

논문 2010-47SP-4-2

# 복합 필터링을 이용한 IPTV-VOD 프로그램 추천 시스템 연구

## ( A Study of IPTV-VOD Program Recommendation System Using Hybrid Filtering )

강 용 진\*, 선 철 용\*, 박 규 식\*\*

( Yong-Jin Kang, Chul-Yong Sun, and Kyu-Sik Park )

### 요 약

본 연구는 IPTV 환경에서 사용자의 취향에 맞는 VOD 프로그램을 추천할 수 있는 시스템을 새로이 제안하였다. 제안 시스템은 내용기반 필터링과 협업 필터링의 장·단점을 상호 보완한 복합 필터링에 의한 IPTV-VOD 프로그램 추천 시스템으로, 각 필터링 기법의 프로그램 선호도(program preference) 값을 단일 지표(single-scale)로 비교·평가할 수 있는 수단을 제공함으로써 실질적인 복합 필터링 추천 시스템을 구축하였다. 사용자의 프로그램 선호 취향을 나타내는 사용자 프로파일(user profile)은 사용자의 과거 프로그램 시청 이력뿐만 아니라 사용자와 유사한 이웃 사용자들의 취향을 1주일 단위로 갱신되는 프로그램 선호도와 중분류 선호도로 표현하였기 때문에 보다 정확한 프로그램 추천이 가능하다. 제안 시스템의 성능평가를 위해 시청률 조사기관인 닐슨리서치의 24주분 지상파 및 케이블 방송 시청 데이터를 IPTV 형식에 맞게 재구성하여 사용하였으며, 다양한 실험을 통해 그 실용성을 입증하였다.

### 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 hybrid filtering method that can cooperatively complements the shortcomings of the content-based filtering and collaborative filtering. For a user program preference, a single-scaled measure is designed so that the recommendation performance between content-based filtering and collaborative filtering is easily compared and reflected to final hybrid filtering procedure. In order to provide more accurate program recommendation, we use not only the user watching history, but also the user program preference and sub-genre program preference updated every week as a user preference profile. System performance is evaluated with modified IPTV data from real 24-weeks cable TV watching data provided by Nilson Research Corp. and it shows quite comparative quality of recommendation.

**Keywords :** IPTV, Recommendation System, Hybrid Filtering, Content-Based Filtering, Collaborative Filtering.

## I. 서 론

IPTV(Internet Protocol TV)는 초고속 인터넷망을 통한 방송·통신 융합 서비스로 2009년 말 약 200만 명

이상이 가입할 것으로 기대되는 차세대 대표적인 뉴미디어이다<sup>[1]</sup>. IPTV는 사용자가 언제 어디서든 자신의 취향에 맞는 프로그램 콘텐츠를 VOD(Video On Demand) 서비스를 통해 제공받을 수 있는 장점으로 향후에도 지속적인 발전이 기대되는 서비스이며, 최근에는 IPTV의 환경에 맞는 프로그램 검색 및 추천에 대한 많은 연구가 진행되고 있다.

일반적으로 프로그램 검색은 사용자가 찾고자 하는 프로그램의 키워드를 입력하고 검색 결과로 제시된 프로그램 중에서 자신이 원하는 것을 선택하는 단계를 거친다. 반면, 자동 프로그램 추천 시스템의 경우에는 사

\* 학생회원, 단국대학교 컴퓨터과학과  
(Dept. of Computer Science, Dankook University)

\*\* 정회원, 단국대학교 컴퓨터학부  
(Division of Computer Science & Engineering,  
Dankook University)

※ 이 연구는 2008년도 단국대학교 대학연구비의 지원으로 연구되었음.

접수일자: 2009년12월28일, 수정완료일: 2010년6월9일

용자의 과거 시청 기록이나, 다른 이웃 사용자들의 시청 패턴을 참조하여 사용자가 선호할만한 프로그램을 예측하여 추천하기 때문에 다양한 콘텐츠의 흥미로운 프로그램을 추천할 수 있다는 장점이 있다.

최근까지의 추천 시스템은 주로 자동화된 정보 필터링(Information Filtering) 기술을 이용해 사용자 취향에 맞는 프로그램을 추천하는 방식으로 크게 내용기반 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 프로그램들을 추천할 수 없다는 단점이 있다. 반면, 협업 필터링은 사용자의 취향과 유사한 다른 사용자들의 취향을 참고하여 프로그램에 대한 선호도를 예측함으로써 내용기반 필터링이 제공할 수 없는 다양하고 유익한 프로그램들을 추천할 수 있는 장점이 있지만, 사용자에게 충분한 데이터가 수집되어야만 정상적인 추천서비스가 가능하다는 점과 새로운 프로그램이 추가 되었을 때 이에 대한 추천이 즉각 불가능하다는 단점이 있다. 복합 필터링(HF)은 협업 필터링과 내용기반 필터링의 장, 단점을 상호 보완한 시스템이며, 인구통계 필터링은 사용자의 나이, 성별, 국적 등의 인구 통계학 정보를 기준으로 프로그램을 추천하는 방법이다<sup>[2]</sup>.

