스템의 필요성을 제기하였다. 2장에서는 본 연구의 가설과 밀접한 연관을 갖고 있는 관련 연구들을 소개한다. 3장에서는 본 연구에서 제안한 추천 시스템의 구조와 시스템에 사용된 알고리즘에 대해 설명하고 4장에서는 본 연구에서 제안한 시스템을 이용한 실험과 성능 평가에 대해 설명한다. 마지막으로 5장에서는 제안한 시스템에 대한 결론을 내리고 향후 연구를 제안한다.

2. 관련 연구

2.1 음악 추천 시스템

추천 시스템은 추천에 사용하는 데이터의 종류에 따라 두 가지로 분류된다. 사용자간 취향의 유사도를 이용하는 추천 시스템은 협업 필터링(Collaborative Filtering) 기법으로 분류되고, 추천 대상인 아이템의 유사도를 이용하는 추천 시스템은 내용 기반 필터링(Content-based Filtering) 기법으로 분류된다. 내용 기반 필터링 기법을 사용한 대표적 음악 추천 시스템으로 판도라의 추천 시스템이 있다. Last.fm의 추천 시스템은 협업 필터링 기법을 사용한다. 국내 주요 음악 서비스의 경우도 대부분 협업 필터링 기법을 사용하는 것으로 알려져 있다.

기존 추천 기법에는 장단점이 존재하는데, 내용 기반 필터링 기법은 음악을 요소로 구분하는 단계에서 어려움이 발생한다. 판도라의 경우 장르에 따라 최대 450개의 요소로 노래를 표현하는데, 이 450개 요소가 정답이 아니기 때문에 전문가에 따라 500개 혹은 900개의 요소가 더 좋다고 주장할 수 있다. 하지만 아이템에 기초하기 때문에 분석만 되어 있다면 아이템의 인기도에 상관없이 추천 대상이 될 수 있다는 장점이 있다. 그러나 여전히 뿌리(Root)가 되는 아이템과 비슷한 유형의 아이템만 추천한다는 한계를 갖고 있다.

협업 필터링 기법은 사용자를 벡터로 표현한다. 벡터의 차원은 각 아이템이 되며 벡터의 값은 '좋다/나쁘다'의 이진 값이나 별점 등을 나타낸다. 내용 기반 추천 기법과 달리 사용자를 표현하는 방법이 명확하다는 장점을 갖고 있지만 만약 n 명의 사용자가 있고 m 개의 아이템이 있다면 추천 시스템은 $O(mn)$ 의 연산을 수행해야 하기 때문에 사용자와 아이템의 규모가 커질수록 연산의 복잡성이 빠르게 증가한다. 이와 더불어 사용자와 아이템 모두에 콜드 스타트(Cold Start)라는 문제를 지니

고 있다. 신규 사용자가 시스템에 들어왔을 때 이 사용자의 취향에 대한 기록을 가지고 있지 않기 때문에 적절한 추천을 할 수 없다. 또한 사용자들에게 인기가 없는 아이템은 대부분의 사용자에게 벡터 값을 갖지 않기 때문에 추천 대상으로 선정될 수 없다.

음악 추천 시스템의 경우에는 협업 필터링 기법과 내용 기반 추천 기법 방식 이외에 맥락에 반영한 맥락 기반(Context-aware) 추천 기법을 활용하기도 한다[3]. 맥락이란 사용자가 음악을 듣는 환경, 음악을 들으면서 하는 행동, 음악을 함께 듣고 있는 사람 등을 의미한다. 일부 연구자들은 체온이나 심장박동, 기온 등을 맥락에 포함시키기도 하였다[4].

[3]에서는 가장 중요한 맥락 정보로 감정을 꼽았다. 본 연구에서는 맥락에 맞는 음악 추천을 위해서 노래에 드러난 감정을 추출하였으며 감정 추출하기 위해 노래의 가사에서 나타나는 감정과 노래의 음악적 구조를 분석하여 유사도 계산에 사용한다.

2.2 감정 분석

감정(Emotion)과 분위기(Mood)는 의미상 작은 차이가 있다. 분위기는 감정의 잔재라고 볼 수 있는데, 감정을 유발한 자극에 대한 집중이 흩어진 후에 감정은 분위기로 전환된다. 즉, 분위기는 좀 더 맥락적인 자극에 의해 유발되는 감정 반응이라고 얘기할 수 있다[5].

본 연구에서는 [5]와 마찬가지로 감정과 분위기를 구분하지 않고 감정이라는 단어로 함께 사용하였다.

