

사용자 청취 습관과 태그 정보를 이용한 하이브리드 음악 추천 시스템

김 현 희 *, 김 동 건 *, 조 진 남 *

A Hybrid Music Recommendation System Combining Listening Habits and Tag Information

Hyon Hee Kim *, Donggeon Kim *, Jinnam Jo *

요 약

본 연구에서는 소셜 음악 사이트에서 사용자가 음악 아이템을 청취한 횟수와 생성한 태그 정보를 혼합하여 음악을 추천하는 시스템을 제안한다. 현재, 상용화된 음악 추천 시스템들은 주로 사용자의 청취 습관과 외부적인 선호도 입력값을 기반으로 음악을 추천하고 있다. 그러나 이 방식은 아직 음악을 청취한 사용자가 많지 않은 새로운 음악이나 청취 정보가 없는 새로운 사용자의 경우 추천하는 데 어려움이 있다. 이 문제를 해결하기 위해서 본 논문에서는 사용자가 선정한 키워드를 아이템에 부여하는 협업 태그로 생성된 태그 정보를 활용하였다. 태그의 의미를 파악하여 감정 표현의 정도에 따라 가중치를 부여한 뒤, 태그 점수와 청취 횟수를 혼합하여 음악 아이템의 선호도를 산출하였다. 이를 기반으로 사용자 프로파일을 생성하고 협업 필터링 알고리즘을 수행하였다. 제안하는 추천 방법의 효율성을 평가하기 위해서, 청취 습관 기반 추천, 태그 점수 기반 추천, 하이브리드 추천 방법의 세 가지 추천 방법에 대해서 정확도, 재현율, 그리고 F-measure를 계산하였다. 실험 결과에 대해 통계적 검증을 시행한 결과, 하이브리드 추천 방법이 다른 두 가지 방식보다 통계적으로 유의한 차이를 보여 성능이 우수한 것으로 나타났다.

▶ Keywords : 하이브리드 음악 추천, 청취 습관, 태그 정보

Abstract

In this paper, we propose a hybrid music recommendation system combining users' listening habits and tag information in a social music site. Most of commercial music recommendation systems recommend music items based on the number of plays and explicit ratings of a song. However, the approach has some difficulties in recommending new items with only a few ratings or recommending items to new users with little information. To resolve the problem, we use tag

• 제1저자 : 김현희

• 투고일 : 2013. 1. 30, 심사일 : 2013. 2. 8, 게재확정일 : 2013. 2. 12.

* 동덕여자대학교 정보통계학과(Dept. of Statistics and Information Science, Dongduk Women's University)

information which is generated by collaborative tagging. According to the meaning of tags, a weighted value is assigned as the score of a tag of an music item. By combining the score of tags and the number of plays, user profiles are created and collaborative filtering algorithm is executed. For performance evaluation, precision, recall, and F-measure are calculated using the listening habit-based recommendation, the tag score-based recommendation, and the hybrid recommendation, respectively. Our experiments show that the hybrid recommendation system outperforms the other two approaches.

▶ Keywords : Hybrid music recommendation, Listening habits, Tag Information

I. 서 론

인터넷이 창출한 새로운 시장에서 음반 산업은 가장 큰 변화를 겪은 산업 중의 하나이다. 많은 사람들이 last.fm[1]과 같은 소셜 음악 사이트에 가입하여 음악을 듣거나 음원을 구입하고, 같은 음악적 취향을 가진 사용자와 친구를 맺고 음악 아이템에 자신이 선정한 키워드를 이용하여 태그를 부여한다. 대부분의 소셜 음악 사이트의 경우, 많은 양의 음악 아이템을 보유하고 있으므로, 사용자에게 적절한 음악 아이템을 추천하는 추천 시스템을 필수적으로 요구하고 있다.

음악 추천은 책이나 영화와 같은 다른 상품 추천과는 다른 특성이 있다. 책이나 영화는 구매 후 사용자들이 직접 점수를 주는 방법으로 상품에 대해 평가를 하지만, 음악은 직접 음악에 점수를 주기보다는 반복적으로 음악을 듣는 경우가 많다. 즉, 청취 횟수와 같은 사용자의 암시적인 정보가 상품에 대한 선호도를 반영한다. 따라서 사용자 프로파일을 생성할 때, 음악 추천 시스템은 사용자의 청취 습관을 기반으로 생성되는 것이 일반적이다[2].

사용자의 청취 습관을 기반으로 사용자 프로파일을 생성할 때 두 가지 문제점이 발생할 수 있다. 먼저 전형적인 cold-start 문제이다. 아직 청취한 사용자가 많지 않은 새로운 음악은 거의 추천이 되지 않고, 또한 청취한 음악이 많지 않은 새로운 사용자에게는 음악을 추천하기가 어렵다. 또한, 사용자가 청취한 음악은 전체 음악 아이템의 크기에 비해 매우 낮은 비율을 차지하므로, 추천 정확도가 떨어지는 경향을 보인다.

본 연구에서는 이러한 문제점을 해결하기 위해 사용자가 음악에 부여한 태그 정보를 활용하였다. 협업 태깅은 미리 정의된 단어나 단어의 계층 구조를 사용하지 않고 직접 선정한

단어를 아이템에 입력하는 기능으로 대부분의 소셜 네트워킹 사이트에서 사용되고 있다. 사용자는 자신이 원하는 방식으로 아이템을 분류하기 위해 태그를 입력하고, 아이템을 검색 및 브라우징하기 위해 다른 사용자가 제공하는 태그를 사용하기도 한다.