기존 연구로 Amazon.com<sup>[3]</sup>은 아이템 기반 협업 필터링을 이용해 웹상에서 도서 추천 시스템을 구축하였으나, 사용자의 구매 이력이나 평점 기록들이 적을 때 추천 성능이 떨어지며, 사용자가 구매 내역에 대해 일일이 평점을 기록해야하는 불편함을 가지고 있다. Christakou<sup>[4]</sup>는 내용기반과 협업 필터링을 이용해 영화 프로그램을 추천할 수 있는 시스템을 제안하였다. 각 영화의 장르, 배우, 시놉시스(synopsis) 등에 대해 신경망 모델을 적용한 다음 내용기반 추천을 하고, 피어슨 상관 계수(pearson correlation coefficient)를 이용해 협업 필터링을 적용하는 방법으로 추천이 이루어진다. DTV(Digital TV) 프로그램 추천에 관한 연구로 Zhiwen<sup>[5]</sup>은 사용자의 시청 패턴에 따라 자동으로 갱신

되는 사용자 프로파일(user profile)과 코사인 유사도(cosine similarity)를 이용해 프로그램 선호도(preference)를 계산하였으며 내용기반 필터링을 적용하였다. 반면, Zimmerman<sup>[6]</sup>은 베이지안과 결정트리를 이용해 선호정보를 학습하는 암시적 추천과 사용자의 입력 정보를 이용한 명시적 추천으로부터 최종 프로그램 선호도를 계산하는 복합추천 기법을 제안하였다. 이연정<sup>[7]</sup>은 사용자의 시청 패턴 변화를 감안하기 위해 최근 성과 시청 횟수를 고려한 시청 패턴 가중치를 프로그램 선호도 연산에 반영함으로써 DTV 프로그램 추천 시스템의 효율을 향상 시켰다. 반면, KT의 정하용, 김문식<sup>[8]</sup>은 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 추천 시스템은 주로 DTV에 국한된 연구가 대부분으로 이는 최근까지도 IPTV-VOD에 대한 시청 데이터가 공개되지 않고 있기 때문이다. 프로그램 추천 또한 사용자 프로그램 선호도(user program preference) 연산에 대한 연구나, 개별적인 내용기반 필터링이나 협업 필터링 적용 방법에 관한 것이 대부분으로 실용적인 측면에서의 IPTV-VOD 프로그램 추천에 대한 연구는 아직 미진하다 할 수 있다. 현재, 국내에서 IPTV를 제공하고 있는 업체들 또한 사람들이 많이 본 프로그램을 일률적으로 추천해 주거나 최신 갱신된 프로그램만을 추천하기 때문에, 사용자 개인의 성향을 고려한 IPTV-VOD의 개인화된 서비스(personalized service) 장점을 제대로 살리지 못하고 있는 실정이다.

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

① 제안 시스템은 기존의 내용기반 필터링(CBF)과 협업 필터링(CF)의 장·단점을 상호 보완한 복합 필터링(HF)기반의 추천 시스템으로 다양한 콘텐츠의 프로그램을 추천할 수 있을 뿐만 아니라 새로운 프로그램이 추가 되었을 때도 그 즉시 추천이 가능한 장점이 있다.

② 기존 연구와 같이 프로그램 추천시 내용기반 필터링과 협업 필터링의 일률적인 비율에 따른 추천이 아닌, 각 필터링의 결과로 주어진 프로그램 선호도(program preference) 값을 단일 지표로 비교·평가할

수 있는 수단을 제공함으로써 실질적인 복합 필터링 엔진을 구축하였다.

③ 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.

④ 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.

본 논문은 다음과 같이 구성된다. 먼저 II장에서는 제안한 프로그램 추천 시스템의 전체 구성과 세부 모듈을 살펴보고, III장에서는 제안 시스템의 실험 환경 및 결과 분석을, 마지막으로 IV장에서는 결론과 향후 연구에 대한 고찰로 끝을 맺는다.

## II. 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

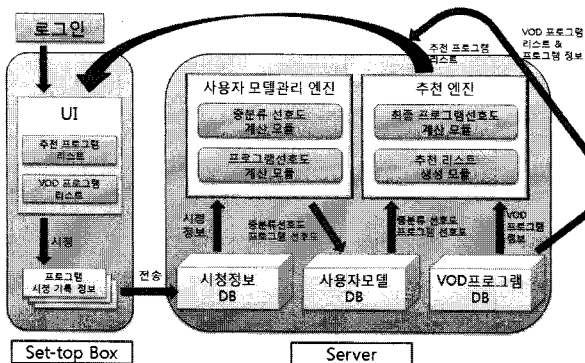


그림 1. 제안 IPTV-VOD 추천 시스템 구조  
Fig. 1. Proposal of IPTV-VOD recommended system architecture.

Interface)를 통해 이들을 사용자에게 제공하고, 이에 대한 사용자의 시청 기록을 서버로 전해주는 역할을 한다. 서버는 시청정보 DB, 사용자 모델 DB, VOD 프로그램 DB, 사용자 모델 관리 엔진, 추천 엔진으로 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 VOD 프로그램 정보들이 저장되며 이 또한 일주일 단위로 갱신된다. 실제 프로그램 추천은 사용자 모델 DB의 중분류와 프로그램 선호도에 기반한 복합 필터링(Hybrid Filtering) 기법을 이용해 VOD 프로그램 DB를 대상으로 최종 프로그램 선호도가 가장 높은 5개 프로그램만을 선정해 셋톱박스로 전송한다.

