

협업 필터링을 이용한 IPTV-VOD 프로그램 추천 시스템에 대한 연구

선철용[†], 강용진^{**}, 박규식^{***}

요 약

본 연구는 IPTV 환경에서 사용자의 취향에 맞는 VOD 프로그램을 추천할 수 있는 시스템을 새로이 제안하였다. 추천 알고리즘으로는 협업 필터링 기법을 사용하였다. 사용자의 프로그램 선호 취향을 나타내는 사용자 프로파일(user profile)은 사용자와 유사한 취향의 이웃 사용자들의 프로그램 선호도와 중분류 선호도 그리고 사용자 유사도를 감안하여 1주 단위로 갱신하였다. 제안 시스템의 성능평가를 위해 시청률 조사기관인 닐슨 리서치의 24주분 지상파 및 케이블 방송 시청 데이터를 IPTV 형식에 맞게 재구성하여 사용하였으며, 다양한 실험을 통해 그 실용성을 입증하였다. 실험결과 사용자 유사도 가중치를 사용하며, 그룹 크기가 5명 그리고 추천 프로그램 수가 5개 일 때 최적의 성능을 나타내었다.

A Study of IPTV-VOD Program Recommendation System using Collaborative Filtering

Chul-Yong Sun[†], Yong-Jin Kang^{**}, Kyu-Sik Park^{***}

ABSTRACT

In this paper, a new program recommendation system is proposed to recommend user preferred VOD program in IPTV environment. A proposed system is implemented with collaborative filtering method. For a user profile which describes user program preference, a program preference, sub-genre preference, and US(user similarity) weight of the user neighborhood is averaged and updated every week. In order to evaluate system performance, real 24-weeks cable TV watching data provided by Nilson Research Corp. are modified to fit for IPTV broadcasting environment and the simulation result shows quite comparative quality of recommendation. The experimental results optimum performance when user similarity based weighting, five person per group and five recommendation programs are used.

Key words: IPTV, Recommendation System(추천 시스템), Collaborative Filtering(협업 필터링)

1. 서 론

IPTV(Internet Protocol TV)는 초고속 인터넷망을 통한 방송·통신 융합 서비스로 사용자가 언제

어디서든 자신의 취향에 맞는 프로그램 콘텐츠를 VOD(Video On Demand) 서비스를 통해 제공받을 수 있는 장점으로 향후에도 지속적인 발전이 기대되는 서비스이다.

※ 교신저자(Corresponding Author): 박규식, 주소: 경기도 용인시 수지구 죽전동 126번지 단국대학교 미디어센터 414호(448-701), 전화: 031-8005-3252, FAX: 031-8005-3227, E-mail: kspark@dankook.ac.kr
접수일: 2009년 12월 29일, 수정일: 2010년 4월 24일
완료일: 2010년 7월 12일

[†] 정회원, 단국대학교 컴퓨터학부 컴퓨터과학과 석사과정 (E-mail: suchyo@dankook.ac.kr)

^{**} 정회원, 단국대학교 컴퓨터학부 컴퓨터과학과 석사과정 (E-mail: rkddywls@dankook.ac.kr)

^{***} 종신회원, 단국대학교 공과대학 컴퓨터학부 교수

※ 이 연구는 단국대학교 대학원 연구보조장학금의 지원으로 이루어진 것임

최근까지 추천 시스템에 관련된 연구는 크게 내용 기반 필터링(CBF, Content-Based Filtering), 협업 필터링(CF, Collaborative Filtering), 그리고 이 둘을 결합한 복합 필터링(HF, Hybrid Filtering), 인구 통계학 정보를 이용한 인구통계 필터링(DF, Demographic Filtering) 등으로 구분할 수 있다.

내용기반 필터링(CBF)은 사용자가 과거에 선호 하였던 프로그램들을 미래에도 선호할 가능성이 높다는 사실을 이용한 방식으로 사용자가 과거에 관심을 보였던 것과 유사한 프로그램을 추천할 때는 우수한 성능을 보이지만 사용자의 잠재적 흥미를 유발할 수 있는 다른 프로그램을 추천할 수 없는 단점이 있다. 협업 필터링(CF)은 사용자의 취향과 유사한 다른 사용자들의 취향을 참고하여 프로그램에 대한 선호도를 예측함으로써 내용기반 필터링이 제공할 수 없는 다양하고 유익한 프로그램들을 추천할 수 있는 장점이 있지만 사용자에 대한 충분한 데이터가 수집되어야만 정상적인 추천서비스가 가능하다는 단점이 있다. 이외에도 협업 필터링과 내용기반 필터링의 장, 단점을 상호 보완한 복합 필터링(HF) 시스템과 사용자의 나이, 성별, 국적 등의 인구 통계학 정보를 기준으로 프로그램을 추천하는 인구통계 필터링(DF) 방법 등이 있다[1].

기존 연구로 Amazon.com[2]은 아이템 기반 협업 필터링을 이용해 웹상에서 도서 추천 시스템을 구축 하였으나, 사용자의 구매 이력이나 평점 기록들이 적을 때 추천 성능이 떨어지며, 사용자가 구매 내역에 대해 일일이 평점을 기록해야하는 불편함을 가지고 있다. DTV(Digital TV) 프로그램 추천에 관한 연구로 Zimmerman[3]은 베이지안과 결정트리를 이용해 선호정보를 학습하는 암시적 추천과 사용자의 입력 정보를 이용한 명시적 추천으로부터 최종 프로그램 선호도를 계산하는 복합추천 기법을 제안하였다. 논문 [4]에서는 사용자의 시청 패턴 변화를 감안하기 위해 최근성과 시청 횟수를 고려한 시청 패턴 가중치를 프로그램 선호도 연산에 반영함으로써 DTV 프로그램 추천 시스템의 효율을 향상 시켰다. 반면, 논문 [5]에서 KT는 실제 자사 IPTV 서비스(QOOK TV)의 시청 데이터를 이용하여 VOD 추천 시스템의 기초 성능을 실험, 평가하였으나, 모든 프로그램에 대한 추천이 아닌 영화에 국한된 추천 시스템만을 다루고 있다.

