

협동적 여과를 기반으로 하는 개인화된 디지털 음악 추천

김준태*, 김형일**

요약

본 논문에서는 개인의 취향에 맞는 음악을 자동으로 추천해주는 음악 추천 시스템을 소개한다. 본 논문에서 소개하는 추천 시스템은 각 음악 사이의 유사도를 그래프로 저장하는 그래프 기반 협동적 여과 방식을 사용하여 사용자의 암시적 선호 정보를 바탕으로 빠른 추천을 할 수 있으며, 또한 사용자의 정적인 성향뿐 아니라 시간에 따라 달라지는 동적인 성향에 맞는 추천도 가능하도록 설계되었다. 추천 서버는 자바로 구현되었으며 독립된 서버로서 클라이언트와 정해진 프로토콜에 따라 통신하도록 구현되었다. 구현된 추천 서버와 실제 사용자들의 음악 다운로드 데이터를 이용하여 추천 데모 사이트를 구축하였으며, 실험을 통하여 추천 결과의 정확도를 측정하였다.

Personalized Digital Music Recommendation Based on the Collaborative Filtering

Juntae Kim*, Hyungil Kim**

Abstract

In this paper, we introduce a music recommendation system that automatically recommends music according to users' musical tastes. The recommendation system uses a graph-based collaborating filtering in which similarities between musics are saved as a graph, and so it can perform fast recommendation based on the implicit preference information. It also has capability of recommending music according to users' dynamically changing preferences as well as users' static preferences. The recommendation server is implemented as an independent server using Java, and communicates with clients according to a specified protocol. A demo web site has been built by using the server and music download data from actual users, and the accuracy of recommendation has been measured through experiments.

Keywords : Digital Music, Recommendation System, Collaborative Filtering, Personalization, Graph

1. 서론

인터넷의 발전은 디지털 콘텐츠를 고부가가치 산업으로 변모시켰다. 다양한 디지털 콘텐츠 중에서도 특히 디지털 음악 관련 산업은 MP3 플레이어의 확산, 모바일 기기의 컨버전스, 국내에서의 저작권 보호 등에 따라 지속적으로 성장할

것으로 전망되고 있으며, 이러한 가운데 개인 취향 분석 및 개인화된 콘텐츠 추천을 위한 추천 시스템 기술에 대한 관심이 높아지고 있다[1].

추천 시스템은 사용자의 선호도를 분석하고, 아이템들에 대한 사용자의 선호도를 예측하여 책, 영화, 음악, 기사, 웹 페이지 등과 같은 아이템들을 추천하는 시스템이다. 여러 가지 추천 방법 중 가장 성공적으로 적용되어온 추천 방법은 협동적 여과(Collaborative Filtering, CF) 기반 추천이다[3][6][9][13]. 협동적 여과 기반 추천 시스템은 아이템들에 대한 각 사용자들의 평가 정보를 이용한다. 일반적인 추천 방식은 사용자들 사이의 평가 정보를 비교하여 특정 사용자의 유사 사용자들을 추출하고, 유사 사용자들의 선호도를 기반으로 특정 아이템에 대한 특정 사용자

※ 제일저자(First Author) : 김준태

접수일자:2007년07월27일, 심사완료:2007년08월17일

* 동국대학교 컴퓨터공학과

jkim@dongguk.edu

** 나사렛대학교 멀티미디어학과

▣ 서울시 보유기술사업화 지원사업의 연구 결과임(과제번호:10711)

의 선호도를 예측하는 것이다. 협동적 여과 기법은 아이템의 내용 정보를 필요로 하지 않기 때문에 내용을 분석하기 어려운 음악이나 영화 같은 아이템을 추천할 수 있다.

그러나 이러한 일반적인 협동적 여과 방법은 실제 그 적용에 있어서 사용자 사이 혹은 아이템 사이의 유사도 계산에 총 사용자 수와 총 아이템 수에 비례하여 많은 시간이 걸리게 되는 확장성(scalability) 문제와, 사용자들의 아이템들에 대한 명시적인 선호도를 필요로 하므로 선호도(평가) 정보가 적을 경우 추천이 어려운 희소성(sparseness) 문제를 가지고 있다. 또한 사용자의 정적 성향만을 사용하여 아이템을 추천하므로, 환경에 따라 동적으로 변화하는 취향을 반영할 수 없다는 한계점이 있다.

본 논문에서는 이러한 문제들을 해결할 수 있는 그래프 기반 음악 추천 시스템을 소개한다. 그래프 기반 협동적 추천 기법은 사용자들의 암시적 선호 정보들로부터 음악 사이의 연관 관계를 가중치 그래프로 표현하고, 이 그래프를 이용하여 사용자에게 음악을 추천하는 것이다. 사용자의 암시적 선호 정보가 그래프 구조에 누적됨으로써 추천 실행 시간에 유사도 계산 등의 과정을 수행할 필요가 없이 빠른 추천을 할 수 있으며, 동시에 특정 사용자의 선호 정보가 희소하더라도 추천이 가능하게 된다.