### 1. IPTV-VOD 프로그램 DB

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.

표 1. IPTV-VOD 프로그램 DB의 메타데이터 예  
Table 1. Example of metadata of IPTV-VOD program DB.

프로그램 ID	프로그램 이름	방영 시간(분)	대분류	중분류
3145	빵빵 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

### 2. 시청 정보 DB와 사용자 모델 DB

서버로부터 전송받은 VOD 프로그램 리스트나 추천 리스트를 대상으로 한 사용자의 실제 시청 정보는 해당 사용자와 시청 내역, 프로그램 정보를 기반으로 그림 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.

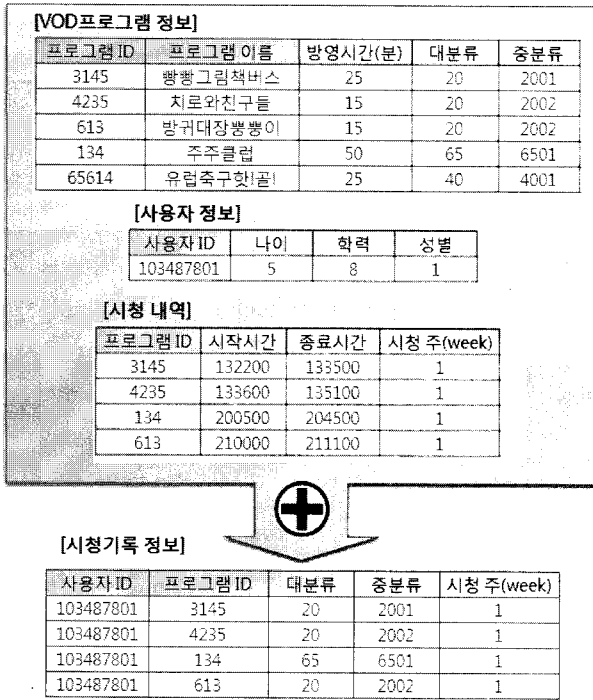


그림 2. 사용자 시청기록 정보 추출 예  
Fig. 2. Example of extracted user viewing history information.

반면, 사용자 모델 DB는 사용자 모델 관리엔진에서 시청 정보 DB를 이용해 1주일 단위로 갱신되는 프로그램 선호도와 중분류 선호도 값(2.3절에서 설명)을 저장한다.

3. 사용자 모델 관리 엔진

사용자 모델은 사용자의 프로그램 시청 패턴을 파악할 수 있는 가장 중요한 단서로서 그림 2의 시청 기록을 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 선호도 또한 1주 단위로 갱신된다.

가. 프로그램 선호도 계산 및 갱신

사용자  $u_i$ 의 프로그램 선호도는 최근 2주 동안 시청한 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,

표 2. 사용자  $u_i$ 의 프로그램 선호도 계산 예  
Table 2. Example of the program preference rating calculation of user  $u_i$ .

프로그램 ID	시청빈도수 ( $p_j$ )	연속시청주 가중치 ( $w_j$ )	프로그램 선호도 ( $PPS_j^{u_i}$ )
1	14	1	1
2	4	1/2	0.142
3	6	1/2	0.214
4	3	1	0.214
5	7	1	0.5
6	8	1	0.571
7	2	1/2	0.071
8	6	1/2	0.428

2주 동안 프로그램 시청 연속성에 대한 가중치를 반영하기 위해 만약 사용자가 2주 연속 해당 프로그램을 시청하였다면 가중치를  $w_j = 1$ 로, 2주중 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.

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

예를 들어, 2주일간 시청 프로그램의 개수가 총 8이고, 시청 빈도수와 연속 시청 가중치가 표 1과 같을 때 최대 시청 빈도수는  $\max_j(p_j) = 14$ 가 되며 각 프로그램에 대한 선호도는 표와 같이 계산될 수 있다. 프로그램 ID가 1인 경우,  $PPS_1^{u_i} = \frac{p_1 \times w_1}{\max_j(p_j)} = \frac{14 \times 1}{14} = 1$ 이 되고, 프로그램 ID가 2인 경우에는  $PPS_2^{u_i} = \frac{p_2 \times w_2}{\max_j(p_j)} = \frac{4 \times (1/2)}{14} = 0.142$ 가 된다.

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 수식 (1)의  $m$ 주차( $m^{th}$  week)의 프로그램 선호도와  $m-1$ 주차( $(m-1)^{th}$  week) 프로그램 선호도의 평균값으로 주어지며 이를 수식 (2)와 같이 표현할 수 있다.

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

수식 (2)에서  $PPS_j^{u_i}(week_{m-1}), PPS_j^{u_i}(week_m)$ 은 각각 수식 (1)을 이용해 계산된 사용자  $u_i$ 의 프로그램  $j$ 에 대한  $m-1$ 주차와  $m$ 주차 선호도를 나타내며  $PPS_{update_j}^{u_i}(week_m)$ 는 전 주의 프로그램 선호도를 반영해 갱신된  $m$ 주차 프로그램 선호도를 나타낸다.

나. 중분류 선호도 계산 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. 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 수식 (3)의 현재  $m$ 주차( $m^{th}$  week)의 중분류 선호도와  $m-1$ 주차( $(m-1)^{th}$  week) 중분류 선호도의 평균값으로 수식 (4)와 같이 계산할 수 있다.