2.2.1 노래의 감정 분류

노래의 감정을 분류하려는 연구들이 다각도에서 이루어졌다. [6]에서는 오디오 정보와 가사를 이용해 노래의 감정을 분류하였다. 네 가지 수준의 감정 angry, happy, sad, relaxed를 선정하여 실험을 진행하였는데 가사만 사용하였을 때보다 오디오와 가사를 모두 사용하였을 때가 더 좋은 성능을 보였다.

[7]에서도 오디오 정보와 가사를 이용해 노래의 감정을 분류하고자 하였다. 태그 정보를 이용하여 감정 범주를 생성한 후에 [6]의 연구와 마찬가지로 오디오, 가사, 오디오+가사의 세 가지 방법에 대한 비교 실험을 수행하였다. 범주에 따라 다른 성능을 나타냈는데 happy나 calm과 같은 범주에서는 오디오를 사용한 분류가 가장 좋은 성능을 보인 반면 romantic이나 angry 같은 범주에서는 가사를 사용하였을 때 가장 좋은 성능을 나타냈다.

본 연구에서는 감정 범주의 선정에 따라 다른 성능을 보인 기존 연구의 문제점을 고려하여 감정을 범주가 아닌 벡터로 표현하여 벡터 연산을 이용해 감정을 비교하였다.

2.2.2 Emolex

텍스트에서 감정을 추출할 수 있다면 여러 연구에 유

용하게 쓰일 수 있다. 텍스트에서 감정 추출을 할 수 있도록 Emolex, ANEW³⁾, SentiWordNet⁴⁾과 같은 감정 단어 사전이 제시되어 왔다. 이 중 Emolex는 심리학자 Plutchik이 제시한 인간의 8가지 기본 감정 anger, anticipation, disgust, fear, joy, sadness, surprise, trust에 긍정과 부정을 더해 10차원 수준의 감정 단어 사전을 제공한다[8]. 현재 최신판의 Emolex에는 14,177개의 단어가 들어있으며 각 단어는 10차원의 이진 벡터로 표현된다. 텍스트의 감정은 텍스트 내부의 단어들의 감정벡터의 합으로 계산된다.

어떤 노래의 가사 Lyrics를 Bag-of-word의 형태로 보면 단어 w의 집합이 된다. Lyrics의 감정벡터는 이 w의 감정벡터들의 합으로 계산된다. 이를 수식으로 나타내면 아래와 같다.

$$\overrightarrow{E_{Lyrics}} = \sum_{w \in Lyrics} \overrightarrow{E_w}$$

[9]는 유튜브(YouTube) 동영상에 달린 댓글을 이용해 영상에 대한 감정을 분석하였다. Emolex를 이용해 댓글에 나타난 감정을 계산하고 댓글이 달려있는 동영상에 나타난 감정과 코사인 유사도를 계산하였을 때 0.75의 높은 수준에서 일치하는 결과를 보였다. 본 연구에서는 [8]에서처럼 노래의 감정을 벡터로 나타낸 후 코사인 유사도를 사용해 노래간의 감정을 비교하였다. 현재 Emolex에서 제공하는 14,177개의 단어는 다소 부족하다고 판단하여 4.1.3에서 설명하는 알고리즘을 통해 단어 사전을 확장하였다.

2.3 노래와 노래 가사

노래는 다양한 메타 데이터를 가지고 있다. 발매년도나 노래 제목, 가수의 이름, 노래가 수록된 앨범은 모두 노래의 메타 데이터라고 할 수 있다. 특히 노래 가사는 웹 검색을 통한 수집의 용이성, 객관성, 풍부함과 같은 장점을 갖고 있는 좋은 메타 데이터라고 할 수 있다.

이런 장점 때문에 가사를 이용한 다양한 연구들이 행해졌다. 가사를 이용해 가수들의 유사도를 계산한 연구 [10]와 가사를 이용해 노래의 장르를 분류하려는 연구 [11]가 그 중 하나이다. 2.2.1에서 소개한 것과 같이 노래의 감정을 분류하려는 연구 또한 이루어졌다[6,7].

본 연구에서도 가사를 매우 중요한 메타 데이터로 보고 가사에서 감정을 추출해 노래간 유사도 계산에 사용하였다. [10]에서는 아티스트 수준의 유사도를 계산하였으나 본 연구에서는 노래간 유사도에 대한 정답 셋을 이용해 노래 수준에서 유사도를 검증하였다. 또한 2.4에서 설명할 가사의 특징을 이용해 가사에서 노래의 구조

를 유도하여 유사도 분석에 사용한다는 점에서 차별성을 갖는다.