또한 본 논문에서 소개하는 추천 시스템은 사용자의 정적 성향뿐만 아니라 시간과 장소에 따라 변경될 수 있는 동적 성향에 따른 추천 기능도 제공한다. 정적 성향이란 어떤 사용자가 가지고 있는 음악에 대한 기본적인 취향을 나타내고, 동적 성향이란 특정한 상황이나 분위기에 따라 유동적으로 변하는 성향을 의미한다. 본 논문에서 소개한 추천 시스템은 가장 최근에 선호한 N개의 음악을 바탕으로 추천 결과가 계속 변화하도록 하는 기능을 구현함으로써 동적 성향을 따라갈 수 있도록 설계되었다.

추천 서버는 자바로 개발되었으며 독립된 서버로 구성되어 클라이언트와 정해진 프로토콜에 따라 통신하도록 구현되었다. 이 추천 시스템을 이용하여 실제 사용자들의 음악 다운로드 데이터를 기반으로 정적인 추천과 동적인 추천을 수행하는 음악 추천 데모 사이트를 구축하였고, 실험을 통하여 추천의 정확도를 측정하였다.

2. 관련연구

협동적 여과 기반 추천 시스템은 많은 사용자로부터의 아이템들에 대한 평가를 이용하는 기법이다. 사용자 기반(user-based) 협동적 여과는 사용자들 사이의 평가 정보를 비교하여 유사 사용자들을 추출하고, 유사 사용자들의 선호도를 기반으로 사용자의 아이템 선호도를 예측한다.

사용자 유사도는 피어슨(Pearson) 상관관계나 벡터 유사도에 의해 계산된다. 유사도가 계산되고 나면 식 1에서와 같이 아이템에 대해 유사한 사용자들의 평가에 대한 가중 평균을 계산하여 사용자에게 대한 아이템 선호도를 예측한다. 식 1에서 $P_{a,i}$ 는 아이템 i 에 대한 사용자 a 의 선호도 예측 값이고, $s_{a,u}$ 는 사용자 a 와 사용자 u 사이의 유사도, $r_{u,i}$ 는 사용자 u 의 아이템 i 에 대한 평가 값이다. n 은 유사 사용자들의 총 수이다.

$$P_{a,i} = \frac{\sum_{u=1}^n (s_{a,u} \cdot r_{u,i})}{\sum_{u=1}^n s_{a,u}} \quad (1)$$

GroupLens[9]와 같은 영화 추천을 위한 협동적 여과 시스템에서는 사용자의 선호도 예측에 상관관계기반 기법이 사용되었고, 다양한 변형된 기법들이 추천 시스템의 정확도 향상을 위해 제안되었다. Breese[3], Herlocker 등[6]은 다양한 유사도 계산 방식과 유사도 가중치 방법에 대한 실험을 수행한 바 있으며, Billsus와 Pazzani[2], Sarwar 등[12]은 충분한 정보가 없을 때 사용자의 선호도 예측을 위해 속성 추출을 적용하는 방법과, SVD(Singular Value Decomposition)를 이용하여 사용자-아이템 선호도 행렬의 차원을 줄이는 방법 등을 제안하였다.

사용자 사이의 유사도를 비교하는 방법 이외에 Deshpande[5], Sarwar[13], Linden 등[10]에 의해 아이템 사이의 유사도를 이용하여 사용자의 증가에 따른 계산 복잡도 문제를 해결하면서 추천의 질을 높일 수 있는 아이템 기반(item-based) 협동적 여과 알고리즘이 제안되었고, Condliff[4], Popescul[11], Hoffman[7] 등에 의해 확률 이론을 바탕으로 하는 다양한 모델

기반(model-based) 추천 방법에 대한 연구도 수행되었다.

본 논문에서 소개하는 음악 추천 시스템은 아이템 기반 협동적 여과와 유사한 그래프 기반 협동적 추천 기법을 사용한다. 그래프 기반 협동적 추천은 사용자의 아이템에 대한 선호도 정보를 누적하여 아이템들 사이의 연관관계를 그래프로 표현함으로써 대용량 데이터에서 실시간 추천이 가능한 장점이 있다.

3. 그래프 기반 추천 알고리즘

이 장에서는 본 논문에서 사용하는 음악 추천 방식의 특징과 데이터 표현 방법 및 그래프기반 추천 알고리즘에 대하여 설명한다.

3.1 추천 방식의 특징

일반적인 음악 서비스 사이트에서 모든 사용자로부터 풍부한 명시적 선호도(평가) 정보를 기대하기는 어렵다. 따라서 원활한 음악 추천 서비스를 구현하기 위해서는 음원의 구매나 실행 횟수 등과 같은 암시적인(implicit) 정보만을 이용하여 실시간으로 음악을 추천할 수 있는 기법이 필요하다. 또한 수백만의 사용자와 아이템을 다루게 되므로 일반적인 사용자 유사도 기반, 혹은 아이템 유사도 기반의 협동적 여과 기법을 적용하기 어렵다.