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

4. IPTV-VOD 프로그램 추천 엔진

본 연구의 프로그램 추천 엔진은 내용기반 필터링(CBF)과 협업 필터링(CF)을 상호 보완한 복합 필터링(HF) 기법을 사용한다. IPTV-VOD 프로그램 DB를 대상으로 2.3절에서 설명한 프로그램 선호도와 중분류 선호도의 곱으로 최종 프로그램 선호도를 계산한 후, 각 필터링 기법을 이용해 5개씩의 프로그램을 추천한 다음, 추천된 10개 프로그램 중 최종 프로그램 선호도가 가장 높은 상위 5개 프로그램만을 사용자에게 추천하

며, 이를 1주일 단위로 갱신한다.

가. 최종 프로그램 선호도 계산

(1) 내용기반 필터링(CBF)

내용기반 필터링은 사용자가 과거에 선호하였던 프로그램들을 미래에도 선호할 가능성이 높다는 사실을 이용한 방식으로, 본 연구에서는 3절의 프로그램 선호도와 중분류 선호도의 곱으로 최종 프로그램 선호도를 계산한 후 상위 5개의 프로그램을 추천하게 된다.  $m$ 주차( $m^{th}$  week)에 사용자  $u_i$ 의 프로그램  $j$ 에 대한 최종 프로그램 선호도는 수식 (5)와 같으며, 1주일 단위로 갱신된다.

$$CBF_j^{u_i}(week_m) = PPS_{update_j}^{u_i}(week_m) \times CPS_{update_k}^{u_i}(week_m) \quad (5)$$

수식에서  $CBF_j^{u_i}(week_m)$ 은  $m$ 번째 주에서 VOD 프로그램 DB의 프로그램  $j$ 를 대상으로 내용기반 필터링에 의해 계산된 최종 프로그램 선호도 값을,  $PPS_{update_j}^{u_i}(week_m)$ 은  $m$ 번째 주에 갱신된 프로그램  $j$ 의 선호도 그리고  $CPS_{update_k}^{u_i}(week_m)$ 은  $m$ 번째주 프로그램  $j$ 에 해당하는 중분류  $k$ 에 대한 갱신된 선호도를 나타낸다.

(2) 협업 필터링(CB)

협업 필터링은 사용자의 취향과 유사한 다른 사용자들의 프로그램 선호 성향을 참고하여 추천하는 방식으로, 본 연구의 협업 필터링기반 프로그램 추천은 크게 2단계로 구성된다. 첫 단계에서는 3절의 중분류 선호도를 이용해 유사 시청 패턴을 갖는 사용자들을 그룹핑하고, 두 번째 단계에서는 같은 그룹에 속해있는 이웃들의 중분류 선호도와 프로그램 선호도를 이용해 최종 프로그램 선호도를 계산한다. 프로그램 추천은 최종 프로그램 선호도가 가장 높은 5개의 프로그램만이 선정된다.

수식 (6)은 사용자와 유사한 시청 성향을 갖는 이웃 사용자들을 그룹핑하기 위한 사용자 유사도(US, User Similarity)로, US가 큰 순서대로 4명의 이웃 사용자들을 선정하여 5명의 사용자 그룹을 구성한다.

$$US(u_i, u_j) = 1 - \sum_{l=1}^G (|GR_l^{u_i} - GR_l^{u_j}|), \quad i \neq j \quad (6)$$

$$GR_i^{u_i} = \frac{CPS_{update_i}^{u_i}(week_m)}{\sum_{n=0}^{60} CPS_{update_n}^{u_i}(week_m)}$$

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

협업 필터링에 의한 최종 프로그램 선호도는 위의 사용자 그룹에서 사용자  $u_i$ 를 제외한 나머지 4명의 사용자별 중분류 선호도와 프로그램 선호도 그리고 사용자 유사도( $US$ )의 곱으로 구한 최종 프로그램 선호도를 평균하여 구한다. 수식 (7)은 사용자  $u_i$ 의 프로그램  $j$ 에 대한 최종 프로그램 선호도를 나타내며 1주일 단위로 갱신된다.

$$CF_j^{u_i}(week_m) = \frac{\sum_{n=0(n \neq i)}^3 PPS_{update_j}^{u_n}(week_m) \times CPS_{update_k}^{u_n}(week_m) \times US_{week_m}(u_i, u_j)}{4} \quad (7)$$

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.

나. 제안 복합 필터링(HF)기반 추천 시스템

본 연구에서 제안한 IPTV-VOD 프로그램 추천 엔진은 사용자의 프로그램 선호도와 중분류 선호도를 이용해 기존의 내용기반 필터링과 협업 필터링의 장·단점을 상호 보완하는 방식으로 동작하기 때문에 다양한 콘텐츠의 프로그램을 추천할 수 있을 뿐만 아니라 새로운 프로그램이 추가 되었을 때도 그 즉시 추천이 가능한

장점이 있다.

