

Automatic Tag Classification from Sound Data for Graph-Based Music Recommendation

Taejin Kim[†] · Heechan Kim^{††} · Soowon Lee^{†††}

ABSTRACT

With the steady growth of the content industry, the need for research that automatically recommending content suitable for individual tastes is increasing. In order to improve the accuracy of automatic content recommendation, it is needed to fuse existing recommendation techniques using users' preference history for contents along with recommendation techniques using content metadata or features extracted from the content itself. In this work, we propose a new graph-based music recommendation method which learns an LSTM-based classification model to automatically extract appropriate tagging words from sound data and apply the extracted tagging words together with the users' preferred music lists and music metadata to graph-based music recommendation. Experimental results show that the proposed method outperforms existing recommendation methods in terms of the recommendation accuracy.

Keywords : Music Recommendation, Automatic Tag Classification, Sound Data

그래프 기반 음악 추천을 위한 소리 데이터를 통한 태그 자동 분류

김 태 진[†] · 김 희 찬^{††} · 이 수 원^{†††}

요 약

콘텐츠 산업의 꾸준한 성장에 따라 수많은 콘텐츠 중에서 개인의 취향에 적합한 콘텐츠를 자동으로 추천하는 연구의 필요성이 증가하고 있다. 콘텐츠 자동 추천의 정확도를 향상시키기 위해서는 콘텐츠에 대한 사용자의 선호 이력을 바탕으로 하는 기존 추천 기법과 더불어 콘텐츠의 메타 데이터 및 콘텐츠 자체에서 추출할 수 있는 특징을 융합한 추천 기법이 필요하다. 본 연구에서는 음악의 소리 데이터로부터 태그 정보를 분류하는 LSTM 기반의 모델을 학습하고 분류된 태그 정보를 음악의 메타 데이터로 추가하여, 그래프 임베딩 시 콘텐츠의 특징까지 고려할 수 있는 KPRN 기반의 새로운 콘텐츠 추천 방법을 제안한다. 카카오 아레나 데이터 기반 실험 결과, 본 연구의 제안 방법은 기존의 임베딩 기반 추천 방법보다 우수한 추천 정확도를 보였다.

키워드 : 음악 추천, 태그 자동 분류, 소리 데이터

1. 서 론

콘텐츠 산업의 꾸준한 성장에 따라 수많은 콘텐츠 중에서 개인의 취향에 적합한 콘텐츠를 자동으로 추천해 주는 연구의 필요성이 증가하고 있다. 특히 음악은 현대 사회에서 가장 많이 유통되는 콘텐츠 중의 하나로 자동 추천 연구가 활발히 이루어지는 대표적인 도메인 중 하나이다.

기존의 음악 추천 관련 연구로 [1]에서는 음악에 대한 사용자의 평점을 이용하여 사용자의 음악 취향을 추론하고 추론된 취향에 따라 새로운 음악을 추천하는 방법을 제안하였다. 또한, [2]와 [3]에서는 협업 필터링(Collaborative Filtering)을 이용한 음악 추천 방법을 제안하였다. 그러나 협업 필터링은 사용자의 선호 이력 또는 소비 이력을 기반으로 추천을 수행하므로 콘텐츠 메타 데이터의 직접적인 이용이 어렵다는 단점이 있다. 이를 극복하기 위하여 [4]에서는 [5]에서 제안한 그래프 임베딩(Graph Embedding) 기법을 이용하여 음악 메타 데이터를 기반으로 음악을 추천하는 방법을 제안하였다. 그러나 소리 데이터(Sound Data)와 같은 비정형 데이터로부터 그래프 임베딩을 수행하기 위해서는 해당 데이터를 범주화하여 정형화하는 일이 필요하다.

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학CT연구센터 지원사업의 연구결과로 수행되었음(IITP-2021-2018-0-01419).

[†] 준 회 원 : 송실대학교 융합소프트웨어학과 석사과정

^{††} 준 회 원 : 송실대학교 융합소프트웨어학과 박사과정

^{†††} 정 회 원 : 송실대학교 소프트웨어학부 교수

Manuscript Received : February 8, 2021

First Revision : June 8, 2021

Accepted : June 18, 2021

* Corresponding Author : Soowon Lee (swlee@ssu.ac.kr)

본 연구에서는 소리 데이터와 음악 메타 데이터 및 사용자별 선호 음악 정보를 함께 이용하여 음악 추천의 정확도를 향상시킬 수 있는 새로운 그래프 임베딩 기반 추천 방법을 제안한다. 본 연구에서 사용하는 그래프 임베딩 방법은 KPRN (Knowledge-aware Path Recurrent Network)[6]에 기반하고 있다. KPRN은 사용자의 선호 음악 정보와 음악 메타 데이터를 그래프로 임베딩한 후 음악간 유사도 계산을 통하여 사용자의 선호에 적합한 새로운 음악을 추천한다. 본 연구에서는 소리 데이터를 추천에 이용하기 위하여 [7]의 연구를 기반으로 소리 데이터로부터 해당 음악에 적합한 범주(Category) 정보를 자동으로 생성하는 분류 모델을 학습한다. 생성된 범주 정보는 메타 데이터에 추가로 태그되어 사용자의 선호 음악 정보와 함께 유사도 계산에 사용된다.

