



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

가사의 감정 분석과 구조 분석을 이용한
노래간 유사도 측정

Popular Music Similarity Evaluation
using Emotion and Structure Analysis on Lyrics

2016 년 2월

서울대학교 대학원
컴퓨터공학부
이재환

가사의 감정 분석과 구조 분석을 이용한
노래간 유사도 측정

지도교수 김형주

이 논문을 공학석사 학위논문으로 제출함

2015년 10월

서울대학교 대학원
컴퓨터공학부
이재환

이재환의 석사학위논문을 인준함

2015년 12월

위 원 장 문 봉 기 (인)

부 위 원 장 김 형 주 (인)

위 원 이 상 구 (인)

요약(국문초록)

기술의 발달로 우리는 공간과 시간의 제약 없이 음악을 들을 수 있다. 음악에 대한 높아진 접근성은 반어적으로 음악 선택의 어려움을 가져오기도 한다. 이런 어려움을 극복하기 위해 음악을 정보의 하나로 간주하고 정보 탐색 기법을 적용한 음악 정보 탐색 연구 분야가 성장하였고, 특히 음악 정보 탐색 분야 중 음악 추천 시스템에 대한 여러 연구가 진행되었다. 기존의 추천 시스템 방식인 협업 필터링 기법과 내용 기반 추천 기법은 연산의 복잡성과 쿨드 스타트, 맥락에 대한 무시와 같은 한계점을 지니고 있다.

사람들은 그들의 행동, 주변 환경이라는 맥락에 맞춰 음악을 소비한다. 따라서 음악 추천 시스템은 맥락을 반영하여야 한다. 기온, 심장 박동, 감정 등이 맥락 정보로 고려되는데, 본 연구에서는 가장 중요한 맥락 정보로 꼽히는 감정을 이용하여 노래간 유사도를 측정하는 방법을 제안한다. 노래에서 감정을 추출하기 위하여 수집이 쉽고 연산이 복잡하지 않은 가사를 분석에 사용하였다. 또한 감정을 풀어내는 방법인 구조를 가사에서 추출하여 유사도 계산에 사용하였다. 제안한 모델의 성능을 평가하기 위하여 가사에서 추출한 감정과 가사에서 추론된 노래 구조를 반영한 시스템과 반영하지 않은 시스템의 추천 성능을 비교 실험하였다.

가사에서 놓칠 수 있는 오디오 정보를 보완하기 위해 빠르기, 높낮이, 음계를 구조적 정보로 사용하였다. 실험을 통해 가사를 통한 감정과 구조 분석이 가능함을 보였고 감정과 노래 구조와 같은 의미적 정보가 음악 추천 시스템의 성능 향상에 보탬이 됨을 확인하였다.

주요어 : 음악 추천, 맥락, 가사, 추천 시스템, 감정

학 번 : 2014-21797

목 차

1. 서 론	1
2. 관련연구	6
2.1. 음악 추천 시스템	6
2.2. 감정 분석	8
2.2.1. 노래의 감정 분류	8
2.2.2. 감정 단어 사전(Emolex)	9
2.3. 노래와 노래 가사	10
2.4. 노래의 구조	11
3. 가사의 감정분석과 구조분석을 이용한 노래간 유사도 측정 시스템	13
3.1. 오디오 정보	14
3.2. 가사 감정 분석	14
3.3. 노래 구조 분석	15
3.4. 유사도 계산 함수	16
4. 실험	18
4.1. 실험 환경 및 데이터	18
4.1.1. Million Song Dataset	18
4.1.2. 가사 데이터 수집	19
4.1.3. 감정 단어 사전 확장(e-Emolex)	20

4.1.4. 노래 구조 분석기 성능 평가	23
4.2. 성능 평가 기준	24
4.3. 성능 평가 방법	26
4.4. 성능 평가 결과	27
5. 결론 및 향후 연구	29
참고 문헌	31
부록 A. e-Emolex Example	34
부록 B. 가사 구조 측정을 위한 문장 반복 횟수 측정의 예	35
Abstract	36

표 목 차

표 1. 주요 글로벌 음악 스트리밍 서비스 업체별 특징[각 사 발표]	2
표 2. 노래 가사에서 가장 빈번하게 등장하는 단어 순위와 그 단어	22
표 3. 감정 단어 사전의 성능 비교	23
표 4. 노래 구조 분석 결과	24
표 5. Syntax Only와 Syntax+Semantic 시스템 성능 비교	27

그림 목 차

그림 1. 전 세계 디지털 음악 시장과 스트리밍 음악 시장 규모 추이	2
그림 2. 시스템 구성도	13
그림 3. Million Song Dataset의 데이터 형태(HDF5 Format)	19
그림 4. Thesaurus.com에서 confused의 동의어를 검색한 화면 ...	21
그림 5. Syntax Only와 Syntax+Semantic 시스템 성능 비교	26

제 1 장

서론

기술의 발달로 사람들이 시간과 장소에 구애 받지 않고 음악을 들을 수 있게 되면서 연구자들은 음악을 재생하고 저장하는 장치뿐만 아니라 음악을 하나의 정보로 여기고 정보 검색(Information Retrieval)기법들을 유사하게 적용하는 것에 관심을 갖게 되었다.

음악 정보 검색(Music Information Retrieval)이라고 불리는 이 연구 분야는 음악학, 심리학, 기계 학습 등 여러 학문이 결합되어있고 실생활에 다양하게 적용이 가능하다는 장점 덕에 빠르게 성장하고 있다. 음악 정보 검색 기술은 음악 추천 시스템, 트랙 추출, 악기 재인, 자동 악보 추출, 장르 분류 등에 적용할 수 있을 것이라 기대되며 특히 음악 추천은 음악 시장의 변화와 맞물려 음악 정보 검색의 중요한 이슈로 주목받고 있다.

세계 음악 시장은 빠르게 변화하고 있다. LP판, 테이프를 거쳐 CD와 같은 저장장치에 의해 소비되던 음악은 다운로드 방식의 소비로 변화했다. 국제 음반산업협회(IFPI)에서는 ‘디지털 음악 보고서(Digital Music Report) 2011’에서 이러한 변화를 음악을 이용하는 방식이 다운로드와 같은 ‘소유’에서 스트리밍과 같은 ‘접근’으로 패러다임 변화를 이루고 있다고 지적하기도 하였다.

그림 1은 전 세계 디지털 음악 시장 규모 추이(左)와 스트리밍 음악 시장 규모의 추이(右)를 나타내었다. 디지털 음악 시장의 성장률은 2011년 10.9%에서 2012년 9.8%, 2013년 5.4%로 성장세가 둔화되는 양상을 보인다. 반면 스트리밍 음악 시장은 2011년 39.0%, 2012년 63.1%, 2013년 51.4%로 빠른 성장세를 보이고 있다.[1]



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

음악 스트리밍 서비스의 성장은 인터넷 네트워크 속도의 향상과 스마트폰, 태블릿, PC 등 여러 단말을 동시에 사용하며 음악 서비스를 이용하는 소비자들의 이용 행태가 영향을 미친 것으로 풀이되기도 한다.

전 세계 주요 음악 스트리밍 서비스 사업자로는 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. 스트리밍 방식의 장점은 다운로드라는 행위 없이 이 모든 음원에 접근할 수 있다는 것인데 이는 사용자들에게 선택의 어려움을 야기한다. 이에 전 세계 주요 스트리밍 서비스는 음악 추천 서비스를 제공한다. 특히 미국의 판도라는 차별화된 추천 서비스로 다른 서비스에 비해 적은 양의 음악 보유에도 불구하고 많은 사용자를 확보하고 있다. 세 글로벌 스트리밍 서비스의 특징은 표 1에 정리되어 있다.