내용기반 필터링은 현재 방영되고 있는 IPTV-VOD 프로그램 DB를 대상으로 각 프로그램별 최종 선호도를 1주 단위로 갱신되는 프로그램 선호도와 중분류 선호도와 곱으로 구한 후, 상위 5개의 프로그램만을 추천한다. 반면, 협업 필터링은 사용자와 유사한 시청 패턴을 갖는 사용자들을 5명 단위로 그룹핑하고, 같은 그룹에 속해있는 4명 이웃들의 최종 프로그램 선호도를 평균해 선호도가 가장 높은 5개의 프로그램만을 추천한다.

2.4.1 절에서 보았듯이 각 필터링 기법에 따른 최종 프로그램 선호도 연산 과정은 약간씩 다르지만, 두 필터링 기법 모두 최종 프로그램 선호도를 프로그램 선호도와 중분류 선호도의 곱으로 표현하기 때문에 단일 지표(single-scale)로 비교가 가능하며, 이러한 단일 지표를 이용해 각 필터링 기법에 의해 추천된 10개 추천 리스트 중 상위 5개의 프로그램만을 최종적으로 사용자에게 추천할 수 있게 된다. 즉, 기존의 복합 필터링 연구 대부분이 내용기반 필터링과 협업 필터링의 일률적 비율에 따른 프로그램을 추천을 하는 반면, 본 연구의 복

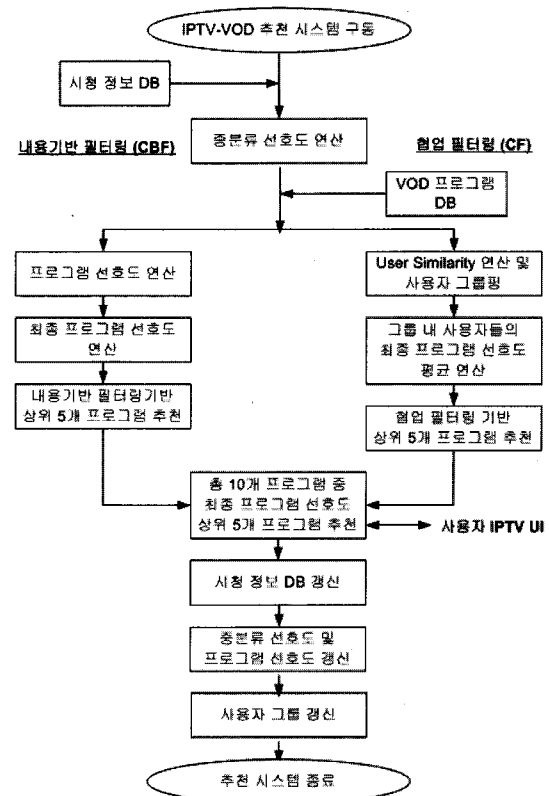


그림 3. 제안 IPTV-VOD 프로그램 추천 시스템의 순서도  
Fig. 3. Proposition IPTV-VOD recommended system flowchart.

표 3. 수학 기호와 용어 정리  
Table 3. Mathematical symbols and terminology arrangement.

기 호	의 미
CBF, CF, HF	내용기반 필터링(CBF), 협업 필터링(CF), 복합(HF) 필터링
$PPS_j^{u_i}$	사용자 $u_i$ 의 프로그램 $j$ 에 대한 선호도 점수 PP(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
$CPS_k^{u_i}$	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
$CBF_j^{u_i}(week_m)$	내용기반 필터링- $m$ 주차( $m^{th}$ week) 사용자 $u_i$ 의 프로그램 $j$ 에 대한 최종 프로그램 선호도
$US(u_i, u_j), i \neq j$	사용자 $u_i$ 의 그룹핑을 위한 사용자 유사도
$GR_{g_l}^{u_i}$	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

합 필터링 시스템은 필터링 방식에 상관없이 최종 프로그램 선호도라는 단일 지표를 이용해 프로그램을 추천하기 때문에 주어진 시스템 조건에 따라 유연하게 프로그램을 추천할 수 있는 장점이 있다.

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

### III. 실험 환경 및 결과 분석

#### 1. 실험 환경

최근까지도 IPTV-VOD 시청 기록 데이터는 공개된 것이 없기 때문에 본 연구에서는 제안 시스템의 성능평가를 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 재구성한 후 1~4주차는 트레이닝 데이터로, 5~24주차는 테스트 데이터로 실험을 수행하였다. 트레이닝 데이터는 내용기반 필터링에서 각 사용자의 초기 프로그램 선호도와 중분류 선호도를 계산할 때, 그리고 협업 필터링에서 초기 사용자 그룹을 생성하는데 사용되었다. 반면, 테스트 기간의 데이터는 제안 시스템에 의해 추천된 프로그램과 비교하는 방식으로 시스템 성능을 평가하는데 사용하였다.

