

# P2P 환경에서 협업 필터링을 이용한 음악 추천 시스템에 대한 연구

원희재<sup>†</sup>, 박규식<sup>\*\*</sup>

## 요 약

본 연구에서는 P2P 네트워크 기반의 음악 추천 시스템을 새로이 제안하였다. 제안 시스템은 기존 Client-Server 환경의 추천 시스템과 달리 P2P 구조에서 사용자 peer들 간에 실시간으로 선호 음악 정보를 공유함으로써 보다 정확한 음악 추천을 할 수 있는 장점을 가진다. 추천 알고리즘으로는 협업 필터링 기술을 사용하였다. Peer 사용자의 음악 선호도 특성을 나타내는 사용자 프로파일(user profile)은 모든 합법적 음악 파일이 포함하고 있는 고유 KID 장르 색인을 사용하였기 때문에 기존 연구와 같이 부정확한 특징벡터 추출로 인한 시스템 성능 저하와 연산 부하를 줄일 수 있는 장점이 있다. 제안 시스템의 성능 평가는 국내 음악 포털 회사인 S사의 약 16주 분량 실제 사용자 음악 청취 기록과 유료 다운로드 기록을 바탕으로 다양하게 수행하였으며 그 실용성을 입증하였다.

## A Study of Music Recommendation System in P2P Network using Collaborative Filtering

Hee-Jae Won<sup>†</sup>, Kyu-Sik Park<sup>\*\*</sup>

## ABSTRACT

In this paper, we propose a new P2P-based music recommendation system. In comparison with previous system in client-server environment, the proposed system shows higher quality of music recommendation through real-time sharing of music preference information between peers. A collaborative filtering is implemented as a recommendation algorithm. As a user preference profile, we use the inherit KID music genre index contained in all legitimate music file instead of music feature vectors as in previous research so that the proposed system can mitigate the performance degradation and high computational load caused by feature inaccuracy and feature extraction. The performance of the proposed system is evaluated in various ways with real 16-weeks transaction data provided by Korean music portal, S company and it shows comparative quality of recommendation with only small amount of computational load.

**Key words:** MRS, Music Recommendation System(음악 추천 시스템), Collaborative Filtering(협업 필터링)

## 1. 서 론

추천 시스템(RS, Recommendation System)은 웹

상에서 사용자의 취향에 맞는 상품을 추천하는 시스템으로 영화, 뉴스, 도서, 음악, TV 프로그램 등 여러 분야에 널리 응용되고 있다. 이러한 추천 시스템은

※ 교신저자(Corresponding Author): 박규식, 주소: 경기도 용인시 수지구 죽전동 126(448-701), 전화: (031)8005-3252, FAX: (031)8005-3228, E-mail: kspark@dankook.ac.kr

접수일: 2008년 6월 24일, 완료일: 2008년 8월 8일

<sup>†</sup> 정회원, (주)소리바다 기술연구소 선임연구원

(E-mail: hjwon@soribada.com)

<sup>\*\*</sup> 정회원, 단국대학교 컴퓨터학부 교수

※ 본 논문은 중소기업청이 주관하는 산학협력실 지원사업인 산학연 공동기술개발 컨소시엄사업 참여로 인한 결과물입니다.

1차적인 검색에 의한 구매 이외에도 사용자들에게 보다 많은 상품을 노출시킴으로써 더 많은 구매 기회를 유도할 수 있는 장점으로 최근에 많은 연구가 이루어지고 있다. 현재 추천 시스템은 해외의 Amazon.com 같은 유명 서점이나 PTV 등의 TV 프로그램 가이드, MyBestBets 등의 쇼핑 사이트 등에 적용되어 그 실효성이 입증되어 왔지만, 국·내외적으로 음악 추천 시스템에서는 뚜렷한 성공 사례를 찾아볼 수 없는 실정이다.

최근까지의 추천 시스템은 주로 자동화된 정보 필터링(Information Filtering) 기술을 이용해 사용자 취향에 맞는 상품을 추천하는 방식으로 크게 내용기반 필터링(CBF, Content-Based Filtering)과 협업 필터링(CF, Collaborative Filtering), 그리고 이 둘을 결합한 복합 필터링(HF) (Hybrid Filtering), 인구 통계학 정보를 이용하는 인구통계 필터링(DF, Demographic Filtering) 시스템 등으로 구분할 수 있다.

내용기반 필터링(CBF)은 사용자가 과거에 선호하였던 상품이나 항목 부류들을 미래에도 선호할 가능성이 높다는 사실을 이용한 방식이다. 이 방식은 사용자가 과거에 접근이나 구매하였던 상품이나 항목 데이터(item data)로부터 특징 벡터를 추출해 사용자 선호도 프로파일(user preference profile)을 구축한 후 사용자 프로파일과 DB 간의 유사도를 계산하여 가장 밀접한 항목들만을 추천한다. 내용기반 필터링을 이용한 대표적인 추천 시스템으로는 NewsWeeder[1], Infofinder[2], News Dude[3] 등이 있다. 이러한 추천 시스템에서는 각 상품이나 항목을 대표할 수 있는 특징벡터를 정확하게 선정 및 추출하는 것이 무엇보다 중요하며 특징벡터의 정확도에 따라 시스템의 성능이 좌우된다. 내용기반 필터링 추천(CBF-R) 시스템은 개념적으로 이해하기 쉽고 계산 과정이 간단한 반면, 사용자가 과거에 관심을 나타낸 상품이나 항목들에 대해서만 추천이 가능하기 때문에 사용자의 잠재적인 흥미를 유발할 수 있는 다른 항목들을 추천할 수 없는 단점이 있다.