판도라의 추천 시스템은 전문적인 음악 교육을 받은 음악 전문가들이 장르에 따라 차이가 있지만 노래마다 최대 450개에 해당하는 특성 값을 직접 분류한다. 분류된 450개의 특성을 450차원의 좌표로

	스포티파이	판도라	디저
가격	\$4.99 / 月	\$9.99 / 月	\$4.99 / 月
보유 음원	2,000만 곡	100만 곡	3,069만 곡
액티브 이용자	2,400만 명	7,300만 명	1,200만 명
오프라인 서비스	무제한 모바일 다운로드	-	무제한 다운로드

표 1. 주요 글로벌 음악 스트리밍 서비스 업체별 특징[각 사 발표자료]

치환하여 좌표간 거리를 노래간의 유사도로 사용한다. 판도라에서는 유사도를 계산함에 있어 단순 유클리드 거리를 사용하지 않고 사용자의 취향을 반영하여 특성 간 다른 가중치를 둔 유클리드 거리를 사용한다.[2]

2002년에 서비스를 시작한 영국의 Last.fm¹⁾도 음악 추천 서비스를 제공하고 있다. 노래에 초점을 맞춘 판도라의 추천 시스템과 달리 Last.fm은 노래를 듣는 사용자에게 초점을 맞추는데 Audioscrobbler이라고 불리는 플러그인을 이용해 데이터를 수집한다. 사용자가 어떤 음악 플레이어를 사용하든 Last.fm에서 제공하는 플러그인을 사용하면 사용자의 청취 기록을 관리할 수 있다. 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. Last.fm에서는 홈페이지를 통해 2003년부터 지금까지 Audioscrobbler를 통해 재생된 노래들의 누적 재생 횟수는 974억 회 이상이라고 공개했다.

이러한 상업적 서비스 외에도 음악 정보 검색 학회인 ISMIR (International Society for Music Information Retrieval)에서는 매년 MIREX(Music Information Retrieval Evaluation eXchange)라는 음

1) <http://www.last.fm/>

악 유사도 측정과 검색에 대한 대회를 개최하여 활발한 연구가 진행되도록 하고 있다.

텍스트를 이용해 음악을 분석한 연구들이 다방면에서 이루어졌다. 텍스트의 종류에 따라 크게 두 가지로 분류할 수 있는데, 노래를 나타내는 중요한 정보라고 할 수 있는 가사를 사용한 연구와 가사 이외의 다른 텍스트를 사용한 연구가 있다.

가사를 이용해 음악 분석을 시도한 연구들에는 가사에 언어 모델을 적용하여 아티스트간의 유사도를 연구한 [9]와 가사 안에 드러난 라임(Rhyme)을 이용하여 노래의 장르 분류를 시도한 [10], 노래의 분위기를 분류한 [5][6], 노래의 감정 분석을 수행한 [9] 등이 있다.

가사가 아닌 다른 텍스트를 이용한 연구들에서는 사용자들이 노래에 달아놓은 태그나 라디오 신청곡 사연 등을 이용하였다. [16]에서는 사용자들이 노래마다 달아놓은 태그를 수집하고 태그에서 감정을 추출하여 감정 분석을 사용한 추천 시스템을 제안하였다. [3]에서는 국내 라디오 프로그램의 특징 중 하나인 신청곡과 사연을 이용한 음악 추천 시스템을 제안하였다. 비슷한 내용의 사연에는 서로의 신청곡을 추천해도 잘 맞을 것이라는 가설 하에 사연들에 대하여 텍스트 분석을 수행하였다.

[7]은 텍스트에서 감정을 추출할 수 있는 감정 단어 사전을 제안하였다.

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 집중하고자 하여 특별한 언급이 없으면 ‘노래’라는 용어를 대중가요로 한정하여 사용하고자 한다. 본 연구에서 제안하는 추천 시스템에서는 노래의 빠르기와 높낮이, 음계, 가사에 드러난 감정, 음악적 구조를 주요소로 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 설명한다. 4장에서는 본 연구에서 제안한 시스템을 이용한 실험과 성능 평가에 대해 설명한다. 마지막으로 5장에서는 제안한 시스템에 대한 결론을 내리고 향후 연구를 제안한다.

제 2 장

관련연구

2.1. 음악 추천 시스템

추천 시스템은 추천에 사용하는 데이터의 종류에 따라 일반적으로 두 가지 종류로 분류된다. 추천 시스템에서 주로 사용하는 데이터가 사용자인 경우에는 협업 필터링(Collaborative-filtering)기법으로 분류되고, 주로 사용하는 데이터가 추천하고자 하는 대상-아이템인 경우 내용 기반 필터링(Content-based filtering)기법으로 분류된다.

음악 추천 시스템도 다른 추천 시스템과 마찬가지로 협업 필터링 기법과 내용 기반 필터링 기법의 두 가지 기법이 주로 사용된다. 내용 기반 필터링 기법을 사용한 대표적 추천 시스템으로 서론에서 설명했던 판도라의 추천 시스템이 있다. 판도라의 추천 시스템은 음악 게놈 프로젝트(Music Genome Project)라고 하며, 노래가 갖고 있는 각 요소들이 마치 유전자(Genome)처럼 그 노래를 표현하고 유전자간의 비교로 노래간의 유사도를 비교한다. 판도라와 함께 서론에서 설명한 Last.fm의 추천 시스템은 협업 필터링 기법을 사용한다. 국내 주요 음악 서비스의 경우도 대부분 협업 필터링 기법을 사용하는 것으로 알려져 있다.

음악에 대한 내용 기반 필터링 기법과 협업 필터링 기법은 서로 다른 장단점을 갖고 있다.

내용 기반 필터링 방식의 추천 시스템은 음악을 요소로 구분하는 단계에서 어려움이 발생한다. 판도라의 경우 장르에 따라 최대 450

개의 요소로 노래를 표현하는데, 이 450개 요소가 정답이 아니기 때문에 전문가에 따라 500개 혹은 900개의 요소가 더 좋다고 주장할 수 있다. 하지만 철저하게 아이템에 기초하기 때문에 분석만 되어 있다면 아이템의 인기 유무에 상관없이 추천 대상에 꼽힐 수 있다는 장점이 있다. 하지만 뿌리(Root)가 되는 아이템과 비슷한 유형의 아이템만 추천할 수 있다는 한계와 좋아하는 아이템과 비슷하다면 그 아이템 역시 좋아할 것이라는 가설을 문제점으로 갖고 있다.

협업 필터링 기법을 사용하는 추천 시스템은 사용자를 벡터로 표현한다. 벡터의 차원은 각 아이템이 되며 벡터의 값은 ‘좋다/나쁘다’의 이진 값을 나타내기도 하고 별점을 나타내기도 한다. 내용 기반 추천 기법과 달리 사용자를 표현하는 방법이 명확하다는 장점을 갖고 있다. 하지만 만약 n 명의 사용자가 있고 m 개의 아이템이 있다면 추천 시스템은 $O(mn)$ 의 연산을 수행해야하기 때문에 사용자와 아이템의 규모가 커질수록 연산의 복잡성이 빠르게 증가한다. 이와 더불어 사용자와 아이템 양쪽 측면에서 ‘콜드 스타트(Cold Start)’ 문제도 지니고 있다. 신규 사용자가 시스템에 들어왔을 때 이 사용자는 ‘좋다/나쁘다’ 혹은 별점에 대한 기록을 가지고 있지 않기 때문에 적절한 추천을 할 수 없다. 또한 사용자들에게 인기가 없는 아이템은 대부분의 사용자에게 벡터 값으로 0을 갖게 되기 때문에 추천 대상으로 선정될 수 없다.

