

# Life-Log기반 협업 필터링을 이용한 음악정보검색 시스템

양영욱\*, 박기남\*\*, 임희석\*\*

\*한신대학교 컴퓨터공학부, \*\*고려대학교 컴퓨터교육학과  
e-mail:spknn@korea.ac.kr

## Music Information Retrieval System Using Collaboration Filtering based on User Life-Log

Yangyoung Wook\*, Kinam Park\*\*, Heuseok Lim\*

\*Div. of computer Engineering, \*\*Dept. of Computer Education

### 요 약

최근 디지털 음악 시장의 비약적인 발전과 더불어 컴퓨터 시스템을 이용한 음악정보검색 시스템의 연구가 활발히 진행되고 있다. 하지만 대부분의 연구가 시스템 구현에 목적이 있고, 시스템에서 행해지는 사용자의 행위에 초점을 맞추기 있기 때문에 사용자들이 정보요구에 대한 만족도가 크지 못한 실정이다. 본 논문에서는 사용자의 상황인지를 통해 구축된 Life-Log를 기반으로 비슷한 성향의 사용자들을 필터링하는 협업필터링 기법을 사용한 음악정보검색 시스템을 제안한다. 실험결과 음악과일에 대한 메타데이터 검색 시스템에 비해 더 나은 사용자 만족도를 보였다.

### 1. 서론

최근 디지털 음악 시장의 비약적인 발전과 웹 기술 및 관련 하드웨어 발전으로 인해 컴퓨터 시스템을 이용한 음악정보검색(Music Information Retrieval) 시스템의 필요성이 대두되고 있다. 하지만 시스템 사용자들은 상업적인 목적의 시스템은 물론이고, 비교적 음악 자료의 조직과 검색이 체계적으로 구현되어있는 디지털 대학 도서관 시스템을 이용하는 것에도 어려움을 느끼고 있다[1]. 이는 대부분의 음악정보검색 시스템 연구가 사용자 중심(user-centered)이 아닌 시스템 중심(system-centered)의 검색 및 적합성 평가 위주로 이루어져 왔으며[2], 주로 음악 검색을 위한 시스템 구현 자체에 목적이 있기 때문이다. 또한 현재 까지도 사용자의 정보요구와 상황을 고려한 사용자 중심의 음악정보검색 시스템에 대한 연구는 전무한 상태이며, 대부분의 연구들이 시스템의 성능비교와 시스템에서 행해지는 사용자의 행위에 초점을 맞추기 있기 때문에 사용자의 정보요

구에 대한 충실한 서비스를 제공해 주지 못한다. 이에 대한 해결방안으로 사용자의 정보요구를 충족시키고, 사용자 성향을 분석하여 사용자가 원하는 정보를 서비스 할 수 있는 추천 시스템(Recommendation system)을 제시할 수 있다.

추천시스템은 사용자에게 사용자의 잠재적인 정보 요구에 부합하는 자료를 자동으로 검색, 제공하는 시스템을 말한다. 즉, 사용자의 패턴이나 기호 등의 사용자 프로파일(user profile)정보를 통해서 사용자에게 맞는 정보를 유추해 서비스하는 것이다. 하지만 사용자의 정보요구에 부합하는 정확한 정보를 유추하기 위해서는 Lifelog 활용, 집단지성(Collective Intelligence)그리고 상황인지와 같은 기술의 활용이 필요하다.

이에 본 연구는 사용자 상황인식(Context-Aware)를 통한 생성된 Life-Log 기반의 협업 필터링(Collaboration Filtering) 기법을 사용한 음악정보검색 시스템을 구현하였다.

## 2. 관련연구

### 2.1 음악정보검색 시스템

음악정보검색 시스템이란 사용자가 컴퓨터 시스템을 이용하여 수많은 노래들 중에 원하는 노래를 검색하는 시스템을 말한다. 음악정보는 일반적으로 텍스트, 멀티미디어로 구성된다. 텍스트 정보는 가수 이름, 작곡가, 노래제목 등과 같은 텍스트 형태로 구성된 정보이며, 멀티미디어 정보는 선율, 장르, 템포 등과 같은 멀티미디어형태의 정보이다. 따라서 일반적인 검색방법을 통해 텍스트 검색과 멀티미디어 메타데이터를 이용한 검색[3]이 가능하다.