이상에서 살펴본 바와 같이 기존의 프로그램 추

천 시스템은 주로 DTV에 국한된 연구가 대부분으로 이는 최근까지도 IPTV-VOD에 대한 시청 데이터가 공개되지 않고 있기 때문이다. 프로그램 추천 또한 프로그램 선호도(program preference) 연산이나 개별 추천 필터링 알고리즘의 적용 방법 등에 관련된 것이 대부분으로 실용적인 측면에서의 IPTV-VOD 프로그램 추천에 대한 연구는 아직 미진하다할 수 있다.

본 연구는 IPTV-VOD 방송 환경에서 사용자 개인의 프로그램 선호 성향을 반영한 개인화된 프로그램 추천 시스템을 제안하였으며 다음과 같이 기존 연구와 차별성을 갖는다.

(1) 논문[4]에서는 내용 기반 필터링을 사용하여 사용자에게 방송 콘텐츠의 프로그램을 추천하였다. 제안 시스템은 협업 필터링(CF) 기반의 추천 시스템으로 사용자와 유사한 취향의 이웃 사용자들을 참조하여 다양한 콘텐츠의 프로그램을 추천할 수 있다.

(2) 추천 시스템의 핵심이랄 수 있는 프로그램 선호도 연산 시 사용자의 선호도 변화와 연속성을 감안하기 위해 최근 2주 동안 연속시청 프로그램에 대한 가중치를 부여하였으며 이를 1주 단위로 갱신하였다. 논문[6]에서는 같은 협업 필터링 기반의 시스템으로 사용자의 전체 시청한 프로그램에 대해 이웃 사용자들과의 그룹을 형성하였다. 그래서 사용자의 선호도 변화와 연속성을 감안하는 점이 부족하였다.

(3) 본 연구에서는 최근까지도 공개된 IPTV-VOD 방송 및 시청 데이터가 없기 때문에 시청률 조사기관인 닐슨리서치의 지상파 및 케이블 방송 24주분 시청 데이터를 IPTV-VOD형식에 맞게 재구성하여 제안 시스템의 성능평가에 사용하였다. 닐슨리서치의 방송 프로그램은 총 9개 대분류에 48개 중분류와 138개 소분류로 구분되어 있으며, 본 연구에서는 뉴스, 날씨, 교통 정보 등 일회성 방송을 제외한 나머지 프로그램들을 총 13개 대분류와 60개 중분류로 재구성하여 거의 모든 장르의 프로그램에 대한 추천이 가능하도록 시스템을 구축하였다. 논문[5]에서는 영화 VOD의 한정된 데이터로 실험하였고, 논문[6]에서는 닐슨에서 제공한 데이터를 전혀 없이 사용하여 실시간 방송과 같은 환경에서 실험하였다. 논문[7]에서는 영화 대역 데이터를 사용하였다.

본 논문은 다음과 같이 구성된다. 먼저 2장에서는 제안한 프로그램 추천 시스템의 전체 구성과 세부

표 1. 프로그램 DB의 메타데이터 예

프로그램 ID	프로그램 이름	방영시간(분)	대분류	중분류
3145	뽕뽕 그림책버스	25	20	2001
4235	치로와 친구들	15	20	2002
613	방귀대장 뽕뽕이	15	20	2002
134	주주클럽	50	65	6501
65614	유럽축구 핫! 골!	25	40	4001

표 2. 시청정보 DB의 구축 예

사용자ID	프로그램ID	대분류	중분류	시청주(week)
103487801	3145	20	2001	1
103487801	4235	20	2002	1
103487801	134	65	6501	1
103487801	613	20	2002	1

모듈을 살펴보고, 3장에서는 제안 시스템의 실험 환경 및 결과 분석을, 마지막으로 4장에서는 결론과 향후 연구에 대한 고찰로 끝을 맺는다.

2. 제안 IPTV-VOD 추천 시스템 구조

그림 1은 본 연구에서 제안한 IPTV-VOD 프로그램 추천 시스템의 구조를 나타낸 것으로 크게 셋톱박스(Set-Top Box)와 서버(Server)로 이루어진다. 그림의 셋톱박스는 서버로부터 VOD 프로그램 리스트와 관련 정보, 그리고 추천 프로그램 리스트를 받아

UI(User Interface)를 통해 이들을 사용자에게 제공하고, 이에 대한 사용자의 시청 기록을 서버로 전해 주는 역할을 한다. 서버는 시청정보 DB, 사용자 모델 DB, VOD 프로그램 DB, 그룹DB, 사용자 모델 관리 엔진, 협업 필터링 추천 엔진, 사용자 가중치 계산 모듈로 구성된다. 시청정보 DB는 셋톱박스로부터 전송된 프로그램 시청 기록을 사용자 별로 저장하며, 사용자 모델 관리 엔진은 이 시청 기록을 바탕으로 사용자 별 중분류 선호도와 프로그램 선호도를 일주일 단위로 갱신해 사용자 모델 DB에 저장한다. 한편, VOD 프로그램 DB에는 현재 방영되고 있는 모든