2.4 노래의 구조

노래의 구조를 이해하는 것은 노래를 의미 단위로 구분할 수 있게 해준다. 노래의 구조를 분석하면 노래 전체를 듣지 않고 노래의 핵심 부분을 잡을 수 있고, 후렴구를 이용해 비슷한 노래를 찾는 것도 가능하다[12].

노래의 구조를 분석하는 단위는 그 역할과 특징에 따라 Intro, Verse, Chorus, Bridge, Outro로 구성된다. [13]에 따르면 일반적인 노래는 아래 세 가지 구조를 가진다.

- 1) Intro, Verse, Chorus, Verse, Chorus, Outro
- 2) Intro, Verse, Chorus, Verse, Chorus, Bridge, Chorus, Outro
- 3) Intro, Verse, Chorus, Verse, Bridge, Chorus, Outro

Intro와 Outro의 경우 대부분 가사가 없어 가사로 찾아내는 것은 어렵다. 따라서 가사를 이용해 구조를 분석하는 연구에서는 Verse와 Chorus, Bridge를 주로 찾아낸다.

본 연구에서는 [12,14]의 연구를 참고하여 가사에서 구조를 분석하는 알고리즘을 설계했다. 웹에서 수집된 가사 데이터는 문단 구분에 통일성이 없고 가사가 아닌 내용이 포함된 경우가 있어 전처리를 수행했다. 구조 분석 알고리즘은 3.3에서 자세히 설명한다.

3. 가사의 감정 분석과 구조 분석을 이용한 노래간 유사도 측정 시스템

본 연구에서는 일부 오디오 정보와 가사에서 추출된 감정과 구조를 이용해 노래간 유사도를 계산하는 시스템을 제안한다. 오디오 정보는 노래의 빠르기와 같은 음악적 구성을 제공한다는 측면에서 문법적(Syntax) 정보라고 구분할 수 있다. 반면에 노래 가사에 담긴 감정이나 노래의 구조는 주제나 의미를 표현한다는 측면에서 의미적(Semantic) 정보로 구분할 수 있다.

그림 2는 전체 시스템 구성 나타낸다. 먼저 각 노래들은 문법적 분석과 의미적 분석 과정을 거쳐 데이터베이스에 저장된다. 시스템에 입력으로 두 노래가 들어오면 데이터베이스를 확인하여 두 노래의 유사도를 계산하여 반환한다.

3.1 오디오 정보

[6,7]의 연구에서 가사만 사용하였을 때보다 오디오 정보를 함께 활용하였을 때 더 좋은 성능을 보였다. [15]의 연구에서도 가사 분석을 주요소로 사용하였으나 춤 출수 있는 정도(Danceability)나 에너지 등의 오디오 요소를 함께 사용하였다. 따라서 본 연구에서도 음악 전문가 세 명의 의견을 받아 노래를 대변할 수 있으면서

3) Affective Norms for English Words

4) <http://sentiwordnet.isti.cnr.it/index.php>

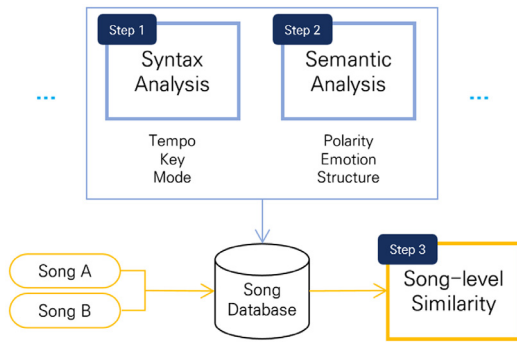


그림 2 시스템 구성도
Fig. 2 System overview

측정하기 쉬운 요소인 빠르기(tempo, BPM)와 높낮이(key), 음계(mode)를 활용하였다. 빠르기는 일반적으로 0에서 250까지의 정수값이 사용된다. 따라서 BPM간의 차이로 유사도를 계산할 수 있다. 높낮이는 음악 전문가의 도움을 받아 화성학에 기초한 유사도를 사용하였는데, 예를 들어 C코드와 G코드는 화성학적으로 매우 유사하므로 동일한 코드와 함께 가장 높은 유사도로 계산된다. 음계는 장조와 단조의 두 가지 값을 가지므로 일치하면 1점, 일치하지 않으면 0점을 부여하였다.