음악을 대상으로 하는 추천 시스템의 경우에는 협업 필터링 기법과 내용 기반 추천 기법 방식 이외에 다른 접근 방식이 필요하다는 의견이 있다.[3] 맥락 기반 추천(Context-aware)기법이라고 불리는 새로운 접근 방식은 음악 추천에 있어서 맥락이 중요하다고 주장한다.

맥락이란 사용자가 음악을 듣는 환경, 음악을 들으면서 하는 행동, 음악을 함께 듣고 있는 사람 등을 의미한다. 일부 연구자들은 체온이나 심장박동, 기온 등을 맥락에 포함시켰다.[4]

[3]의 연구에서는 국내 라디오 프로그램에 온 사연과 신청곡을 이

용한 추천 시스템을 제안하였다. 사연과 신청곡은 맥락에서 일치할 것이라 가정하고 사연에 대하여 pLSA 기법을 사용하여 문서 간 유사도를 측정하였다. 유사도가 높은 다른 사연을 찾고 찾아진 사연에 해당하는 신청곡을 추천 곡으로 선정하였다. 사연과 신청곡 간의 밀접한 상관관계를 관찰하였으나 제안한 시스템의 추천 성능은 통계적으로 유의미한 결과를 나타내지는 못하였다.

[3]에서는 가장 중요한 맥락 정보로 감정을 뽑았다. 따라서 본 연구에서는 맥락에 맞는 음악 추천을 위해서 노래의 분위기를 추출하였으며 노래에서 분위기를 추출하기 위해 노래의 가사에서 나타나는 감정과 노래의 음악적 구조를 분석하여 유사도 계산에 사용한다.

2.2. 감정 분석

감정(Emotion)과 분위기(Mood)는 비슷해 보이지만 다른 의미를 갖고 있다. 분위기는 감정의 잔재라고 볼 수 있다. 감정을 유발한 자극에 대한 집중이 흩어진 후에 감정은 분위기로 전환된다. 반대로 이야기하자면 분위기는 좀 더 맥락적인 자극에 의해 유발되는 감정 반응이라고 얘기할 수 있다.[16]

본 연구에서는 [16]과 마찬가지로 감정과 분위기를 구분하지 않고 감정이란 단어로 함께 사용한다. 2.2절에서 소개하는 일부 연구에서는 노래를 분위기로 분류하는 작업을 수행하였지만, 노래에서 나타나는 분위기는 듣는 사람의 감정을 일으키는 자극이므로 분위기와 감정을 함께 사용하여도 무리가 없을 것이다.

2.2.1. 노래의 감정 분류

노래의 감정을 분류하려는 연구들이 다각도에서 이루어졌다. [5]에서는 오디오 정보와 가사를 이용해 노래의 감정을 분류하였다. 네 가지 수준의 감정 화남(Angry), 행복(Happy), 슬픔(Sad), 편안(Relaxed) 선정하여 실험을 진행하였는데 가사만 사용하였을 때보다 오디오만 사용하였을 때 정확도가 더 높게 나타났고 오디오와 가사

를 모두 사용하였을 때가 가장 좋은 성능을 나타냈다.

[6]에서도 오디오 정보와 가사를 이용해 노래의 감정을 분류하고자 하였다. 먼저 감정 범주(Category)에 대해 통일된 기준이 없어 감정 분류를 연구하는 것이 어렵다는 점을 지적하였고 기준이 될 수 있는 감정 범주를 만들기 위해 태그 정보를 수집해 비슷한 단어를 묶어가며 18개의 감정 범주를 제안하였다. [5]의 연구와 마찬가지로 오디오, 가사, 오디오+가사의 세 가지 방법에 대한 비교 실험을 수행하였는데 범주에 따라 다른 성능을 나타냈다. 행복(happy)이나 차분(calm)과 같은 범주에서는 오디오를 사용한 분류가 가장 좋은 성능을 보인 반면 낭만(romantic)이나 화남(angry) 같은 범주에서는 가사를 사용하였을 때 가장 좋은 성능을 나타냈다.

본 연구에서는 노래의 감정 분류에 한정된 범주로 감정을 표현할 때 범주의 선정이 시스템의 성능에 현저한 영향을 미친다는 문제점을 고려하여 감정을 범주가 아닌 벡터로 표현하여 벡터 연산을 이용해 감정을 비교하고자 한다.

2.2.2. 감정 단어 사전(Emolex)

텍스트에서 자동으로 감정을 추출하는 것은 유용하다. 텍스트에서 감정 추출을 할 수 있도록 Emolex, ANEW²⁾, SentiWordNet³⁾과 같은 감정 단어 사전이 여러 연구를 통해 제시되었다. 이 중 Emolex는 심리학자 Plutchik이 제시한 인간의 여덟 가지 기본 감정 anger, anticipation, disgust, fear, joy, sadness, surprise, trust에 긍정과 부정을 더해 10차원 수준의 감정 단어 사전을 제공한다. [7] 현재 최신판의 Emolex에는 14,177개의 단어가 들어있으며 각 단어는 10차원의 이진 벡터로 표현된다. 감정을 추출하고 싶은 텍스트를 단어들의 집합(Bag-of-Word)로 간주하고 텍스트 내의 모든 단어들의 감정 벡터를 구하여 구해진 모든 감정 벡터의 합으로 텍스트의 감정

2) Affective Norms for English Words

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

을 나타낸다.

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

(where *Lyrics* = set of words)

[8]은 Emolex를 기반으로 유튜브(YouTube) 동영상에 달린 댓글을 이용해 해당 영상에 대한 감정을 분석하였다. 유튜브에 올라온 동영상을 선택하여 해당 동영상의 댓글을 모아 하나의 텍스트로 구성한 후 Emolex를 이용해 감정 벡터로 나타냈다. 댓글에 나타난 감정은 댓글이 달려있는 동영상에 나타난 감정과 코사인 유사도 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 단어 사전을 확장한 후에 실험을 진행하였으며 단어 사전을 확장하는 알고리즘은 3장에서 자세히 설명한다.

2.3. 노래와 노래 가사

노래는 다양한 메타 데이터를 가지고 있다. 발매년도나 노래 제목, 가수의 이름, 노래가 수록된 앨범은 모두 노래의 메타 데이터라고 부를 수 있다. 가사도 메타 데이터로 사용될 수 있는데 가사는 메타 데이터로써 몇 가지 장점을 가진다.

- 많은 노래의 가사를 웹 검색을 통해 찾을 수 있다.
- 가사는 주관적이지 않은 정보를 담고 있다.
- 가사는 제목이나 가수 이름과 같은 메타 데이터에 비해 더 풍부한 묘사를 담고 있다.