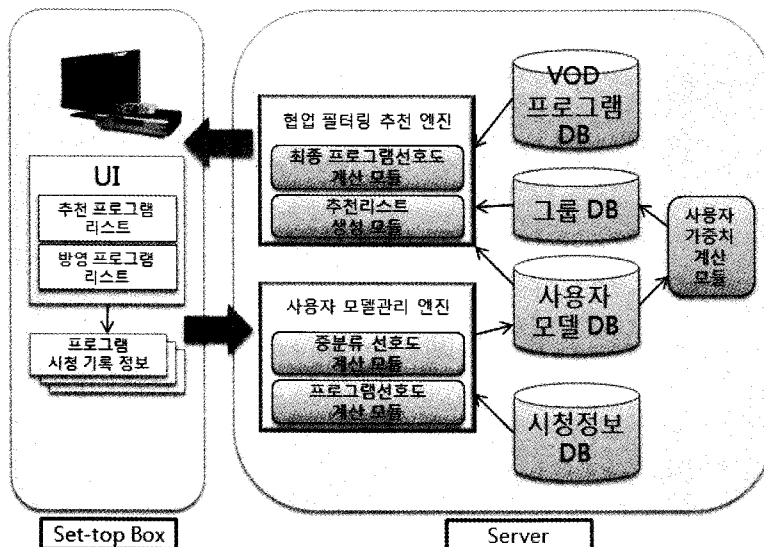


그림 1. 제안 추천 시스템 구조

VOD 프로그램 정보들이 저장되며 이 또한 일주일 단위로 갱신된다. 실제 프로그램 추천은 사용자 모델 DB의 중분류와 프로그램 선호도에 기반한 협업 필터링(CF) 기법을 이용해 VOD 프로그램 DB를 대상으로 최종 프로그램 선호도가 가장 높은 5개 프로그램만을 선정해 셋톱박스로 전송한다.

2.1 서버 DB

서버 DB는 프로그램 DB, 시청정보 DB, 사용자 모델 DB, VOD 프로그램 DB, 그룹 DB 등으로 구성된다. 프로그램 DB는 표 1과 같이 IPTV-VOD 방송 프로그램을 프로그램 ID, 이름, 방영시간, 대분류, 중분류로 구성된 메타데이터(Metadata)로 저장한다. 본 연구에서는 총 13개의 대분류와 60개 중분류로 프로그램들을 구분하였다.

서버로부터 전송받은 VOD 프로그램 리스트나 추천 리스트를 대상으로 한 사용자의 실제 시청 정보는 해당 사용자와 시청 내역, 프로그램 정보를 기반으로 표 2와 같이 재구성된다. 최종 시청기록 정보는 사용자가 방영시간 중 총 10% 이상 시청한 프로그램만을 대상으로 사용자 ID, 프로그램 ID, 대분류, 중분류, 시청 주(week)로 재구성되어 서버의 시청정보 DB에 저장된다.

반면, 사용자 모델 DB는 사용자 모델 관리엔진에서 시청 정보 DB를 이용해 1주일 단위로 갱신하는 프로그램 선호도와 중분류 선호도 값(2.2절에서 설명)을 저장한다.

2.2 사용자 모델 관리 엔진

사용자 모델은 사용자의 프로그램 시청 패턴을 파악할 수 있는 가장 중요한 단서로서 표 2의 시청 정보 DB를 이용해 시청 프로그램에 대한 선호도와 중분류 선호도를 계산한다. 사용자의 프로그램 선호도는 최근의 시청 변화 패턴과 특정 프로그램에 대한 시청 연속성을 반영하기 위해 시청 정보 DB 중 최근 2주 동안의 기록만을 대상으로 연속 시청에 대한 가중치를 부여하여 선호도를 계산하고 이를 1주일 단위로 갱신하였다. 반면, 시청 프로그램의 특성을 나타내는 중분류 선호도는 1주 단위로 계산되어 갱신된다.

2.2.1 프로그램 선호도 연산 및 갱신

사용자 u_i 의 프로그램 선호도는 최근 2주 동안 시

청한 프로그램을 대상으로 프로그램 j 의 시청 빈도수 p_j ($j=1,2,3,\dots,N$)(N =총 프로그램 개수)를 구한 후 이를 최대값 $\max_j(p_j)$ 으로 정규화하여 계산한다. 또한, 2주 동안 프로그램 시청 연속성에 대한 가중치를 반영하기 위해 만약 사용자가 2주 연속 해당 프로그램을 시청하였다면 가중치를 $w_j=1$ 로, 2주 중 1주만 시청하였을 경우에는 가중치를 $w_j=1/2$ 로 하였다. 수식(1)은 사용자 u_i 의 프로그램 j 에 대한 선호도 점수 PPS(Program Preference Score) $PPS_j^{u_i}$ 를 나타낸 것으로 0~1 사이의 값을 갖는다.

$$PPS_j^{u_i} = \frac{p_j \times w_j}{\max_j(p_j)}, \quad j=1,2,3,\dots,N \quad (1)$$

프로그램 선호도 갱신은 시간에 따른 사용자 시청 패턴 변화를 반영하기 위해 해당 주 바로 1주 전의 프로그램 선호도를 반영하여 계산된다. 즉, 현재 m 주차(m^{th} week)에 사용자 u_i 의 프로그램 j 에 대한 선호도는 m 주차(m^{th} week)의 프로그램 선호도와 $m-1$ 주차($(m-1)^{th}$ week) 프로그램 선호도의 평균 값으로 주어지며 이를 다음 수식(2)과 같이 표현할 수 있다.

$$PPS_{update_j}^{u_i}(week_m) = \frac{PPS_j^{u_i}(week_{m-1}) + PPS_j^{u_i}(week_m)}{2} \quad (2)$$

$PPS_j^{u_i}(week_{m-1}), PPS_j^{u_i}(week_m)$ 은 각각 $PPS_j^{u_i}$ 를 이용해 계산된 사용자 u_i 의 프로그램 j 에 대한 $m-1$ 주차와 m 주차 선호도를 나타내며 $PPS_{update_j}^{u_i}(week_m)$ 은 전 주의 프로그램 선호도를 반영해 갱신된 m 주차 프로그램 선호도를 나타낸다.