3.2 가사의 감정 분석

가사의 감정은 가사를 하나의 텍스트로 간주하여 각 단어들의 감정 벡터의 합을 이용해 계산하였다. 계산된 감정 벡터 값은 가사 내의 감정 단어의 개수에 따라 큰 편차를 보이므로 유사도 계산에 앞서 정규화 하였다. 먼저 긍정과 부정에 해당하는 벡터 값의 합이 1이 되도록 두 값을 정규화 한 후, 긍정과 부정을 제외한 8가지 기본 감정 벡터 값의 합이 1이 되도록 정규화 하였다. 정규화된 벡터간의 코사인 유사도가 가사의 감정간 유사도로 사용되었다.

3.3 노래 구조 분석

[12,14]의 연구를 참고하여 아래와 같이 구조 분석에 대한 근사 규칙을 설립하였다. Bridge를 구분하는 것이 노래 구조를 이해하는 데 있어 중요하지 않다고 판단하여 본 연구의 노래 구조 분석기는 Chorus와 Verse만 구분하였다.

한 번 이상 반복된 문장은 Chorus가 될 가능성이 있다. 두 문장의 각 첫 n 개의 단어 중 m 개 이상의 단어가 겹칠 때 두 문장은 반복되었다고 판정한다. 한 번 이상 반복되는 문장 문치는 Chorus일 확률이 매우 높으므로 가사에서 먼저 Chorus를 찾아낸다. 그 후에 분류되지 않은 문장을 Verse로 분류하면 노래의 Verse와 Chorus를 구분할 수 있다. 두 노래 구조의 유사도는 아래와 같은 규칙으로 정의하였다. 가장 높은 유사도와 높은 유사

도는 1점, 보통 유사도는 0.5점, 마지막으로 낮은 유사도는 0점으로 계산되었다.

- ㄱ. 가장 높은 유사도 : 완전히 동일한 구조
- ㄴ. 높은 유사도 : 시작과 끝이 일치
- ㄷ. 보통 유사도 : 시작이나 끝 하나만 일치
- ㄹ. 낮은 유사도 : 그 외

3.4 유사도 계산 함수

노래간 유사도는 크게 구조적 분석과 의미적 분석으로 나뉜다. 문법적 분석은 다시 빠르기, 높낮이, 음계로 나뉘고 의미적 분석은 극성(Polarity)과 감정, 구조로 나뉜다.

각각의 요소에 대한 유사도 함수를 정의하고 각각의 가중치를 합산하는데 가중치를 부여하여 합산하였다. 즉, 두 노래의 유사도는 다음과 같이 정의된다.

$$\begin{aligned} \text{유사도}(\text{노래 } A, \text{노래 } B) \\ = \sum_{\text{요소}} \left(\text{가중치}_{\text{요소}} \times \text{유사도}_{\text{요소}}(\text{노래 } A, \text{노래 } B) \right) \end{aligned}$$

4. 실험

4.1 실험 환경 및 데이터

본 연구에서는 총 2대의 PC를 사용해 실험을 수행했다. 데이터베이스의 구축 및 운용을 위해 CPU Intel(R) Core(TM) i5-4670 3.40GHz, RAM 8GB, CentOS 6.7의 PC를 사용하였다. DBMS는 MySQL 14.14버전을 사용하였고 PHP는 5.3.3 버전을 사용하였다. Java로 작성된 추천 시스템을 구동한 PC는 CPU Intel(R) Core(TM) i5-4670 3.40GHz, RAM 8GB, 운영체제 Windows 8 Enterprise K 64bit의 PC를 사용하였다.

4.1.1 Million Song Dataset

실험에는 Million Song Dataset이 사용되었다. Million Song Dataset은 총 100만곡에 대한 정보를 포함하고 있다. 각각의 곡에 대하여 제목, 아티스트 이름, 발매년도와 같은 메타데이터와 BPM(tempo), 높낮이, 오디오 분석과 같은 음악 정보가 함께 입력되어 있다[16].

본 연구에서는 문법적 분석에 필요한 빠르기, 높낮이, 음계 정보와 Last.fm에서 제공하는 유사도 정보를 추출해 사용하였다.

4.1.2 가사 데이터 수집

LyricsWikia⁵⁾의 API를 사용하여 가사를 검색하였다. LyricsWikia에서 제공하는 API를 사용해 가사의 보유 여부를 확인한 후에 HTML DOM Parser를 이용하여 실제로 가사가 존재하는 URL에 접근하여 가사를 추출하였다. 실험데이터로 사용한 Million Song Dataset의 100만곡 중 LyricsWikia를 통해 206,191곡의 가사가 수집되었다.