협업 필터링(CF)을 이용한 추천 시스템은 사용자의 취향과 유사한 다른 사용자의 취향을 참고하여 상품이나 항목에 대한 선호도를 예측하여 추천하는 방식이다. 사용자 수가 많을 때는 유사한 사용자 선호도 프로파일을 갖는 사용자들만을 따로이 그룹핑(grouping)한 다음, 같은 그룹에 속해있는 이웃 사용

자(neighbor user)에 의해 가장 높은 선호도를 보인 항목을 사용자에게 추천한다. 이 방식은 실생활에서 주위의 친구들이나 다른 사용자들의 경험이나 추천을 참조해 상품을 구매하는 것과 같이 “입소문(Word of Mouth)”의 개념을 자동화 한 것이라 할 수 있다. 협업 필터링을 이용한 대표적인 추천 시스템으로는 Ringof[4], SiteSeer[5] 등이 있다. 협업 필터링은 내용기반 필터링이 제공할 수 없는 다양하고 유익한 항목들을 추천할 수 있는 장점이 있지만 사용자에 대한 충분한 데이터가 수집되어야만 정상적인 추천 서비스가 가능하다는 점과 새로운 상품이나 항목이 추가되었을 때 이에 대한 추천이 즉각 불가능하다는 단점이 있다.

이외에도 내용기반 필터링과 협업 필터링 기술의 장·단점을 보완해 결합한 복합 필터링(HF) 시스템으로 Tapestry[6], GroupLens[7]이 있으며, 사용자의 나이, 성별, 국적 등의 인구 통계학 정보를 기준으로 항목을 추천하는 인구 통계 필터링(DF)[8] 방식 등이 있다.

이상에서 살펴본 대부분의 추천 시스템은 Client-Server 환경에서 동작하는 시스템으로 Server에서 각 Client 정보들을 수집한 다음 이를 분석해 추천하기 때문에 사용자 수가 늘어날수록 Server에서의 연산 및 네트워크 부하가 증가될 수밖에 없으며, 방대한 상품에 대한 사용자의 선택권이 제한될 수밖에 없다. 이러한 이유로 최근에는 Client-Server 환경 대신 P2P 구조에 대한 연구가 활발해 지고 있으나, P2P 또한 사용자 수가 늘어날수록 공유되는 콘텐츠 수 또한 방대해지기 때문에 보다 손쉬운 검색 및 구매 의사 결정을 위해 추천 시스템에 대한 요구가 더욱 높아질 수밖에 없는 실정이다.

P2P 기반 추천 시스템에 대한 대표적인 연구로 Olsson[9]은 P2P 구조에서 뉴스 헤드라인 추천 시스템을 제안하였고, Peng[10]은 추천 시스템의 연산량을 감소시키기 위해 분산 해쉬 테이블(hash table)을 이용한 분산 협업 필터링 알고리즘을 제안하였다. 한편, Kim[11]은 협업 필터링 알고리즘을 이용한 이미지 추천 시스템을 구현하였다. 이러한 P2P 기반 추천 시스템은 대부분의 연산을 분산 처리함으로써 Server의 연산 부하를 줄일 수 있으며 시간에 따른 각 peer 사용자들의 선호도 변화를 즉시 반영할 수 있는 장점을 가질 수 있다.

본 연구에서는 협업 필터링기반의 P2P 음악 추천 시스템(MRS, Music Recommendation System)을 제안하였다. 제안 시스템은 P2P 구조에서 사용자 peer들 간에 실시간으로 음악 선호 정보를 공유함으로써 모든 참여자가 서비스 제공자(Server)인 동시에 서비스 수요자(Client)가 되어 보다 정확한 음악 추천을 할 수 있는 장점을 가진다. 본 연구는 다음과 같은 점에서 기존의 연구와 차별성을 갖는다.

(1) 추천 시스템의 핵심이랄 수 있는 사용자 선호도 프로파일(user preference profile) 구축 시 기존 연구에서처럼 음악에 대한 특징벡터를 추출해 사용하는 대신, 모든 합법적 음악 파일이 포함하고 있는 고유의 음악 ID와 KID 장르 분류 색인을 사용하였다. 즉, peer 사용자의 음악 선호도 특성을 선호 음악 장르로 표현하였기 때문에 부정확한 특징벡터 추출로 인한 시스템 성능 저하와 연산 부하를 줄일 수 있다.

(2) 사용자 선호도 프로파일 구축 시 Server에 자동으로 기록되는 peer 사용자의 음악 청취나 유료 음악 다운로드 이력(transaction)을 사용함으로써 기존 협업 필터링 시스템에서 사용자가 직접 음악 취향에 대한 정보를 제공해야 하는 불편함을 제거하였다.

(3) Peer 사용자의 시간에 따른 음악 선호도 변화를 반영하여 가장 최근의 선호도를 갖는 음악 파일만을 대상으로 음악을 추천함으로써 보다 정확한 추천을 할 수 있다.

