

# 시간 기반의 전역 선호도와 지역 선호도를 고려한 개인화된 TV 프로그램 추천

오성탁<sup>o</sup>, 이지형<sup>\*</sup>

<sup>o</sup>\* 상균관대학교 전자전기컴퓨터공학과

e-mail: {ohsm014, john}@skku.edu<sup>o\*</sup>

## Personalized TV Program Recommendation Considering Time-based Global and Local Preference

Suntak Oh<sup>o</sup>, Jee-Hyong Lee<sup>+</sup>

<sup>o+</sup>Dept. of Electrical and Computer Engineering, Sungkyunkwan University

### ● 요약 ●

TV는 타 도메인과 달리, 사전에 정해진 시간에 콘텐츠가 방영된다. 그러므로 TV 프로그램 추천 시스템은 시청자의 현재 시각(time-context)을 고려해야 한다. 시간 기반의 TV 프로그램 추천 방법이 다수 연구되었지만, 대부분의 기존 연구는 특정 시간대(timeslot)에서의 시청자의 선호도를 계산하는 데에만 집중되어 있고, 시청 내역 전체기간에서의 선호도를 고려하지 않은 문제점이 있다. 이러한 문제를 해결하기 위해, 시청자의 지역 선호도와 전역 선호도를 모두 고려한 시간 기반의 TV 프로그램 추천 기법을 제안한다. 이를 위해 제안 방법에서는 시간대의 길이에 따라 여러 가지 선호도 모델을 사용한다. 여러 개의 선호도 모델로부터 산출된 선호도를 병합하여 가장 선호도가 높은 TV 프로그램을 추천한다. 실 데이터를 이용한 실험을 통해 기존방식과 비교함으로써, 제안 방법의 효용성을 검증하였다.

**키워드:** TV 프로그램 추천(TV program recommendation), 내용 기반 필터링(content-based filtering), 상황 기반 추천(context-based recommendation), 추천 시스템(recommender system)

## I. 서론

디지털 방송이 활성화되면서 시청자가 선택할 수 있는 TV 채널 및 콘텐츠의 수가 다수 증가함에 따라, 시청자가 원하는 TV 프로그램을 탐색하는데 많은 시간이 소요된다. 따라서 시청자의 시청 히스토리를 기반으로 시청자가 선호하는 TV 프로그램을 추천하는 시스템이 필요하다. TV 프로그램 추천을 위한 많은 연구가 있었지만, 기존의 개인화된 추천 기법은 TV의 특성을 고려하지 않은 문제점이 있다 [1]. TV는 영화, 음악, 웹 페이지 등 타 도메인과 달리 모든 콘텐츠가 사전에 정해진 요일과 시각에 방영된다. 시청자가 선호하는 TV 프로그램이 방영되는 시각에 TV를 켜 줄 경우, 해당 TV 프로그램을 볼 가능성이 높다. 그러므로 시청자의 현재 시각(time-context)을 고려하여 TV 프로그램을 추천해야 한다. 기존에도 시청자의 현재 시각을 고려한 추천 연구가 다수 있었다. 하지만 기존의 시간 기반 추천 방법 [2,3,4]은 특정 시간대(timeslot)에서의 시청자의 선호도를 계산하는 데에만 초점이 맞추어져 있고, 시청자의 일상적인 선호도를 고려하지 않은 문제점이 있다.

이러한 문제를 해결하기 위해 본 연구에서는 시청자의 현재 시각에 기반하여 TV 프로그램을 추천한다. 또한 시간 범위 상에서의 전역 선호도와 지역 선호도를 함께 고려한 추천 방법을 제안한다. 전역

선호도란, 요일이나 시간과 상관없이 시청자가 평소에 선호하는 TV 프로그램의 속성(feature)을 뜻한다. 지역 선호도란, 시청자가 특정 요일이나 시간에 선호하는 TV 프로그램의 속성을 뜻한다. 시청자의 TV 프로그램에 대한 추천 점수를 계산하기 위해, TV 프로그램의 채널과 장르에 대한 선호도를 각각 산출하였다. 이를 실 데이터에 적용하여, 특정 시각에 시청자가 시청할 TV 프로그램을 예측하고 예측 정확도를 측정하였다.