본 논문의 구성은 다음과 같다. 2장에서 관련 연구에 대해 소개하고 3장에서는 제안 방법을 제시한다. 4장에서는 제안 방법과 기존의 추천 방법에 대한 비교 실험 결과를 분석한다. 마지막 5장에서는 결론 및 향후 연구에 대해 기술한다.

2. 관련 연구

음악 추천 관련 대표적인 기존 연구로는 협업 필터링 기반 연구들과 딥러닝 기반 연구들이 있다.

협업 필터링은 사용자의 아이템에 대한 선호도를 기반으로 추천하는 방법으로 사용자 기반(User-based) 협업 필터링[8], 아이템 기반(Item-based) 협업 필터링[9], 잠재 요인(Latent Factor) 협업 필터링[10] 등으로 구분된다. 사용자 기반 협업 필터링은 사용자의 아이템에 대한 평점으로부터 사용자 간의 유사도를 계산한 후 이를 바탕으로 새로운 아이템에 대한 평점을 예측하여 추천하는 방법이다. 사용자 간 유사도는 평점을 벡터로 표현한 후 코사인 유사도(Cosine Similarity)나 피어슨 상관관계수(Pearson Correlation Coefficient) 등을 통하여 계산된다. 아이템 기반 협업 필터링은 아이템 간의 유사도를 통하여 사용자의 아이템에 대한 평점을 예측하고 추천을 하는 방법이다. 마지막으로 잠재 요인 협업 필터링은 행렬 분해(Matrix Factorization) 등을 이용하여 사용자와 아이템 간의 관계 정보를 분해하거나 축소하여 잠재 요인을 찾는 방법이다. 협업 필터링은 사용자 선호 정보 외에 메타 데이터 등 다양한 정보를 사용하기에는 구조적으로 힘들다는 문제가 있으며 콜드 스타트(Cold Start) 문제도 존재한다.

딥러닝 기반 추천 방법은 협업 필터링에 비하여 구조적으로 좀 더 유연하게 다양한 정보를 이용할 수 있으므로 최근 많이 연구되고 있다. 자연어 처리 분야 연구에서 좋은 성능을 보인 Word2vec[11]을 추천 시스템에 사용한 연구로 Item2vec이 제안되었다[12]. Item2vec은 선호 리스트에 존재하는 아이템 목록을 이용하여 학습되기 때문에, 선호 리스트 내의 아이템 동시 존재 여부가 해당 아이템을 벡터화할 때 가장 큰 영향을 미친다. Node2vec[13]은 그래프를 벡터로

Title	Singer	Preferred User	Genre
Eight	IU	A, B	Dance
Night letter	IU	A, B	Ballad
Bye for a moment	Mc The Max	A	Ballad

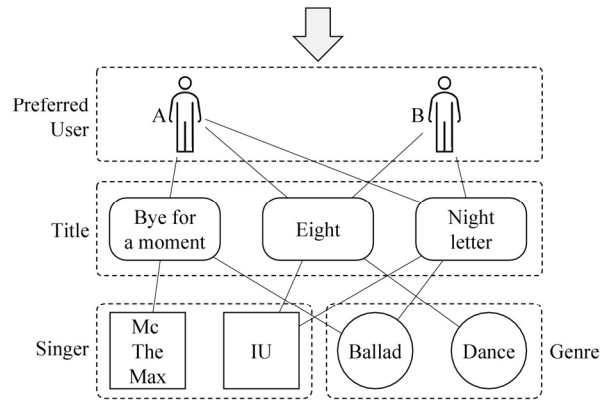


Fig. 1. An example of a graph for Node2vec

표현하는 방법 중 하나로, 이 Node2vec을 이용한 음악 추천 방법들은 벡터화된 음악 노드를 기반으로 벡터 간의 유사도를 계산하여 추천 음악을 결정한다. Fig. 1은 상단의 음악 정보로부터 구성된 Node2vec 그래프의 예시이다. Fig. 1에서 선호 사용자는 사람 모양으로, 음악은 둥근 사각형으로, 가수는 정사각형으로, 장르는 원형으로 표현되었다. Node2vec은 그래프 내 주변 노드와의 연결 관계를 기반으로 노드를 벡터화하므로, 주변 노드 정보가 벡터화에 큰 영향을 준다. Node2vec은 일반적으로 Item2vec에 비하여 좋은 성능을 보이지만 생성해야 하는 노드가 많은 경우에 메모리 부족 문제가 발생할 수 있다는 단점이 있다.

고차원의 데이터를 저차원으로 임베딩하여 잠재 요인을 학습하고자 할 때 오토 인코더(Autoencoder)를 이용한다[14]. 선호 리스트는 각 음악의 포함 여부를 0과 1로 표현하는 벡터로 정의되고, 이 벡터와 오토 인코더를 통하여 해당 리스트들은 벡터화된다. 오토 인코더를 이용한 음악 추천에 대한 연구로 EASE(Embarrassingly Shallow Autoencoders for Sparse Data)라는 모델이 제안되었다[15]. EASE는 오토 인코더를 이용하여 음악 추천 성능을 향상시켰으나 다른 오토 인코더 기반 추천 방법과 마찬가지로 메타 데이터 사용이 구조적으로 힘들다는 문제점이 있다.