본 논문에서 제안하는 그래프 기반 추천 기법은 다수 사용자들의 음악 구매 기록(download), 음악 실행 횟수(play count), 개인 음악 목록(play list) 등 암시적 선호 정보들로부터 음악 사이의 연관 관계를 가중치 그래프로 표현하여 저장하고, 이러한 그래프를 이용하여 추천 대상 사용자에게 대한 음악의 추천을 빠르게 수행하는 방법이다.

그래프 기반 추천 기법은 사용자의 암시적 선호 정보가 그래프 구조에 누적됨으로써 추천 실행 시간에 유사도 계산 등의 과정을 수행할 필요가 없어 추천 계산 시간이 사용자의 수나 아이템의 수에 비례하여 늘어나지 않는다. 또한 암시적인 선호 정보만으로 계산되는 아이템 사이의 연관 관계를 이용하므로 특정 사용자의 선호 정보가 희소하더라도 추천이 가능하게 되어, 종

래의 협동적 여과 방식의 단점인 확장성 문제와 희소성 문제를 해결할 수 있다.

그래프 모델을 사용하여 음악을 추천하는 과정은 크게 네 단계로 나눌 수 있다. 첫째, 각 음악 사이의 연관 관계를 그래프 모델로 표현하고 저장한다. 둘째, 추천 대상 사용자의 취향에 맞을 가능성이 있는 음악들을 저장된 그래프를 이용하여 찾아낸다. 셋째, 찾아진 음악들에 대하여 해당 사용자의 선호도 예측 값을 그래프 데이터를 이용하여 계산한다. 넷째, 산출된 예측 값을 이용하여 추천 대상 음악에 순위를 정하여 사용자에게 적합한 음악을 추천한다.

3.2 데이터의 표현

그래프에서 각 정점 M_i 는 각 음악을 나타내고, 정점사이의 간선 (M_i, M_j) 은 각 음악 사이의 연관성을 나타낸다. 각 정점에는 각 음악에 대한 사용자들의 선호도 총합 $C(M_i)$ 가 저장된다. $C(M_i)$ 는 M_i 에 대한 총 구매 수, 다운로드 수, 실행 회수 또는 이들의 조합 등이 될 수 있다. 각 간선에는 각 간선이 연결하는 두 정점에 해당하는 음악의 동일 사용자에서의 동시 출현 빈도 총합 $C(M_i, M_j)$ 가 저장된다. 이때 동시 출현 빈도는 동일 사용자에서 발생한 음악들의 쌍을 나열하여 이 쌍들의 수로서 계산한다.

어떤 사용자 u 가 새로운 음악 M_i 를 다운로드하거나, 선곡 리스트에 넣거나, 실행하게 되면 사용자 u 의 변경된 음악 선호도 값 $\Delta C_u(M_i)$ 을 결정한 다음 그래프 데이터의 해당 음악에 대한 선호도 총합 $C(M_i)$ 를 $C(M_i) + \Delta C_u(M_i)$ 로 변경하며, 해당 사용자의 선호 음악 리스트에 있는 다른 모든 음악들 M_j 에 대하여 그래프 데이터의 $C(M_i, M_j)$ 와 $C(M_j, M_i)$ 를 각각 변경한다. 이러한 작업이 사용자가 음악을 다운로드하거나 실행할 때마다 수행되어 음악들 사이의 연관관계를 표현하는 그래프가 만들어지고 지속적으로 갱신된다.

그래프에서 $C(M_i)$ 값이 클수록 많은 사용자로부터 선호되어지는 인기있는 음악이며, $C(M_i, M_j)$ 값이 클수록 두 음악은 동시에 선호될 가능성이 높은 음악이 된다. 이러한 그래프는 무방향 가중치 그래프(undirected weighted graph)로서 총 음악 수에 해당하는 개수의 리스트로 저장될 수 있다.

3.3 추천 알고리즘

추천 대상 사용자 u 가 선정되면 해당 사용자의 선호 음악 리스트를 읽어 온다. 추천 대상 사용자가 $M_1 \dots M_n$ 의 n 개 음악을 선호하는 것으로 알려졌다며, 각 음악에 대한 선호도가 $C_u(M_1) \dots C_u(M_n)$ 이라고 하자. 추천 대상 음악은 이 사용자가 이미 선호하는 것으로 알려진 음악들과 높은 연관성을 가지는 것으로서 이들은 연관관계 그래프로부터 얻을 수 있다. 즉, 추천 대상 음악은 $M_1 \dots M_n$ 과 간선으로 연결된 인접 정점에 해당하는 음악들로서, 정점 $M_1 \dots M_n$ 각각으로부터 인접 정점 리스트를 읽으면 얻을 수 있다. 이들을 $A_1 \dots A_m$ 이라 하자.