메타 데이터로써 가사의 이런 장점 때문에 가사를 이용한 다양한 연구들이 행해졌다. 가사를 이용해 가수들의 유사도를 계산한 연구 [9]와 가사를 이용해 노래의 장르를 분류하려는 연구[10]가 그 중 하나이다. 2.2.1에서 소개한 것과 같이 노래의 감정을 분류하려는 연구 또한 이루어졌다.[5][6]

본 연구에서도 기존 연구들과 마찬가지로 가사를 매우 중요한 메타 데이터로 보고 가사에서 감정을 추출해 노래간 유사도 계산에 사용하였다. [9]의 연구에서는 노래간 유사도를 검증하기 위한 자료가 없어 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 구조를 유도하여 유사도 분석에 사용한다는 점에서 차별성을 갖는다.

2.4. 노래의 구조

노래의 구조를 이해하는 것은 노래를 의미 단위로 구분할 수 있게 해준다는 점에서 중요하다. 노래의 구조를 분석하면 노래 전체를 듣지 않고 노래의 핵심 부분을 잡을 수 있고, 짧은 후렴구를 이용해 비슷한 노래를 찾는 것도 가능해진다.[11]

노래의 구조를 분석하는 단위는 그 역할과 특징에 따라 Intro, Verse, Chorus, Bridge, Outro로 구성된다. [12]에 따르면 일반적인 노래는 아래 세 가지 구조를 가진다. 편의상 연속된 같은 단위는 생략하였다. 예를 들어 Verse1, Verse2가 이어지는 경우 단순히 Verse로 표기하였다.

ㄱ. Intro, Verse, Chorus, Verse, Chorus, Outro

ㄴ. Intro, Verse, Chorus, Verse, Chorus, Bridge, Chorus, Outro

ㄷ. 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.

[11]의 연구에서는 가장 많이 반복되는 문단을 Chorus로, 반복이 없는 가장 짧은 문단을 Bridge로, 남은 문단을 Verse로 구분하였다.

[13]은 Verse, Chorus를 구분하는 제약조건으로 모든 노래는 최소 2개의 Verse와 3개의 Chorus를 가져야 하며 Verse에 속하는 가사들은 서로 매우 다르며 Chorus에 속하는 가사들은 비슷하다고 보았다.

본 연구에서는 [11][13]의 연구를 참고하여 가사에서 구조를 분석하는 알고리즘을 설계하였다. 웹을 통해 수집된 가사 데이터는 문단 구분에 통일성이 없고 가사가 아닌 내용이 포함된 경우가 있어 구조 분석 알고리즘에 전처리 과정을 포함시켰다. 구조 분석 알고리즘은 3장에서 자세히 설명한다.

제 3 장

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 문법적 정보를 제공한다는 측면에서 구조적(Syntax) 정보라고 구분할 수 있다. 반면에 노래 가사에 담긴 감정이나 노래의 구조는 노래에 담긴 주제, 의미를 표현한다는 측면에서 의미적(Semantic) 정보로 구분할 수 있다.

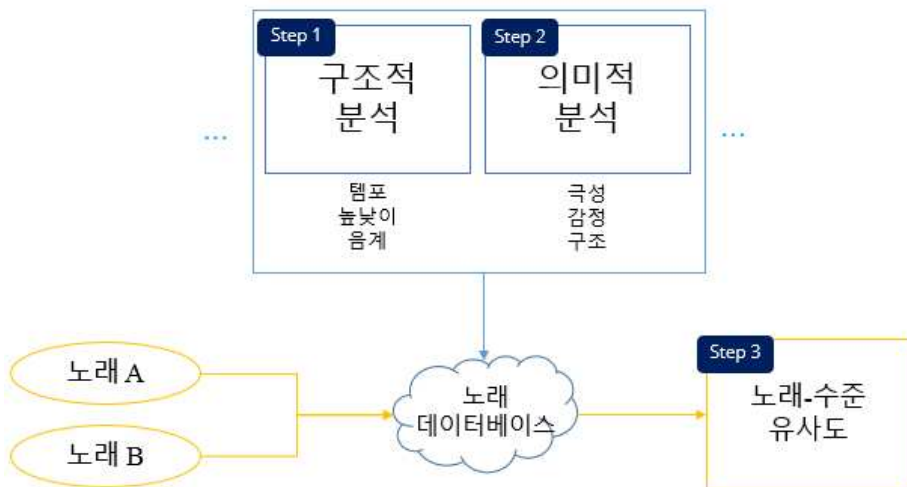


그림 2. 시스템 구성도

그림 2는 전체 시스템 구성 나타낸다. 먼저 각 노래들은 구조적 분석과 의미적 분석 과정을 거쳐 데이터베이스에 저장된다. 시스템에 입력으로 두 노래가 들어오면 데이터베이스를 확인하여 두 노래의 유사도를 계산하여 반환한다. 입력으로 들어온 노래가 데이터베이스에 존재하지 않으면 유사도의 계산 전에 분석을 거쳐 데이터베이스를 갱신한 후에 유사도를 계산한다.

3.1. 오디오 정보

[5][6]의 연구에서 가사만 사용하였을 때보다 오디오 정보를 함께 활용하였을 때 더 좋은 성능을 보였다. [14]의 연구에서도 가사 분석을 주요소로 사용하였으나 춤 출수 있는 정도(Danceability)나 에너지 등의 오디오 요소를 함께 사용하였다. 따라서 본 연구에서도 일부 오디오 정보를 활용하고자 하였다.

4년 이상의 음악 정규 교육을 받은 음악 전문가 세 명의 의견을 수렴하여 오디오 정보 중 노래를 대변할 수 있으면서 측정하기 쉬운 요소인 빠르기(tempo, BPM)와 높낮이(key), 음계(mode)를 사용했다. 빠르기는 필요한 경우 songbpm⁴⁾에서 검색을 하거나 TempoTap⁵⁾을 이용해 간편하게 측정할 수 있다. 높낮이와 음계도 노래 전체를 듣지 않고도 충분히 판단할 수 있다.

3.2. 가사 감정 분석

가사의 감정은 2.2.2에서 설명한 것과 같은 방법으로 추출하였다. 가사를 하나의 텍스트로 간주하여 각 단어들의 감정 벡터의 합을 계산하였다. 계산된 감정 벡터 값은 가사 내의 감정 단어의 개수에 따라 큰 편차를 보이므로 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 정규화하

4) <https://songbpm.com/>

5) <http://www.tempotap.com/>

여 사용하였다. 따라서 모든 노래의 감정 벡터는 아래 두 조건을 만족한다.

$$\neg. \text{Positive} + \text{Negative} = 1$$

$$\neg. \sum_{\text{element} \in e} \text{element} = 1, \text{ where } e = \{e_{\text{anger}}, e_{\text{anticipation}}, \dots, e_{\text{trust}}\}$$

3.3. 노래 구조 분석

[11][13]의 연구를 참고하여 아래와 같이 구조 분석에 대한 근사 규칙을 설립하였다. 기존의 연구는 단락이 구분되어 있었기 때문에 본 연구에 적용하는 것은 무리가 있다. 또한 Bridge를 구분하는 것이 노래 구조를 이해하는 데 있어 중요하지 않다고 판단하여 본 연구의 노래 구조 분석기는 Chorus와 Verse만 구분하였다.

노래 구조 분석기에서는 먼저 각 문장이 몇 번씩 반복되는 지를 세어 가사를 숫자 배열로 표현한다. 배열의 값은 각 위치에 해당하는 문장이 노래에서 몇 번 반복된 문장인지를 나타내는데, 한 번 이상 반복된 문장은 Chorus가 될 가능성이 있다. 이 때 문장 반복의 기준은 유사한 문장으로 확인한다. 두 문장의 각 첫 n 개의 단어 중 m 개 이상의 단어가 겹칠 때 두 문장은 유사하다고 판정한다.

