

동시출현정보분석을 이용한 음원 협력추천 서비스에 대한 연구

Collaborative Recommendation for Songs Based on Co-Occurrence Analysis Method

최상희, 대구가톨릭대학교, shchoi@cu.ac.kr

Sanghee Choi, Catholic University of Daegu

협력추천은 이용자가 가지고 있던 지식과 경험 또는 정보를 활용하는데 사용하였던 지식을 토대로 다른 이용자들이 효율적으로 정보를 획득할 수 있도록 지원하는 것을 목적으로 하는 서비스이다. 이 연구에서는 음원 서비스의 기존 이용자들이 구축해놓은 공개앨범에 나타난 정보를 분석하는 과정에 동시인용분석기법을 적용하여 음원을 찾고자 하는 이용자들에게 선호할 만한 음악을 추천해주는 방안을 제시하였다. 동시출현한 정보를 기반으로 구축된 가수 네트워크에서는 연관 가수 집단이 폭 넓게 표현될 수 있었고, 동시출현한 빈도가 높은 상위 곡은 이용자에게 직접적으로 유사정보를 추천하는 방안으로 활용될 수 있는 것으로 나타났다.

1. 서론

대용량 데이터 속에서 이용자들이 원하는 정보를 획득하는데 겪는 어려움을 해소하고자 하는 것은 정보학을 연구하는 많은 연구자들이 관심을 갖는 기본 과제이다. 특히, 이용자들이 원하는 정보를 스스로 찾아내는 것이 점차 달성하기 힘든 과제가 되어 가면서 유사한 정보를 이용한 다른 이용자들의 의견이나 사례를 활용하여 이용자에게 정보를 추천하는 것이 이와 같은 문제점을 해결하는 대안으로 활발히 연구되고 있다.

이용자가 정보를 이용한 기록이나 정보에 대해 내린 평가 또는 이용자가 습득한 정보를 활용하여 생산한 정보물들이 체계적으로 축적되고, 쌍방향으로 이용자들이 소통할 수 있는 정보채널이 다양해짐으로써 이용자의 취향에 맞는 정보를 제공할 수 있는 방법이 활발하게 모색되고 있다. 이와 같은 방안 중 하나가 협력추천시스템(Collaborative recommender system)이다. 협력추천시스템은 다수의 이용자들이 정보

시스템이나 전자상거래시스템, 소셜네트워크 등에서 정보활용 과정을 공유하고 서로 협력하면서 축적한 정보를 기반으로 새로운 이용자에게 개인화된 추천내용을 제공하는 것을 목적으로 하는 시스템이다. 협력추천시스템은 이용자가 원하는 정보를 활용했던 기존 이용자들이 정보를 평가한 결과를 기반으로 정보를 추천해주는 경우와 기존 이용자들이 정보를 활용한 기록을 분석하여 의미 있는 패턴을 도출하여 이용자에게 유사한 정보를 추천해주는 경우가 있을 수 있다(Adomavicius 2005).

이 연구에서는 후자의 경우를 택하여 특정 가수를 중심으로 하여 그 가수를 좋아하는 이용자가 선호할만한 음악이나 가수를 추천하는 방법을 제시하고자 하였다. 이 연구에서 활용한 것은 이용자들이 여러 곡을 선택하여 다른 이용자들을 위해 만든 공개앨범인데, 공개앨범은 앨범생산자가 특정 테마 나 장르, 주제 등 다양한 기준으로 같이 들으면 좋을 만한 음악을 선곡한 곡 목록이다. 이용자에게 음원을 추천하는 기준은 가수이며 특정 가수가 포

함된 공개앨범정보를 수집한 후 각 앨범 내에 그 가수와 동시출현한 가수 및 곡을 분석하여 그 가수를 좋아하는 이용자들에게 선호할만한 가수 또는 분야를 추천할 수 있는 방안을 제시하고자 한다.