지상파 및 케이블의 방송 환경은 다음과 같은 측면에서 IPTV(VOD)와 다른 방송 환경을 갖는다. 첫째, 프로그램 선택이 수동적이며 시간 접근성에 제한이 있다는 점이다. 지상파 및 케이블 방송은 정해진 편성표에 의해 프로그램이 방영되기 때문에 시청자는 프로그램의 선택 폭이 채널수로 한정된다. 또한, 지상파 및 케이블 방송은 일정한 편성표에 따라 프로그램을 방영하기 때문에 IPTV-VOD 서비스처럼 사용자가 원하는 시간에 선택 시청할 수 있는 방송 환경을 제공할 수 없다. 둘째, 방송 특성상 제공되는 두 매체간의 프로그램의 종류와 장르 분류에 약간의 차이점이 있다. 예를 들어, IPTV-VOD는 케이블 방송과 달리 뉴스나 날씨 정보 등 일회성의 프로그램을 보유하고 있지 않다. 또한, 지상파 및 케이블 방송은 110개의 채널(T-broad 기준)을 서비스하고 있는 반면, IPTV-VOD 프로그램(QOOK TV 기준, 2009년 5월)은 4개의 지상파 채널과 영화, 시리즈, 애니메이션, 스포츠, 다큐멘터리, 취미/레저, 음악, 종교, 연애/오락, 어린이 등의 대분류로 구성되며 각 대분류는 다시 세부 중분류로 나누어진다.

본 연구에서는 이러한 방송 매체별 환경차이를 극복하기 위해 다음과 같이 닐슨리서치의 지상파 및 케이블 방송의 시청 데이터를 IPTV-VOD 형식에 맞게 재구성하여 사용하였다.

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 정규 방송 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-VOD 형식으로 수정시 가능한 한 KT사 QOOK IPTV-VOD의 프로그램 분류

체계를 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.

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.

2. 실험 결과 및 분석

제안 추천 시스템의 성능 평가 방법은 Hit ratio, Ranking point average, Percentage of contents, Precision, Recall등 다양한 기법들이 존재하나 본 연구에서는 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 (8)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.

$$Precision = \frac{N_k}{N} = \frac{(시스템\ 추천\ 프로그램) \cap (실제\ 사용자\ 시청\ 프로그램)}{총\ 추천\ 프로그램\ 수 (=5)} \quad (8)$$

제안 추천 시스템 평가는 다음과 같이 5 가지 측면에서 비교 분석하였다.

- ① 개별 협업 필터링과 내용기반 필터링의 추천 성능 비교
- ② 내용기반 필터링과 협업 필터링의 일률적 추천 비율에 따른 복합 필터링 시스템의 성능
- ③ 제안 복합 필터링 시스템의 성능 비교
- ④ 최종 프로그램 선호도 연산 방식에 따른 성능 비교

⑤ 추천 프로그램 수에 따른 제안 시스템의 성능 비교

(1) 개별 협업 필터링과 내용기반 필터링의 추천 성능 비교

그림 4는 협업 필터링과 내용기반 필터링을 개별적으로 적용해 각각 5개 프로그램을 추천한 경우에 대한 성능을 주(week)별로 비교한 것이다.

그림에서 보듯이 협업 필터링은 평균 0.68, 그리고 내용기반 필터링은 평균 0.679로서 거의 비슷한 평균 성능을 보이고 있다. 그러나 주 후반으로 갈수록 추천 성능이 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-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 해당 사용자의 과거 시청 기록에 기초해 최종 프로그램 선호도를 연산하는 과정에 많은 영향을 끼치기 때문이다.

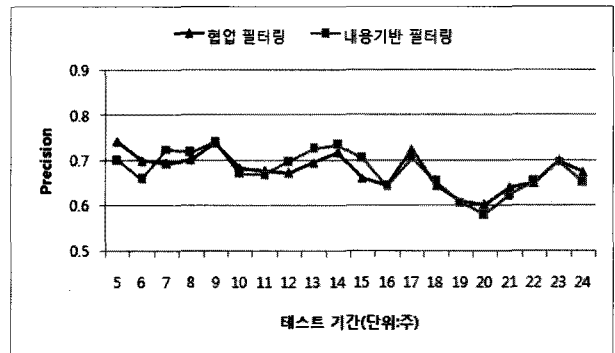


그림 4. 협업 필터링과 내용기반 필터링 성능 비교  
Fig. 4. Comparison performance between Collaborative filtering and content-based filtering.

(2) 내용기반 필터링과 협업 필터링의 일률적 추천 비율에 따른 복합 필터링 시스템의 성능

그림 5는 개별 필터링 기법이 최종 시스템 성능에 미치는 영향을 알아보기 위해, 내용기반 필터링과 협업 필터링의 추천 비율을 4:1(내용기반 필터링 추천의 상



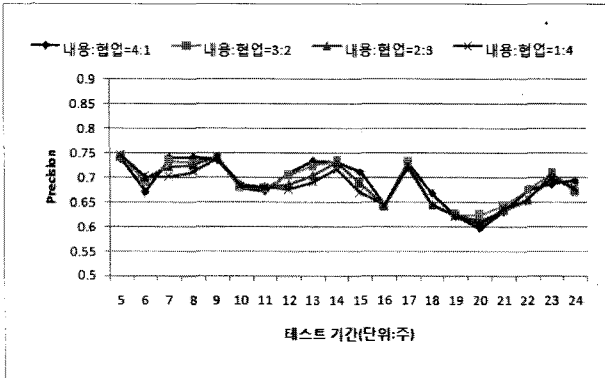


그림 5. 내용기반 필터링과 협업 필터링의 추천 비율 변화에 따른 최종 성능 비교

Fig. 5. Final performance comparison based on recommendation percentage change of Content-based filtering and recommendation percentage change of Collaborative filtering.