지금까지 대부분의 음악정보검색 시스템은 사용자 중심이 아닌 시스템 중심의 검색 및 적합성 평가 위주로 이루어져 왔으며 주로 음악 검색을 위한 시스템 구현을 강조해 왔다. 하지만 음악정보를 검색하는 사용자의 특성상 사용자 정보요구 의도가 중요하기 때문에 이런 식의 연구는 사용자의 검색동기, 검색 결과에 대한 성공, 실패의 정의 그리고 사용자들이 선택한 정보 검색 전략에 대한 정보는 제공해주지 못한다.

### 2.2 음악 추천 시스템

음악 추천 시스템이란 콘텐츠 중심의 음악 검색이 아니라 사용자의 성향이나 상황을 고려해서 사용자 기호에 맞는 음악을 동적으로 제시해 주는 시스템을 말한다[4]. 즉, 사용자에게 사용자의 잠재적인 정보요구에 부합하는 정보를 자동으로 검색하고, 제공하는 시스템이다. 하지만 사용자의 정보요구에 부합하는 정확한 정보를 유추하기 위해서는 Life-Log 활용, 집단지성(Collective Intelligence) 그리고 상황인지와 같은 기술의 활용이 필요하다.

### 2.3 Life-Log

상황인식이란 응용프로그램이 작동되고 있는 환경에 대한 정보를 활용하여 그에 적절한 정보 및 서비스를 사용자에게 제공하는 컴퓨팅기술을 말한다[4]. 이러한 컴퓨팅 기술 및 개인화 장비의 기술이 발전함에 따라 다양한 형태의 로그 수집이 가능해졌고, 이를 통해 사용자는 컴퓨터와 웹을 이용해 자신의 정보를 저장하려는 시도가 많아 졌다. 최근에는

이러한 로그, 즉 컴퓨터 시스템 환경에서 수집한 개인의 라이프로그를 분석하여 사용자의 기호나 성향을 분석하려는 연구가 활발히 진행되고 있다[5].

### 2.4 협업 필터링

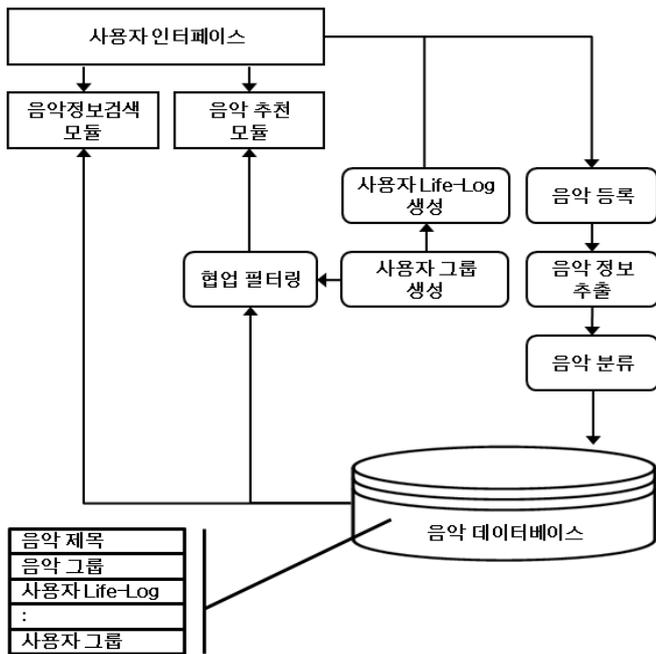
시스템 사용자들의 정보요구에 적합한 정보를 자동적으로 예측하기 위한 방법으로 현재 웹기반 추천 시스템에서 많이 사용하는 방법 중에 하나가 협업 필터링 이다.