분석한 이용자 공개앨범은 음원서비스 멜론에서 제공하고 있는 공개앨범이며 데이터 수집 기준으로 선정한 가수는 ‘정기고’로 가수 ‘정기고’가 포함된 최신 앨범 100개를 수집하여 각 앨범 내에 수록된 가수 및 곡정보를 분석하였다. 따라서 도출된 추천정보는 가수 ‘정기고’를 좋아하는 이용자가 선호할만한 가수와 곡에 해당한다. 이 가수를 선정한 이유는 일반인에게 잘 알려지지 않아 사전 정보가 부족하며 노래의 장르가 모호한 가수이므로 이용자에게 협력추천의 필요성이 대두되는 대표적인 사례라고 판단되었기 때문이다.

분석을 위하여 수집한 데이터는 ‘정기고’가 출현한 최신 공개앨범(2013년 7월 20일 기준) 100개에 포함된 11,384개 곡이며 이중 중복되지 않는 곡 6,113개 곡을 대상으로 분석을 수행하였다.

2. 동시인용기법을 활용한 이용자 공개앨범 분석

특정 가수를 중심으로 연관 가수와 음원을 추천하기 위하여 적용한 기법은 동시인용기법을 적용한 분석기법 중 White(2000)가 제안한 자아중심 인용분석 기법이다. White는 사회네트워크 분석분야 기법을 학술정보네트워크에 적용하여 개인연구자를 자아(ego)로 설정해서 분석하는 네 가지 방법을 제안하였는데 이 중 개인 연구자의 연구에 대하여 분석할 때 가장 직접적으로 정보를 제공하는 방안으로 연구자가 인용한 저자집합을 기반으로 하는 인용 정체성과 연구자와 동시인용된 저자집합을 분석하는 인용이미지가 있다(이재윤 2012). 인용 정체성은 연구자 본인이 생각하

는 자신의 모습으로 주관적 측면이 반영된 것이라면 인용 이미지는 타인의 눈에 비친 연구자의 모습이므로 해당 연구자의 사회적 평가가 반영될 것이라 할 수 있다.

이 연구에서는 위 방법 중 후자인 인용이미지 기법을 적용하여, 공개앨범에 동시 출현한 가수와 곡을 분석한 결과를 기반으로 한 가수에 대해 여러 사람들이 생각하는 이미지를 도출함으로써 이용자가 가수를 기점으로 유사한 음악을 추천받을 수 있도록 시도하였다. 학술문헌의 동시인용분석 단위를 공개앨범에 대응한 사항은 아래 <표 1> 학술문헌 동시인용분석 VS. 공개앨범 동시출현정보분석과 같다.