음악 사이트의 경우, 약 70%의 태그가 음악 장르와 음악 가를 나타내는데[3], 이는 곧 사용자의 음악적 취향을 직접 반영하고 있다. 또한, 약 10%의 태그는 음악에 대한 의견, 긍정적 혹은 부정적 감정을 나타내는 감정 태그로 이 또한 특정 음악에 대한 선호도를 포함하는 정보로 활용될 수 있다. 따라서 최근에 태그 정보를 활용하여 음악 추천 시스템에 적용하는 연구가 활발히 이루어지고 있다[4].

태그 정보를 기반으로 사용자 프로파일을 생성하면, 태그가 나타내는 의미를 고려할 수 있으므로 단순한 음악 청취 횟수를 사용하여 음악을 추천하는 경우보다 정확도를 향상할 수 있다[5]. 그러나 이 경우 실제 음악을 청취하지 않고도 태그를 입력할 수 있기 때문에 재현율이 낮은 경향을 보인다. 높은 정확도를 유지하면서 재현율을 유지하는 방법으로 두 가지 방식을 혼합해 볼 수 있다. 즉 한 음악 아이템에 대한 사용자의 청취 횟수와 태그 정보를 동시에 고려하여 사용자 프로파일을 생성하는 것이다.

본 연구에서는 사용자의 청취 습관과 태그 정보를 기반으로 사용자 프로파일을 생성하고 협업 필터링 알고리즘을 적용하여 음악을 추천하는 하이브리드 음악 추천 시스템을 개발하였다. 먼저 사용자의 청취 정보와 태그 정보를 모두 수집한 다음 음악을 청취도 하고 태그도 입력한 아이템들만 제한하여 선정하였다. 다음으로 청취 횟수와 태그 점수를 계산하고 두 값을 혼합하여 사용자의 음악 아이템에 대한 선호도를 산출하였다. 마지막으로 협업 필터링 알고리즘을 수행하여 정확도, 재현율, 그리고 F-measure 값을 측정하였다.

시스템의 성능을 평가하기 위해 전통적인 청취 습관 기반

음악 추천 방법과 태그 정보를 사용한 음악 추천 방법의 정확도, 재현율, 그리고 F-measure 값을 측정하였다. 정확도, 재현율, 그리고 F-measure값의 비대칭성을 완화하기 위하여 로그변환하여 일원 분산 분석을 시행하였다. 분석 결과는 제안하는 하이브리드 추천 시스템이 다른 두 시스템에 비해 통계적으로 유의하게 높은 정확도, 재현율, 그리고 F-measure를 보여 우수한 것으로 나타났다.

본 연구의 공헌은 사용자의 청취 횟수와 태그 정보를 고려한 하이브리드 음악 추천 시스템을 개발하고 검증한 것이다. 태그 집합을 감정 태그와 사실 태그로 분류하고, 감정 태그들에 대해서는 감정의 정도에 따른 가중치를 부여하여 태그 점수를 생성하였다. 또한, 청취 횟수를 기반으로 청취 점수를 생성한 뒤, 두 점수를 혼합하여 사용자 프로파일을 생성하였다. 제안한 방법의 효율성을 측정하기 위해서 프로토타입 시스템을 구현하였으며, last.fm으로부터 1,000명의 정보를 추출하여 전통적인 방법과 비교 실험을 수행하였다. 실험 결과는 제안한 추천 방법이 전통적인 청취 습관 기반 추천 방법이나 태그 정보만을 고려한 추천 방법보다 정확도, 재현율, 그리고 F-measure의 평가 모델에 대해 효율적임을 보여주었다.

본 논문은 다음과 같이 구성되어 있다. 2장에서 관련 연구를 살펴보고, 3장에서 본 연구에서 제안하는 하이브리드 음악 추천 방법을 자세히 설명한다. 4장에서는 실험 방법 및 성능 평가 결과를 서술하고 마지막으로 5장에서 결론 및 향후 연구를 제시한다.

II. 관련 연구

일반적인 추천 시스템과 같이 음악 추천 시스템도 내용 기반 추천 방식과 협업 필터링 기반 추천 방식, 그리고 하이브리드 추천 방식으로 나눌 수 있다(6). 내용 기반 추천은 아이템의 내용을 파악하여 사용자가 과거에 선호했던 아이템과 유사한 아이템을 추천하는 방식이다. 음악 아이템은 책이나 영화와는 달리 텍스트를 기반으로 내용을 파악하기 어렵고, 음악을 세그먼트화 하여 특징을 추출하거나 음향 시그널을 분석하는 등의 작업이 필요하다. 따라서 음악 추천 분야에서는 다른 분야보다 널리 사용되지 않았으나 최근 음악을 듣는 환경과 관련하여 연구가 활발히 이루어지고 있다(7).

협업 필터링 기반 추천은 사용자와 유사한 성향이 있는 사용자들을 선정한 다음, 유사한 사용자들이 선호한 아이템을 추천하는 방식이다. 협업 필터링 방식은 효율적인 사용자 프로파일을 생성하는 것이 필수적이다. 외부적으로 1에서 5가

지 선호도를 입력하는 다른 추천 시스템과 달리, 음악 추천 시스템은 사용자가 음악을 청취한 횟수나 구매 내역과 같은 암시적인 정보를 주로 활용한다.