앞에서 검색한 각 음악 A_j 에 대한 추천 대상 사용자의 선호도 예측 값 $P_u(A_j)$ 를 해당 사용자가 이미 선호하는 것으로 알려진 음악 M_i 에 대한 선호도 및 M_i 와 A_j 의 연관도로부터 식 2 또는 3과 같이 계산할 수 있다. 식 2는 M_i 와 A_j 의 상호정보(Mutual Information) $MI(M_i, A_j)$ 를 이용하는 것이고, 식 3은 베이스의 정리(Bayes' Rule)를 이용하여 $P_u(A_j)$ 를 추정하는 것이다. 이러한 계산은 모두 그래프에 저장된 데이터로부터 계산될 수 있으며, 검색된 인접 음악의 수 m 만큼 수행된다.

$$P_u(A_j) = \sum_{i=1}^n C_u(M_i) MI(M_i, A_j) \quad (2)$$

$$P_u(A_j) = P(A_j | M_1, M_2, \dots, M_n) \quad (3)$$

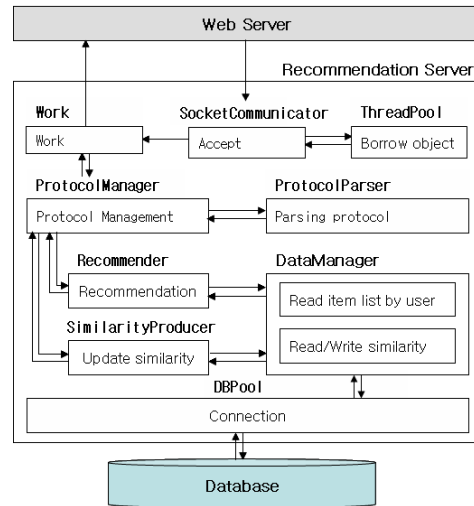
위의 식에 따라 $P_u(A_1) \dots P_u(A_m)$ 이 얻어지면, 이 값들을 내림차순으로 정렬하고, 추천하고자 하는 상위 K 개의 음악을 선택함으로써 최종적인 추천 음악 리스트가 완성된다. 이러한 추천의 계산 시간은 각 사용자의 평균 선호 아이템 수와 그래프의 정점 당 평균 간선 수에 의해 결정되며, 총 사용자 수나 총 아이템 수에는 영향 받지 않는다.

4. 음악 추천 시스템의 구성

본 논문에서 소개하는 음악 추천 시스템은 추천 계산을 수행하는 추천 서버와, 사용자의 추천 요청을 추천 서버에 전달하고 추천 결과를 사용

자에게 제공하는 웹 서버(추천 서버의 클라이언트)로 구성된다. 추천 서버는 자바로 구현되었으며, 클라이언트인 웹서버와 정해진 프로토콜을 이용하여 통신한다.

4.1 시스템 구성



(그림 1) 추천 시스템 구성도

추천 서버는 사용자의 추천 요청에 따라 아이템을 추천하며, 사용자의 평가 정보를 이용하여 그래프의 유사도 정보를 갱신시킨다. 추천 서버는 소켓 통신기(Communicator), 프로토콜 관리자(ProtocolManager), 데이터 관리자(DataManager), 추천기(Recommender), 유사도 계산기(SimilarityProducer)로 구성된다. (그림 1)은 추천 시스템 구성도를 나타낸다.

소켓 통신기는 웹 서버와 통신하는 모듈이며, 스레드 풀은 멀티-스레드 방식을 사용하여 여러 개의 추천 요청을 처리하기 위한 모듈이다. 프로토콜 관리기는 요청 내용이 추천을 위한 것인지 유사도 갱신을 위한 것인지 등을 판단하는 모듈로서, 그 결과에 따라 추천기 또는 유사도 계산기가 호출된다. 추천기는 사용자의 요청 내용에 따라 추천 알고리즘을 수행하여 아이템을 추천하는 모듈이고, 유사도 계산기는 사용자의 평가 정보를 이용하여 아이템들 간의 유사도를 갱신시키는 모듈이다. 데이터 관리기는 데이터베이스에 접근하기 위한 모듈로써 접속의 부하를 줄이

기 위해서 데이터베이스 풀을 사용한다. 데이터 베이스는 사용자 ID, 성별, 나이와 같은 사용자 정보와 곡명, 아티스트명, 앨범명 등의 음악 정보를 저장하며, 추천 알고리즘에 이용되는 유사도 정보를 그래프 형태로 저장한다.

4.2 통신 프로토콜

추천 서버는 별도로 정의한 프로토콜에 의하여 추천 요청이나 선호도 갱신 요청을 받도록 설계되었다. (그림 2)에 통신 프로토콜의 기본 구조와 아이템 추천, 선호도 갱신 요청, 그리고 동적 추천을 위한 N개 아이템 기반 추천 요청 프로토콜을 나타내었다.