한 번 이상 반복되는 문장들의 문치는 노래의 Chorus일 확률이 매우 높으므로 먼저 Chorus를 찾아낸다. 이 후에 분류되지 않은 문장을 Verse로 분류하면 노래의 Verse와 Chorus를 구분할 수 있다.

노래 구조 분석기는 아래와 같은 알고리즘으로 작동한다.

StructureAnalyzer(String lyric)

```
{  
    Array lyrics = lyric.split("\n");  
    Array count = new int [lyrics.length];  
    for (i = 0; i < lyrics.length; i++)
```

```

    for (j = i; j < lyrics.length; j++)
        if (IsSameString(lyrics[i], lyrics[j]) count[i]++; count[j]++;
String structure = count.toString();
while (hasNonZeroLongestRepeatedSubstring(structure))
    structure.replaceAll(LongestRepeatedSubstring, "Chorus");
sturcture.replaceAll(0, "Verse");
return structure;
}

function IsSameString(String a, String b)
{
    MultiSet sa = a.split(" ").subarray(0, n);
    MultiSet sb = b.split(" ").subarray(0, n);
    return Intersection(sa, sb).length ≥ m;
}

```

분석된 구조는 ‘Verse, Chorus’, ‘Chorus’와 같은 형태로 표현된다. 따라서 두 구조에 대한 유사도를 정의할 필요가 있다. 두 노래 구조의 유사도는 아래와 같은 원칙으로 정의하였다.

- ㄱ. 가장 높은 유사도 : 완전히 동일한 구조
- ㄴ. 높은 유사도 : 노래의 시작과 끝의 구조 단위가 모두 일치
- ㄷ. 보통 유사도 : 노래의 시작이나 끝의 구조 단위 중 하나가 일치
- ㄹ. 낮은 유사도 : 노래의 시작과 끝의 구조 단위가 모두 불일치

3.4 유사도 계산 함수

노래간 유사도는 크게 구조적 분석과 의미적 분석으로 나뉜다. 구조적 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.

$$Sim_{song}(Song_A, Song_B) = \sum_{e \in E} (Weight_e \times Sim_e(Song_A, Song_B))$$

$$E = \{key, mode, tempo, polarity, emotion, structure\}$$

제 4 장

실험

4.1. 실험 환경 및 데이터

본 연구에서 제안한 음악 추천 시스템을 실험하기 위해 사용된 시스템은 총 2대이다. 데이터베이스의 구축 및 운용에 사용된 PC의 사양은 다음과 같다. CPU는 Intel(R) Core(TM) i5-4670 3.40GHz, RAM은 8GB, 운영체제는 CentOS 6.7이다. 데이터베이스는 MySQL 14.14버전을 사용하였고 PHP는 5.3.3 버전을 사용하였다. Java로 작성된 시스템을 운용하여 실험을 수행한 PC의 시스템 사양은 다음과 같다. CPU는 Intel(R) Core(TM) i5-4670 3.40GHz, RAM은 8GB, 운영체제는 Windows 8 Enterprise K 64bit를 사용하였다.

4.1.1. Million Song Dataset

실험에는 Million Song Dataset이 사용되었다. Million Song Dataset은 총 100만곡에 대한 정보를 포함하고 있다. 그림 3처럼 각각의 곡에 대하여 곡의 제목, 아티스트 이름, 발매년도와 같은 메타데이터와 BPM(tempo), 높낮이, 오디오 분석과 같은 음악 정보가 함께 입력되어 있다. 각 곡은 고유한 트랙 ID로 구별된다.[15]

Million Song Dataset의 핵심데이터는 Echo Nest⁶⁾에 의해 제공되었다. Echo Nest는 MIT의 미디어 연구소에서 파생되어 2005년 설

6) <http://the.echonest.com/>

The image shows a file explorer on the left with a tree view containing 'analysis', 'songs', 'metadata', and 'musicbrainz'. The main window displays a table titled 'songs at /analysis/ [msd_summary_file.h5 in F:\data\summary]'. The table has columns: mode, mode_conf., start_of_fa..., tempo, time_signa..., time_signa..., and track_id. The data rows are as follows:

	mode	mode_conf.	start_of_fa...	tempo	time_signa...	time_signa...	track_id
0		0.688	236.635	87.002	4	0.94	TRMMYQ...
1		0.355	148.66	150.778	1	0.0	TRMMKD...
2		0.566	138.971	177.768	4	0.446	TRMMRX...
3		0.451	138.687	87.433	4	0.0	TRMMCH...
4		0.29	506.717	140.035	4	0.315	TRMMWA...
5		0.581	811.799	90.689	4	0.158	TRMMXN...
6		0.627	206.629	101.45	1	0.96	TRMLLR...
7		0.35	212.12	98.02	4	0.982	TRMMBB...
8		0.523	130.479	115.427	1	0.324	TRMMHY...
9		0.48	104.489	124.339	4	1.0	TRMMML...
10		0.59	68.963	62.294	4	0.425	TRMMNS...
11		0.656	165.413	87.319	3	0.953	TRMMXJ1...
12		0.52	291.405	85.34	5	0.581	TRMMCJ...

그림 3. Million Song Dataset의 데이터 형태(HDF5 Format)

립된 음악 지능 및 데이터 플랫폼이다. 개발자나 미디어 서비스 회사에게 유용한 데이터를 제공해주고 있다.

Million Song Dataset은 다른 음악 커뮤니티들의 협력으로 더 풍성한 정보와 트랙 ID를 공유한다. musiXmatch⁷⁾는 각 노래의 가사 정보를 제공한다. 저작권상의 문제로 가사의 원본 형태를 제공하지는 못하지만 가사에 가장 많이 사용되는 5,000개의 단어를 이용해 단어 집합 형태로 노래의 가사 정보를 나타냈다. Last.fm은 사회적 추천 방식으로 계산된 노래간 유사도 정보를 제공한다.

본 연구에서는 Syntax 분석에 필요한 빠르기, 높낮이, 음계 정보만 Million Song Dataset에서 추출하여 사용하였다. 시스템의 성능평가를 위해서 Last.fm에서 제공하는 유사도 정보와 겹치는 노래의 데이터만 수집하였다.

4.1.2. 가사 데이터 수집

Million Song Dataset에서 제공하는 메타 데이터 테이블에서 노래의 제목과 아티스트 이름을 추출하여 가사를 수집하였다. 가사는 LyricsWikia⁸⁾의 API를 사용하여 검색하였다. LyricsWikia에서 제공

7) <https://www.musixmatch.com/>

하는 API를 사용하면 LyricsWikia 페이지 내에 해당 노래의 가사가 존재하는지 아닌지를 알 수 있다. API를 통해 가사의 존재 유무를 확인한 후에 HTML DOM Parser를 이용하여 실제로 가사가 존재하는 URL에 접근하여 가사를 추출하였다.

수집된 가사에는 단락이 구분되어 있는 노래도 있고 구분되어 있지 않은 노래도 있었다. Chorus, x2, Instrumental, Written by XXX와 같이 가사를 축약하거나 가사와 관계없는 단어들이 포함된 가사에 대한 전처리를 수행하였다.

Million Song Dataset의 100만곡 중 LyricsWikia를 통해 206,191곡의 가사가 수집되었다.