최근 협업 태깅 시스템이 많은 소셜 네트워킹 사이트에 도입됨에 따라, 사용자 프로파일을 생성할 때 태그 정보를 활용하는 연구가 활발히 이루어지고 있다. [8]에서 태그 정보를 기반으로 사용자 프로파일을 생성하면 기존의 청취 횟수 기반 사용자 프로파일보다 유사 사용자를 효율적으로 찾아낼 수 있다는 것을 보여주었다. MusicBox(4)는 소셜 태그를 기반으로 한 개인화된 음악 추천 시스템이다. 사용자, 태그, 그리고 음악 아이템 간의 관계를 찾아내기 위해서, 3-order tensor 모델을 사용하였으며, 추천의 성능이 향상됨을 보여주었다.

하이브리드 추천 시스템은 내용 기반 추천 방법과 협업 필터링 기반 추천 방법을 혼합한 추천 시스템이다. [9]는 오디오 시그널을 기반으로 유사한 음악 아이템을 추출한 다음, 사용자, 그룹, 태그, 그리고 자원 간의 관계를 고려하여 협업 필터링 알고리즘을 수행하였다. Tan의 실험 결과에 따르면 두 가지 방법을 혼합한 Hypergraph 모델을 사용할 경우, 사용자 기반 협업 필터링 추천 방법보다 뛰어난 성능을 보인다.

Foafing the music 시스템(10)은 시맨틱 웹 환경에서 음악에 대한 콘텐츠를 수집하여 추천하는 시스템이다. 즉, 사용자 프로파일, 사용자 청취 습관, 새로운 음악 정보, 콘서트 정보 등을 수집하여 음악가와 밴드와 같은 음악 아이템을 선별한다. 수집된 정보를 기반으로 유사한 음악가를 계산하여 사용자의 선호도를 고려하여 추천하는 방식을 취한다. 음악 아이템이 기본적으로 텍스트 정보가 부족하다는 점을 고려하면, 텍스트 정보를 최대한 활용하였다는 장점을 가지고 있다. 상용화된 음악 추천 시스템들이 널리 사용되고 있다.

Pandora(11)는 잘 알려진 내용 기반 음악 추천 시스템이다. 먼저 사용자가 좋아하는 음악이나 음악가의 이름을 제공하면 시스템은 유사한 음악이나 음악가를 추천한다. 이때, 유사한 음악이나 음악가는 Music Genome 데이터베이스로부터 트랙의 사운드 속성, 악기 배열과 같은 음악적 속성을 사용한다.

반면에 last.fm은 협업 필터링 방식을 사용한다. 사용자가 음악을 들은 횟수, 좋아하는 음악으로 선정한 음악 등을 기반으로 사용자 프로파일을 생성한다. 이 사용자 프로파일을 기반으로 유사한 사용자들을 선정한 다음 유사한 사용자들이 좋아하는 음악 아이템을 추천한다. last.fm은 비교적 좋은 추천 성능을 보이는데, 이는 사용자 커뮤니티와 사용자 생성 데이터가 풍부하기 때문이다(12). 그러나 이 방식은 새롭게 잘 알려지지 않은 음악들은 적절하게 추천되지 않는다는 한계가 있다.

최근 소셜 웹 사이트에서 추천 시스템에 관한 연구가 활발히 이루어지고 있다. 특히 협업 태깅으로 생성된 태그 정보는 사용자가 직접 선택한 단어를 태그로 사용하므로 사용자의 선호도를 구체적으로 반영한다. 따라서 유사한 음악 성향이 있는 사용자를 찾아내는데 유용한 정보로 활용될 수 있고[8], 많은 음악 추천 시스템에서 태그 정보를 다양하게 활용하고 있다[3].

선행 연구에서 태그 정보를 기반으로 사용자 프로파일을 생성하여 음악을 추천할 경우 청취 빈도를 기반으로 사용자 프로파일을 생성하는 경우보다 정확도가 향상됨 보였다[5]. 그러나 재현율에 있어서는 청취 빈도를 기반으로 추천한 경우보다 성능이 떨어지는데 이는 청취 빈도는 반드시 음악을 들은 경우에만 가능하고 태그는 음악을 듣지 않고도 입력이 가능한 데서 기인한 것으로 보인다.

III. 하이브리드 음악 추천 방법

본 장에서는 먼저 제안하는 추천 시스템의 구조를 살펴보고, 사용자 프로파일 생성 방식에 대하여 살펴본 다음, 추천 알고리즘을 제시한다.

1. 시스템 구조

사용자가 특정 음악을 들은 횟수 그리고 음악에 부여한 태그 정보를 웹 사이트로부터 추출하여 사용자 데이터베이스에 저장한다. 사용자 프로파일을 생성하는 데 필요한 정보는 청취 습관 정보와 태그 정보이다. 그림 1의 왼쪽에 보이는 사용자 청취 습관 수집기와 태그 수집기가 각 정보를 수집한다. 특히 태그 정보는 태그의 의미에 따라 점수를 부여하는데 태그 점수 생성기가 이 기능을 담당한다.