2.2.2 중분류 선호도 연산 및 갱신

중분류 선호도는 사용자가 어느 분류의 프로그램을 선호하는 가를 측정하는 모듈로 최근 1주일 동안의 시청 기록만을 대상으로 프로그램 중분류 선호도와 갱신이 이루어진다. 사용자 u_i 의 시청 프로그램에 대한 중분류 선호도는 최근 1주 동안 시청한 프로그램을 대상으로 각 프로그램에 해당하는 중분류 k 의 시청 빈도수 g_k ($k=1,2,3,\dots,L$)(L =총 중분류 개수=60)를 계산한 후 이를 최대값 $\max_k(g_k)$ 으로 정규화하여 구한다. 다음 수식(3)은 사용자 u_i 의 프로그램

중분류 k 에 대한 선호도 점수 CPS(Class Preference Score) $CPS_k^{u_i}$ 를 나타낸 것으로 0~1사이의 값으로 표현된다.

$$CPS_k^{u_i} = \frac{g_k}{\max_k(g_k)}, \quad k=1,2,3,\dots,L \quad (3)$$

한편, 사용자 u_i 의 프로그램 중분류 k 에 대한 선호도 갱신은 프로그램 선호도 갱신과 마찬가지로 해당 주 바로 1주 전의 중분류 선호도를 반영하여 현재 m 주차(m^{th} week)의 중분류 선호도와 $m-1$ 주차($(m-1)^{th}$ week) 중분류 선호도의 평균값으로 다음 수식(4)과 같이 계산할 수 있다.

$$CPS_{update_k}^{u_i}(week_m) = \frac{CPS_k^{u_i}(week_{m-1}) + CPS_k^{u_i}(week_m)}{2} \quad (4)$$

3. 협업 필터링 기반 추천 엔진

본 연구의 프로그램 추천 엔진은 협업 필터링(CF)을 기법을 사용한다. IPTV-VOD 프로그램 DB를 대상으로 2.2절에서 설명한 프로그램 선호도와 중분류 선호도의 곱으로 최종 프로그램 선호도를 계산한 후 이웃 사용자와의 유사도를 감안하여 최종 프로그램 선호도가 가장 높은 상위 5개 프로그램만을 사용자에게 추천하며, 이를 1주일 단위로 갱신한다.

3.1 최종 프로그램 선호도 계산

협업 필터링은 사용자의 취향과 유사한 다른 사용자들의 프로그램 선호 성향을 참고하여 추천하는 방식으로, 본 연구의 협업 필터링기반 프로그램 추천은 크게 2단계로 구성된다. 첫 단계에서는 2.2.2절의 중분류 선호도를 이용해 유사 시청 패턴을 갖는 사용자들을 그룹핑하고, 두 번째 단계에서는 같은 그룹에 속해있는 이웃들의 중분류 선호도와 프로그램 선호도 그리고 사용자 유사도를 이용해 최종 프로그램 선호도를 계산하여 그룹DB에 저장된다. 프로그램 추천은 최종 프로그램 선호도가 가장 높은 5개의 프로그램만이 선정된다.

사용자와 유사한 시청 성향을 갖는 이웃 사용자들을 그룹핑하기 위한 사용자 유사도(US, User Similarity)로, US가 큰 순서대로 4명의 이웃 사용자

들을 선정하여 5명의 사용자 그룹을 구성한다.

수식(5)에서 u_i 는 추천 대상의 사용자, u_j 는 이웃 사용자를 말하며, G 는 총 중분류 개수로 본 연구에서는 60개의 중분류를 사용하였다. $GR_{g_l}^{u_i}$ 은 사용자 u_i 의 중분류 l 에 대한 중분류 선호 비율로, 사용자 유사도 US는 사용자 u_i 와 이웃 사용자 u_j 의 중분류 선호도가 유사할수록 1에 가까워진다. 사용자 그룹 또한 1주일 단위로 갱신되어 새로운 그룹이 생성된다.

$$US(u_i, u_j) = 1 - \sum_{l=1}^G (|GR_l^{u_i} - GR_l^{u_j}|), \quad i \neq j$$

$$GR_l^{u_i} = \frac{CPS_{update_l}^{u_i}(week_m)}{\sum_{n=0}^{60} CPS_{update_n}^{u_i}(week_m)} \quad (5)$$

협업 필터링에 의한 최종 프로그램 선호도는 위의 사용자 그룹에서 사용자 u_i 를 제외한 나머지 4명의 사용자별 중분류 선호도와 프로그램 선호도 그리고 사용자 유사도(US) 가중치 곱으로 구한 최종 프로그램 선호도를 평균하여 구한다. 다음 수식(6)은 사용자 u_i 의 프로그램 j 에 대한 최종 프로그램 선호도를 나타내며 1주일 단위로 갱신된다.

$$CF_j^{u_i}(week_m) = \frac{\sum_{n=0}^3 PPS_{update_j}^{u_n}(week_m) \times CPS_{update_k}^{u_n}(week_m) \times US_{week_m}(u_i, u_j)}{4} \quad (6)$$

수식에서 $CF_j^{u_i}(week_m)$ 은 m 번째 주에서 VOD 프로그램 DB의 프로그램 j 를 대상으로 CF에 의해 계산된 최종 프로그램 선호도 값을, $PPS_{update_j}^{u_n}(week_m)$ 은 사용자 u_i 를 제외한 나머지 4명의 m 번째 주에 갱신된 프로그램 j 의 선호도 그리고 $CPS_{update_k}^{u_n}(week_m)$ 은 m 번째주 해당 프로그램 j 에 해당하는 중분류 k 에 대한 갱신된 선호도를 나타낸다. 반면, $US_{week_m}(u_i, u_j)$ 는 m 번째주에 갱신된 사용자 u_i 와 이웃 사용자 u_j 의 사용자 유사도 값을 나타낸다.