5) http://lyrics.wikia.com/wiki/Lyrics_Wiki

4.1.3 감정 단어 사전 확장(e-Emolex)

감정 단어 사전 Emolex를 확장하기 위하여 [5]의 방법을 기반으로 하여 동의어를 수집하였다. [5]는 실험을 통해 감정 단어의 동의어를 이용하여 자연 언어 처리를 수행하여도 감정 분포가 흐트러지지 않음을 확인하였다.

본 연구에서는 Thesaurus.com⁶⁾을 이용해 동의어를 수집하였다. Common읍선을 활성화하여 사전 상 의미는 비슷하지만 활용 면에서 동 떨어지는 동의어는 제거하여 총 6,101개의 동의어를 수집하였다.

각 동의어는 [5]의 연구에서와 마찬가지로 TF-IDF에 따라 가중치를 설정하였다. DF는 Emolex에 포함된 단어를 문서로 간주하여 계산하였다. 만약 단어 A가 Emolex의 감정 단어 4개의 동의어로 선택되었다면 단어 A의 TF-IDF 값은 1/4이 된다. 단어 A의 감정 벡터 값은 동의어 4개의 감정 벡터의 합에 가중치인 TF-IDF 1/4을 곱한 값을 갖게 된다.

예를 들어 단어 freedom은 Emolex의 감정 단어 19개의 동의어로 나타났으므로 1/19의 TF-IDF 값을 갖게 된다. freedom의 동의어로 뽑힌 단어는 다음과 같다.

{abandon, autonomy, democracy, dismissal, escape, familiarity, immunity, leave, leisure, liberalism, license, lifeboat, passage, safety, sanctification, scope, securities, security, visa}

19개 단어의 감정 벡터를 합하면 {5, 4, 1, 3, 1, 3, 3, 3, 3} 이 된다. 이 감정 벡터에 TF-IDF 값인 1/19을 곱하면 아래와 같이 freedom의 감정 벡터 값을 구할 수 있다.

freedom = {0.2632, 0.2105, 0.0526, 0.1579, 0.0526, 0.1579, 0.1579, 0.1579, 0.1579, 0.1579}

따라서 확장된 감정 단어 사전 e-Emolex는 0과 1을 벡터 값으로 갖던 이진 벡터의 성질을 잃게 된다.

Emolex는 위의 방법을 통해 1,968개의 단어가 추가 되어 총 16,145개의 단어의 e-Emolex로 확장되었다. Emolex와 e-Emolex를 이용하여 가사에 대한 감정 분석을 실시하였을 때 Emolex는 노래별 평균 43.60개의 단어와 일치하였으나 e-Emolex를 사용하였을 때는 평균 91.30개의 단어와 일치하여 성능 향상을 보였다. 가사에 포함된 평균 단어의 수는 247.12개, 표준편차는 149.23이다.

4.1.4 노래 구조 분석기 성능 평가

노래 구조 분석기에서 문장의 반복을 판단하기 위해 두 문장의 각 첫 n 개의 단어 중 m 개 이상의 단어를 확인한다. 실험에서는 문장의 반복 여부를 $n = 4$, $m = 3$ 으로 설정하여 판단하였다. 즉 첫 4개의 단어 중 3개 이상

일치하면 반복되는 문장으로 판정한다.

수집된 노래 가사에 대한 구조 분석결과 가장 높은 빈도를 나타내는 구조는 VCVCVC(15.44%)와 VCVC(14.49%)이다. 본 연구에서 Bridge를 따로 구분하지 않는 것을 고려하면 VCVCVC와 VCVC의 두 구조는 [13]에서 주장한 전형적인 노래 구조와 일치한다.

일부 노래에서 VCVCVCVCVCVCVCVC와 같은 복잡한 패턴이 관측되었으나 15분 이상의 긴 노래에 많은 양의 가사를 가진 협합 곡에 해당하였고 빈도가 매우 낮아 실험 결과에는 큰 영향을 미치지 않았다.

4.2 성능 평가 기준

추천 시스템의 성능을 평가하는 기준으로 Last.fm에서 제공하는 노래간 유사도 점수를 사용하였다. Last.fm은 협업 필터링 기법을 사용하여 음악간 유사도를 측정한다.

Last.fm에서 제공하는 유사도 데이터는 하나의 시드(Seed)곡에 대하여 평균 43.75개(표준편차 35.13)의 노래와 유사도를 가지고 있다. 데이터를 살펴본 결과 일부 곡은 0.9에서 1에 육박하는 매우 높은 유사도 수치를 가진 반면 다른 곡은 0.005보다 낮은 유사도 수치를 가져 큰 편차를 보였다. 높은 유사도를 나타내는 곡들은 시드 곡과 동일한 가수의 곡이었다. 따라서 시드 곡과 동일한 아티스트의 곡은 제외하고 실험을 수행하였다.