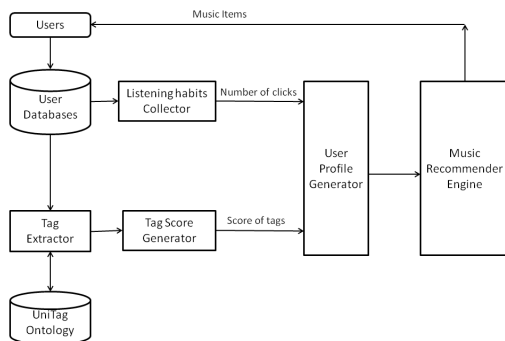


그림 1. 시스템 구조
Fig. 1. System Architecture

태그 점수를 생성하기 위해서는 그림 1의 가장 하단에 보이는 UniTag 온톨로지에 정의된 점수를 사용한다. UniTag 온톨로지는 음악 사이트에서 자주 사용되는 태그들에 대해 긍정적 표현과 부정적 표현으로 나누고, 각 표현의 정도에 따라 강, 중, 약으로 분류하여 가중치 값을 정의한다. 이렇게 생성된 사용자 청취 습관 정보와 태그 정보는 사용자 프로파일 생성기로 입력된다. 사용자 프로파일 생성기는 두 정보에 가중치를 부여하여 두 정보를 모두 고려한 사용자 프로파일을 생성한다.

생성된 사용자 프로파일은 음악 추천 엔진에 입력되고 음악 추천 엔진은 사용자 프로파일을 바탕으로 유사한 성향이 있는 사용자를 추출한 뒤, 사용자가 듣지 않은 음악 아이템 중에서 유사 사용자들이 가장 좋은 점수를 준 음악을 추천한다. 성능 평가를 위해서, 사용자 청취 습관만을 고려한 사용자 프로파일과 태그 점수만을 고려한 사용자 프로파일을 동시에 생성하여 동일한 알고리즘을 적용하여 음악을 추천하는 실험을 하였다.

2. 사용자 프로파일 생성

본 연구에서는 세 가지 종류의 프로파일을 생성하여 추천 엔진에 적용하였다. 첫째는 사용자 청취 습관을 바탕으로 한 사용자 프로파일이고, 둘째는 사용자 태그 정보를 바탕으로 한 사용자 프로파일이며, 셋째는 사용자 청취 습관과 태그 정보를 혼합한 프로파일이다. 각 프로파일을 생성하는 방법을 자세히 알아보자.

2.1 청취 습관 기반 사용자 프로파일

집합 $U1 = \{u_1, u_2, \dots, u_m\}$ 를 모든 사용자 집합이라고 하고, 집합 $I1 = \{i_1, i_2, \dots, i_n\}$ 를 사용자가 한번이라도 청취한 적이 있는 모든 아이템 집합이라고 하자. 각 사용자 u_i 는 본인이 청취한 음악 아이템 리스트 I_{u_i} 를 가지고 있고, 각 아이템에 대한 선호도는 청취 습관 즉 음악 아이템을 청취한 횟수 n 으로 표현된다. 따라서 사용자 u 가 특정 아이템 i 를 청취한 횟수는 삼진 관계 $\langle u, i, n \rangle$ 으로 표현된다.

청취 횟수는 시스템 구조에서 설명한 바와 같이 사용자 데이터베이스로부터 청취 습관 수집기가 추출한다. 청취 횟수는 실수로써 0회에서 최대 150,000회까지 범위가 넓고 숫자가 태그 점수를 바탕으로 한 경우보다 크다. 청취 횟수와 태그 점수를 모두 표준화하여 사용하였다.

2.2 태그 점수 기반 사용자 프로파일

집합 $U2 = \{u_1, u_2, \dots, u_m\}$ 를 모든 사용자 집합이라

고 하고, 집합 $I2 = \{i_1, i_2, \dots, i_n\}$ 를 사용자가 한번이라도 태그를 부여한 모든 아이tem 집합이라고 하자. 각 사용자 u_i 는 본인이 태그를 부여한 음악 아이tem 리스트 I_{u_i} 를 가지고 있고, 각 아이tem에 대한 선호도는 부여한 태그의 점수 s 로 표현된다. 따라서 사용자 u 가 특정 아이tem i 에 부여한 태그 점수는 삼진 관계 $\langle u, i, s \rangle$ 로 표현된다.

여기서 태그 점수는 UniTag 온톨로지에 정의된 태그에 대한 점수를 활용한다. UniTag 온톨로지는 음악 사이트에서 자주 사용되는 태그들에 대해 가중치 점수를 정의한 온톨로지로서 자세한 온톨로지 정의 내용은 [5]에 정의되어 있다. 기본적으로 태그들을 사실 태그와 감정 태그로 분류하고 감정 태그의 경우 감정의 정도에 따라 가중치 값을 정의하였다. 감정 태그는 다시 긍정적 태그와 부정적 태그로 나뉘고, 감정의 정도에 따라 강, 중, 그리고 약으로 분류하여 가중치 값을 적용하였다. 이 가중치 값 역시 청취 습관 기반 프로파일과 함께 사용하기 위해 표준화하여 사용하였다.

2.3 하이브리드 사용자 프로파일

하이브리드 사용자 프로파일은 청취 습관 기반 사용자 프로파일과 태그 점수 기반 사용자 프로파일을 혼합한 프로파일이다. 먼저 집합 $U3 = \{u_1, u_2, \dots, u_m\}$ 를 모든 사용자 집합이라고 하고, 집합 $I3 = \{i_1, i_2, \dots, i_n\}$ 를 아이tem 집합이라고 하면, 집합 $I3 = I1 \cap I2$ 로 구성된다. 즉, 음악 아이tem을 청취도 하고 동일한 음악 아이tem에 태그도 부여한 경우의 아이tem만을 선택한다.