4.1.3. 감정 단어 사전 확장(e-Emolex)

감정 단어 사전인 Emolex는 14,117개의 단어를 포함하고 있다. 이를 확장하기 위하여 [16]의 방법을 기반으로 하여 동의어를 수집하였다. [16]의 연구에서는 실험을 통해 감정 단어와 감정 단어의 동의어를 이용하여 자연 언어 처리를 수행하여도 감정 분포가 흐트러지지 않음을 보였다. 핵심 감정을 16개로 구분한 Circumplex모델을 기초로 하여 16개 감정에 대한 동의어를 수집하고 이를 TF-IDF 벡터로 표현한 후에 주성분 분석(Principal Component Analysis)을 수행한 결과 처음의 Circumplex모델과 유사한 2차원 형태의 분포를 얻을 수 있었다.

본 연구에서는 Thesaurus.com⁹⁾을 이용해 동의어를 수집하였다. Thesaurus.com에서 단어를 검색할 때 'Common' 옵션을 활성화하였다. Common 옵션을 활성화하면 두 단어의 사전 상 의미가 비슷하여도 활용 면에서 동 떨어지는 의미의 단어를 제거할 수 있다. Common 옵션을 활성화하고 단어 'confused'의 동의어를 검색하면 distracted, puzzled, thrown의 세 단어가 출력되는데, 이는 그림 4에

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

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

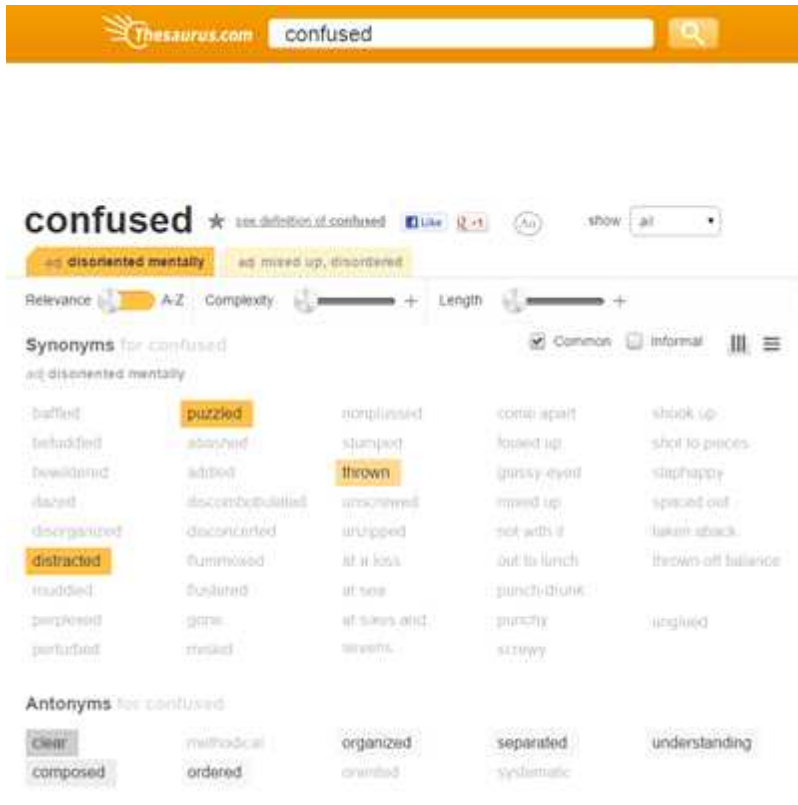


그림 4. Thesaurus.com에서 confused의 동의어를 검색한 화면

나타나있다. 이러한 과정을 통해 6,101개의 동의어를 수집하였다.

각 동의어는 [16]의 연구에서와 마찬가지로 TF-IDF에 따라 가중치를 설정하였다. DF를 계산할 때는 Emolex에 포함된 단어를 문서로 간주하였다. 예를 들어 단어 A가 Emolex 안의 단어 4개의 동의어로 선택되었다면 단어 A의 TF-IDF 값은 $1/4$ 이 된다. 이 때 단어 A의 감정 벡터 값은 동의어 4개의 감정 벡터의 합에 가중치 TF-IDF인 $1/4$ 을 곱한 값을 갖게 된다.

예를 들어 단어 'freedom'은 Emolex의 단어 중 19개 단어의 동의어로 나타났는데, 이 경우 위의 정책에 따라 $1/19$ 의 TF-IDF 값을 갖게 된다. 'freedom'의 동의어로 뽑힌 단어는 다음과 같다.

순위	단어	순위	단어	순위	단어
1	i	11	my	4991	writer
2	the	12	is	4992	motivo
3	you	13	of	4993	bake
4	to	14	your	4994	insist
5	and	15	that	4995	wel
6	a	16	do	4996	santo
7	me	17	on	4997	pe
8	it	18	are	4998	gee
9	not	19	we	4999	colleg
10	in	20	am	5000	kad

표 2. 노래 가사에서 가장 빈번하게 등장하는 단어 순위와 그 단어

{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, 3}이 된다. 이 감정 벡터에 TF-IDF 값인 1/19을 곱하면 ‘freedom’의 감정 벡터 값이 된다.

‘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 1만을 벡터 값으로 갖던 이전 벡터의 성질을 잃게 된다.

총 14,177개의 단어를 포함하고 있는 Emolex는 위의 방법을 통해 1,968개의 단어가 추가되어 총 16,145개의 단어의 e-Emolex로 확장되었다. musiXmatch에서 제공하는 데이터를 이용해 두 감정 사전의 포용력을 비교해 보았다. musiXmatch는 표 2와 같이 노래 가사에서

	Emolex	e-Emolex
평균 일치 단어수 (표준편차)	43.600개 (28.962)	91.303개 (60.535)

표 3. 감정 단어 사전의 성능 비교

가장 빈번하게 등장하는 5,000개 단어의 리스트를 제공한다. Emolex는 5,000개의 단어 중 1,817개를 포함한다. e-Emolex는 5,000개의 단어 중 2,121개를 포함한다. 304개 단어가 추가되어 약 6.08%만큼 포용력이 증가하였다.

e-Emolex를 사용하여 수집된 206,191의 노래 가사에 대하여 감정 분석을 수행하였다. 수행 결과 감정 벡터 값이 0 혹은 NaN 등으로 충분한 감정 단어를 가지지 못한 노래 46,643곡을 제외하고 159,548곡을 실험에 사용하였다.

Emolex와 e-Emolex를 이용하여 가사에 대한 감정 분석을 실시하였을 때 Emolex는 노래별 평균 43.600개의 단어와 일치하였으나 e-Emolex를 사용하였을 때는 평균 91.303개의 단어와 일치하는 성능 향상을 보였다. 표 3에 두 감정 단어사전의 성능을 비교한 결과를 나타냈으며 가사에 포함된 평균 단어의 수는 247.117개, 표준편차는 149.233이다.

4.1.4 노래 구조 분석기 성능 평가

3.3에서 노래 구조 분석기에서 두 문장의 유사성을 판정하기 위하여 아래와 같은 근사 규칙을 설정하였다.

ㄴ. 두 문장에서 각 첫 n 개의 단어 중 m 개 이상의 단어가 겹칠 때 두 문장은 유사하다고 판정한다.