또한 노래에 대한 유사도 수치 간 차이가 소수점 둘째 자리 이하로 매우 작은 차이를 가지는 경우가 많았다. 따라서 Last.fm에서 제공하는 유사도 수치가 아닌 유사도의 랭킹을 비교하여 성능을 평가하였다.

4.3 성능 평가 방법

성능 평가를 위해 정보 검색 분야에서 검색 알고리즘 간의 성능 비교를 위해 사용하는 방법 중 하나인 NDCG(Normalized Discounted Cumulative Gain)[17]를 이용하였다.

NDCG는 두 알고리즘에서 결과로 출력한 랭킹의 성능차이를 계산할 수 있는데 높은 순위에 있는 아이템에는 높은 가중치를 부여하고 낮은 순위에 있는 아이템에는 낮은 가중치를 부여한다. 이를 식으로 나타내면 다음과 같다.

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

$IDCG_p$ 는 가장 이상적인 점수를 의미한다. DCG_p 를 $IDCG_p$ 로 나눈 값은 최댓값이 1인 정규화 값을 갖게 된다. rel_i 는 i 번째 항목의 가중치를 나타내는데 본 연구에서는 Last.fm의 정답 순서에 비례하게 가중치를 갖도

6) <http://www.thesaurus.com/>

록 $rel_i = p - i$ 로 정의하여 계산하였다. 본 연구에서 제안한 시스템은 모든 노래 쌍에 대하여 유사도를 계산할 수 있다. 하지만 Last.fm에서 제공하는 유사도 정보가 없으면 계산된 유사도에 대하여 성능을 평가할 수 없다. 실험을 위해 Last.fm에서 제공하는 유사도 테이블을 먼저 읽어 존재하는 유사도 쌍에 대해서만 유사도 계산을 수행하였다.

본 연구에서는 가사를 이용한 의미적 분석이 유사도 계산에 있어 유의미하다는 것을 보이고자 한다. 따라서 문법적 요소만 사용한 시스템과 문법적 요소와 의미적 요소를 함께 사용한 시스템에 대한 비교 실험을 수행하였다. 시뮬레이션 결과 가장 성능이 좋았던 가중치를 실험에 사용하였다. 문법적 분석과 의미적 분석은 1:1의 가중치로 합산되었고 각 요소의 상대적 가중치는 {빠르기, 높낮이, 음계, 극성, 감정, 구조}의 순으로 {2,1,1,2,1,1}로 사용되었다.

4.4 성능 평가 결과

두 시스템을 이용하여 가사와 유사도 정보를 모두 갖고 있는 159,548개의 시드 곡에 대하여 유사도 분석을 수행하였다. 성능 평가 척도인 NDCG는 시드 곡마다 계산되므로 실험의 결과로 각 시스템은 159,548개의 NDCG 값을 갖는다. 시스템의 평균적인 성능을 평가하기 위하여 각 시스템의 NDCG 값들의 평균과 표준편차를 이용해 두 시스템의 성능을 비교하였다.

그림 3은 두 시스템의 NDCG 값의 분포를 나타내는 상자 그림이다. 그림 3의 하단에 있는 빨간색 실선은 임의로 랭킹을 뽑았을 때의 평균 NDCG의 값으로 0.1751을 값으로 갖는다. 두 시스템 모두 기준선 보다 나은 성능을 나타냈다. 문법적 분석만 사용한 시스템은 NDCG

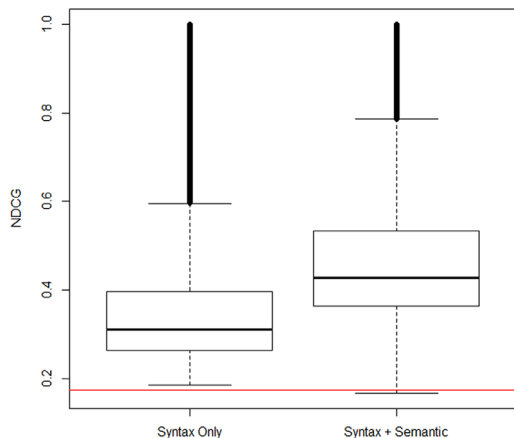


그림 3 Syntax Only와 Syntax+Semantic 성능 비교
Fig. 3 Performance comparison between Syntax Only and Syntax+Semantic