(4) 본 연구에서 제안한 P2P-MRS 시스템의 성능 평가는 국내 음악 포털사인 S사에서 구축된 약 16주 분량의 실제 사용자 음악 청취 기록과 유료 다운로드 기록을 바탕으로 수행되었다.

본 논문은 다음과 같이 구성된다. 먼저 2장에서는 제안한 P2P-MRS 시스템의 전체 구성과 세부 모듈을 살펴보고, 3장에서는 제안 시스템의 추천 과정을 설명하며, 4장에서는 실제 사용자 데이터를 이용한 다양한 실험을 통해 제안 시스템의 성능을 입증하였다. 마지막으로 5장에서는 결론과 향후 연구에 대한 고찰로 끝을 맺는다.

## 2. 제안 P2P-MRS 시스템

그림 1은 본 연구에서 제안한 P2P-MRS 에이전트(agent)의 블록도를 나타낸다. 그림의 P2P-MRS 에이전트는 네트워크상에서 상호 연결된 각 peer 사용

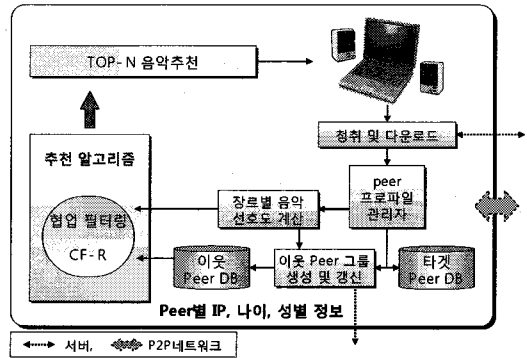


그림 1. 제안 P2P-MRS 에이전트

자의 컴퓨터에 내장되며, 모든 peer가 실시간으로 음악 선호 정보를 공유함으로써 다른 peer들로부터 음악을 추천 받을 수도 있고, 동시에 다른 peer들에게 음악을 추천할 수도 있다.

본 연구의 P2P-MRS 시스템은 크게 음악 추천에 사용되는 협업 필터링(CF-R) 알고리즘, Top-N 음악 추천 모듈, 사용자 peer 프로파일 관리자, 협업 필터링 추천을 위한 이웃 peer 그룹 생성 및 갱신, 장르별 음악 선호도 계산 등으로 구성된다. 그림 1에서 각 peer 사용자의 청취 및 유료 다운로드 기록은 개인별 peer 트랜잭션(transaction)으로 기록되고 이웃 peer 그룹(neighborhood peer group)을 생성하거나 음악 장르별 선호도를 계산하는데 사용된다. 따라서 본 연구에서의 모든 추천 관련 연산은 peer 에이전트에서 담당하게 되고, Server는 각 peer들의 청취나 유료 다운로드 기록만을 갱신해서 각 peer에 이 정보를 제공하는 역할만을 하게 된다.

### 2.1 Peer 프로파일 관리자

Peer 프로파일(peer profile)은 각 peer의 음악적 취향 변화를 파악할 수 있는 가장 중요한 단서로서 협업 필터링 음악 추천에서 핵심적인 역할을 한다. 본 연구의 peer 프로파일은 peer에 자동으로 기록되는 peer 사용자의 음악 청취나 유료 다운로드 기록(transaction)을 사용함으로써 기존 협업 필터링 시스템에서와 같이 사용자가 직접 음악 취향에 대한 정보를 제공해야 하는 불편함을 제거하였다. 또한 기존 연구에서처럼 음악에 대한 특징벡터를 추출하는 대신 음원 파일에 포함된 29개 KID 장르 분류체계를 이용해 각 peer의 시간대 별 음악 청취 특성이나 유

로 다운로드 특성을 트랜잭션으로 표현하기 때문에 부정확한 특징벡터 추출로 인한 시스템 성능 저하와 연산 부하를 줄일 수 있는 장점이 있다.

본 연구의 peer 프로파일은 peer 사용자의 최근 음악 청취 이력을 이용하는 방법과 유료 다운로드 이력을 이용하는 방법 2가지가 가능하다. Peer의 음악 청취 프로파일은 추천받은 Top-N 음악 리스트 중 사용자 peer가 직접 선택하여 1분 이상 청취한 음악에 대한 음악 ID, 음악 KID 장르, 트랜잭션으로 구성된다. 하나의 트랜잭션은 사용자 peer가 연속으로 10분 이상 60분 이내에 청취한 음악파일에 대한 기록으로, 본 연구의 peer 프로파일은 사용자가 가장 최근에 청취한 5개 이상 10개 이하의 청취 트랜잭션으로 구성된다. 이러한 peer 프로파일 구성은 사용자 peer의 최근 음악적 취향 선호도를 우선 반영함으로써 보다 정확한 음악 추천을 가능하게 한다. 다음의 표1은 5개 트랜잭션으로 구성된 peer 프로파일 예를 보이고 있다. 유료 다운로드 이력을 이용한 프로파일 구성은 표 1과 같은 형식을 취하며, 다만 청취 기록대신 다운로드 기록을 사용한다는 점이 다르다.