KPRN(Knowledge-aware Path Recurrent Network) [6]은 음악 메타 데이터와 사용자의 선호 정보를 지식 그래프(Knowledge Graph)로 표현한 후 음악 간의 지식 경로를 기반으로 음악 간 유사도를 학습한다. 이때 두 음악 간의 경로 설정 시 제한이 없다면 모든 음악끼리 연결될 수 있으므로 최대 경로 길이를 설정하여 일정 거리까지의 경로만 연결한다. 이 경로는 유사도가 1로 정의되는 긍정 경로(Positive Path)로 정의되며 각 경로의 양 끝은 음악 노드이다. 반대로 서로 연결되지 않은 경로는 부정 경로(Negative Path)로 정의된

다. 두 음악 간 유사도는 이 두 경로를 임베딩하여 다층 LSTM을 기반으로 선호 여부에 대한 이진 분류 문제를 학습하는 형태로 학습된다.

3. 제안 방법

3.1 데이터

본 연구는 카카오 아레나(Kakao Arena)¹⁾에서 제공하는 플레이리스트(Playlist) 데이터와 음악 메타 데이터를 기반으로 제안 모델을 설계하였다. 카카오 아레나 데이터는 총 115,071명의 사용자에게 대하여 사용자별 한 개의 플레이리스트, 그리고 플레이리스트별 최대 200곡, 총 707,989곡에 대한 데이터로 구성되어 있다.

Table 1은 카카오 아레나 플레이리스트의 예시이다. Table 1에서 User Id는 사용자를 구분하기 위한 고유값이며, Playlist Title는 해당 사용자의 플레이리스트 제목이다. Tag List는 해당 플레이리스트에 어울리는 단어를 사용자가 해시 태그(#) 형태로 기입한 단어 목록이며, 태그의 총 개수는 29,160이다. Music List는 플레이리스트에 포함된 음악 ID의 목록이며, Like Count는 해당 플레이리스트를 다른 사용자가 좋다고 평가한 횟수이다.

Fig. 2는 플레이리스트별 수록된 음악의 개수 분포이다. Fig. 2의 가로축은 플레이리스트 내 곡의 수를 의미하고 세로축은 해당 수의 곡을 가진 플레이리스트의 수를 의미한다. 예를 들어 20곡이 들어있는 플레이리스트의 수는 7,346개이며, 200곡이 들어있는 플레이리스트의 수는 3,811개이다. 플레이리스트 당 평균 46개의 음악이 수록되어 있다.

Table 2는 Kakao Arena 음악 데이터의 예시이다. Music Id는 음악을 구분하기 위한 Index이고, Song Name과 Artist는 해당 음악에 대한 제목과 가수이며, Album과 Genre는 해당 음악이 속한 앨범과 장르이다. 제목, 가수, 앨범, 장르 등의 메타 데이터 외에 스펙트로그램(Spectrogram)은 해당 음악에 대한 주파수 데이터이다.

각 음악의 태그는 플레이리스트에 할당된 태그 리스트를 이용하여 설정된다. 예를 들어 Table 1의 'music like traveling' 플레이리스트에 포함되는 525514, 129701 등의 음악은 Rock 태그를 가진다. 사용할 태그는 태그의 수를 기준으로 상위 k 개의 태그이다. 한 음악은 여러 플레이리스트에 속할 수 있는데, 이르면 주어진 여러 태그 중에 그 등장 횟수가 많은 태그를 해당 음악의 태그로 할당한다.

3.2 제안 모델의 개요

본 연구에서는 메타 데이터를 사용하기에 적합한 그래프 임베딩 방법을 기반으로 하여 소리 데이터로부터 해당 음악에 대한 태그 정보를 추출할 수 있는 음악 자동 분류 모듈을

Table 1. Example of Kakao Arena Playlists

User Id	Playlist Title	Tag List	Music List	Like Count
0	music like traveling	Rock	525514, 129701, 13954	71
1	These days, you.	Memory	432406, 675945, 24427	1
...
115070	Favorite Popsong!!	Song Recommendation	26008, 456354, 143548	17

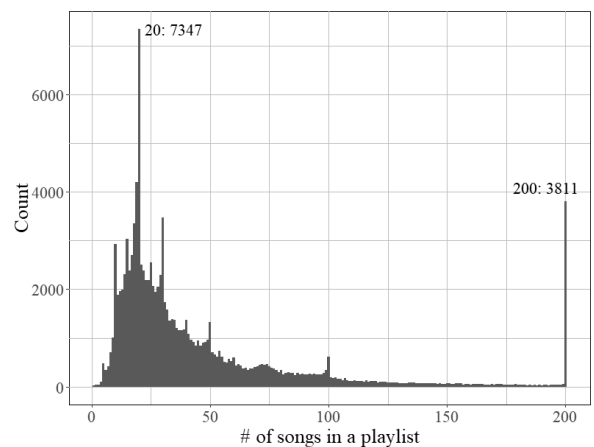


Fig. 2. Distribution of # of Songs in Kakao Arena Playlists

Table 2. Example of Kakao Arena Music Data

Music Id	Song Name	Artist	Album	Genre	Spectrogram
440160	On rainy days	Beast	Fiction And Fact	Ballad	Frequency data over time
221986	Dali, Van, Picasso	Beenzino	12	Rap	Frequency data over time
...