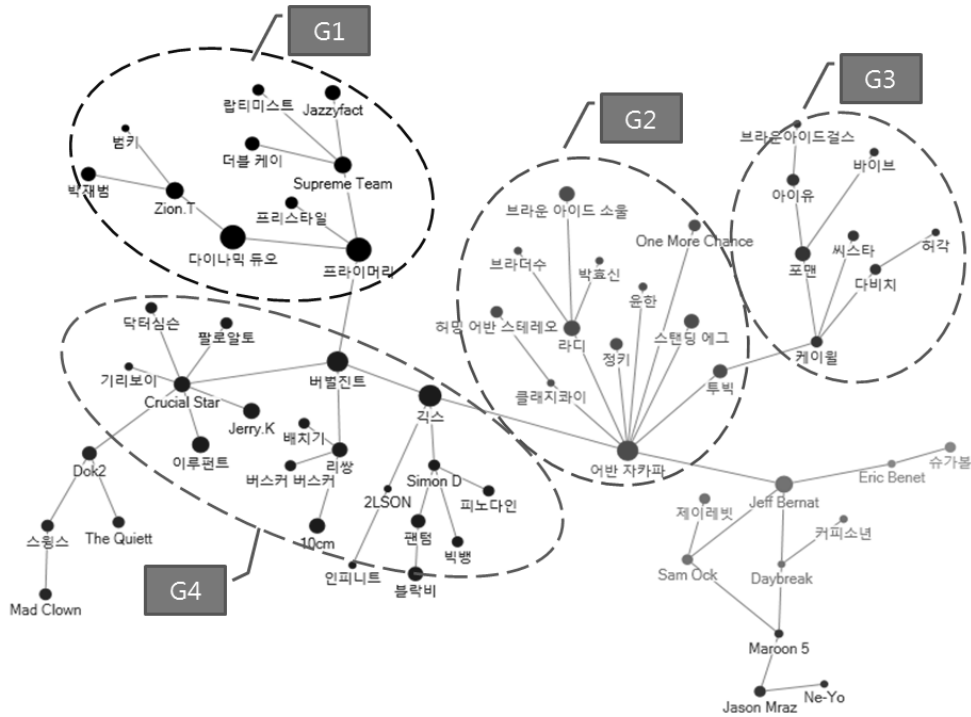
<표 1> 학술문헌 동시인용분석 VS. 공개앨범 동시출현정보분석

구분	학술저자 - 인용 이미지	가수 - 선호이미지
데이터 단위	학술저자가 포함된 문헌	가수가 출현한 공개앨범
분석 단위	저자 - 문헌	가수 - 곡
분석결과로 도출된 주요 정보	특정 저자와 함께 인용되는 연구자	특정 가수와 같이 선곡되는 가수
	특정 저자와 함께 인용되는 문헌	특정 가수와 같이 선곡되는 곡

3. 동시출현 가수 네트워크 분석 및 연관 곡 협력추천

가수 ‘정기고’가 포함된 공개앨범에 동시출현한 가수들의 현황을 분석하기 위하여 총 11,384개 곡을 분석한 결과, 가수 총 875명이 ‘정기고’를 선호하는 사람들에 의하여 함께 선택되어진 것으로 나타났다. 전체 곡 수 대비 가수비율을 분석해보면 한 가수 당 평균 약 13개의 곡이 포함되어 있는 것으로 조사되어 특정 가수 집단에 대한 선호 성향이 나타날 수 있는 것으로 판단되었다.

가장 많이 출현한 가수는 ‘다이나믹 듀오’로 총 235번 출현했으며 다음 순으로 ‘프라이어리’



<그림 1> 빈도 30 이상 동시출현한 가수 네트워크 분석

(227), ‘긱스’(182), ‘버벌진트’ (146), ‘어반 자카파’(143)로 100번이 넘는 가수는 총 5명에 해당하였다.

함께 출현한 주요 가수들을 기반으로 가수 ‘정기고’의 이미지를 도출하기 위하여 동시출현 빈도가 30번이 넘는 가수를 추출하여, 가수들 간의 관계를 Pathfinder network와 PNNC 알고리즘(이재운 2006)을 이용하여 분석한 결과는 <그림 1> 빈도 30 이상 동시출현한 가수 네트워크 분석과 같다.

동시출현 30회를 넘은 가수는 총 63명으로 PNNC 알고리즘으로 클러스터링 한 결과 총 8개의 군집이 생성되었다. 이 중 할당된 수가 5를 초과한 군집을 주요 집단으로 살펴보았을 때 총 4개의 주요 가수집단이 생성되었으며 각 가수들의 동시출현빈도는 노드 크기에 반영하여 중요도를 표현하였다.

각 집단의 특성을 살펴보자면 그룹1(G1)은 힙합 발라드가 중심을 이루고 있었으며 그룹

2(G2)는 R&B와 소울 발라드로 보컬이 강조되는 음악을 하는 가수들이 중심인 것으로 분석되었다. 그룹 3(G3)은 K-POP성향의 가수들로 형성되었으며 그룹 4(G4)은 어반(Urban) 랩과 인디풍의 발라드 음악을 하는 가수들이 중심이 되는 것으로 나타났다. 또한 각 그룹을 연결하는 노드들의 특성을 살펴보면 G1과 G2를 연결하는 노드인 ‘프라이머리’와 ‘버벌진트’는 랩 발라드라는 공통 특성을 가지고 있다. G4와 G2를 연결하는 ‘긱스’와 ‘어반자카파’는 대중적인 인디 밴드라는 측면으로 연결된 것으로 해석될 수 있으며 밴드의 멤버들간 협력 음악활동 사례도 많은 것으로 나타났다. G2와 G3 연결 노드인 ‘투빅’과 ‘케이월’은 남성 파워보컬이라는 측면에서 연결성을 나타내고 있는 것으로 분석되었다. 이와 같이 네트워크로 표현된 각 그룹간의 관계를 훑어나가는 과정에서 가수 ‘정기고’와 연관된 가수들을 추천 받을 수 있고 각 가수간의 연관관계를 유추하는 과정에서 유사 장르로 정보획득 과정을 확장시켜