표 1에서 트랜잭션  $T_i$ 는 peer 사용자가 가장 최근에 청취한 기록으로 최초 P2P 네트워크 접속 후 10분 이내에 청취한 2개 음악 파일로 구성된다. 반면, 2008/02/01에 청취한 14개 음악들은 비록 같은 날에 청취하였다더라도 M0014부터 M0003의 12개 음악 파일은 60분 이내 청취한 기록이기 때문에 트랜잭션

표 1. Peer 프로파일 예

접속 시간	음악 ID	음악 KID 장르	트랜잭션
2008/02/03 23:45	M0013	KID2	T1
2008/02/03 23:41	M0037	KID3	T1
2008/02/01 21:25	M0014	KID4	T2
2008/02/01 21:16	M0013	KID3	T2
:	:	:	:
2008/02/01 20:27	M0003	KID1	T2
2008/02/01 20:22	M0002	KID4	T3
2008/02/01 20:18	M0001	KID3	T3
2008/01/31 12:11	M0103	KID3	T4
2008/01/31 12:08	M0101	KID1	T4
2008/01/31 07:45	M0122	KID2	T5
2008/01/31 07:40	M0121	KID1	T5

그림 2. KID 음악 장르 분류 체계

$T_2$ 로 기록되고, 나머지 2개 음악파일은 별도의  $T_3$  트랜잭션으로 기록된다. 한편, 표 1에서 사용된 음악 장르는 산업체에서 통용되는 150 여개 음악 장르 분류를 그림 2와 같이 29개로 재분류하여 정리한 것이다.

### 2.2 장르별 음악 선호도 계산

그림 1의 장르별 음악 선호도 계산 모듈은 peer 사용사가 어느 장르의 음악을 선호하는 지 장르별 음악 청취 비율(MGLR, Music Genre Listening Ratio),  $MGLR_i$ ,  $i = 1, 2, \dots, 29$ ,을 계산하는 과정으로 표 1의 peer 프로파일로부터 계산되며 협업 필터링 음악 추천에서 peer들 간의 유사도를 계산해 이웃 peer 그룹을 생성할 때 사용된다.

각 peer의 장르별 음악 청취 비율  $MGLR_i$ 는 다음과 같이 계산된다.

$$MGLR_i = \frac{GLC_i}{\sum_{g=1}^G GLC_g}, \quad i = 1, 2, 3, \dots, 29 \quad (1)$$

수식 (1)에서  $GLC_i$ 는 음악 장르  $i$ 에 대한 peer 사용자의 청취 음악 개수를 나타내며  $G$ 는 전체 음악 장르 수, 즉 29개 KID 장르를 나타낸다.

한편 수식 (1)의 장르별 음악 청취 비율은 표 1의 peer 프로파일에 있는 트랜잭션별 가중치를 부과해 수식 (2)와 같이 수정될 수 있다.

$$TWGR_i = \frac{\sum_{t=1}^T TW_t X MGLR_{t,i}}{\sum_{g=1}^G \sum_{t=1}^T TW_t X MGLR_{t,g}}, \quad t = 1, 2, \dots, T; \quad i, g = 1, 2, 3, \dots, 29 \quad (2)$$

수식 (2)는 사용자가 가장 최근에 들었던 음악 트

랜잭션에 좀 더 많은 가중치를 주는 방법으로  $TW_t$ ,  $t=1,2,\dots,T$ 는  $T$ 개 트랜잭션에 대한 가중치 그리고  $MGLR_{t,i}$ 는  $t$ 번째 트랜잭션에서 각 음악 장르  $i$ 에 대한 청취 음악 비율을 나타낸다. 예를 들어,  $T_2$  트랜잭션에서 peer가 총 30개의 음악을 청취하였는데 그 중 Rock 장르가 25곡, Classic이 5곡이라면 이 때  $MGLR_{2,Rock} = 25/30$ 이 되고,  $MGLR_{2,Classic} = 5/30$ 이 되며 나머지 장르에 대한  $MGLR$  값은 0이 된다.

본 연구에서는 다양한 실험을 통해 트랜잭션 가중치로  $TW_t = 1 \times (0.9)^{t-1}$ 를 사용하였다. 즉, 트랜잭션  $T_1$ 에 대한 가중치는 0.9,  $T_2$ 에 대한 가중치는 0.81 등으로 가장 최근에 청취한 트랜잭션에 더 많은 가중치를 주게 된다. 한편, 하나의 음악 장르에 대한 청취 비율 값은 8bit로 표현되며 각 peer의 음악 장르별 청취 정보는 총 29Byte 배열로 저장된다.

2.3 이웃 peer 그룹 생성/갱신과 타겟 peer 그룹

협업 필터링 추천 시스템은 유사한 peer 프로파일을 갖는 peer 들을 그룹핑 한 다음, 같은 그룹에 속해 있는 이웃 peer(neighborhood peer)들의 프로파일을 비교해 가장 높은 선호도를 보인 음악을 추천한다. 따라서 협업 필터링 추천에서는 유사한 peer 프로파일을 갖는 peer들을 어떻게 잘 그룹핑 하느냐에 따라 추천 시스템의 성능이 좌우된다.

본 연구에서는 다음과 같이 2단계를 거쳐 이웃 peer 그룹을 결정하게 된다.

1 단계) 먼저 표 2와 같이 각 peer들의 나이와 성별 정보를 표현한 6bit 벡터를 비교해 호스트 peer와 가장 유사한  $L$ 개의 peer들을 선정한다.