선정된 아이tem에 대한 선호도 $m = \alpha \times n + (1 - \alpha) \times s$ 로 정의되며, 사용자 u 가 특정 아이tem i 에 부여한 혼합 선호도 점수는 삼진 관계 $\langle u, i, m \rangle$ 으로 표현된다. 여기서 임계치 α 는 0.3에서 0.7로 그 값을 변경하여 실험하여 가장 성능이 좋은 0.5로 선정하였다.

3. 음악 추천 알고리즘

기본적인 협업 필터링 방식의 음악 추천 알고리즘의 첫 단계는 유사 사용자 그룹을 선정하는 것이다. 유사한 사용자 그룹을 선정하기 위해 사용자 간의 거리를 구하는 방법은 여러 가지가 있으나 본 연구에서는 두 사용자 간의 공분산과 표준편차 간의 비율을 계산하는 Pearson 상관 계수 유사성 계산법을 사용한다. 사용자 X 와 Y 사이의 거리는 Pearson 상관 계수 계산법을 사용하면 식 (1)과 같다.

$$r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2} \sqrt{\Sigma(Y - \bar{Y})^2}} \quad (1)$$

(1)의 계산법에 따라 사용자 X 와 가장 유사한 사용자를 2명, 4명, 8명, 16명, 32명, 그리고 64명으로 선정한 다음, 사용자 X 가 듣지 않은 음악 아이tem들 중에서 유사한 사용자들이 가장 높은 점수를 준 음악 아이tem을 순서대로 추천한다. 사용된 음악 추천 알고리즘은 알고리즘 1과 같다.

Algorithm 1. 음악 추천 알고리즘

Input: 사용자 프로파일 집합 UP

$\{ \langle u_1, i_1, n \rangle, \dots, \langle u_m, i_n, m \rangle \}$

Output: 추천 음악 아이tem 집합 $RI \{i_1, i_2, \dots, i_n\}$

1. **for all** $y_i \in U$
 compute a similarity s between x and y_i
2. **sort** by similarity
3. **select** top n neighbors
 $U_n \langle - \{u_k, \dots, u_l\} \rangle$
4. $I_c = I_t - I_x$
5. **for all** $u_i \in U_n$
 compute a similarity t between x and $u_i \in U_n$
 for all $i_j \in I_c$
 $preference += t^*pref$
6. **rank by preference**
7. **select** top n items

먼저 음악 아이tem을 추천하고자 하는 목표 사용자 X 와 다른 사용자들의 유사성을 계산한 다음, X 와 가장 유사한 n 명의 사용자를 선정한다(1-3). 음악 아이tem도 전체 음악 아이tem I 중에서 사용자 X 가 청취한 음악 아이tem I_x 을 제외한 음악 아이tem I_c 만을 선정한다(4). 태그 점수 기반 프로파일의 경우는 태그를 부여하지 않은 음악 아이tem만을 선정하고, 하이브리드 사용자 프로파일의 경우에는 청취도 하고 태그도 부여한 음악 아이tem들을 제외한 음악 아이tem을 선정한다.

이제 유사 사용자 그룹으로 선정된 사용자들에 대해서 목표 사용자 X 와의 유사성을 계산한 다음, 이 값을 가중치로

하여 음악 아이템에 대한 선호도를 계산한다(5). 마지막으로 각 음악 아이템에 대한 선호도 값대로 정렬하여 상위 n개 아이템을 추천한다(6-7).

IV. 성능 평가

본 장에서는 먼저 제안하는 추천 시스템의 성능을 평가하기 위한 실험 방법과 결과를 제시하고, 실험 결과에 대해 통계적 검증을 시행하여 세 가지 추천 방법의 성능을 비교하였다.

1. 평가 모델

시스템의 성능을 평가하기 위해서, 각 프로파일 기반 추천 결과의 정확도, 재현율, 그리고 F-measure를 계산하였다. 실험을 위해 무작위로, 1,000 사용자 정보를 last.fm으로부터 수집하였으며, 대상이 된 사용자들이 청취한 약 18,700 음악가와 사용자가 부여한 약 12,600 태그 집합을 수집하였다. 전체 데이터 집합의 70%를 트레이닝 데이터로, 나머지 30%를 테스트 데이터로 사용하였다.

추천 시스템은 사용자 데이터베이스로 MySQL 5.0을 사용하였으며, 청취 습관 수집기와 태그 점수 생성기는 Apache 웹 서버와 PHP를 이용하여 구현하였다. 추천 엔진과 UniTag 온톨로지, 그리고 태그 추출기는 JDK 6.0을 사용하였다.

추천 시스템의 성능은 추천 아이템의 수와 유사 사용자의 수에 따라 민감하게 달라진다. 이 두 가지 변수의 다양한 효과를 살펴보기 위해서, 유사 사용자 수를 2, 4, 8, 16, 32, 그리고 64명으로 변화시키면서 성능을 측정하고, 추천 아이템의 수는 5, 10, 15, 그리고 20개로 증가시키면서 성능을 측정하였다.

시스템 성능 비교를 위해 선정된 정확도, 재현율, 그리고 F-measure는 다음과 같이 정의된다. 먼저 추천 엔진에 의해 추천된 음악 아이템은 사용자가 실제로 관심이 있거나 (True Positive, TP) 관심이 없는 (False Positive, FP) 두 가지 경우로 나뉜다. 또한, 추천 엔진에 의해 추천되지 않은 음악 아이템도 역시 사용자가 실제로 관심이 있거나 (False Negative, FN) 관심이 없는 (True Negative, TN) 두 가지 경우로 나눌 수 있다.