표 4. 내용기반 필터링과 협업 필터링의 추천 비율에 따른 최종 시스템의 평균 성능

Table 4. Average performance of the final system based on the Content-based filtering recommendation rate and Collaborative filtering recommendation rate.

추천 비율	4:1	3:2	2:3	1:4
추천 성능	0.705	0.690	0.686	0.681

위 4개, 그리고 협업 필터링의 상위 1개- 중복의 경우에는 차 순위 추천)과 3:2, 2:3, 1:4로 지정하였을 경우에 대한 복합 필터링 시스템의 성능을 나타낸 것이다. 표 4는 각 추천 비율에 따른 평균 성능을 요약한 것이다.

그림과 표에서 보듯이 전반적으로 비슷한 평균 성능을 보이나, 내용기반 필터링과 협업 필터링의 추천 비율이 4:1일 경우가 평균 0.705로 가장 높아 최종 시스템 성능에 내용기반 필터링 기법이 더 많은 영향을 주는 것으로 나타났다. 비록 미미한 차이지만 추천 비율이 4:1부터 1:4까지 모든 경우에 있어 결과 성능이 실험 (1)의 그림 5에서 개별 필터링 기법을 이용한 결과인 0.68(협업 필터링 경우)보다 높은 것으로 보아(추천 비율 4:1에서 최대 약 2.6% 향상된 성능), 개별 필터링 기법을 사용할 때 보다 각 필터링 기법을 복합한 추천 시스템이 더 효율적이라는 사실을 알 수 있다.

(3) 제안 복합 필터링 시스템의 성능 비교

그림 6은 실험 (1)의 협업 필터링 시스템, 실험 (2)의 내용기반 필터링과 협업 필터링의 추천 비율이 4:1인 경우, 그리고 본 연구에서 제안한 복합 필터링 시스템의 추천 성능을 비교한 것이다. 그림에서 보듯이 제안

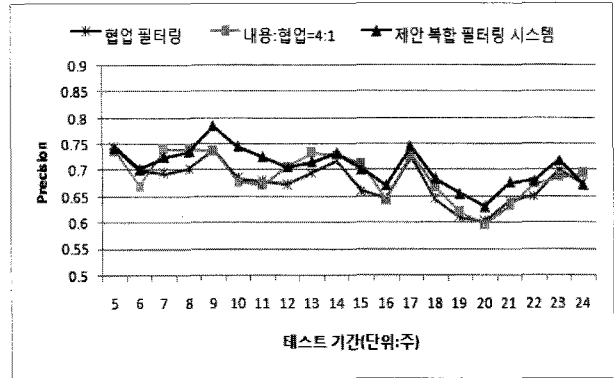


그림 6. 제안 복합 필터링 시스템의 성능 비교

Fig. 6. Performance comparison of the proposed complex filtering system.

한 복합 필터링기반 추천 시스템이 평균 0.72로 실험 (1)보다 약 4.3% 그리고 실험 (2)보다 약 1.7% 정도의 높은 추천 성능을 보이고 있음을 알 수 있다.

이는 본 연구에서 제안한 복합 필터링기반 추천 시스템이 각 내용기반 필터링과 협업 필터링으로부터 추천 결과를 최종 프로그램 선호도라는 단일 지표로 비교한 다음 최종적으로 상위 5개 프로그램만을 추천하기 때문에 주어진 시스템 조건에 따라 유연하게 프로그램을 추천할 수 있기 때문으로 분석된다.

(4) 최종 프로그램 선호도 연산 방식에 따른 성능 비교

그림 7은 각 필터링 기법에서 최종 프로그램 선호도를 계산할 때 중분류 선호도와 프로그램 선호도와와의 곱만을 사용했을 경우와 이에 덧붙여 대분류 선호도를 반영한 경우에 대한 성능을 비교한 것이다.

그림에서 보듯이 중분류와 프로그램 선호도만을 이용했을 때 평균 0.722 성능을 그리고 부가적으로 대분

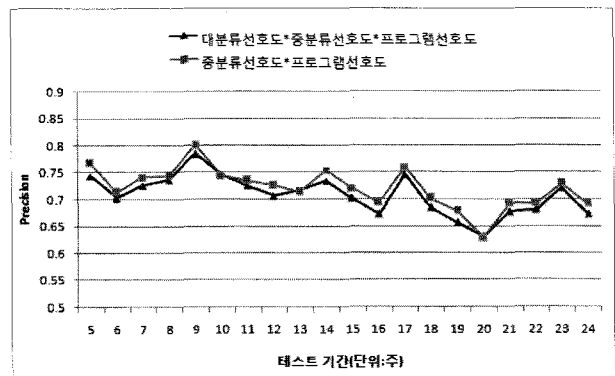


그림 7. 최종 프로그램 선호도 연산 방식에 따른 성능 비교

Fig. 7. Final program for rating algorithms of performance comparisons.

류 선호도를 반영했을 때 평균 0.709로 중분류와 프로그램 선호도만을 이용할 경우가 약 1.3%정도 높음을 알 수 있다. 이는 사용자의 대분류 선호도가 최종 추천 시스템의 성능에 거의 영향을 미치지 않음을 의미한다.