접목한 새로운 음악 추천 모델을 제안한다. Fig. 3은 본 연구에서 제안하는 음악 추천 모델의 구조도이다.

Fig. 3에서 제안 음악 추천 모델의 입력은 사용자의 플레이리스트 정보와 음악 데이터이다. 음악 태그 분류기(Music Tag Classifier)는 스펙트로그램 데이터로부터 주어진 음악의 태그를 분류한다. 분류된 태그는 주어진 음악의 새로운 메타 데이터로 추가되며, 지식 그래프 생성기(Knowledge Graph Generator)는 이 메타 데이터를 기준으로 지식 그래프를 생성한다. 생성된 지식 그래프를 기반으로 KPRN 기반 추천기(KPRN based Recommender)는 KPRN으로 음악 노드 간 유사도를 계산하고, 이 유사도를 기반으로 주어진 플레이리스트의 음악에 대한 추천 음악을 결정한다.

1) arena.kakao.com

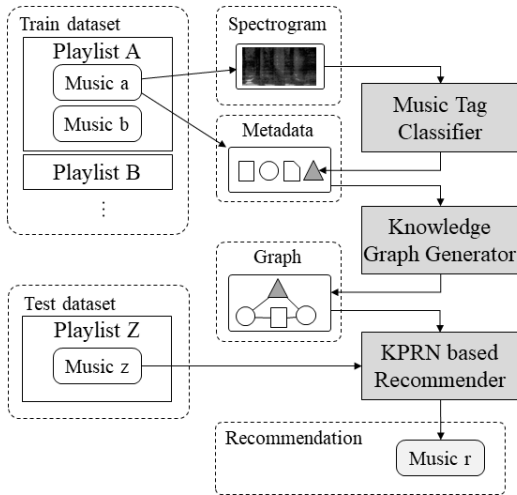


Fig. 3. A Structure of the Proposed Music Recommendation Model

3.3 음악 태그 분류기

음악 데이터를 이용한 음악의 속성 분류를 위하여 [7]에서는 스펙트로그램 데이터로부터 CNN(Convolutional Neural Network) 기반의 분류기를 제시하였다. 스펙트로그램은 시간축에 따른 각 주파수대별 크기를 나타내는 데이터이다. 본 연구에서는 각 주파수 대역은 일종의 특징 차원으로 볼 수 있으며, 주파수의 특성상 각 특징 차원과 시간 차원의 지역 의존성을 동시에 고려하는 CNN보다 시간 차원에 대한 의존성만을 이용하는 LSTM이 분류기에 더 적합할 수 있다고 판단하였다. 실증적인 근거를 위하여 사용할 태그의 수를 50개 ($k=50$)로 고정한 후 전체 실험을 진행해 본 결과, CNN 기반 분류기를 사용한 경우 $NDCG_{10}$ 은 0.4762, LSTM 기반 분류기를 사용한 경우 $NDCG_{10}$ 은 0.4956의 정확도를 기록하였다. 이 결과를 기반으로 본 연구에서는 LSTM 기반 음악 태그 분류기를 제안한다. 이 분류기는 스펙트로그램의 특징을 추출하기 위한 두 개의 LSTM 층과 태그 분류를 위한 단일 완전 연결 층(Fully Connected Layer)으로 구성된다. 완전 연결 층의 입력으로는 LSTM의 마지막 은닉 상태가 주어진다. 스펙트로그램으로부터 음악의 태그가 분류되면 해당 음악의 메타 데이터에 태그 정보가 추가된다.

Fig. 4는 각 태그 단어를 포함하는 Playlist 개수 분포이다. 가장 많이 등록된 단어들은 '기분전환', '감성', '휴식' 등이 있다.

3.4 지식 그래프 생성기

음악 간의 유사도를 계산하기 위하여, 사용자의 선호 리스트와 음악의 메타 데이터 정보를 이용하여 지식 그래프를 생성한다. Fig. 5는 지식 그래프의 예시이며, KPRN을 적용하기 위하여 음악 노드를 중심으로 표현되었다.

Fig. 1의 그래프와 다르게, Fig. 5의 지식 그래프는 지식 정보가 포함되어 있으므로 방향 그래프(Directed Graph)이다.