정확도는 추천 아이템 중에서 사용자가 관심이 있는 아이템의 비율이다. 즉, 식 (2)와 같이 계산된다.

$$\text{정확도 } P = \frac{TP}{TP + FP} \tag{2}$$

재현율은 사용자가 관심이 있는 아이템 중에서 추천 시스템이 실제 추천한 아이템의 비율이다. 즉, 식 (3)과 같이 계산된다.

$$\text{재현율 } R = \frac{TP}{TP + FN} \tag{3}$$

마지막으로 F-measure는 정확도와 재현율을 모두 고려한 가중 조화 평균을 사용하며, 식 (4)와 같다.

$$\text{F-measure } F = \frac{2 \cdot P \cdot R}{P + R} \tag{4}$$

먼저 추천 아이템 수에 따른 정확도 변화를 알아보기 위해 유사 사용자 그룹을 2, 4, 8, 16, 32, 64명 일 때로 나누어 세 가지 모델에 대해서 각각 실험한 다음, 정확도의 평균을 나타낸 실험 결과는 그림 2와 같고, 같은 실험 방법으로 재현율을 나타낸 실험 결과는 그림 3과 같다.

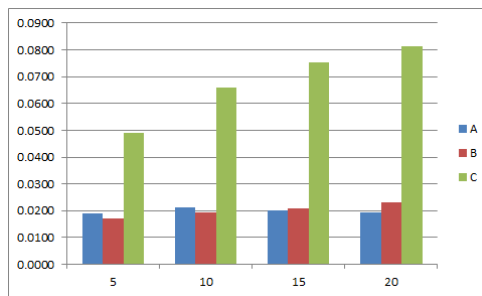


그림 2. 아이템 수의 증가에 따른 정확도
Fig. 2. Precisions for number of recommended items

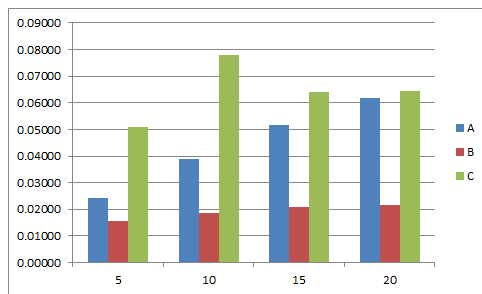


그림 3. 아이템 수의 증가에 따른 재현율
Fig. 3. Recalls for number of recommended items

그림에서 A는 청취 습관 기반 추천, B는 태그 점수 기반 추천, 그리고 C는 하이브리드 추천 방식을 나타낸다. 정확도와 재현율 모두 제안하는 추천 방법 C가 다른 두 가지 방법보다 우수함을 알 수 있다. 또한, 추천 아이템의 수를 증가시키기에 따라 하이브리드 추천 방식의 경우 정확도는 함께 증가하는 경향을 보이지만, 재현율은 추천 아이템의 수에 큰 영향을 받지 않는다. 다른 두 가지 방식은 추천 아이템의 수에 따른 정확도와 재현율이 특별한 경향성을 보이지 않는다고 할 수 있다.

다음으로, 유사 사용자의 수에 따른 정확도와 재현율을 보기 위해 추천 아이템 수를 5, 10, 15, 그리고 20 개로 나누고 유사 사용자 그룹을 2, 4, 8, 16, 32 그리고 64명으로 증가시키면서 실험을 시행하였다. 그림 4는 유사 사용자 수에 따른 정확도를 그림 5는 유사 사용자 수에 따른 재현율을 보여준다.

정확도의 경우는 A, B, 그리고 C 세 가지 방식 모두가 유사 사용자 수를 증가시키면 정확도는 감소하는 경향을 보인다. 반면 재현율의 경우는 각각 극대치를 갖는 사용자 수가 있는데 A와 B의 경우는 8명일 경우, C는 4명일 경우 재현율이 극대치를 갖는다. 또한 정확도와 재현율 모두가 제안하는 C 방법이 다른 두 가지 방법보다 효율적임을 알 수 있다. F-measure의 경우는 정확도와 재현율의 조화 평균이므로 생략하기로 한다.

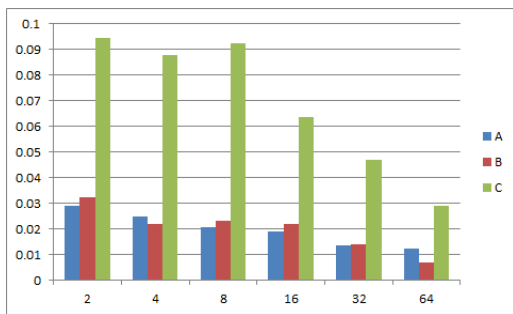


그림 4. 유사 사용자 수의 증가에 따른 정확도
Fig. 4. Precisions for number of similar users

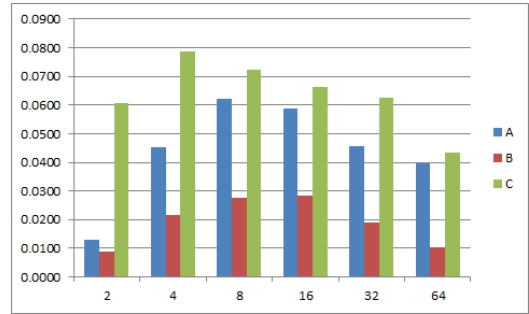


그림 5. 유사 사용자 수의 증가에 따른 재현율
Fig. 5. Recalls for number of similar users

2. 통계적 검증

실험 결과를 토대로 정확도와 재현율 그리고 F-measure에서 유의한 차이가 있다고 판단할 수 있는지를 알기 위하여 통계적 검증을 시행하였다. 토집단을 청취 습관 기반 추천 방법 A, 태그 점수 기반 추천 방법 B, 하이브리드 추천 방법 C의 세 가지로 보고 일원 분산 분석법 (one-way ANOVA)을 실시하였다.