통신 프로토콜은 크게 헤더와 몸통 부분으로 구성된다. 프로토콜의 헤더 부분은 프로토콜에 대한 기본 정보를 정의하는 부분으로 각각의 프로토콜에 대하여 공통된 구조를 가지며 몸통 부분은 프로토콜의 종류에 따라 그 구조가 달라진다. 헤더의 첫 3바이트는 프로토콜 이름을 저장하고, 다음 2바이트는 프로토콜의 길이를 저장한다. 다음 3바이트는 해당 프로토콜이 요청을 위한 프로토콜('REQ')인지 응답을 위한 프로토콜('RES')인지 프로토콜의 타입을 저장한다. 다음 2바이트는 해당 프로토콜이 아이템 추천 프로토콜('RI')인지 선호도 갱신 프로토콜('PS')인지 해당 프로토콜의 행위를 명시한다. 다음 18바이트는 추천기 이름을 적는 공간으로 추천 엔진이 어떤 추천 방식을 사용하는지를 명시한다. 몸통 부분은 각 프로토콜의 종류에 따라 해당 항목들이 저장된다.

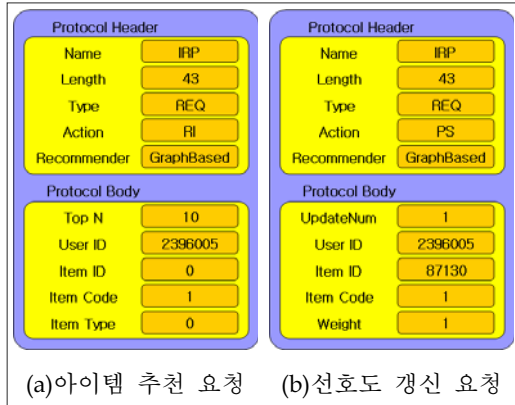
아이템 추천 프로토콜의 몸통 부분은 추천 받을 아이템 수를 의미하는 TopN, 사용자 ID, 아이템 ID, 추천 받을 Item Code(곡, 앨범, 또는 아티스트), 추천 받을 Item Type(팝, 클래식 등)을 저장한다. 이때 사용자 ID에 데이터가 있다면 해당 사용자의 선호도에 기반한 추천을 수행하고 아이템 ID에 데이터가 있다면 해당 아이템과 유사한 아이템들을 추천한다. 선호도 갱신 프로토콜은 갱신할 아이템 수를 의미하는 UpdateNum, 사용자 ID, 아이템 ID, Item Code, 갱신할 가중치 값인 Weight를 저장한다. 동적 성향에 기반한 추천을 위한 N개 아이템 기반 추천 요청 프로토콜은 추천 받을 아이템 수와 Item Code, Item Type, 아이템 수, 최근 N개 아

이템의 ID와 평가값을 저장한다.



(그림 2) 프로토콜 구조

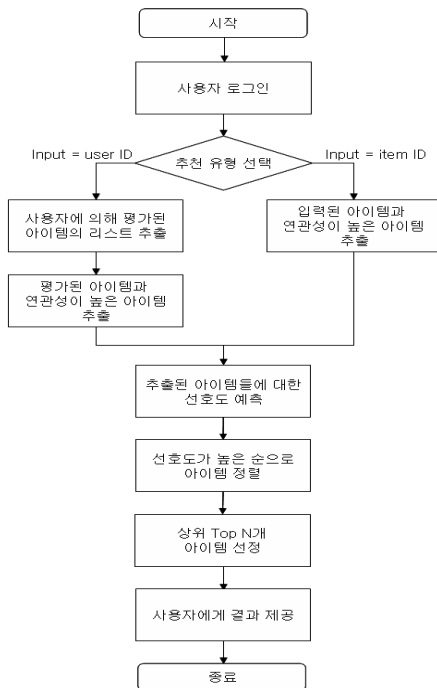
(그림 3)은 프로토콜의 사용 예이다. 그림 3-(a)는 ID가 2396005인 사용자에게 장르 제한 없이(Item Type=0) 곡을(Item Code=1) 10(TopN=10)곡 추천해달라고 요청하는 것이고, 그림 3-(b)는 ID가 2396005인 사용자가 ID가 87130인 곡(Item Code=1)을 가중치 1만큼 선호했다는 내용으로 사용자 선호도와 그래프 데이터를 갱신하도록 요청하는 것이다.



(그림 3) 프로토콜 사용 예

4.3 정적 성향에 기반한 추천

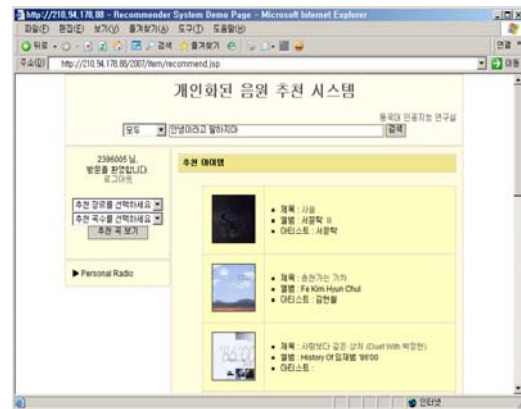
본 논문에서 소개하는 추천 시스템은 정적 성향에 기반한 추천과 동적 성향에 기반한 추천을 모두 제공한다. 정적 성향에 기반한 추천에는 크게 두 가지가 있다. 하나는 특정 사용자에게 그 사용자의 선호도 정보를 이용하여 음악을 추천하는 것이고, 다른 하나는 특정 음악에 대하여 그 음악과 유사한 음악을 추천하는 것이다.