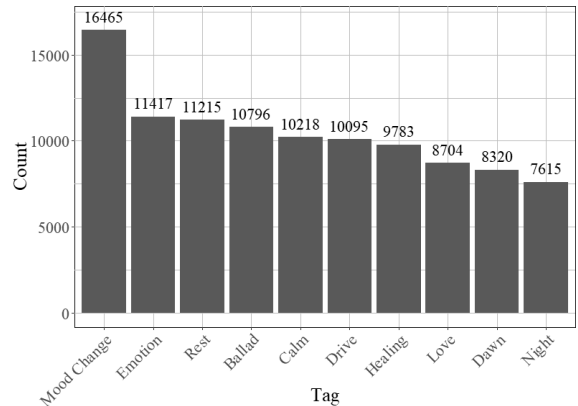


Fig. 4. Distribution of Playlist Counts by Tag

User Preference			
User	Creep	Thinking about you	One more night
Alice	O	O	-
Tony	O	-	O

Music Metadata			
Music	Genre	Artist	Album
Creep	Rock	Radiohead	Pablo Honey
Thinking about you	Rock	Radiohead	Pablo Honey
One more night	Rock	Maroon 5	Overexposed

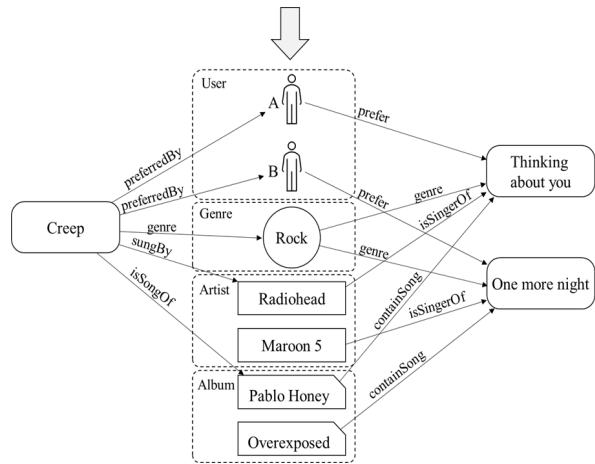


Fig. 5. An Example of a Knowledge Graph for KPRN

3.5 KPRN 기반 추천기

주어진 플레이 리스트와 음악 메타 데이터를 이용하여 음악 간 유사도를 학습하기 위하여 본 연구에서는 KPRN 모델을 사용한다. 음악 간 유사도 학습을 위하여 두 음악 사이의 긍정 경로를 탐색한다. Fig. 5에서 지식 그래프는 'Creep' 음악으로부터 두 음악 'Thinking about you'와 'One more night' 사이의 긍정 경로의 예시이다. 이때 두 음악 사이에 존재할 수 있는 경로는 다수가 존재한다. 긍정 경로가 아닌 모든 음악 간 경로는 부정 경로로 정의된다. KPRN 모

델은, 기존 연구[6]와 동일하게, 단일 LSTM 층과 두 층의 완전 연결 층을 통하여 입력받은 경로들이 긍정 경로인지 부정 경로인지 학습하며, 이때의 확률값은 유사도로 사용된다. 두 음악 사이에는 여러 경로가 존재할 수 있으므로 최종 분류 확률은 경로들의 확률에 대한 합의 형태로 Equation (1)과 같이 정의된다.

$$y_{ab} = \sigma \left(\log \left(\sum_{k=1}^K \exp \left(\frac{s_k}{\gamma} \right) \right) \right) \quad (1)$$

Equation (1)에서 두 음악 a 와 b 의 최종 유사도 y_{ab} 는 각 경로 $path_k$ 에 대한 점수 s_k 의 합으로 정의된다. s_k 는 완전 연결 층의 출력 로짓이며, γ 는 경로 합산 시 각 경로의 가중치를 나타내는 매개변수로 0부터 1사이의 값을 가진다. γ 의 값이 작을수록 큰 값의 영향력이 커지는 Max Pooling 효과를 가지며, 반대로 γ 의 값이 클수록 Min Pooling 효과를 가진다. σ 는 시그모이드 함수를 나타낸다. 본 연구에서는 γ 를 1로 설정하였다.

이렇게 학습된 음악 간 유사도를 기반으로 주어진 음악에 대하여 이 유사도가 가장 높은 음악은 사용자의 선호 리스트와 음악의 소리 데이터 및 메타 데이터를 기반으로 가장 적합한 추천 음악이며, 본 연구에서는 이 음악을 사용자에게 추천한다.

4. 실험

4.1 실험 방법 및 평가 척도

본 연구에서는 제안 방법의 타당성을 평가하기 위하여 카카오 아레나의 플레이리스트 데이터를 이용하여 추천 정확도를 실험적으로 비교하였다. 총 115,071건의 플레이리스트에 대하여 학습, 검증 및 평가 데이터를 60%, 20%, 20%로 구분하여 실험을 진행하였다. 제안 모델을 학습시키기 위하여 KPRN 기반 추천기의 LSTM은 256차원의 은닉 상태를 가지며, KPRN의 입력 경로 임베딩을 위하여 노드 객체는 64차원으로, 노드 타입 및 관계 임베딩은 각각 32차원으로 임베딩되었다. 두 층의 완전 연결 층은 256차원의 은닉 상태를 가지고, 두 개의 확률값을 출력하며, 활성화함수로 ReLU(Rectified Linear Unit)가 사용되었다. 최종 경로의 최대 길이는 6으로 설정되었다. 모델은 Adam 최적화 알고리즘[16]을 통하여 10주기까지 학습되었고, 학습률과 L2 정규화의 영향력은 검증 데이터에 대한 그리드 탐색을 통하여 좋은 성능을 보이는 값으로 결정되었다. 후보 학습률은 0.001, 0.002, 0.01, 0.02이며, 후보 L2 정규화의 영향력은 $1e-5$, $1e-4$, $1e-3$, $1e-2$ 이다.