표 2 Syntax Only와 Syntax+Semantic 성능 비교 통계치
Table 2 The statistics of the performance comparison between Syntax Only and Syntax+Semantic

	Syntax Only	Syntax + Semantic
Average	0.3593	0.4702
SD	0.1462	0.1529
Median	0.3118	0.4287

값의 평균이 0.3593으로 나타났다. 문법적 분석에 의미적 분석을 더한 시스템의 평균 NDCG는 0.4702로 구조적 분석만 사용한 시스템보다 0.1109만큼 높게 나타났다. 표준편차와 중앙값을 포함한 통계치는 표 2에 나타났다. 두 시스템에서 출력한 NDCG 값에 대하여 대응 t-검정을 실시하였을 때 유의확률 p는 2.2e-16로 나타났다(t = -284.14). 따라서 두 시스템은 유의수준 0.05에서 통계적으로 유의미한 차이를 보인다.

5. 결론 및 향후 연구

본 연구에서는 협업 필터링 기법과 내용 기반 추천 기법의 단점을 보완하기 위하여 가사를 이용한 음악 추천 시스템을 제안하였다. 가사만을 사용했을 때 놓칠 수 있는 오디오 정보를 보완하기 위하여 빠르기, 높낮이, 음계도 유사도 분석에 사용하였다. 본 논문에서 제안한 음악 추천 시스템의 요소들은 비교적 쉽게 분류할 수 있다는 장점이 있다. 웹에서 가사와 필요한 오디오 정보를 찾으면 어떤 노래든 노래 간의 유사도를 계산할 수 있다.

문법적 분석 기법은 임의의 랭킹 알고리즘 보다 나은 성능을 보였으며, 의미적 분석 기법을 문법적 분석 기법과 함께 사용하였을 때 가장 좋은 성능을 보였다. 따라서 의미적 분석 기법에 사용된 극성, 감정, 구조가 노래 간 유사도에 고려되어야 하는 요인이라고 할 수 있다.

특히 음악 추천에 있어 맥락의 중요성이 대두되고 있는 시기에 맥락을 반영할 수 있도록 노래에 드러난 감정이 음악 추천에 중요한 의미를 지니고 있음을 실험을 통해 확인하였다는 점에서 고무적이라 할 수 있다.

다만 실험 결과를 해석하는 데는 주의가 필요하다. 본 연구에서 제시한 모델은 기존 추천 시스템이 갖고 있던 콜드 스타트와 높은 계산 복잡성이라는 단점을 보완하며 기존 협업 필터링 방식의 추천 시스템인 Last.fm과 유사한 성능을 나타냈다는 데에 의의가 있다. 실험의 결과는 이런 부분에서 해석되어야 하며 실제 두 시스템간의 추천 성능 비교는 실험 참가자 등을 통한 추가 실험으로 측정되어야 할 것이다.

또한 음악 추천 시스템의 일부로 구현된 노래 구조 분석기는 기대 이상의 성능을 나타냈다. 따라서 노래 구조 분석기를 발전시켜 순수한 가사에서 신뢰도 높은 노

래 구조를 추론할 수 있다면 다양한 분야에 적용이 가능할 것으로 보인다.

감정 단어 사전을 한 단계 더 확장하는 것도 가능할 것으로 보인다. Plutchuk이 분류한 8개의 기본 감정은 각각 대응되는 다른 감정을 하나씩 가진다. 예를 들어 joy는 sadness와 대응되고 fear는 anger와 대응된다. anticipation은 surprise에 대응되고 disgust는 trust와 대응된다. 따라서 Emolex 이진 벡터의 값을 반대로 사용한다면 반의어를 활용한 감정 단어 사전의 확장도 가능할 것이라 기대된다.

마지막으로 본 연구에서는 문법적 요소와 의미적 요소들의 가중치를 몇 차례의 시뮬레이션을 통한 최적 값으로 설정하였다. 데이터의 크기가 약 15만 곡으로 많기 때문에 Training Set과 Test Set으로 구분하기 충분하다. 따라서 기계 학습 기법을 이용해 최적의 가중치 계수를 찾는 것도 추천 시스템의 성능을 향상시킬 수 있다.