2 단계) 1단계에서 구한 총  $L$ 개의 peer 중 수식

표 2. Peer의 나이, 성별 정보 표현 (나이 24세 남자 예)

Bit 필드	필드 의미
1	$age \leq 18$
2	$18 < age \leq 29$
3	$29 < age \leq 49$
4	$50 \leq age$
5	남자
6	여자

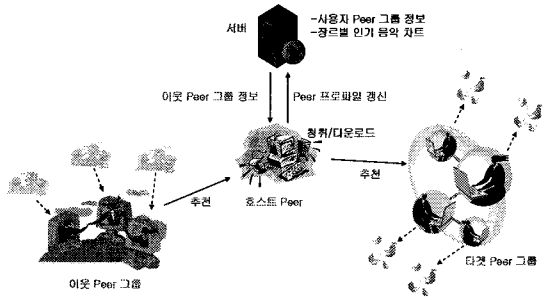
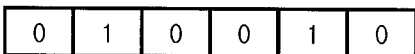


그림 3. 호스트 peer, 이웃 peer 그룹, 타겟 peer 그룹간의 음악 추천

3.3의 이웃 유사도(NS, Neighbor Similarity)를 계산해 NS가 큰 순으로 최종  $C(L)$ 개의 이웃 peer들을 선정한다.

$$NS(hp, np) = 1 - \sum_{g=1}^G (|TWGR_g^{hp} - TWGR_g^{np}|) \tag{3}$$

수식 (3)에서  $hp$ 는 호스트 peer 그리고  $np$ 는 1단계에서 구한  $L-1$ 개의 peer들을 말한다.  $TWGR_g$ 는 각 음악 장르  $g$ 에서의 peer 사용자의 청취 음악 장르별 중요도 비율을 나타내며  $G=29$ 는 전체 음악 장르 수를 나타낸다. 위 수식에서 NS는 호스트 peer와 이웃 peer 간의 유사도가 낮을수록 0에 가까우며, 유사도가 높을수록 1에 가까운 값이 된다.

음악 추천은 위의 2 단계 과정으로 구해진 이웃 peer 그룹에서 호스트 peer를 제외한 나머지  $C-1$ 개 peer의 프로파일로부터 가장 높은 선호도를 보인 음악을 추천받게 되며, 만약 호스트 peer가 추천 받은 음악을 1분 이상 청취하거나 유료 다운로드를 하였다면, 호스트 peer의 프로파일을 새롭게 갱신하고 이에 따른 이웃 peer 그룹도 새롭게 갱신되어야 한다. 이와 동시에 호스트 peer를 이웃 peer로 두고 있는 타겟 peer(target peer)들 또한 같은 음악 파일을 추천 받아 청취나 구매 의사 결정을 내리게 된다. 그림 3은 이러한 과정을 나타낸다.

3. 협업 필터링 기반 P2P-MRS 시스템

제안 P2P-MRS 시스템은 크게 음악 추천에 사용되는 추천 알고리즘 모듈, peer 프로파일 관리자, 이웃 peer 그룹 생성/갱신, 장르별 음악 선호도 계산 등으로 구성된다.

음악 추천은 먼저 훈련 데이터로부터 초기 이웃 peer 그룹을 생성해 호스트 peer에 음악을 추천하고, 이 중 호스트 peer의 프로파일에 포함되지 않은 음악만을 대상으로 peer 선호 점수인 PPS를 계산해 점수가 높은 순으로  $N$ 개의 음악 추천 리스트를 제공한다. 이를 위해 협업 필터링 추천 알고리즘은 2.3장에서 구한  $C$ 개 이웃 peer 그룹에서 호스트 peer를 제외한 나머지  $C-1$ 개 peer의 프로파일로부터 추천 받은 음악으로부터 가장 높은 선호도를 갖는  $N$ 개 음악을 선정한다. 선호도 계산은 각 이웃 peer들로부터 가장 많은 추천을 받은 음악을 대상으로 호스트 peer와의 이웃 유사도(NS)를 감안해 수식 (4)과 같이 Peer 선호 점수(PPS, Peer Preference Score)를 계산한 다음 이 점수가 높은 순으로 추천될 음악의 순위를 결정한다.

$$PPS(hp, i) = \sum_{np=1}^{C-1} P_{npi} \times NS(hp, np) \quad (4)$$

수식 (4)에서,  $PPS(hp, i)$ 는 호스트 peer  $hp$ 에게 추천된 음악  $i$ 에 대한 Peer 선호 점수이며,  $P_{npi}$ 는 이웃 peer  $np$ 가 음악  $i$ 를 추천 했는지에 대한 여부로 0과 1로 표현된다.

이때 호스트 peer가 추천된 음악 중 1분 이상 청취하거나 유료 다운로드 구매를 한 음악 파일이 있다면 호스트 peer 프로파일 새로이 갱신되고 이에 따라 이웃 peer 그룹이 새로이 갱신되는 동시에 해당 음악 파일에 대한 정보가 타겟 peer에도 전송된다.

그림 4는 제안한 P2P-MRS 시스템의 음악 추천 순서도를 나타낸 것이다.