평가는 평가 데이터로 주어진 각 플레이리스트 내의 앞 50% 음악들을 이용하여 음악을 추천하고, 추천된 음악이 해당 리스트의 나머지 50% 음악 중 하나인지를 판단하는 것으로 진행되었다. 평가 지표로는 NDCG(Normalized Discounted

Cumulative Gain)를 사용하였고, 추천의 순서에 가중치를 두어 추천 정확도를 평가하는 지표이며 Equation (2)와 같이 정의된다.

$$NDCG_p = \frac{DCG_p}{IDCG_p} \quad (2)$$

$$IDCG_p = \sum_{i=1}^p \frac{1}{\log_2(i+1)}$$

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

Equation (2)에서 re_i 는 i 번째 항목의 관련도 점수(Relevance Score)로, 추천 결과가 일치한다면 1, 일치하지 못한다면 0의 값을 가진다. p 는 추천 항목의 개수이다. DCG_p (Discounted Cumulative Gain)는 상위 p 개의 추천 결과들에 대한 관련도 점수의 가중합이며, 추천 순위가 후순위일수록 페널티를 주는 방식이다. $IDCG_p$ (Ideal DCG)는 추천한 항목이 모두 정확히 맞춘 경우의 DCG_p 점수를 의미한다. $NDCG_p$ 는 DCG_p 를 $IDCG_p$ 로 나눈 값으로 정의되며, 0에서 1사이의 값을 가진다.

4.2 모델별 비교 실험 결과

Table 3은 모델별 사용 변수에 따른 추천 정확도 비교 실험 결과이며, 이때 추천 항목의 개수 p 는 10으로 설정되었다. 비교 모델은 사용자 및 아이템 기반 협업 필터링 모델(User-based, Item-based CF) 이외에도 Genre Popular 모델, Item2vec, 오토 인코더 모델이다. Genre Popular 모델은 해당 장르에서 가장 인기 있는 음악을 추천해 주는 Baseline 모

Table 3. Experimental Results by Various Recommendation Models

Model	Variables used	$NDCG_{10}$	Learning time (h:m)
User-based CF	Music List	0.0124	0:13
Item-based CF	Music List	0.0078	0:12
Genre Popular	Genre	0.0536	0:05
Item2vec (Skipgram)	Music List, Artist	0.0575	7:26
Item2vec (Cbow)	Music List, Artist	0.0661	7:51
Autoencoder	Music List, Artist	0.3047	6:42
KPRN-Artist	Music List, Artist	0.4382	5:30
KPRN-Artist/Album	Music list, Artist, Album	0.4622	5:30
KPRN-Artist/Album /Genre	Music list, Artist, Album, Genre	0.4295	5:30
KPRN-Artist/Album /Tag	Music list, Artist, Album, Music Tag	0.4956	5:30

델이다. Item2vec은 학습 방법에 따라 Cbow방법과 Skip-Gram 방법으로 구분하여 실험을 진행하였다. Item2vec (Cbow)은 플레이리스트에 있는 주변 음악들로부터 중심 음악을 유추하는 방식으로 학습하는 반면 Item2vec(Skip-Gram)은 중심 음악을 통해 주변 음악을 유추하는 방식이다. 본 논문의 제안 모델은 KPRN-Artist/Album/Tag로 표현되었다.

Table 3에 따르면 제안 모델은 $NDCG_{10}$ 이 0.4956으로 비교 모델 대비 가장 좋은 추천 정확도를 보였다. KPRN-Artist는 오토 인코더 보다 약 0.1335 정도 더 좋은 결과를 보였다.

KPRN-Artist/Album/Genre는 KPRN-Artist/Album보다 장르 변수를 추가적으로 이용하였지만 오히려 $NDCG_{10}$ 가 약 0.0327 정도 떨어졌다. 이 실험 결과로부터 카카오 아레나 데이터의 장르 정보는 KPRN 기반의 추천에 적합하지 않다는 것을 알 수 있다. 반면 KPRN-Artist/Album/Tag는 장르 변수 대신 음악 자동 분류 모듈에서 추출된 태그 정보를 이용한 모델이다. 이 모델은 장르 변수를 이용한 KPRN-Artist/Album/Genre 보다 약 0.0661 정도 높은 $NDCG_{10}$ 값을 보였다. 그 이유는 카카오 아레나 데이터에서 장르의 총 개수는 224 개로 같은 발라드 음악이더라도 출시연도에 따라 다른 장르로 태그되는 등 지나치게 세분화되어 추천에 적합한 일반화(Generalization)에 오히려 부정적으로 작용하는 것으로 사료된다.

Table 4는 제안 모델의 우수성을 확인하기 위하여 KPRN-Artist/Album/Genre 모델과 제안 모델인 KPRN-Artist/Album/Tag 모델의 카카오 아레나 플레이리스트에 가장 많이 등록된 세 개의 음악에 대한 추천 결과이다. 이 비교는 장르 정보 대신 소리 데이터 기반 태그 정보가 얼마나 타당한지 확인하기 위한 것이다.