References

- [1] "Main Music Streaming Services and Technology Analysis Changing Digital Music Market," *KOCCA (Korea Creative Content Agency)*, Jul. 2014.⁷⁾
- [2] W. T. Glaser, T. B. Westergren, J. P. Stearns, and J. M. Kraft, "Patent {US7003515}: Consumer item matching method and system," *Pandora Media, Inc.*, Vol. 1, No. 12, 2006.
- [3] Z. Hyung, K. Lee, and K. Lee, "Music recommendation using text analysis on song requests to radio stations," *Expert Syst. Appl.*, Vol. 41, No. 5, pp. 2608-2618, 2014.
- [4] J. Su, H. Yeh, P. Yu, and V. Tseng, "Music recommendation using content and context information mining," *Intell. Syst. IEEE*, Vol. 25, No. 1, pp. 16-26, 2010.
- [5] I. Fernández-Tobías, I. Cnatador, and L. Plaza, "An Emotion Dimensional Model based on Social Tags: Crossing Folksonomies and Enhancing Recommendations," *Proc. of the 14th International Conference on Electronic Commerce and Web Technologies*, pp. 88-100, 2013.
- [6] C. Laurier, J. Grivolla, and P. Herrera, "Multimodal music mood classification using audio and lyrics," *Proc. - 7th Int. Conf. Mach. Learn. Appl. ICMLA 2008*, pp. 688-693, 2008.
- [7] X. Hu, J. S. Downie, and A. F. Ehmman, "Lyric text mining in music mood classification," *Am. Music*, Vol. 183, No. Ismir, pp. 411-416, 2009.
- [8] S. M. Mohammad and P. D. Turney, "Emotions Evoked by Common Words and Phrases: Using

Mechanical Turk to Create an Emotion Lexicon," *Proc. of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 26-34, 2010.

- [9] C. Orellana-rodriguez, E. Diaz-aviles, and W. Nejdl, "Mining Emotions in Short Films: User Comments or Crowdsourcing?," *Proc. of the 22nd international conference on World Wide Web companion*, pp. 69-70, 2013.
- [10] B. Logan, a. Kositsky, and P. Moreno, "Semantic analysis of song lyrics," *2004 IEEE Int. Conf. Multimed. Expo (IEEE Cat. No.04TH8763)*, Vol. 2, pp. 27-30, 2004.
- [11] R. Mayer, R. Neumayer, and A. Rauber, "Rhyme and Style Features for Musical Genre Classification By Song Lyrics," *Science. Ismir*, pp. 337-342, 2008.
- [12] H. T. Cheng, Y. H. Yang, Y. C. Lin, and H. H. Chen, "Multimodal structure segmentation and analysis of music using audio and textual information," *Proc. IEEE International Symposium on Circuits and Systems*, pp. 1677-1680, 2009.
- [13] "Ten Minute Master No 18: Song Structure," *MUSIC TECH magazine*, pp. 62-63, 2003.
- [14] N. C. Maddage, C. Xu, M. S. Kankanhalli, and X. Shao, "Content-based music structure analysis with applications to music semantics understanding," *Proc. 12th Annu. ACM Int. Conf. Multimed.*, pp. 112-119, 2004.
- [15] A. Jamdar, J. Abraham, K. Khanna, and R. Dubey, "Emotion Analysis of Songs Based on Lyrical and Audio Features," *Int. J. Artif. Intell. Appl.*, Vol. 6, No. 3, pp. 35-50, 2015.
- [16] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, "The Million Song Dataset," *Ismir*, pp. 591-596, 2011.
- [17] K. Järvelin and J. Kekäläinen, "IR Evaluation Methods for Retrieving Highly Relevant Documents," *Proc. of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '00*, pp. 41-48, 2000.



이 재 환

2014년 서울대학교 컴퓨터공학부 학사
2016년 서울대학교 컴퓨터공학부 석사
관심분야는 소셜 네트워크 분석, 데이터 마이닝

7) 이 문헌은 한글로 작성되어 한국콘텐츠진흥원에서 발간한 간행물로, 제목은 한국정보과학회 논문 투고 규정에 따라 영어로 번역하였음. 원제는 "디지털 음악시장의 판도를 바꾸는 주요 음악 스트리밍 서비스와 기술분석".



임 혜 원

2008년 숙명여자대학교 컴퓨터과학과 학사. 2008년~현재 서울대학교 컴퓨터공학부 석박사통합과정. 관심분야는 소셜 네트워크 분석



김 형 주

1982년 서울대학교 전산학과 학사. 1985년 Univ. of Texas at Austin 석사 1988년 Univ. of Texas at Austin 박사. 1988년~1990년 Georgia Institute of Technology 부교수. 1991년~현재 서울대학교 컴퓨터공학부 교수. 관심분야는 데이터베이스, XML, 시맨틱 웹, 빅데이터