#### (5) 추천 프로그램 수에 따른 제안 시스템의 성능 비교

그림 8은 최종 추천 프로그램 수에 따른 제안 시스템의 추천 성능을 비교한 것이다. 그림에서 보듯이 추천 수가 늘어날수록 오히려 추천 성능이 떨어짐을 볼 수 있다. 최종 추천 프로그램 수가 5일 경우 실험 (3)에서와 같이 평균 0.72, 프로그램 수가 10일 때는 평균 0.625 그리고 15일 때는 평균 0.556으로 지속적인 하향 성능을 보이고 있다, 이에 본 연구에서는 최종 추천 프로그램 수를 5로 설정하여 모든 실험을 진행하였다.

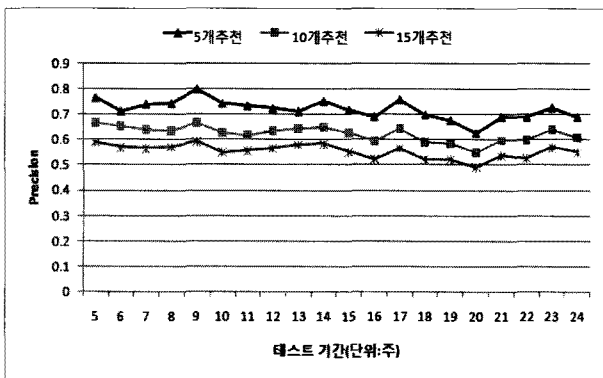


그림 8. 추천 프로그램 개수에 따른 성능 비교  
Fig. 8. Recommend program based on the number of performance comparisons.

## IV. 결 론

본 연구는 IPTV-VOD 방송 환경에서 사용자 개개인의 프로그램 선호 성향을 반영한 개인화된 프로그램 추천 시스템을 제안하였다. 제안 시스템은 기존의 내용 기반 필터링과 협업 필터링의 장·단점을 상호 보완한 복합 필터링기반의 추천 시스템으로 다양한 콘텐츠의 프로그램을 추천할 수 있을 뿐만 아니라 새로운 프로그램이 추가 되었을 때도 그 즉시 추천이 가능하다. 또한, 제안 시스템은 기존 연구와 달리 각 내용기반 필터링과 협업 필터링으로부터 추천 결과를 최종 프로그램 선호도라는 단일 지표로 비교한 다음 최종적으로 상위 5개 프로그램만을 추천하기 때문에 주어진 시스템 조건에 따라 유연하게 프로그램을 추천할 수 있는 장점

이 있다.

추천 프로그램 선정에 핵심적인 역할을 하는 최종 프로그램 선호도는 내용기반 필터링의 경우 1주 단위로 갱신되는 프로그램 선호도와 중분류 선호도의 곱으로 구하여 추천 리스트를 생성하였으며, 협업 필터링의 경우 사용자와 유사한 시청 패턴을 갖는 사용자들을 그룹핑 한 다음 같은 그룹의 이웃 사용자들의 최종 프로그램 선호도를 평균해 추천 리스트를 생성하였다. 최종적으로는 각 필터링 기법으로 추천된 프로그램 리스트 중, 동적으로 상위 5개의 최종 프로그램 선호도 값을 갖는 프로그램만을 추천함으로써 기존의 일률적 추천 비율에 따른 복합 필터링 시스템 보다 더욱 효율적임을 확인할 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] 이재신, "IPTV 서비스의 발전방향", 동서언론, Vol. 11, pp. 297-328, 2008.
- [2] 원희재, "P2P 네트워크상의 사용자 프로파일에 기반한 음악 추천 시스템", 멀티미디어학회 논문집, Vol. 11, No 10, pp. 1338-1346, 2008.
- [3] Greg Linden, Brent Smith, Jeremy York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, Vol.7, No.1, pp. 76-80, 2003.
- [4] Christina Christakou, Andreas Stafylopatis, "A Hybrid Movie Recommender System Based on Neural Networks", Intelligent Systems Design and Applications, pp. 500-505, 2005.
- [5] Zhiwen Yu, Xingshe Zhou, "TV3P: An Adaptive Assistant for Personalized TV", IEEE Transactions, Vol. 50, pp. 393-399, 2004.
- [6] John Zimmerman, Kaushal Kurapati, Anna L. Buczak, Dave Schaffer, Srinivas Gutta, ,

Jacquelyn Martino, "TV Personalization System", *Kluwer Academic Publishers*, pp. 27-51, 2004.

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

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

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

[10] Robin Van Meteren, Maarten Van Someren, "Using Content-Based Filtering for Recommendation", *Proceedings of the Machine Learning and Population Age: MLnet/ECML2000 Workshop*, 2000.

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

저 자 소 개



강 용 진(학생회원)  
2009년 단국대학교 정보컴퓨터  
학부 컴퓨터과학과  
학사 졸업.  
2009년~현재 단국대학교  
컴퓨터학부 컴퓨터과학과  
석사과정

<주관심분야 : 음성 및 음향 신호처리, 멀티미디어 신호처리, DSP>



선 철 용(학생회원)  
2009년 단국대학교 정보컴퓨터  
학부 컴퓨터과학과  
학사 졸업.  
2009년~현재 단국대학교  
컴퓨터학부 컴퓨터과학과  
석사과정

<주관심분야 : 음성 및 음향 신호처리, 멀티미디어 신호처리, DSP>



박 규 식(정회원)  
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>