(그림 4) 정적 성향에 기반한 추천과정

특정 사용자에게 대한 추천은 해당 사용자의 아이템에 대한 다운로드, 스트리밍 등 암시적 선호 정보를 활용하여 아이템을 추천하는 방식으로서, 먼저 사용자에게 의해 평가된 아이템의 리스트를 추출하고, 이 아이템들을 이용하여 그래프 모델에서 연관성이 높은 추천 대상 아이템들을 추출한 다음, 추천 대상 아이템들에 대하여 선호도를 예측하여 선호도 예측값에 따라 상위 N개의 아이템을 추천한다. 특정 음악에 대한 추천은 그래프 모델에서 해당 아이템과 연관성이 있는 추천 대상 아이템들을 추출하고, 연관도에 따라 상위 N개의 아이템을 추천한다.

(그림 4)는 이러한 정적 성향에 기반한 추천 과정을 나타내며, (그림 5)는 추천 데모 시스템에서 특정 사용자에게 대한 음악 추천을 수행한 결과 화면을 나타낸다.



(그림 5) 특정 사용자에게 대한 추천 데모 화면

4.4 동적 성향에 기반한 추천

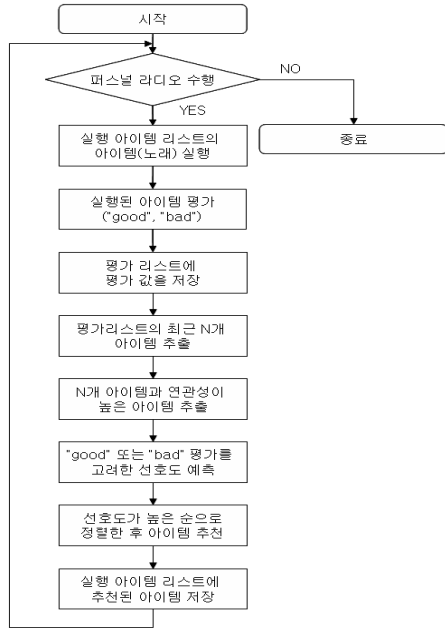
동적 성향이란 특정한 상황에 따라 유동적으로 변화하는 사용자의 성향을 뜻한다. 예를 들어 평소에 댄스 음악을 즐겨 듣는 사용자일지라도 때로는 감상적인 기분 때문에, 혹은 조용한 음악이 어울리는 시간 또는 장소이기 다른 분위기의 음악을 원할 수도 있을 것이다. 동적 성향에 기반한 추천은 사용자의 변화하는 현재 취향에 맞춰 아이템을 추천하는 기능이다.

본 논문에서 소개하는 추천 시스템은 '퍼스널 라디오'라는 기능을 통하여 사용자에게 동적 성향에 기반한 추천을 수행한다. 퍼스널 라디오는 실시간 노래 추천 시스템으로서 최근에 선호한

N개 아이템을 기반으로 사용자의 현재 취향에 따라 아이템을 추천한다. 이를 위해 퍼스널 라디오는 임시 선호도 평가 리스트와 실행 아이템 리스트를 유지한다. 임시 선호도 평가 리스트는 퍼스널 라디오에서 실행된 아이템에 대한 사용자의 평가 정보를 저장한 리스트로 사용자의 동적 성향을 나타내며, 실행 아이템 리스트는 이러한 성향에 따라 추천된 앞으로 실행될 아이템을 저장한 리스트이다.

임시 선호도 평가 리스트에 저장된다. 현재 곡이 끝나거나 사용자에게 의해 중단되면 실행 아이템 리스트에 있는 다음 곡이 순서에 따라 실행되며, 퍼스널 라디오는 임시 선호도 평가 리스트에 있는 최근 N개의 아이템들을 기반으로 선호도 예측값이 가장 높은 1개의 새로운 곡을 추천 받아 실행 아이템 리스트에 다음 곡으로 포함시킨다.

(그림 6)은 이러한 동적 성향에 기반한 추천 과정을 나타내며, (그림 7)은 추천 데모 시스템에서 퍼스널 라디오를 실행하는 화면을 나타낸다. 퍼스널 라디오를 들으면서 현재 실행되는 음악에 대한 평가를 지속적으로 입력하면 이러한 정보에 따라 연속적으로 추천되는 곡들이 사용자의 현재 취향에 맞추어 가게 된다.