3.2 추천 알고리즘

본 연구에서 제안한 IPTV-VOD 프로그램 추천 엔진은 협업 필터링을 이용해 사용자와 유사한 시청 패턴을 갖는 사용자들을 5명 단위로 그룹핑하고, 같은 그룹에 속해있는 4명 이웃들의 최종 프로그램 선호도를 평균해 선호도가 가장 높은 5개의 프로그램

만을 추천한다.

그림 2는 본 연구에서 제안한 협업 필터링기반 IPTV-VOD 프로그램 추천 시스템의 순서도를 나타낸 것이며, 표 3은 본 논문에서 사용한 수학 기호와 의미를 요약한 것이다.

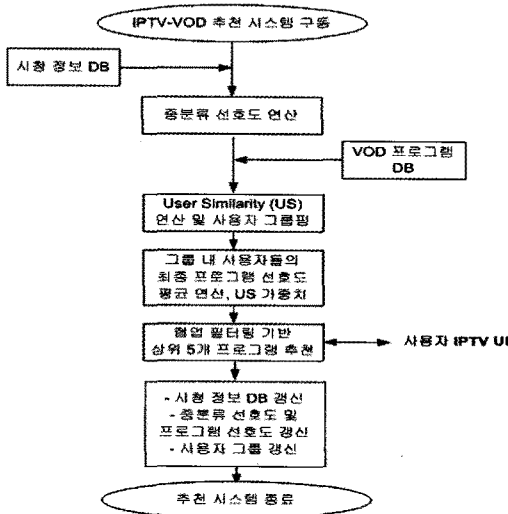


그림 2. 협업 필터링기반 프로그램 추천 시스템의 순서도

4. 실험 환경 및 결과 분석

4.1 실험 환경

본 연구에서는 최근까지도 공개된 IPTV-VOD 시청 기록 데이터가 없기 때문에 제안 시스템의 성능 평가를 위해 시청률 조사기관인 닐슨리서치의 지상파 및 케이블 방송의 시청 데이터를 IPTV-VOD 형식에 맞게 재구성하여 사용하였다. 닐슨 리서치의 데이터는 2008년 1월 1일~2008년 06월 14일까지 24주간

시청자 145명의 지상파 및 케이블 방송 시청 기록으로서, IPTV-VOD 형식에 맞추어 재구성된 후 1~4 주차는 트레이닝 데이터로, 5~24주차는 테스트 데이터로 실험을 수행하였다. 트레이닝 데이터는 협업 필터링에서 초기 사용자 그룹을 생성하는데 사용된 반면, 테스트 기간의 데이터는 제안 시스템에 의해 추천된 프로그램과 비교하는 방식으로 시스템 성능을 평가하는데 사용하였다.

닐슨리서치의 시청 데이터는 IPTV와는 다른 방송 환경을 갖기 때문에 다음과 같이 IPTV-VOD 형식에 맞게 재구성하였다.

첫째, IPTV-VOD 방송 특성상 닐슨 리서치의 시청 데이터 중 뉴스, 날씨, 교통 정보등 일회성 방송에 해당하는 프로그램은 모두 제거하고, 나머지 프로그램만으로 IPTV-VOD 방송 DB를 구축하였다. 그러나 정규방송 외 특집방송 등은 일회성이긴 하지만 IPTV-VOD 서비스로 제공하는 사례가 많아 실험 DB에 포함시켰다.

둘째, 프로그램 장르 분류 문제로, 닐슨리서치의 지상파 및 케이블 프로그램을 IPTV-VOD형식으로 수정시 가능한 KT사 QOOK IPTV-VOD의 프로그램 분류체계를 따르도록 하였다. 닐슨리서치의 방송 프로그램은 총 9개 대분류에 48개 중분류와 138개 소분류로 구분되어 있으나, 본 연구에서는 뉴스, 날씨, 교통 정보 등 일회성 방송을 제외한 나머지 프로그램들을 총 13개 대분류와 60개 중분류로 재구성하여 거의 모든 장르의 프로그램에 대한 추천이 가능하도록 하였다. 특히, 케이블 방송에서 OCN, CGV 같이 채널 이름에 따라 분류 특성이 뚜렷한 경우에는 대분류 영화로 편성을 하였고 이를 다시 IPTV-VOD에 맞게 중분류로 세분화하였다.

표 3. 수학 기호와 용어 정리

기 호	의 미
PPS_j^u	사용자 u_i 의 프로그램 j 에 대한 선호도 점수 PPS(Program Preference Score)
$PPS_{update_j}^{u_i}(week_m)$	m 주차(m^{th} week) 사용자 u_i 의 프로그램 j 에 대한 선호도 갱신
CPS_k^u	사용자 u_i 의 프로그램 중분류 k 에 대한 선호도 점수 CPS(Class Preference Score)
$CPS_{update_k}^{u_i}(week_m)$	m 주차(m^{th} week) 사용자 u_i 의 중분류 k 에 대한 선호도 갱신
$US(u_i, u_j), i \neq j$	사용자 u_i 의 그룹핑을 위한 사용자 유사도
GR_g^u	사용자 u_i 의 중분류 l 에 대한 중분류 선호 비율
$CF_j^{u_i}(week_m)$	협업 필터링- m 주차(m^{th} week) 사용자 u_i 의 프로그램 j 에 대한 최종 프로그램 선호도

셋째, 본 연구에서는 사용자의 정확한 시청 선호도를 파악하기 위해 사용자가 해당 프로그램을 10% 이상 시청한 프로그램만을 대상으로 사용자의 프로그램 선호도와 중분류 선호도 프로파일을 구성하였으며 이를 1주일 단위로 갱신하여 사용자 모델 DB를 구축하였다.