<표 2> 동시출현 상위 10개 곡

순위	빈도	곡명	가수
1	30	The Lady (Feat. Bumkey & Dok2)	2LSON
2	30	Love (Feat. Bumkey, Paloalto)	프라이머리
3	27	뻘한 멜로디 (Feat. Crush)	Zion.T
4	24	BAAAM (Feat. Muzie Of UV)	다이내믹 듀오
5	24	자니 (Feat. Dynamic Duo)	프라이머리
6	23	Officially Missing You	빅스
7	22	그냥 가요 (Feat. 조현아 Of 어반자카파)	빅스
8	22	만나 (Feat. Zion.T)	프라이머리
9	22	멘트 (Feat. 개코 Of Dynamic Duo)	더블 케이
10	21	입장정리 (Feat. 최자, Simon D)	프라이머리

나갈 수도 있다. 이와 같은 점에서 이용자들이 생산한 공개앨범에 동시출현한 정보를 분석하여 제공하는 것은 특정 가수를 기점으로 연관된 정보를 폭넓게 찾고자 하는 이용자에게 필요한 정보를 협력추천할 수 있는 방안이 될 것이다.

<표 2> 동시출현 상위 10개 곡은 이용자의 공개앨범에 수록된 곡을 분석한 결과로, 가수 '정기고'가 선택된 앨범에 포함된 다른 가수들의 곡 중 가장 많이 출현한 곡이다. 1위를 한 '2LSON'의 경우 네트워크 상에서는 중요한 위치를 차지하지 못하였지만 곡 빈도 분석에서는 '정기고'와 함께 가장 많이 선택된 가수로 나타났다. 이와 같은 결과는 이용자들이 '2LSON'의 곡 중 한 곡을 집중적으로 선택하고 있는 것으로 나타나 곡 단위의 유사성은 높으나 가수 단위의 유사성은 상대적으로 낮은 것으로 분석된다. 반면 몇몇 가수는 네트워크 분석에서의 중요도와 곡 순위 분석의 순위가 비례하고 있는 것으로 나타나 일관된 연관성을 보여주고 있었다. 상위 10곡의 경우 1곡을 제외하고 대부분의 곡이 '정기고'의 대표곡과 유사한 풍의 발라드로 나타나 연관곡으로 추천할 수 있는 것으로 판단되었다.

4. 결론

이 연구에서는 동시인용분석기법을 음원서비스에서 생산된 이용자들의 공개앨범 내 출현한 정보분석에 적용하여 특정 가수를 중심으로

음원을 검색하는 이용자에게 연관된 정보를 추천하는 방안을 제시하고자 하였다. 동시출현한 정보를 기반으로 구축된 가수 네트워크에서는 해당 가수의 이미지가 반영된 가수 집단이 폭넓게 표현되어 연관가수 집단을 훑어 볼 수 있는 것으로 나타났다. 또한 공개앨범에 동시출현한 빈도가 높은 상위 곡은 해당 가수의 대표곡과 유사성이 높은 것으로 분석되어 이용자에게 직접적으로 유사정보를 추천하는 방안으로 활용될 수 있는 것으로 나타났다.

참고문헌

- Adomavicius, G. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734 - 749.
- White, H.D. (2000). Toward ego-centered citation analysis. In B. Cronin, & H.B. Atkins(Eds.). *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 475-496), Medford, New Jersey: Information Today, Inc.
- 이재윤. (2006). 지적구조 분석을 위한 새로운 클러스터링 기법에 관한 연구, *정보관리학회지*, 23(4), 215-231.
- 이재윤. (2012). 폭소노미 연구 문헌에 대한 자아 중심 주제 인용 분석, *정보관리학회지*, 29(4), 295-312.