협업 필터링은 크게 사용자 기반 협업필터링(user-based collaborative filtering)과 아이템기반 필터링(item-based collaborative filtering)으로 구분된다. 사용자 기반은 사용자들이 직접적으로 선호도를 부여하여 그 데이터를 가지고 다른 사용자에게 추천을 해주는 방식이다[7]. 이 경우에는 직접적으로 평가를 한 데이터이기 때문에 신뢰가 들어가 있는 정보라고 할 수 있다. 아이템 기반은 아이템 간의 유사성을 고려하여 사용자의 선호도를 분석하여, 예측 후 추천하는 방법이다. 이러한 협업 필터링에서 중요한 점은 사용자와 유사한 사람들을 찾는 것이다. 그 그룹이 평가한 데이터 중에 가장 높은 데이터를 바탕으로 사용자에게 추천을 해주고 또한 자신의 데이터가 그 그룹의 데이터의 일환으로 사용된다.

## 3. Life-Log기반 협업 필터링을 이용한 음악정보 검색 시스템

본 연구는 사용자의 음악선호도, 상황인식에 따른 Life-Log 데이터를 이용하여 음악정보를 검색하고, 사용자들의 간의 유사도 기반의 협업필터링 기법을 사용자에게 적절한 음악을 서비스 할 수 있는 추천 기능을 갖춘 시스템을 구현하다.

시스템 구성도 [그림 1]과 같이 음악정보 검색 시스템은 음악정보검색 모듈, 사용자 Life-Log를 생성하는 모듈, 사용자 그룹 기반의 협업 필터링을 통한 음악추천 모듈과 사용자가 음악을 직접 등록할 수 있는 음악 등록 모듈로 구성된다.



[그림 1] 시스템 구성도

### 3.1 음악정보 검색 모듈

음악정보 검색 모듈은 음악과 관련된 정보를 사용하여 음악검색을 위한 모듈이다. 본 연구에서는 음악정보 검색 방법 중 음악과 관련된 텍스트 메타데이터를 이용한 검색방법을 제안한다.

#### 색인과정

음악정보 검색 모듈에서의 색인과정은 [그림 2]와 같이 음악과 관련된 메타데이터에서 명사를 추출한 후, 명사에 대한 출현 빈도를 추출한다. 추출된 명사와 출현 빈도는 역파일 구조로 저장하게 되는데, 역파일 구조 텍스트 구조 검색의 속도를 고려한 방법이다.



[그림 2] 색인과정

#### 가중치 부여

추출된 색인어는 가중치를 부여하여 색인한다. 가중치는 색인어 용어빈도수(term frequency)와 역문헌빈도수(inverse document frequency)를 고려하여

부여한다. 색인어 출현빈도는 한 음악관련 메타데이터에서 자주 출현하는 색인어에 높은 가중치를 부여하고, 역문헌빈도는 출현한 전체 음악관련 메타데이터의 수가 적은 색인어에 높은 가중치를 부여한다.

본 논문에서는 [수식 1]과 같이 일반적으로 대용량 문서 검색에 있어 양호한 성능을 보이는 tf-idf 가중치 부여 방법을 사용하였다.

$$w_{i,j} = f_{i,j} \times idf_i = \frac{freq_{i,j}}{\max_l freq_{l,j}} \times \log \frac{N}{n_i} \quad \text{[수식 1]}$$

용어빈도수 tf는 음악관련 메타데이터  $d_j$ 에서의 색인어  $k_i$  출현 빈도수  $freq_{i,j}$ 를 음악관련 메타데이터  $d_j$ 에서 빈도수가 가장 큰 색인어  $\max_l$ 로 나눈 값을 가지며, 역문헌빈도수 idf는 전체 문헌 수  $N$ 을 색인어  $k_i$ 가 출현한 문헌 수  $n_i$ 로 나눈 값을 계산한다. 가중치 값은 0~1사이의 실수 값을 가지며, 가중치 값이 클수록 문서에 나타나는 어휘가 차지하는 비중이 크다고 볼 수 있다.