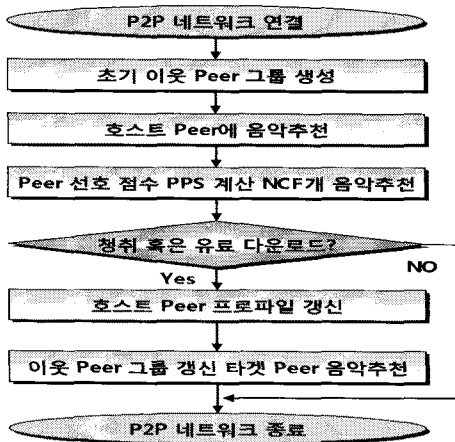


그림 4. 전체 음악 추천 절차

표 3. 수학 기호와 용어 정리

표 기	의 미
P2P MRS	P2P 기반 음악 추천 시스템
$GLC_i$ $i=1,2,3,\dots,29$	장르별 음악 청취 개수
$MGLR_i$ $i=1,2,3,\dots,29$	장르별 음악 청취 비율
$TW_t$ $t=1,2,\dots,T$	트랜잭션 가중치
$GLC_{t,i}$ $t=1,2,\dots,T$ $i=1,2,3,\dots,29$	$t$ 번째 트랜잭션에서 각 음악 장르 $i$ 에 대한 청취 음악 비율
$TWGR_i$ $i=1,2,3,\dots,29$	가중화된 장르별 음악 청취 비율
$NS(hp, np)$	peer 유사도, $hp$ =호스트 peer, $np$ = 이웃 peer
$N$	시스템에 의해 추천되는 총 음악 수
PPS	Peer 선호 점수 (Peer Preference Score)

표 3은 지금까지 본 논문에서 논의 되었던 수학 기호와 용어들을 정리한 것이다.

## 4. 실험 결과 및 분석

### 4.1 실험 방법

본 연구에서는 제안한 시스템의 성능을 평가하기 위해 국내 P2P 음악공유 서비스 업체인 S사로부터 실제 peer 사용자들의 청취 및 유료 다운로드 자료를 제공받아 다양한 실험을 수행하였다. 실험에 사용된 자료는 2008년 1월 7일부터 2008년 4월 27일까지 총 16주간의 데이터중의 일부로 총 124,765명의 사용자와 199,067곡을 대상으로 하였다.

제공받은 사용자 peer 데이터들을 임의로 기간별 트레이닝(training) 데이터와 테스트(test) 데이터로 나누었으며, 트레이닝 기간 동안의 peer 데이터는 협업 필터링 시스템에서 초기 이웃 peer 그룹을 생성하는데 사용되고, 테스트 기간의 peer 데이터는 제안 시스템에 의해 추천된 음악 파일과 비교하는 방식으로 시스템 성능을 평가하는데 사용하였다.

제안 시스템의 추천 성능 평가 방법은 Hit ratio, Ranking point average, Percentage of contents 등 다양한 기법들이 존재하나 본 연구에서는 Hit ratio를 사용하였다. Hit ratio는 제안 시스템의 추천 성능을 평가할 수 있는 척도로 훈련 과정을 거친 제안

시스템의  $N$ 개 추천 음악 중 테스트 기간 동안 해당 peer들이 실제 청취하거나 유료 다운로드한 것과 일치하는 음악 파일 수  $N_h$ 의 비율로 다음과 같이 구할 수 있다.

$$Hit\ ratio = \frac{(\text{실제 청취나 다운로드 음악}) \cap (\text{시스템 추천 음악})}{\text{총 추천 음악 수}} = \frac{N_h}{N} \quad (5)$$

#### 4.2 실험 결과 분석

본 연구에서 제안한 P2P-MRS 시스템의 추천 성능은 트레이닝-테스트 기간 비율, peer 프로파일 구축 방법, 이웃 peer 그룹의 크기, 사용자 Peer 프로파일을 구성하는 트랜잭션 수, 연산 부하 등에 따라 좌우된다. 따라서 본 절에서는 위의 다양한 시스템 파라미터에 의해 제안한 음악 추천 시스템의 성능을 비교 분석한다. 그림 5는 구현된 시스템의 추천 UI를 보이고 있다. 사용자의 프라이버시를 보호하기 위하여 사용자 아이디는 보이지 않게 하였다.

##### 4.2.1 트레이닝-테스트 기간 비율에 따른 시스템 성능

그림 6은 트레이닝-테스트 기간 비율에 따른 시스템 성능을 측정하는 것으로, 총 16주 기간 중 트레이닝 기간을 각각 2주(즉, 테스트 기간은 12주), 4주, 6주, 8주, 10주, 12주, 14주로 변화시켰을 때 그리고 Top- $N$  추천 모듈에 의한 추천 음악 개수를 총  $N=20, 40$ 으로 하였을 때 제안 시스템의 Hit ratio를 비교한 것이다. 그림에서 보듯이 트레이닝-테스트 기간 비율이 전체 시스템 성능에 영향을 미침을 알 수 있다. 처음 트레이닝 기간이 늘어날수록 시스템 성능은 좋아지

지만, 12주를 기점으로 다시 성능이 떨어지는 것을 볼 수 있다. 그림 6의 결과에 따라 본 연구에서는 트레이닝 기간을 12주, 테스트 기간을 4주로 최종 선정하여 나머지 실험에 적용하였다. 한편, 추천 목록 수에 따른 성능은 이전과 같이  $N=20$ 일 때가 더 우수함을 알 수 있다.