논문의 구성은 다음과 같다. 2장에서는 본 연구에서 제안하는 TV 프로그램 추천 방법을 설명한다. 3장에서는 실험을 통하여, 기존 방법과 제안 방법의 비교를 통해 그 효용성을 분석한다. 마지막으로 4장에서는 연구에 대한 결론 및 향후 과제를 제시한다.

## II. 제안 방법

그림 1은 본 논문에서 제안하는 추천 과정의 개괄적인 흐름을 나타낸다. 각 4개의 선호도 모델은 시청자의 TV 프로그램 시청 시간에 기반하여 시청자의 정규화된 채널 선호도와 장르 선호도를 산출한다. 그리고 각 모델로부터 산출된 선호도를 병합하여 시청자의 최종 채널 선호도와 장르 선호도를 산출한다. 최종적으로, 현재 방영중

인 TV 프로그램들 중에서 시청자의 채널 선호도와 장르 선호도가 가장 높은 TV 프로그램 1개를 시청자에게 추천한다.

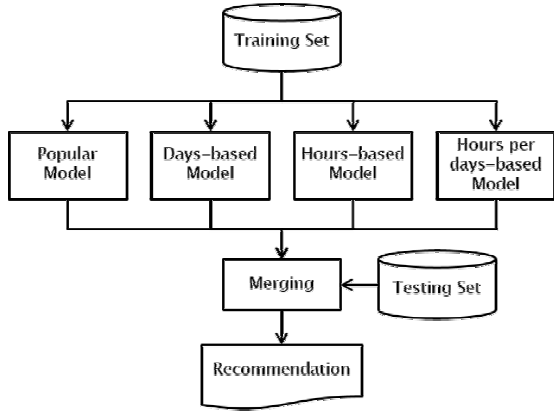


그림 1. TV 프로그램 추천 과정  
Fig. 1. Flows of TV Program Recommendation

### 1. 선호도 계산

시청자의 TV 프로그램에 대한 평가 점수를 계산하기 위해, 채널과 장르에 대한 선호도를 산출한다. 특정 속성에 대한 총 시청시간을 이용하여 선호도를 계산하였다. 예를 들어 시청자가 KBS를 30분 동안 시청했을 경우, KBS에 대한 선호도는 30만큼 증가한다. 시청자  $u$ 가 TV 프로그램  $p$ 를 시청했을 경우, 시청자  $u$ 의 TV 프로그램  $p$ 에 대한 평가 점수  $R_{u,p}$ 는 식 (1)로 계산한다.

$$R_{u,p} = t_{u,p,end} - t_{u,p,begin} \quad (1)$$

$t_{u,p,begin}$ ,  $t_{u,p,end}$ 는 각각 사용자  $u$ 의 TV 프로그램  $p$ 에 대한 시청 시작, 종료 시각이다. 특정 시간대에서의 시청자의 선호도를 파악하기 위해, 시청자  $u$ 의 시간대  $s$  (예: 월요일 13시~14시)에서의 속성  $f$  (예: KBS)에 대한 선호도  $P_{u,f,s}$ 는 식 (2)로 계산한다.

$$P_{u,f,s} = \sum_i R_{u,p_i,f,s} \quad (2)$$

$p_i$ 는 시간대  $s$ 에 시청한 속성  $f$ 를 갖고 있는  $i$ 번째 TV 프로그램이다. 즉, 시간대가  $s$ 일 때 속성  $f$ 에 대해서 시청한 총 시간이다. 그리고 각 속성에 대한 선호도를 정규화한  $NP_{u,f,s}$ 를 식 (3)으로 산출한다.

$$NP_{u,f,s} = \frac{P_{u,f,s}}{\sum_f P_{u,f,s}} \quad (3)$$

시간대  $s$ 는 월요일 화요일과 같은 요일, 그리고 1사~2시, 9사~10시와 같은 시간이 될 수 있다. 제안 방법에서는  $s$ 가 요일, 시간, 요일 및 시각일 때의 3가지 경우를 고려하였다. 시청자는 특정 시간대에 선호하는 지역 선호도가 있지만, 현재 시각과 상관없이 평상시에

좋아하는 지역 선호도도 갖고 있다. 따라서 지역 선호도와 지역 선호도를 모두 고려하기 위해, 각 시간대별 선호도를 가중평균 (weighted average)한다. 시청자  $u$ 의 시간대  $s$ 에서의 속성  $f$ 에 대한 최종적인 선호도  $M_{u,f,s}$ 는 식 (4)로 산출한다.