(그림 6) 동적 성향에 기반한 추천 과정



(그림 7) 퍼스널 라디오의 실행 화면

퍼스널 라디오는 먼저 사용자의 현재 선호 정보를 바탕으로 3 개의 곡을 추천받아 실행 아이템 리스트를 만들고, 첫 번째 곡을 실행한다. 사용자가 실행중인 아이템에 대하여 Good 또는 Bad로 평가를 수행하면, 사용자의 평가 정보가

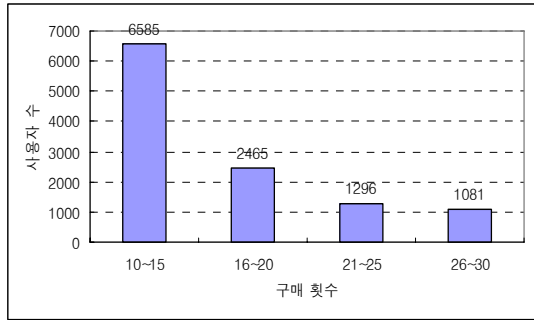
5. 실험

본 논문에서 제안한 그래프 기반 추천 시스템의 추천 정확도를 측정하기 위해 음악 서비스 사이트로 운영되었던 'FunCake' 사이트의 실제 데이터를 사용하여 실험을 수행하였다. FunCake은 온라인 음악 서비스 사이트로서 스트리밍 서비스와 다운로드 서비스를 제공한다. FunCake 사이트에서 실험에 사용할 수 있도록 제공받은 실험용 데이터는 2005년 7월부터 2006년 3월까지 FunCake 사이트에서 판매된 곡들에 대한 데이터로, 사용자의 개인정보를 제외한 사용자 ID, 곡 ID, 구매일자가 나타나 있다. 전체 데이터 집합에 있는 사용자 총수는 68,394명이고, 곡의 총수는 40,605개이며, 판매 건수는 553,118건이다.

본 논문에서는 추천 실험을 위해 구매 횟수가 10개 이상 30개 이하인 사용자 중 1,000명을 랜덤하게 추출하여 데이터 집합을 구성하였다. 실험 데이터 집합에서 사용자의 총 구매 아이템 수는 16,135개이며, 일인당 평균 구매 횟수는 16.14, 표준 편차는 6.47이다. 실험 데이터 집합에서 선호도 정보가 나타난 비율은 0.34%로, 희소도는 매우 높다. 그림 8은 전체 데이터에서 구매 횟수별 사용자 수 분포를 나타낸다. 구매 횟수가 10개 이상 15개 이하인 사람들이 6,505명으로 가장 높은 비율을 차지했다.

추천 정확도 실험은 아이템에 대한 선호도 계산 방법으로 상호정보를 사용한 경우(MI)와 베

이스의 정리를 사용한 경우(Bayesian) 등 두 가지에 대하여 수행하였다. 실험 방법은 다음과 같다.



(그림 8) 구매 횟수에 대한 사용자 분포

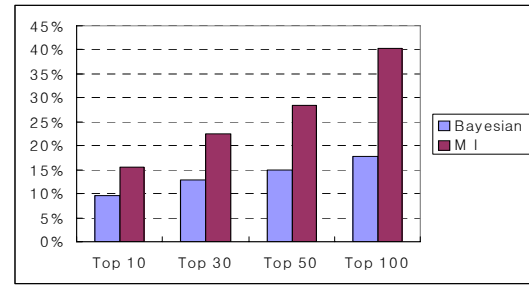
우선 실험 데이터를 훈련 집합(training set)과 검증 집합(test set)으로 나누고, 훈련 집합의 데이터를 이용하여 각 음악 사이의 연관 관계를 그래프 모델로 표현하고 저장한다. 그런 다음 검증 집합의 사용자들에 대하여 각 사용자가 구매한 음악 중 하나를 제외하고 나머지 데이터를 바탕으로 사용자에 대한 추천을 수행한다. 추천된 N개의 음악 리스트 안에 사용자가 실제 구매한(선호하는) 음악이 포함되어 있으면 적중(hit)이라고 하고, 추천 결과에 대한 정확도는 검증 데이터 집합의 전체 구매 음악 총수에 대한 적중 횟수 비율인 적중률(hit ratio)로 나타낸다. 적중률 식은 식 4와 같다.

$$Hit\ Ratio = \frac{total\ number\ of\ hits}{total\ number\ of\ tests} \quad (4)$$

최종 정확도 값을 10-fold cross validation을 사용하여 구하였다. 즉, 전체 실험 데이터 집합을 동일한 크기의 10개 실험 데이터로 나누고, 나누어진 10개의 실험 데이터에서 하나의 실험 데이터 집합을 검증 집합으로 선정한 후, 나머지 9개의 실험 데이터 집합을 훈련 집합으로 사용하여 실험하고, 이와 같은 방법을 서로 다른 10개의 실험 데이터 집합에 모두 적용하여 결과를 평균하였다.