그림 7은 사용자 peer 프로파일을 구축할 때 트랜잭션으로 사용자의 청취 기록을 사용한 경우와 사용자의 유료 파일 다운로드 기록을 사용한 경우를 트레이닝-테스트 기간 비율에 따라 비교한 것이다. 실험 결과는 트랜잭션을 7개로, 이웃 peer 그룹을 20개로 그리고 총 추천 음악의 개수를 20으로 했을 때 얻은 결과이다. 그림에서 보듯이 전체 시스템의 추천 성능에는 큰 차이가 없는 것을 알 수 있다. 이는 음악 매체의 특성상 유료 다운로드 기록 못지않게 실제 사용자가 1분 이상 청취한 기록도 사용자 취향을 파악하기에 충분하다는 것을 말한다.

##### 4.2.2 사용자 Peer 프로파일을 구성하는 트랜잭션 수

그림 8은 사용자 peer 프로파일을 구성하는 트랜

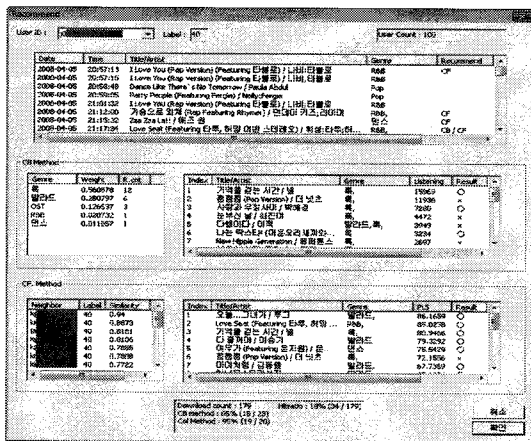


그림 5. 제안 추천 시스템의 UI

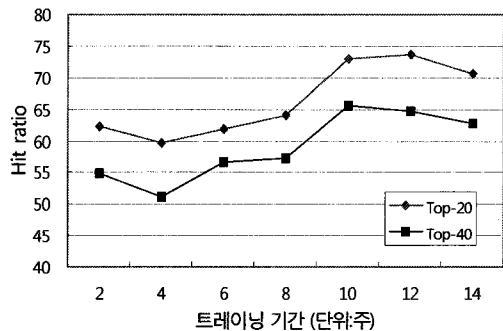


그림 6. 트레이닝-테스트 기간 비율에 따른 시스템 성능 비교

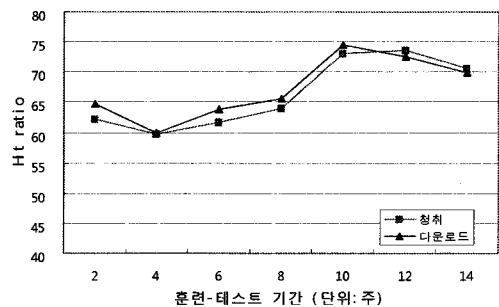


그림 7. 사용자 프로파일 (청취 기록, 다운로드 기록) 구축 방법에 따른 성능 비교

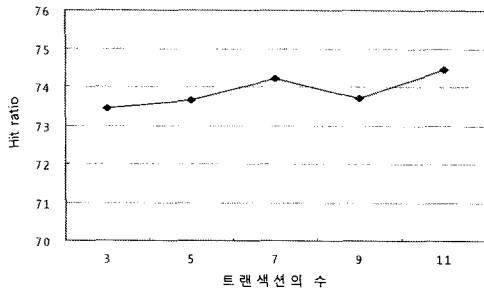


그림 8. 트랜잭션 수에 따른 성능 비교

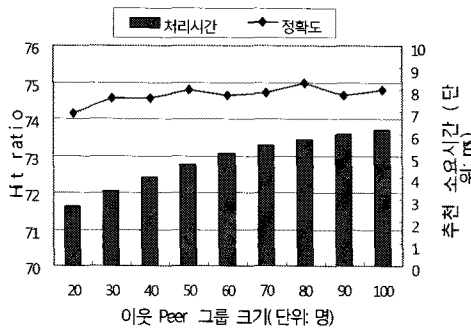


그림 9. 이웃 peer 그룹 크기에 따른 성능 및 연산시간 비교

잭션 수의 변화에 따른 추천 시스템의 성능을 비교한 그림이다. 그림에서 보듯이 트랜잭션 수가 늘어날수록 Hit ratio가 약간 증가하는 결과를 보이나 연산시간을 고려하여 본 연구에서는 트랜잭션 수를 7로 고정하여 사용하였다.

4.2.3 이웃 Peer 그룹의 크기에 따른 시스템 성능

그림 9는 음악 추천 시 추천을 받게 되는 이웃 peer 그룹의 크기에 따른 전체 시스템의 추천 성능과 연산시간을 비교한 것이다. 이웃 peer 그룹의 크기에 따른 성능 변화를 알아보기 위해 이웃 peer 그룹의 크기를 20부터 100까지 변화시켜 Hit ratio와 연산시간을 측정하였다. 그림에서 보듯이 전반적으로 이웃 peer 그룹 크기가 전체 시스템의 성능에 별 다른 영향을 미치지 않는 반면, peer 그룹 크기가 증가할수록 추천에 소요되는 시간이 더욱 늘어나는 것을 볼 수 있다. 따라서 본 연구에서는 이웃 peer 그룹의 크기를 20으로 고정하여 실험에 적용하였다.