Table 4의 결과에 따르면, 두 모델이 같은 음악을 추천하는 경우도 있지만 KPRN-Artist/Album/Tag 모델이 소리 데이터를 이용한 만큼 음향적으로 더욱 유사한 음악을 추천한다는 것을 확인할 수 있다.

Table 4. Example of Classified Tags and Recommended Music

Music	Genre	Tag	KPRN-Artist/Album/Genre	KPRN-Artist/Album/Tag
Night letter	• Ballad • 10-	Ballad	<ul style="list-style-type: none"> • Just like that day • See you on Friday. • It's raining, so... 	<ul style="list-style-type: none"> • Just like that day • See you on Friday • Spring love, not cherry blossoms.
Give me a hug.	• Indi • Rock • 10-	Ballad	<ul style="list-style-type: none"> • In Dream • We could have been friends. • I don't love you. 	<ul style="list-style-type: none"> • I don't love you. • At the same time, you. • How's it going?
Rain	• R&B	Ballad	<ul style="list-style-type: none"> • Every day, every moment. • I don't love you. • accidentally saw 	<ul style="list-style-type: none"> • At the same time, you. • This is the ending. • Stalker

Table 5. Learning Time Comparison Among Various Models

User-based CF	Item-based CF	Genre Popular	Item2vec (Skipgram)	Item2vec (Cbow)
13	12	5	446	471
Auto-encoder	KPRN -Artist	KPRN -Artist/Album	KPRN -Artist/Album/Genre	KPRN -Artist/Album/Tag
402	330	330	330	330

Table 5는 모델별 학습 시간(Learning Time)을 나타낸 표이다.

Table 5의 결과에 따르면, 협업 필터링(User-based, Item-based CF)의 경우 학습 시간은 가장 짧았으나, 추천 정확도는 타 모델에 비하여 많이 떨어지는 것을 확인할 수 있다. Item2vec을 이용한 모델과 오토 인코더를 이용한 모델은 모두 KPRN보다 학습 시간도 길고 추천 정확도도 낮은 것으로 평가되었다. Item2vec, 오토 인코더를 이용한 모델들이 긴 학습 시간을 필요로 하는 이유는 구조적으로 End-to-end 방식으로 한 번의 학습에 많은 데이터가 들어가야만 하기 때문이다.

4.3 음악 자동 분류 실험 결과

음악 자동 분류 과정에서 대상 태그 개수에 따른 추천 결과 정확도는 Fig. 6과 같다. 이 정확도는 모두 KPRN-Artist/Album/Tag 모델을 기반으로 평가되었으며, 가로축은 목표 태그 개수이고 세로축은 $NDCG_{10}$ 값이다.

Fig. 6에 따르면, 실험 결과 대상 태그의 개수가 50개인 경우 $NDCG_{10}$ 값이 0.4956으로 가장 높은 것으로 평가되었다. 태그의 수에 따라 $NDCG_{10}$ 의 값의 변화가 큰 이유는 명확한 음악적 특징을 가지는 태그들이 정확도에 영향을 크게 주기 때문이다. 본 연구에서 제안하는 방법은 태그의 추천 개수에 따라서 음악을 설명하는 메타 데이터가 달라질 수 있다. 즉 태그의 개수에 따라서 다른 메타 데이터를 사용하는 것과

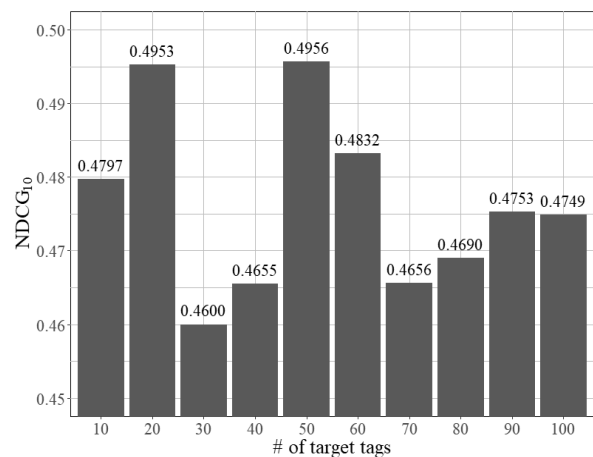


Fig. 6. $NDCG_{10}$ Changes by # of Target Tags

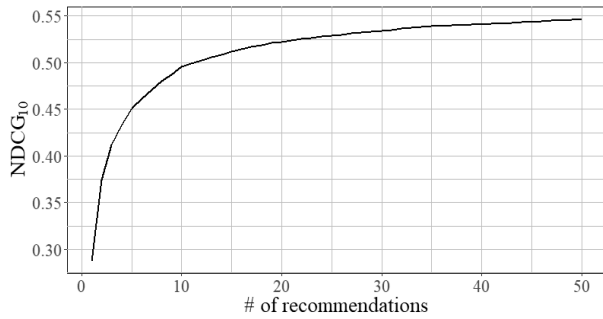


Fig. 7. $NDCG_{10}$ Changes by # of Recommendations

같은 영향을 준다. 따라서 이 표를 통하여 태그의 개수는 $NDCG_{10}$ 에 큰 영향을 준다는 것을 볼 수 있다.