$$M_{u,f,s} = \frac{\alpha \cdot NP_{u,f,all} + \beta \cdot NP_{u,f,d} + \gamma \cdot NP_{u,f,h} + \delta \cdot NP_{u,f,dh}}{\alpha + \beta + \gamma + \delta} \quad (4)$$

$NP_{u,f,all}$ 은  $s$ 가 시청 히스토리의 전체기간일 때의 선호도인 지역 선호도를 의미한다.  $NP_{u,f,d}$ ,  $NP_{u,f,h}$ ,  $NP_{u,f,dh}$ 는 각각  $s$ 가 요일, 시간, 요일 및 시간일 때의 선호도인 지역 선호도를 의미한다.  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ 는 각 선호도에 대한 가중치이다.

### 2. 추천 점수 계산

시청자  $u$ 의 시간대  $s$ 에서 방영중인 TV 프로그램  $p$ 에 대한 추천점수  $S_{u,p,s}$ 는 각 속성에 대한 선호도를 가중합 (weighted sum) 하여 식 (5)로 산출한다.

$$S_{u,p,s} = \sum_i w_{f_i} \cdot M_{u,f_i,s} \quad (5)$$

$f_i$ 는 TV 프로그램  $p$ 의  $i$ 번째 속성 (예: 채널)에 해당하는 속성값 (예: KBS)을 의미한다.  $w_{f_i}$ 는 속성  $f_i$ 에 대한 가중치이다. 이를 통해, 현재 방영하고 있는 TV 프로그램들 중에서 추천점수  $S_{u,p,s}$  값이 가장 높은 1개를 시청자에게 추천한다.

## III. 실험

### 1. 실험 데이터

제안방법의 효용성을 검증하기 위해 시청률 조사 기업인 TNmS의 6개월간 개인 시청 내역 데이터와 TV 프로그램 메타데이터를 사용하였다. 시청 횟수가 너무 적을 경우 정확한 선호도 계산이 어렵기 때문에, 6개월간 시청 기록이 200개 이상인 시청자 2303명을 대상으로 실험을 진행하였다. 데이터를 정제하기 위해, 10분 미만으로 시청한 기록을 사전에 모두 제거하였다. 또한 같은 TV 프로그램에 대한 시청 기록 사이에 다른 TV 프로그램에 대한 시청 기록이 1개 존재한다면, 해당 시청 기록도 제거하였다. TV 프로그램의 채널은 24개, 장르는 43개이다.

### 2. 실험

본 연구에서 제안한 방법을 검증하기 위해 4가지의 추천 방식을 기준 (baseline)으로 사용하여 대조 실험하였다. 식 (4)에서 각 추천 방식에 설정한 가중치 값  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ 는 표 1과 같다. Popular는 현재 시각을 전혀 고려하지 않고, 시청자가 가장 선호하는 채널 또는 장르의 TV 프로그램을 추천하는 방식이다. Days, Hours, Hours per Days는 각각 시청자의 현재 요일/시간/요일 및 시간을 고려한 추천 방식이다.

표 1. 각 추천 방식의 가중치 값

Table 1. The Weight Values of Each Recommendation Policy

속성	가중치		$\alpha$	$\beta$	$\gamma$	$\delta$
	추천 방식					
채널	기준 방식	Popular	1,00	0	0	0
		Days	0	1,00	0	0
		Hours	0	0	1,00	0
		Hours per Days	0	0	0	1,00
	제안 방식		0,25	0,25	0,25	0,25
장르	기준 방식	Popular	1,00	0	0	0
		Days	0	1,00	0	0
		Hours	0	0	1,00	0
		Hours per Days	0	0	0	1,00
	제안 방식		0,10	0,20	0,30	0,40

실험의 평가 척도로서, 예측 정확도 (precision)를 사용하였다. 예측 정확도는 식 (6)으로 계산하였다.

$$Precision = \frac{\text{추천된 TV 프로그램을 시청한 횟수}}{\text{TV 프로그램 추천 횟수}} \quad (6)$$

각 추천 방식의 성능을 평가하기 위해 개인 시청 내역의 80%를 트레이닝 데이터, 20%를 테스트 데이터로 활용하였다. 그리고 각 시청자가 매 정시에 시청한 TV 프로그램을 예측하여 예측 정확도를 산출하였다.