5. 결 론

본 논문에서는 P2P 네트워크 환경에서 Peer들 간

의 음악 선호 정보를 실시간으로 공유하여 음악을 추천할 수 있는 시스템을 제안하였다. 음악 추천 알고리즘으로는 협업 필터링 기술을 사용하였다. Peer 사용자의 음악 선호도 특성을 나타내는 사용자 프로파일(user profile)은 음악 파일이 포함하고 있는 고유 KID 장르 색인을 사용하여 기존의 부정확한 특징 벡터 추출로 인한 시스템 성능 저하와 연산 부하를 줄일 수 있다. 제안 시스템의 성능은 음악 포털회사의 실제 16주 분량의 사용자 음악 청취 기록과 유료 다운로드 기록을 바탕으로 트레이닝-테스트 기간 비율, peer 프로파일 구축 방법, 이웃 peer 그룹의 크기, 사용자 Peer 프로파일을 구성하는 트랜잭션 수, 연산 부하 등 다양한 각도로 평가되었다. 실험 결과 제안 시스템의 추천 성능은 약 12주의 트레이닝 기간이 적절하다는 것을 알 수 있었다. 한편, 사용자 peer 프로파일을 구축할 때 사용자의 유료 다운로드 기록 못지않게 청취 기록도 아주 중요하다는 점을 알 수 있었고, 프로파일을 구성하는 트랜잭션의 수가 7개 그리고 이웃 peer 그룹의 크기가 20일 때 약 75%에 이르는 최적의 결과를 얻을 수 있었다.

향후 연구로는 추천 시스템의 정확도와 안정성을 높이기 위해 기존의 내용기반 필터링과 협업 필터링 알고리즘을 최적으로 조합할 수 있는 방법이나 사용자 peer 프로파일 구축 시 peer 사용자의 음악 선호 변화를 좀 더 세밀하게 감안할 수 있는 방법 등에 대한 연구가 필요할 것으로 예상된다. 또한 제안 시스템의 좀 더 엄밀한 성능평가를 위해서는 더 많은 peer 사용자들을 대상으로 다양한 실험이 필요할 것으로 사료된다.

참 고 문 헌

[1] Lang K. Newsweeder: Learning to Filter Netnews. In Proceedings of International Conference on Machine Learning, pp. 331-339, 1995.  
 [2] Krulwich, B. and Burkey, C. Learning User Information Interests through Extraction of Semantically Significant Phrases. In Proceedings of AAAI Spring Symposium on Machine Learning in Information Access. 1996.  
 [3] Billsus, D. and Pazzani, M. A Hybrid User



Model for News Story Classification. In Proceedings of International Conference on User Modeling, 1999.

[4] Sharadanand, U. and Maes, P. Social Information Filtering: Algorithms for Automating 'Word of Mouth'. In Proceedings of CHI'95 Conference on Human Factors in Computing Systems, pp. 210-217, 1995.

[5] Rucker J. and Polanco, M.J. Personalized Navigation for theWeb. Communications of the ACM, 40, 73-89, 1997.

[6] Goldberg, D., Nichols, D., Oki, B.M., and Terry, D. Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, 35, 61-70, 1992.

[7] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J., GroupLens: Applying Collaborative Filtering to Usenet News, Communications of the ACM, Vol.40, No.3, pp. 77-87, 1997.

[8] M. J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," Artificial Intelligent Review, pp. 394-408, 1999.

[9] Olsson, T. Bootstrapping and decentralizing recommender systems. Ph.D. Thesis, Dept. of Information Technology, Uppsala Univ, 2003.

[10] Peng, H., Bo, X., Fan, Y., & Ruimin, S., "A scalable P2P recommender system based on distributed collaborative filtering. Expert Systems with Applications, 27, 203-210, 2004.

[11] J.K. Kim, H.K. Kim, Y.H. Cho, "A user-oriented contents recommendation system in peer-to-peer architecture," Expert Systems with Applications, 34, 300-312, 2008.



원희재

1998년 충남대학교 농학과 학사 졸업.  
 2008년 단국대학교 컴퓨터학과 석사 졸업  
 1999년~2004년 (주)543미디어텍 멀티미디어연구소 개발 팀장

2004년~현재 (주)소리바다 기술연구소 선임연구원  
 관심분야 : 음성 및 오디오 신호처리, 음원인식, DSP 구현 및 설계, 멀티미디어 정보처리



박규식

1986년 Polytechnic University 전자공학과 학사 졸업.  
 1988년 Polytechnic University 전자공학과 석사 졸업.  
 1993년 Polytechnic University 전자공학과 박사 졸업.  
 1994년~1996년 삼성전자 마이크

로사업부, 선임 연구원  
 1996년~2001년 상명대학교 컴퓨터·정보통신공학부 조교수  
 2001년~현재 단국대학교 정보컴퓨터학부 교수  
 관심분야 : 음성 및 음향신호처리, 멀티미디어 신호처리, DSP 시스템 구현