#### 검색과정

음악관련 메타데이터  $d_j$ 와 질의  $q$ 의 유사도  $sim(d_j, q)$ 는 색인어로 구성된 음악관련 메타데이터 벡터  $\vec{d}_j$ 와 음악관련 질의 벡터  $\vec{q}$ 의 곱을 문헌 벡터의 크기  $|\vec{d}_j|$ 와 질의 벡터의 크기  $|\vec{q}|$ 의 곱으로 나눈 값으로 구할 수 있다[수식 2]. 코사인 유사도를 이용해 유사도를 계산함으로써 질의어와 유사한 색인어를 가진 음악관련 메타데이터에는 비 연관된 색인어를 가진 음악관련 메타데이터 보다 높은 순위를 부여할 수 있다. 또한 질의어와 색인어의 완전 일치가 아닌 부분정합도 가능하기 때문에 사용자 정보 요구에 유연한 대처가 가능하다

$$sim(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} = \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad \text{[수식 2]}$$

### 3.2 음악추천 모듈

#### 음악 Database 구축

음악에 관련된 정보들을 내포하고 있는 문헌을 생성한다. 음악 문헌에 있는 정보로는 음악에 대한 가수, 작곡가, 작사가, 제목, 가사 등이 들어가 있다. 음악에 대한 검색을 할 시에 이 문헌정보를 이용해서 해당 음악을 찾아주게 된다. 단순한 DB검색을 이용하는 것이 아니고, 유사도를 이용해서 검색하기 때문에 보다 좋은 검색결과를 보여 줄 수 있다.

#### 선호도 평가

음악에 대한 선호도는 사용자의 날씨 상황과 연관되어 변하기 때문에 본 논문에서는 음악에 대한 선호도 평가를 날씨정보와 사용자의 감정을 선호도 평가의 기준으로 설정하였다. 사용자는 음악에 대한 선호도 평가를 각 속성을 평가하게 된다.

[표 1] 음악에 대한 사용자 평가 예

사용자	음악 제목	날씨	선호도	적극성	소극성	긍정도	부정도
사용자1	A	비	7	8	0	7	0
사용자2	B	흐림	5	6	0	0	8
사용자3	C	맑음	8	0	5	4	0
:	:	:	:	:	:	:	:

사용자 1은 “A”라는 곡이 ‘비’가 내리는 날에 ‘7’ 정도의 선호도를 보였다. 그리고 적극도가 ‘8’이고 긍정도가 ‘7’이 듣기 좋다고 평가 내렸다. 이와 같은 음악관련 사용자의 상황인식, 감정 정보는 사용자의 Life-log로 저장되어 사용자에게 적합한 음악을 추천하는 데이터로 사용된다.

### 3.3 유사도를 이용한 사용자 군집화

비슷한 선호도를 갖는 사용자 군집화는 사용자 Life-log를 이용하여 유사도 측정을 통해 이루어 졌다. 사용자간 유사도측정 방법은 [식 3]와 같다.

$$sim(Lifelog_j, user) = \frac{\sum_{i=1}^m Lifelog_{i,j} \times user_i}{\sqrt{\sum_{i=1}^m Lifelog^2} \times \sqrt{\sum_{i=1}^m user_i^2}}$$

[수식 3] 사용자들 간 유사도

Lifelog는 현재사용자 이외의 사용자들이 평가한 Life-log에 대한 벡터 표현 값이며, user는 현재 사용자의 Life-log의 벡터 표현 값이다.

### 3.4 음악추천

사용자에게 음악을 추천하기 위해 현재 사용자와 유사한 사용자 그룹에서 정보를 추출한다. 사용자들이 높게 평가를 하고 현재의 사용자가 평가하지 않은 곡들 중에서 현재 날씨와 일치하는 곡들을 추천해준다. 곡들을 추천해줄 때도 곡들에 대한 선호도의 평균을 구해서 높은 곡부터 차례대로 사용자에게 추천한다.