실험에서는 $n = 4, m = 3$ 으로 설정하여 구조 분석을 실시하였다. 수집된 노래 가사에 대하여 구조 분석을 한 결과는 표 3과 같다. 표 3

구조	빈도(개수,%)
VCVVCV	31,829(15.44%)
VCVC	29,875(14.49%)
V	24,222(11.75%)
VCVVCVCV	14,207(6.90%)
VCV	12,688(6.15%)
CVCVC	11,759(5.70%)
VC	10,347(5.02%)
VCVCV	10,267(4.98%)
CVCVCVC	9,617(4.66%)
...(중략)...	
CVCVCVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCV	1
VCVVCVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCV	1

표 4. 노래 구조 분석 결과(V:Verse, C:Chorus)

을 살펴보면 가장 높은 빈도를 나타내는 구조는 VCVVCV와 VCVC이다. 본 연구에서 Bridge를 따로 구분하지 않는 것을 고려하면 VCVVCV와 VCVC의 두 구조는 [12]에서 주장한 전형적인 노래 구조에 속하므로 [12]의 주장과 일치한다고 볼 수 있다.

일부 노래에서 VCVVCVVCVVCVVCVVC와 같이 무의미한 반복 패턴이 관측되었다. 하지만 가사를 직접 확인해본 결과 15분 이상의 긴 노래에 많은 양의 가사를 가진 힙합 곡에 해당하는 등 구조 분석 알고리즘의 품질 평가에 큰 영향을 주지 않는다.

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 측정하고, 측정된 유사도를 기반으로 음악 추천 서비스를 제공한다.

Last.fm에서 제공하는 유사도 데이터는 하나의 시드(Seed)곡에 대

하여 평균 43.75개(표준편차 35.13)의 노래와 해당하는 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 가지는 등 유사도에서 큰 편차를 보였다.

0.9에서 1에 이르는 높은 유사도를 나타내는 곡들을 살펴본 결과 대체로 씨앗곡과 같은 가수의 곡이 비정상적으로 높은 유사도를 갖고 있는 것을 발견하였다. 따라서 시드 곡에 대한 유사도 정보 중 동일한 아티스트의 곡은 제외하고 실험을 수행하였다.

또한 같은 시드 곡에 있는 노래에 대한 유사도 수치도 소수점 둘째자리 이하의 아주 작은 차이를 가지는 경우가 많았다. 따라서 Last.fm에서 제공하는 유사도가 아닌 유사도의 랭킹을 비교하여 성능을 평가하였다.

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

NDCG를 이용하면 두 알고리즘에서 결과로 출력한 랭킹의 성능차이를 계산할 수 있다. NDCG는 DCG를 이용하여 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. 여기서 rel_i 는 i 번째 항목의 가중치를 나타낸다. 본 연구에서는 Last.fm에 의해 순서가 정

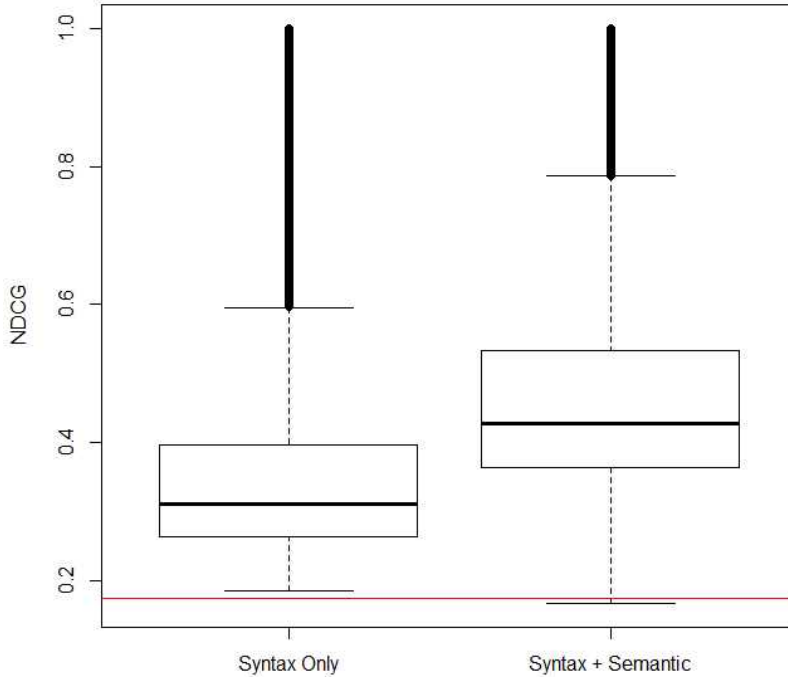


그림 5. Syntax Only와 Syntax+Semantic 시스템 성능 비교

해져있으므로 $rel_i = p - i$ 로 정의하여 계산하였다.

4.3 성능 평가 방법

본 연구에서 제안한 시스템은 159,548곡을 가지고 만들 수 있는 모든 노래 쌍에 대하여 유사도를 계산할 수 있다. 하지만 Last.fm에서 제공하는 유사도 정보가 없으면 계산된 유사도에 대하여 성능을 평가할 수 없다. 따라서 성능을 평가할 수 있는 노래는 매우 제한적이다. 실험을 위해 Last.fm에서 제공하는 유사도 테이블을 먼저 읽어 존재하는 유사도 쌍에 대해서만 유사도 계산을 수행하였다.

본 연구에서 제안하는 추천 시스템은 가사를 이용한 의미적 분석

	Syntax Only	Syntax+Semantic
평균	0.3593	0.4702
표준편차	0.1462	0.1529
중앙값	0.3118	0.4287

표 5. Syntax Only와 Syntax+Semantic 시스템 성능 비교

이 유사도 계산에 있어 유의미하다는 것을 보이는 것이다. 따라서 비교를 위해 구조적 요소만 사용한 시스템과 구조적 요소와 의미적 요소를 모두 사용한 시스템의 두 시스템을 이용해 실험을 수행하였다. 시뮬레이션 결과 가장 성능이 좋았던 가중치를 실험에 사용하였다. 구조적 분석의 경우 빠르기, 높낮이, 음계가 1:1:1로 동일한 가중치를 부여했다. 의미적 분석의 경우 극성, 감정, 구조를 1:1:2의 비율로 가중치를 부여하였고 구조적 분석과 의미적 분석은 1:1의 비율로 가중치를 부여했다.

4.4 성능 평가 결과

구조적 분석만 사용한 시스템과 의미적 분석을 함께 사용한 두 시스템은 각각 159,548개의 시드 곡에 대하여 성능을 실험하였다. 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 시스템을 비교하였다.

그림 5는 두 시스템의 NDCG 값의 분포를 나타내는 상자 그림이다. 그림 5의 빨간색 실선은 기준선을 나타내는데 임의로 랭킹을 뽑았을 때의 평균 NDCG의 값을 의미하며 0.1751을 값으로 갖는다. 두 시스템 모두 기준선 보다 나은 성능을 나타냈음을 알 수 있다.

구조적 분석만 사용한 시스템은 159,548의 시드 곡에 대한 NDCG 값의 평균이 0.3593으로 나타났다. 구조적 분석에 의미적 분석을 더한 시스템의 평균 NDCG는 0.4702로 구조적 분석만 사용한 시스템

에 비해 0.1109만큼 높게 나타났다. 표준편차와 중앙값을 포함한 통계치는 표 5에 나타나있다.

두 시스템에서 출력한 NDCG 값들에 대하여 대응 t-검정을 실시하였을 때 t값은 -284.14, 유의확률 p는 $2.2e-16$ 로 0.05보다 작게 나타나 두 시스템은 유의수준 0.05에서 통계적으로 유의미한 차이를 보였다.