4.1 실험 결과 및 분석

본 연구에서는 제안 추천 시스템의 성능 평가를 위해 Precision을 사용하였다. Precision은 제안 시스템에 의해 최종 추천된 5개 프로그램을 대상으로 테스트 기간 동안 사용자가 실제로 관심 있게 시청한 프로그램 항목의 비율이 얼마나 되는지를 나타내는 것으로 수식(7)과 같이 구할 수 있다. 수식에서 총 추천 프로그램 수는 제안 시스템에 의해 추천되는 프로그램 개수($N=5$)이며, N_k 는 N 개 추천 프로그램 중 테스트 기간 동안 해당 사용자가 실제 시청한 것과 일치하는 프로그램 수를 나타낸다.

$$Precision = \frac{(시스템\ 추천\ 프로그램) \cap (실제\ 사용자\ 시청\ 프로그램)}{총\ 추천\ 프로그램\ 수 (=5)} = \frac{N_k}{N} \quad (7)$$

본 연구에서 제안한 추천 시스템의 성능은 수식(7)의 최종 프로그램 선호도 연산방식, 사용자 그룹의 크기, 추천 프로그램 수 등에 따라 좌우되기 때문에 본 절에서는 다음과 같이 4가지 측면에서 시스템 성능을 비교 분석하였다.

- ① 사용자 유사도(US) 가중치 사용 여부에 따른

추천 시스템의 성능 비교

- ② 최종 프로그램 선호도 연산 방식에 따른 추천 성능 비교
- ③ 사용자 그룹의 크기에 따른 추천 성능 비교
- ④ 추천 프로그램 수에 따른 제안 시스템의 성능 비교

4.1.1 사용자 유사도(US) 가중치 사용 여부에 따른 추천 시스템의 성능 비교

그림 4는 수식 (7)의 최종 프로그램 선호도 연산에 사용자 유사도(User Similarity) 가중치 $US_{week_m}(u_i, u_j)$ 를 적용한 경우와 그렇지 않은 경우의 추천 성능을 비교한 것이다.

그림에서 보듯이 사용자 유사도(US) 가중치를 적용한 경우 평균 0.695, 그렇지 않은 경우가 평균 0.68로 가중치를 적용했을 때 약 1.5% 높은 성능을 보이고 있으나 그 차이는 미미한 것으로 나타났다. 그러나 사용자 유사도 $US_{week_m}(u_i, u_j)$ 는 해당 사용자 이웃 사용자와의 프로그램 취향 유사도를 측정할 수 있는 유일한 수단으로, 향후 실용적인 IPTV-VOD 추천 시스템 구축에는 효율성이 더 높아질 것으로 기대된다.

그림에서 추천 성능은 주(week) 후반으로 갈수록 약간씩 떨어지는 경향을 보이는데, 이는 주 후반으로 갈수록 추천 대상이 되는 IPTV-VOD 프로그램의 개수가 점차 누적되면서 크게 증가하기 때문에 나타나는 현상이다. 예를 들어, 주 전반 100개 IPTV-

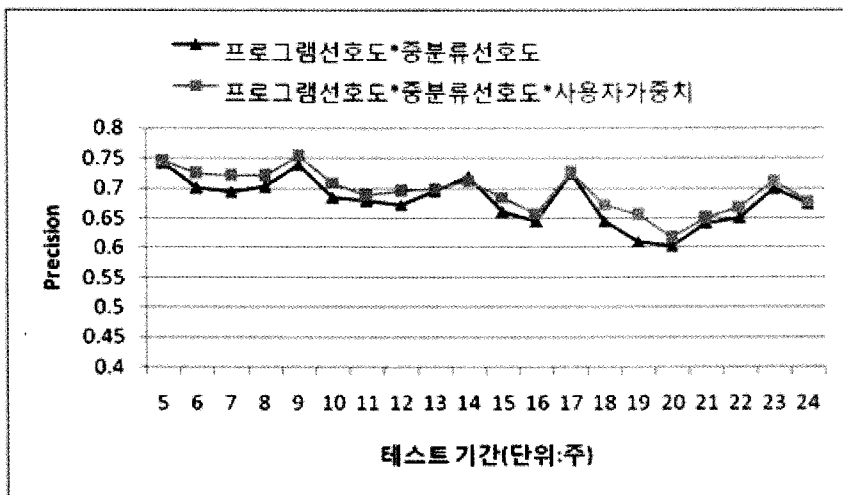


그림 3. 사용자 유사도 가중치 사용 여부에 따른 추천 성능 비교

VOD 프로그램을 대상으로 5개 프로그램을 추천하는 것과, 주 후반에 갈수록 1000개 정도로 누적된 IPTV-VOD 프로그램들을 대상으로 5개 프로그램을 추천하는 경우와는 차이가 있기 때문이다. 또한, 그림과 같이 20주 근처에서 급격하게 추천 성능이 떨어지는 이유는 프로그램 시청 내역이 없거나 적은 사용자들이 해당 주에 많이 발생했기 때문이며, 이로 인해 최종 프로그램 선호도를 연산하는 과정이 많은 영향을 받기 때문이다.

4.1.2 최종 프로그램 선호도 연산 방식에 따른 성능 비교

그림 4는 최종 프로그램 선호도를 계산할 때 중분류 선호도와 프로그램 선호도 그리고 사용자 유사도가중치와의 곱만을 사용했을 경우와 이에 덧붙여 대분류 선호도를 부가적으로 반영한 경우에 대한 성능을 비교한 것이다.