Fig. 7은 $NDCG_{10}$ 이 대상 태그의 개수가 50개일 때 추천 항목의 변화에 따른 $NDCG_{10}$ 값을 비교한 결과이다. 추천 수가 1에서 10까지는 $NDCG_{10}$ 값이 높은 상승폭을 기록하지만 10 이상인 경우에는 상승폭이 급격히 줄어드는 것을 확인할 수 있다.

5. 결론 및 향후 연구

본 연구에서는 그래프 기반 음악 추천을 위한 소리 데이터 기반 태그 자동 분류 방법을 제안하였다. 자세한 소리 데이터를 기반으로 음악의 태그를 자동으로 분류하는 LSTM 기반의 분류기와 이 태그를 통하여 확장된 메타 데이터 기반 KPRN 유사도를 통한 음악 추천기를 제안하였다. 카카오 아레나 데이터를 이용한 비교 실험 결과 제안 방법이 기존의 추천 방법보다 높은 정확도를 보였다.

본 연구에서 제안한 방법은 기존에 존재하는 메타 데이터와 소리 데이터의 태그 분류를 통하여 생성된 태그를 새로운 메타 데이터로 사용하는 방법이다. 실험 결과로부터 이러한 추가적인 메타 데이터가 추천 모델의 성능 향상에 많은 부분 기여한다고 판단할 수 있으며 이 결과로 미루어보아, 음악의 가사 정보 등의 비정형 데이터를 통하여 음악 간 유사도 계산이 더욱 정교해진다면 추가적인 정확도 개선이 가능하다고 사료된다.

아울러 본 연구에서 제안한 방법은 아이টেম에 대한 메타 데이터 또는 비정형 데이터가 풍부한 타 도메인에서도 활용 가능하므로 실험적으로 그 결과를 평가할 필요가 있다.

References

[1] U. Shardanand and P. Maes, "Social information filtering: Algorithms for automating 'word of mouth'," In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp.210-217, 1995.

[2] D. Sánchez-Moreno, A. B. G. González, M. D. M. Vicente, V. F. L. Batista, and M. N. M. García, "A collaborative filtering method for music recommendation using playing coefficients for artists and users," *Expert Systems with Applications*, Vol.66, pp.234-244, 2016.

[3] E. Shakirova, "Collaborative filtering for music recommender system," In *Proceedings of the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering*, pp.548-550, 2017.

[4] D. Wang, X. Guandong, and D. Shuiguang, "Music recommendation via heterogeneous information graph embedding," In *Proceedings of the International Joint Conference on Neural Networks*, pp.596-603, 2017.

[5] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp.1112-1119, 2014.

[6] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T. S. Chua, "Explainable reasoning over knowledge graphs for recommendation," In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp.5329-5336, 2019.

[7] K. Choi, G. Fazekas, and M. Sandler, "Automatic tagging using deep convolutional neural networks," *arXiv preprint arXiv:1606.00298*, 2016.

[8] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pp.43-52, 1998.

[9] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," In *Proceedings of the International Conference on World Wide Web*, pp.285-295, 2001.

[10] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.426-434, 2008.

[11] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Vol.2, pp. 3111-3119, 2013.

[12] O. Barkan and Koenigstein, M. "Item2vec: Neural item embedding for collaborative filtering," In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing*, pp.1-6, 2016.

[13] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.855-864, 2016.

- [14] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," In *Proceedings of the International Conference on Artificial Neural Networks*, pp.52-59, 2011.
- [15] H. Steck, "Embarrassingly shallow autoencoders for sparse data," In *Proceedings of the World Wide Web Conference*, pp.3251-3257, 2019.
- [16] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," In *Proceedings of the International Conference on Learning Representations*, 2015.



김 태 진

<https://orcid.org/0000-0001-9989-2050>
e-mail : dectijay@gmail.com
2021년 송실대학교 융합소프트웨어학과 석사과정
관심분야: (콜론 앞뒤로 반각)



김 희 찬

<https://orcid.org/0000-0002-7564-2230>
e-mail : heathkim@soongsil.ac.kr
2013년 송실대학교 컴퓨터학부(학사)
2015년 송실대학교 컴퓨터학과(석사)
2015년~현 재 송실대학교
융합소프트웨어학과 박사과정

관심분야: 인공지능, 자연어처리, 문서요약



이 수 원

<https://orcid.org/0000-0001-5863-1188>
e-mail : swlee@ssu.ac.kr
1982년 서울대학교 계산통계학과(학사)
1984년 한국과학기술원 전산학과(석사)
1994년 University of Southern
California 전산학과(박사)

2003년 ~ 2004년 한국정보과학회 인공지능연구회 분과위원장
2008년 ~ 2009년 한국정보과학회 논문지(SA) 편집위원장
1995년 ~ 현 재 송실대학교 소프트웨어학부 교수
관심분야: 데이터사이언스, 인공지능