세 그룹의 정확도, 재현율, F-measure의 분포가 비대칭적이어서 로그 변환하여 분산분석을 수행한 후 Tukey의 다중비교법으로 그룹간 차이를 확인하였다. 다중비교법은 Tukey 방법 이외에 Duncan 방법, Scheffe 방법 등이 있는데, 이 중에서 Tukey 방법을 선택한 이유는 귀무가설을 기각할 가능성이 낮은 가장 보수적인 방법으로서 널리 사용되고 있기 때문이다.

표1, 표2, 그리고 표3은 정확도, 재현율, 그리고 F-measure에 따른 Tukey 다중 비교 결과이다. 실험은 유사 사용자 수 2, 4, 8, 16, 32, 그리고 64명 그리고 추천 아이템 수 5, 10, 15, 그리고 20명에 대해 실시하였으며 데이터 수는 각 그룹마다 24개이다. Tukey 다중비교의 결과 유의한 차이를 보이는 그룹들은 다른 알파벳 문자로 표시하였다.

표 1은 세 집단 간의 정확도에 대한 다중비교 결과이다. 방법 1번은 청취 습관 기반 추천, 방법 2번은 태그 점수 기반 추천, 그리고 방법 3번은 하이브리드 추천 방식을 나타낸다. 정확도의 경우 3번이 가장 우수하며 다음으로 1번, 2번 순이다. 1번과 2번은 유의한 차이를 보이지 않고 3번은 1번과 2번에 비해 우수하다고 할 수 있다.

표 1. 정확도에 대한 Tukey 다중 비교 결과
Table 1. Tukey multiple comparison test for Precision

Method	1	2	3	F
Mean of log(perc)	-3.962B	-4.036B	-2.879A	34.27***
Mean precision (SD)	0.020 (0.006)	0.020 (0.009)	0.068 (0.040)	
N	24	24	24	

***: p<0.001

표 2는 세 집단 간의 재현율에 대한 다중 비교 결과이다. 재현율도 정확도와 마찬가지로 제한하는 방법인 3번이 가장 우수하고 다음으로 1번, 2번 순으로 나타나 세 가지 방법이 모두 유의한 차이를 보였다.

표 2. 재현율에 대한 Tukey 다중 비교 결과
Table 2. Tukey multiple comparison test for Recall

Method	1	2	3	F
Mean of log(recall)	-3.285 _B	-4.099 _C	-2.63 _{5^A}	26.80** *
Mean recall (SD)	0.044 (0.023)	0.019 (0.010)	0.093 (0.056)	
N	24	24	24	

***: p<0.001

마지막으로 표 3은 정확도와 재현율의 조화 평균인 F-measure에 대해 다중 비교를 실시한 결과이다. F-measure역시 제한하는 하이브리드 추천 방법이 가장 우수한 성능을 보이며 그 다음이 청취 습관 기반 추천 방법, 마지막이 태그 점수 기반 추천 방법이며 세 가지 방법은 유의한 차이를 보인다.

표 3. F-measure에 대한 Tukey 다중 비교 결과
Table 3. Tukey multiple comparison test for F-measure

Method	1	2	3	F
Mean of log(F-measure)	-3.748 _B	-4.117 _C	-2.894 _A	41.31***
Mean F-measure (SD)	0.024 (0.006)	0.018 (0.008)	0.064 (0.034)	
N	24	24	24	

***: p<0.001

V. 결론 및 향후 연구

본 논문에서는 소셜 음악 사이트에서 사용자의 청취 습관과 태그 정보를 모두 고려하여 음악 아이템을 추천하는 방법을 제안하였다. 대부분의 음악 추천 시스템들이 사용자의 청취 습관을 바탕으로 음악을 추천하고 있으나 사용자 청취 습관을 수집하는데 시간이 걸리고 새로운 사용자에게 음악을 추천하거나 새로운 음악 아이템을 추천하는 데는 어려움이 따른다.

소셜 웹 사이트에서 필수적인 기능으로 자리 잡은 협업 태그 시스템은 사용자가 선정한 단어를 직접 입력하여 음악 아이템을 분류하거나 검색하는데 사용된다. 이 태그 정보를 활용하면 청취 습관 기반 추천 시스템이 갖는 제약점을 극복할 수 있다. 사용자는 직접 음악을 듣지 않고도 태그를 부여할 수 있고, 부여한 태그는 사용자의 음악에 대한 선호도를 구체적으로 반영하기 때문이다.

본 연구에서는 사용자 프로파일을 생성할 때, 청취 습관 정보와 태그 정보를 혼합하여 사용하였다. 청취 습관은 음악 아이템에 대한 청취 횟수를 추출하였으며, 태그 정보는 사용자가 부여한 태그의 의미에 따라 긍정적 태그와 부정적 태그로 분류하고 감정 표현의 강, 중, 그리고 약에 따라 가중치를 부여하였다.