제 5 장

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.

구조적 분석 기법은 임의의 랭킹 알고리즘 보다 나은 성능을 보였으며, 의미적 분석 기법을 구조적 분석 기법과 함께 사용하였을 때 가장 좋은 성능을 보였다. 따라서 의미적 분석 기법에 사용된 극성, 감정, 구조가 노래간 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.

또한 음악 추천 시스템의 일부로 구현된 노래 구조 분석기는 기대 이상의 성능을 나타냈다. 음악의 구조는 음악의 유사도 계산 이외에도 다양한 적용 분야를 가지고 있다. 따라서 노래 구조 분석기를 발전시켜 순수한 가사에서 신뢰도 높은 노래 구조를 추론할 수 있다

면 다른 연구에 보탬이 되는 좋은 연구가 될 것이라 기대한다.

감정 단어 사전을 한 단계 더 확장하는 것도 가능할 것으로 보인다. Plutchuk가 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 사용하면 반의어를 활용한 감정 단어 사전의 확장도 가능할 것이라 기대된다.

마지막으로 본 연구에서는 Syntax와 Semantic 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 향상 시킬 수 있는 방법이 될 것이라 기대한다.

참 고 문 헌

- [1] 디지털 음악시장의 판도를 바꾸는 주요 음악스트리밍 서비스와 기술분석, 한국콘텐츠진흥원, 2014년 7월
- [2] Consumer item matching method and system, US 7003515, B1, 2002.05.16
- [3] Hyung, Ziwon, Kibeom Lee, and Kyogu Lee. "Music recommendation using text analysis on song requests to radio stations." *Expert Systems with Applications* 41.5 (2014): 2608-2618.
- [4] Su, Ja-Hwung, et al. "Music recommendation using content and context information mining." *Intelligent Systems, IEEE* 25.1 (2010): 16-26.
- [5] Laurier, Cyril, Jens Grivolla, and Perfecto Herrera. "Multimodal music mood classification using audio and lyrics." *Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on. IEEE, 2008.*
- [6] Hu, Xiao, J. Stephen Downie, and Andreas F. Ehmann. "Lyric text mining in music mood classification." *American music* 183.5,049 (2009): 2-209.
- [7] Mohammad, Saif M., and Peter D. Turney. "Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon." *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Association for Computational Linguistics, 2010.*
- [8] Orellana-Rodriguez, Claudia, Ernesto Diaz-Aviles, and Wolfgang Nejdl. "Mining emotions in short films: user comments or crowdsourcing?." *Proceedings of the 22nd*

- international conference on World Wide Web companion. International World Wide Web Conferences Steering Committee, 2013.
- [9] Logan, Beth, Andrew Kositsky, and Pedro Moreno. "Semantic analysis of song lyrics." *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*. Vol. 2. IEEE, 2004.
- [10] Mayer, Rudolf, Robert Neumayer, and Andreas Rauber. "Rhyme and Style Features for Musical Genre Classification by Song Lyrics." *ISMIR*. 2008.
- [11] Cheng, Heng-Tze, et al. "Multimodal structure segmentation and analysis of music using audio and textual information." *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*. IEEE, 2009.
- [12] Ten Minute Master No 18: Song Structure. *MUSIC TECH* magazine, October, 2003, pp 62 - 63 www.musictechmag.co.uk
- [13] Maddage, Namunu C., et al. "Content-based music structure analysis with applications to music semantics understanding." *Proceedings of the 12th annual ACM international conference on Multimedia*. ACM, 2004.
- [14] Jamdar, Adit, et al. "Emotion Analysis of Songs Based on Lyrical and Audio Features." *arXiv preprint arXiv:1506.05012* (2015).
- [15] Bertin-Mahieux, Thierry, et al. "The million song dataset." *ISMIR 2011: Proceedings of the 12th International Society for Music Information Retrieval Conference*, October 24-28, 2011, Miami, Florida. University of Miami, 2011.
- [16] Fernández-Tobías, Ignacio, Iván Cantador, and Laura Plaza. "An emotion dimensional model based on social tags: Crossing folksonomies and enhancing recommendations."

E-Commerce and Web Technologies. Springer Berlin Heidelberg, 2013. 88-100.

- [17] Järvelin, Kalervo, and Jaana Kekäläinen. "IR evaluation methods for retrieving highly relevant documents." Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2000.

부록 A.

e-Emolex Example

본 연구에서 동의어를 이용해 확장한 e-Emolex에 추가된 일부 감정 벡터들은 다음과 같다.

words	positive	negative	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
growing	0.230769	0.307692	0.076923	0.230769	0.153846	0.076923	0.153846	0.230769	0.076923	0.076923
propose	0.153846	0.076923	0.076923	0.230769	0	0.076923	0.153846	0.076923	0.076923	0.307692
wage	0.222222	0.333333	0.222222	0.222222	0	0.111111	0.222222	0.111111	0	0.333333
up	0.222222	0	0	0.222222	0	0.055556	0.222222	0	0.055556	0.055556
bored	0.555556	0.333333	0	0.222222	0.111111	0	0.444444	0.333333	0.111111	0.111111
prize	0.434783	0.086957	0.086957	0.217391	0.086957	0.043478	0.391304	0.086957	0.086957	0.26087
arise	0.071429	0	0	0.214286	0	0	0	0	0.071429	0
following	0.052632	0.105263	0.052632	0.210526	0	0.105263	0	0.052632	0	0.052632

부록 B.

가사 구조 측정을 위한 문장 반복 횟수 측정의 예

구조 측정을 위해 가사를 숫자 배열로 나타내어야 한다. James Brown의 I don't mind라는 곡은 아래와 같이 '550425050425050420451'라는 숫자 배열로 바뀌어 진다.

```
I don't mind your love
I don't mind the one
You're thinking of
But I know, I know
You're gonna miss me
I don't mind
Your body and soul
I don't mind your
Love growing cold
But I know, I know
You're gonna miss me
I don't mind
This is my song
I don't mind
Goodbye, so long
I know, I know
You're gonna miss me
Yeah
I know, I know
I know, know, know
Know, know, know, know
```

Paragraphizer!

```
5:I don't mind your love
5:I don't mind the one
0:You're thinking of
4:But I know, I know
2:You're gonna miss me
5:I don't mind
0:Your body and soul
5:I don't mind your
0:Love growing cold
4:But I know, I know
2:You're gonna miss me
5:I don't mind
0:This is my song
5:I don't mind
0:Goodbye, so long
4:I know, I know
2:You're gonna miss me
0:Yeah
4:I know, I know
5:I know, know, know
1:Know, know, know, know
550425050425050420451
```

Abstract

Popular Music Similarity Evaluation using Emotion and Structure Analysis on Lyrics

Jaehwan Lee

Computer Science and Engineering

The Graduate School

Seoul National University

People can listen to almost every music without any possession with music streaming services. Ironically it makes harder to choose what to listen to. Music recommendation system helps people to listen. However previous recommendation systems have high computation complexity and do not care about context information. Emotion is one of the most important context information of music. Lyrics can be easily computed with various language processing techniques and even extract emotion of music from itself. We suggest music-level similarity evaluation method using emotion and structure. By our experiments, it is shown that it is important to consider semantic information when evaluate similarity of music.

**keywords : Music Recommendation, Context, Lyrics,
Emotion, Recommender System**

Student Number : 2014-21797