## 4. 실험 및 결과

본 논문에서 제안한 Life-Log기반 협업 필터링을 이용한 음악정보 검색 시스템의 평가를 위해 기존의 일반적인 방식의 Top N방식(인기도 순위 검색 및 추천)의 일반추천과 비교분석 하였다.

### 4.1 실험 방법 및 실험 과정

#### 실험참가자

실험에 참여한 실험참가자는 한신대학교 재학생 30명을 대상으로 하였다.

#### 실험과정

실험 참가자들에게 랜덤으로 음악리스트를 제고하고, 음악을 듣거나, 사전에 알고 있는 음악을 평가하고 순위화(Top N 방식 추천) 한다. 이후 본 논문에서 제안한 시스템을 이용한 방식을 이용하여 참가자들에게 추천을 하고 사용자 만족도를 비교분석한다.

### 4.2 실험 결과

실험결과 일반 추천을 이용한 결과는 [표 1]과 같고, [표 2]는 본 논문에서 제안한 시스템을 이용한 추천 결과 이다.

[표 2] 일반 추천결과

순번	제목	가수
1	Isn's She Lovely	Stevie Won
2	비밀번호 486	윤하
3	냥만 고양이	체리필터
4	Rain	Siam Shade
5	Lonely Night	부활
6	진달래꽃	마야
7	별 바람 햇살 그리고 사랑	김종국
8	Peace & Rock Roll	체리필터
9	Jump	김동률
10	Revolution A.D	체리필터

[표 3] 제안한 시스템 추천 결과

순번	제목	가수
1	냥만 고양이	체리필터
2	Isn't She Lovely	Stevie Won
3	잊을께	윤도현
4	한번만 더	이승기
5	Marry me	Ellegarden
6	Rain	Siam Shade
7	비밀번호 486	윤하
8	진달래꽃	마야
9	Beat It I	Fall out B
10	Lonely night	부활

평가한 사용자들의 만족도는 0~10까지 값을 이용하여 평가하였다. 평가 결과[표 4]와 같다.

[표 4] 사용자 만족도

실험 참가자	일반 추천 만족도	제안한 시스템 추천 만족도
A	4	9
B	5	6
C	7	7
D	3	5
E	7	5
F	7	9
G	6	7
H	2	8
:	:	:
평균	0.54	0.71

## 6. 실험 결론 및 향후 연구 방향

본 논문은 사용자 상황인식(Context-Aware)를 통한 생성된 Life-Log 기반의 협업 필터링(Collaboration Filtering) 기법을 사용한 음악정보검색 시스템을

을 구현하고 실험하였다. 실험결과 기존의 Top N방식과 본 논문에서 제안한 시스템간의 비교 분석을 통해 사용자 만족도가 향상됐음을 살펴볼 수 있었다. 하지만, 사용자들의 상황인식이나, 감정에 대한 표현이 본 논문에서 사용한 특징들 이외에 다양한 형태로 표현되기 때문에 감정정보나 상황인식 특징 추출에 대한 연구가 더욱 필요할 것이다.

## 참고 문헌

- [1] 정유진, 최두석, “음악자료의 접근점 제어 및 확장에 관한 연구”, 제9회 한국정보관리학회 학술대회 논문집, pp. 191-196, 2002.
- [2] Chai, W. and Verco, B. "Using User Models in Music Information Retrieval Systems". Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000), 2000.
- [3] 지정규, 오해석, “디지털 음악정보 검색 시스템 설계”, 한국데이터베이스학회 국제 학술 대회, pp. 425-437, 1997.
- [4] Kim, J. and Belikin, N. J. "Categories of Music Description and Search Terms and Phrases Used by Non-music Experts". Edited by Michael Fignerhut Proceedings of the Third International Conference on Music Information Retrieval: ISMIR (Paris, France, October 2002): p p. 209-214, 2002.
- [5] Cuddy, S., Katchabaq, M., and H. Lutfiyya, "Context-Aware Service Selection based on Dynamic and Static Service Attributes," IEEE International Conference on Wireless and Mobile Computing, Networking and Communication, 2005.
- [6] "collaborative filtering", <http://www.wikipedia.org/>