그림에서 보듯이 대분류 선호도를 감안하지 않았을 때 평균 0.695, 감안하였을 때 평균 0.679로 사용자의 대분류 선호도가 최종 추천 시스템의 성능에 거의 영향을 미치지 않음을 알 수 있다. 따라서 본 연구에서는 최종 프로그램 선호도 연산시 대분류 선호도를 감안하지 않았다.

4.1.3 사용자 그룹의 크기에 따른 추천 성능 비교

그림 5는 프로그램 추천 시 추천을 받게 되는 이웃

사용자 그룹의 크기에 따른 제안 시스템의 추천 성능을 비교한 것이다. 이웃 사용자 그룹의 크기에 따른 성능 변화를 알아보기 위해 그룹 크기를 5, 10, 15가지 변화시켜 추천 성능을 측정하였다.

그림에서 보듯이 그룹 크기가 5명인 경우 평균 0.695, 10명인 경우 평균 0.656, 15명인 경우 평균 0.635로 이웃 사용자 그룹의 크기가 증가할수록 추천 성능이 떨어지는 것을 알 수 있다. 이러한 이유로 본 연구에서는 이웃 사용자 그룹의 크기를 5명으로 제한하였으며, 이로 인해 사용자 유사도 연산량 또한 감소시킬 수 있는 장점을 얻을 수 있다.

4.1.4 추천 프로그램 수에 따른 제안 시스템의 성능 비교

그림 6은 최종 추천 프로그램 수에 따른 제안 시스템의 추천 성능을 비교한 것이다. 그림에서 보듯이 추천 수가 늘어날수록 오히려 추천 성능이 떨어짐을 볼 수 있다. 추천 프로그램 수가 5일 경우 평균 0.695, 프로그램 수가 10일 때는 평균 0.6 그리고 15일 때는 평균 0.532로 지속적인 하향 성능을 보이고 있다, 이에 본 연구에서는 최종 추천 프로그램 수를 5로 설정하여 모든 실험을 진행하였다.

그림에서 보듯이 5개를 추천하였을 경우 평균 0.695, 10개를 추천하였을 경우 평균 0.600, 15개를 추천하였을 경우 0.532로 추천 개수가 증가할수록 추천 성능이 떨어지는 것을 알 수 있다. 이러한 이유로

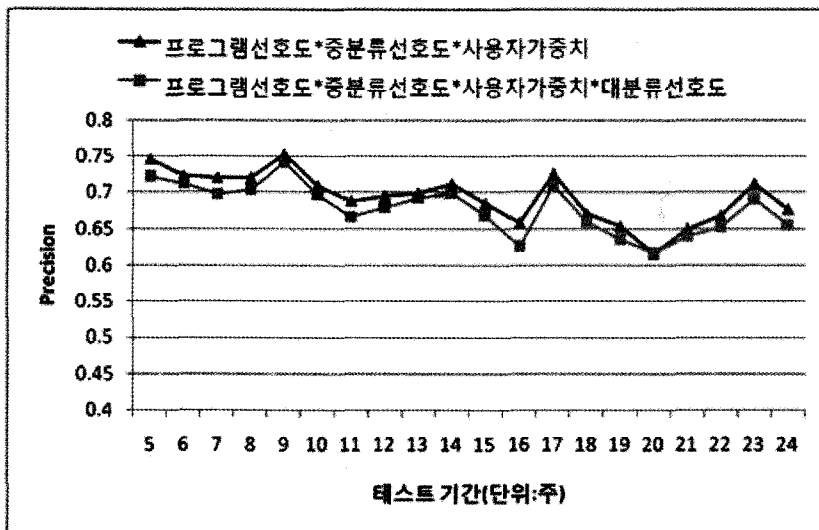


그림 4. 최종 프로그램 선호도 연산 방식에 따른 성능 비교

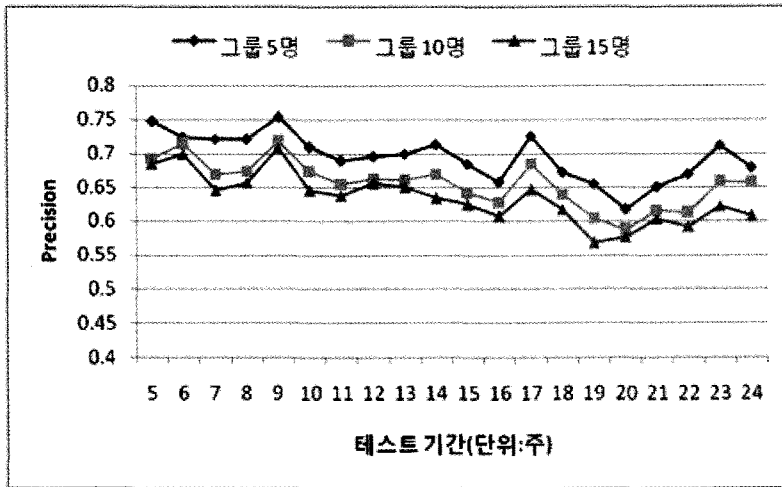


그림 5. 이웃 사용자 그룹 크기에 따른 추천 성능 비교

프로그램 추천 개수를 5개로 제한하였으며, 이로 인해 추천 성능을 향상 시킬 수 있다.

5. 결론

본 연구는 IPTV-VOD 방송 환경에서 사용자 개인의 프로그램 선호 성향을 반영한 개인화된 프로그램 추천 시스템을 제안하였다. 제안 시스템은 협업 필터링 알고리즘을 이용해 해당 사용자와 유사한 시청 취향을 갖는 이웃 사용자들을 참조해 프로그램을 추천하기 때문에 다양한 콘텐츠의 프로그램을 추천할 수 있다.