제한한 추천 방법의 효율성을 평가하기 위해서 청취 습관 기반 사용자 프로파일, 태그 정보 기반 사용자 프로파일, 그리고 하이브리드 사용자 프로파일의 세 종류의 프로파일을 생성한 다음, 사용자 기반 협업 필터링 알고리즘을 적용하여 음악 아이템을 추천하였다. 추천 결과는 정확도, 재현율, 그리고 F-measure의 세 가지 평가 지표를 가지고 평가하였으며, 실험 결과가 유의한 차이를 보이는지 검증하기 위해서 일원 배치 분산 분석 (one-way ANOVA)을 실시하였다.

성능 평가 결과 청취 습관과 태그 정보를 모두 고려한 하이브리드 추천 방법이 정확도, 재현율, 그리고 F-measure의 세 가지 측면에서 다른 두 가지 추천 방법과 비교하여 통계적으로 유의한 차이를 보여 성능이 우수한 것으로 나타났다. 이는 사용자의 태그 정보가 추천 시스템의 전형적인 cold-start 문제를 해결하면서 사용자 선호도를 구체적으로 반영하므로 추천 성능을 향상하는데 기여할 수 있다는 것을 알려준다.

현재 성능 향상을 위한 다양한 실험이 계속되고 있다. 특히 음악 아이템 간의 관계를 소셜 네트워크 분석 방법[13]을 이용하여 추출한 다음 음악 아이템에 가중치를 부여하여 추천하는 연구가 진행 중이다. 또한, 본 논문에서 사용된 성능 평가 모델 외에 다양한 모델[14]에서 성능 테스트를 수행할 예정이다.

참고문헌

- [1] Last.fm, <http://www.last.fm/>
- [2] O. Celma, "Music Recommendation and Discovery", Springer, 2010.
- [3] P. Lamere and E. Pampalk, "Social Tagging and Music Information Retrieval", In Proc. of International Conference on Music Information Retrieval, Journal of New Music Research, Vol. 37, No. 2, pp. 101-114, Nov. 2008.
- [4] A. Nanopoulos et al., "MusicBox: Personalized Music Recommendation based on cubic Analysis of Social Tags", IEEE Trans. on Audio, Speech and Language Processing, Vol. 18, No. 2, pp. 1-7, Feb. 2010.
- [5] H. H. Kim, "A Tag-based Music Recommendation Using UniTag Ontology", Journal of The Korea Society of Computer and Information, Vol. 17, No. 11, pp. 133-140, Dec. 2012.
- [6] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Trans. on Knowledge and Data Engineering, Vol. 17, No. 6, pp. 734-749, June, 2005.
- [7] S. Jun, S. Rho, and E. Hwang, "Music Retrieval and Recommendation Scheme Based on Varying Mood Sequences", Int. J. on Semantic Web & Information Systems, Vol. 6, No. 2, pp. 1-16, April, 2010.
- [8] H. H. Kim, J. Jo, and D. Kim, "Generation of Tag-based User Profiles for Clustering Users in a Social Music Site", In Proc. of Asian Conference on Intelligent Information and Database Systems, pp. 51-61, Kaoshiung, Taipei, April, 2012.
- [9] S. Tan et al., "Using Rich Social Media Information for Music Recommendation via Hypergraph Model", ACM Trans. on Multimedia Computing, Communications and Applications, Vol. 7, No. 3, pp. 1-20, Oct. 2011.
- [10] O. Celma, "Foafing the Music: Bridging the Semantic Gap in Music Recommendation", Web Semantics: Science, Services, and Agents on the World Wide Web, Vol. 6, No. 4, pp.250-256, Sept. 2008.
- [11] Pandora, <http://www.pandora.com/>
- [12] M. Kaminskas and F. Ricci, "Contextual Music Information Retrieval and Recommendation: State of the Art and Challenges", Computer Science Review, Vol. 6, No. 2, pp. 89-119, May, 2012.
- [13] C. H. Choi et al., "An Influence Value Algorithm based on Social Network in Knowledge Retrieval Service", Journal of the Korea Society of Computer and Information, Vol. 14, No. 10, pp. 43-53, Oct. 2009.
- [14] S. J. Yu, "A Comprehensive Performance Evaluation in Collaborative Filtering", Journal of the Korea Society of Computer and Information, Vol. 17, No. 4, pp. 83-90, April, 2012.

저 자 소 개



김 현 희
1996: 이화여자대학교
전자계산학과 공학사.
1998: 이화여자대학교
컴퓨터공학과 공학석사.
2005: 이화여자대학교
컴퓨터공학과 공학박사.
현 재: 동덕여자대학교
정보통계학과 조교수
관심분야: 추천 시스템,
소셜 시맨틱 웹
Email : heekim@dongduk.ac.kr



김 동 건
1986: 연세대학교 경영학과 학사.
1990: Virginia Tech 통계학 석사.
1995: Virginia Tech 통계학 박사.
현 재: 동덕여자대학교
정보통계학과 교수
관심분야: 데이터마이닝, 통계계산
Email : dongg@dongduk.ac.kr



조 진 남
1980: 연세대학교
응용통계학과 경제학사.
1982: 연세대학교
응용통계학과 경제학 석사.
1992: Virginia Tech 통계학 박사
현 재: 동덕여자대학교
정보통계학과 교수
관심분야: 실험계획법, 표본조사
Email : jinnam@dongduk.ac.kr