실험 결과는 (그림 9)와 같다. 추천 아이템 수 N을 10개, 30개, 50개, 100개로 했을 때, 베이지

안 방법의 추천 정확도는 평균 13.8%이었고, 상호정보 방법의 추천 정확도는 평균 26.7%로서 상호정보 방법을 적용한 추천이 베이지안 확률을 적용한 추천 보다 더 우수한 추천 성능을 나타내었다. 특히 상호정보 방법을 적용한 추천은 추천 아이템 수가 많아질수록 적중률이 크게 높아지는 것을 알 수 있다. 본 실험 데이터에서 상호정보 방법을 적용한 음악 추천의 top-10 추천 정확도는 15.5% 이었다.



(그림 9) 음악 추천의 정확도

6. 결론

본 논문에서는 사용자의 개인적 취향에 맞는 음악을 추천할 수 있는 음악 추천 시스템을 소개하였다. 본 논문에서 소개한 시스템은 그래프 기반 협동적 여과 방식을 사용하여 추천을 수행한다. 그래프 기반 추천 방식은 각 음악 사이의 유사도를 그래프로 저장하고 관리함으로써 추천 계산 시간이 사용자 수나 음악 수에 비례하여 늘어나지 않는 추천 방법이다. 또한 이 추천 시스템은 특정 사용자의 정적인 성향뿐만 아니라 동적인 성향에 맞는 추천도 수행할 수 있는 기능을 제공하도록 구현되었다.

추천 시스템은 자바로 개발되었으며 독립된 서버로 구성되어 클라이언트와 정해진 프로토콜에 따라 통신하도록 구현되었다. 이 추천 서버를 이용하여 실제 사용자들의 음악 다운로드 데이터를 기반으로 정적인 추천과 동적인 추천을 수행하는 음악 추천 데모 사이트를 구축하였고, 음악 서비스 사이트의 실제 데이터로 실험을 수행하였다. 실험 결과 상호정보를 이용한 음악 추천 방법이 베이지안 방법보다 높은 추천 정확도를 보였으며 top-10 추천에서의 적중률은 15.5%였

다.

향후 협동적 추천 방식에 음악에 대한 콘텐츠 분석을 조합하는 방법을 연구하면 보다 좋은 추천 결과를 얻을 수 있을 것이다. 또한 본 추천 시스템을 실제 음악 사이트의 개인화 서비스에 활용하여, 추천 결과에 대한 사용자들의 주관적인 평가를 다양한 측면에서 분석해볼 필요가 있다.

참 고 문 헌

[1] G. Adomavicius, A. Tuzhilin, "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, Vol. 17, No. 6, 2005.

[2] D. Billsus and M. J Pazzani, "Learning Collaborative Information Filters," Proceedings of the 15th International Conference on Machine Learning, 1998.

[3] J. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, 1998.

[4] M. Condliff, D. Lewis, D. Madigan, and C. Posse, "Bayesian Mixed Effects Models for Recommender Systems," Proceedings of ACM SIGIR '99 Workshop on Recommender Systems, 1999.

[5] M. Deshpande and G. Karypis, "Item-Based Top-N Recommendation Algorithms," ACM Transaction on Information Systems, Vol.22, No.1, 2004.

[6] J. Herlocker, J. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering," Proceedings of ACM SIGIR '99, 1999.

[7] T. Hoffman, "Latent Semantic Models for Collaborative Filtering," ACM Transaction on Information Systems, Vol. 22, No. 1, 2004.

[8] H. Kim, J. Kim, and J. L. Herlocker, " Feature-Based Prediction of Unknown Preferences for Nearest-Neighbor Collaborative Filtering," Proceedings of ICDM'04, 2004.

[9] J. Konstan, B. Millr, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News," Communications of the ACM, Vol.40, No.3, 1997.

[10] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering", IEEE Internet Computing, February, 2003.

[11] A. Popescul, L. Ungar, D. Pennock, and S. Lawrence, "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments," Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence, 2001.

[12] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Application of Dimensionality Reduction in Recommender System-A Case Study," Proceedings of the ACM WebKDD Workshop, 2000.

[13] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," Proceedings of the 10th International WWW conference, 2001.

[14] J. Schafer, J. Konstan, and J. Riedl, "Recommender System in E-Commerce," Proceedings of the ACM Conference on Electronic Commerce, 1999.

김 준 태

1986년 : 서울대학교 제어계측공학과(공학사)
 1990년 : 미국 Univ. of Southern California, 전기공학과(공학석사)
 1993년 : 미국 Univ. of Southern California, 컴퓨터공학과(공학박사)



1995년~현재 : 동국대학교 교수
 2003년~2004년 : 미국 Oregon Stat Univ. 방문교수
 관심분야 : 인공지능, 기계학습, 데이터마ining, 추천 시스템

김 형 일

2001년 : 동국대학교 컴퓨터공학과(공학석사)
 2004년 : 동국대학교 컴퓨터공학과(공학박사)



1996년~1998년 : (주)경기은행
 2005년~2006년 : 동국대학교 컴퓨터공학과 IT교수 (정보통신부)
 2007년~현재 : 나사렛대학교 멀티미디어학과 교수
 관심분야 : 지능형시스템, 추천시스템, 기계학습, 게임