추천 프로그램 선정에 핵심적인 역할을 하는 최종 프로그램 선호도는 사용자와 유사한 시청 패턴을 갖는 사용자들을 그룹핑 한 다음 같은 그룹의 이웃 사용자들의 최종 프로그램 선호도를 평균해 최종 5개의 추천 리스트를 생성하였다.

제안 추천 시스템의 성능은 사용자 유사도 가중치 사용 여부, 최종 프로그램 선호도 연산 방식, 사용자 그룹 크기, 추천 프로그램 수 등 다양한 각도로 평가되었다. 실험 결과 최종 프로그램 선호도 연산시 사용자 유사도 가중치를 사용하는 것이 바람직한 것으로 나타났으며, 사용자 그룹 크기가 5명 그리고 추천 프로그램 수가 5일 때 평균 69.5%에 이르는 최적의

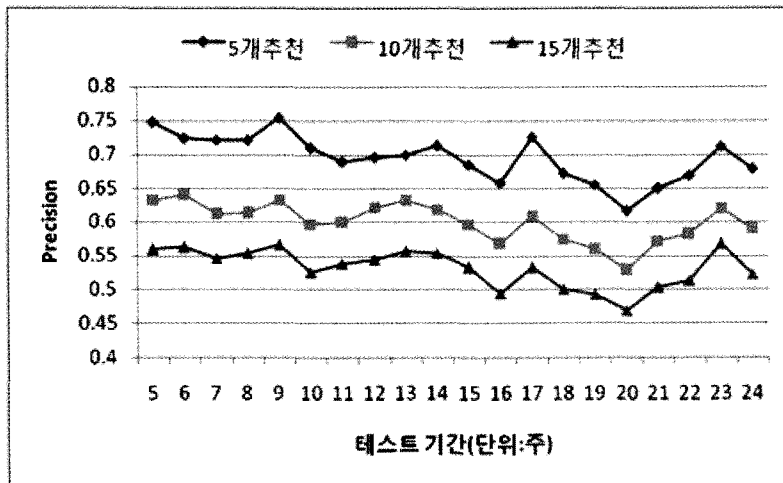


그림 6. 추천 프로그램 개수에 따른 성능 비교

결과를 얻을 수 있었다.

향후 연구로는 추천 시스템의 정확도와 안정성을 높이기 위한 기존 내용기반 필터링과 협업 필터링 알고리즘의 최적 조합 방법이나 사용자의 프로그램 시청 변화를 좀 더 세밀하게 감안할 수 있는 프로그램 선호도 프로파일 구축 방법에 대한 연구가 필요할 것으로 예상된다. 또한, 본 연구에서는 최근까지도 공개된 IPTV-VOD 시청 데이터가 없어 기존의 지상파 및 케이블 시청 데이터를 IPTV-VOD 방송 환경에 맞게 재구성해 실험을 하였지만, 향후 실제 IPTV 시청 데이터가 주어진다면 보다 실용성 있는 연구가 가능해지리라 기대된다.

참 고 문 헌

[1] Miquel Montaner, Beatriz Lopez, and Josep Lluís De La Rosa, "A Taxonomy of Recommender Agents on the Internet," *Artificial Intelligence Review*, Vol.19, No.3, pp. 285-330, 2003.

[2] Greg Linden, Brent Smith, and Jeremy York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, Vol. 7, No. 1, pp. 76-80, 2003.

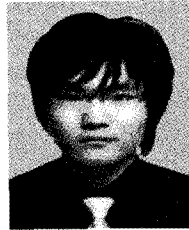
[3] John Zimmerman, Kaushal Kurapati, Anna L. Buczak, Dave Schaffer, Srinivas Gutta, and Jacquelyn Martino, *TV Personalization System*, Kluwer Academic Publishers, pp. 27-51, 2004.

[4] 이연정, "DTV 프로그램 추천 효율 향상을 위한 프로그램 시청 패턴 반영 기법," 석사학위 논문, 숭실대학교, 서울, 2006.

[5] 정하용, 김문식, "KT IPTV 사용자의 시청 이력을 이용한 VOD추천 시스템의 성능 분석," 한국정보과학회 학술발표 논문집, Vol. 36, No. 1, pp. 116-121, 2009.

[6] 김은희, 표신지, 김문철, "협업 필터링을 이용한 순위 정렬 모델 기반 (IP)TV 프로그램 자동 추천," 방송공학회지, Vol. 14, No. 2, pp. 238-252, 2009.

[7] Christina Christakou and Andreas Stafylopatis, "A Hybrid Movie Recommender System Based on Neural Networks," *Intelligent Systems Design and Applications*, 2005. ISDA '05. Proceedings. 5th International Conference on, pp. 500-505, 2005.



선 철 용

2009년 단국대학교 정보컴퓨터학부 컴퓨터과학과 학사 졸업
 2009년~현재 단국대학교 컴퓨터학부 컴퓨터과학과 석사 과정
 관심분야: 음성 및 음향 신호처리, 멀티미디어 신호처리, DSP



강 용 진

2009년 단국대학교 정보컴퓨터학부 컴퓨터과학과 학사 졸업
 2009년~현재 단국대학교 컴퓨터학부 컴퓨터과 학과 석사 과정
 관심분야: 음성 및 음향 신호처리, 멀티미디어 신호처리, DSP



박 규 식

1986년 Polytechnic University 전자공학과 학사 졸업
 1988년 Polytechnic University 전자공학과 석사 졸업
 1993년 Polytechnic University 전자공학과 박사 졸업
 1994년~1996년 삼성전자 마이크로사업부, 선임연구원
 1996년~2001년 상명대학교 컴퓨터·정보통신 공학부 조교수
 2001년~현재 단국대학교 공과대학 컴퓨터학부 교수
 주관심분야: 음성 및 음향 신호처리, 멀티미디어 신호처리, DSP 시스템구현, 감성인식