### 3. 실험 결과

표 2. TV 프로그램 예측 정확도 (%)

Table 2. Precision of TV Program Recommendation

추천 방식	속성	채널	장르	채널+장르
기준 방식	Popular	29,48	16,64	31,21
	Days	30,27	16,90	32,08
	Hours	33,19	19,00	34,57
	Hours per Days	33,96	22,43	34,87
제안 방식		36,09	22,26	37,45

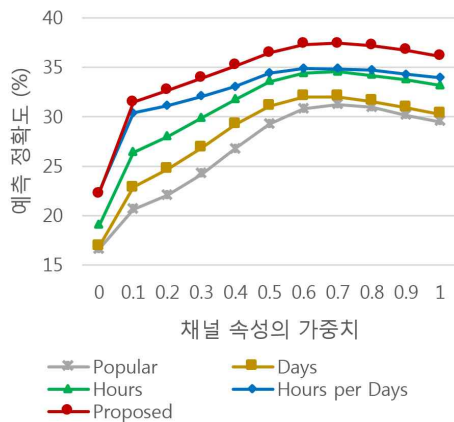


그림 2. 채널 속성의 가중치 값에 따른 예측 정확도  
Fig. 2. Precision According to the Weight Value of Channel Feature

표 2는 추천 방식에 따른 TV 프로그램 예측 정확도를 나타낸다. 속성이 채널+장르인 경우에는, 식 (5)의 가중치 값 w에 따른 가장 좋은 결과를 나타낸다. Popular 방식보다 시간 기반 추천 방식의 성능이 높았다. 이를 통해, 시청자가 선택하는 콘텐츠와 현재 시각이 서로 강한 연관성을 가짐을 알 수 있었다. 채널, 장르 속성을 모두 사용하여 제안 방식을 적용했을 때, 가장 높은 예측 정확도인 37.45%를 나타냈다. 이는 기준 방식에 비하여 높은 성능이다. 이로부터 기준 방식보다 제안 방식이 시청자에게 더욱 만족도가 높은 콘텐츠를 제공함을 알 수 있었다. 그리고 채널과 장르 속성 모두 시청자의 취향을 반영함을 알 수 있었다. 또한 시청자가 콘텐츠를 선택하는데 있어서, 현재 시각에 선호하는 속성뿐만 아니라 일상적으로 선호하는 속성도 영향을 미침을 알 수 있었다.

그림 2는 속성이 채널+장르일 때, 식 (5)의 채널 속성에 대한 가중치 값에 따른 예측 정확도를 나타낸다. 채널에 대한 가중치와 장르에 대한 가중치의 합이 1이 되게 하였다. 추천 방식에 상관없이, 장르보다 채널에 가중치를 더 높게 주었을 때, 예측 정확도가 가장 높다. 이를 통해 시청자가 콘텐츠를 결정하는데 있어서, 채널 속성이 더 중요하게 작용함을 알 수 있었다.

## IV. 결론

본 논문에서는 TV 프로그램의 채널, 장르 속성을 이용해 시청자의 선호도를 산출하였다. 그리고 시청자의 전역 선호도와 지역 선호도를 함께 고려하는 시간 기반의 개인화된 TV 프로그램 추천 방식을 제안하였다. 실 데이터를 이용한 제안 방식과 기준 방식과의 비교 실험을 통하여 제안 기법의 효용성을 검증하였다. 향후 과제로서, 시청자의 시간에 따른 시청기록을 스무딩 (smoothing) 해 준다면 시청기록이 적은 시간대에 추천 성능이 저하되는 문제점을 해결할 수 있을 것이라고 생각한다.

## Acknowledgement

본 연구는 정부(미래창조과학부) 및 한국산업기술평가관리원의 SW컴퓨팅산업융합원천기술개발사업의 일환으로 수행된 연구임 (2014-044-024-002). 또한, 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음(10041244, 스마트TV 2.0 소프트웨어 플랫폼).

## 참고문헌

- [1] Jinoh Oh, Sungchul Kim, Jinha Kim, and Hwanjo Yu, "When to recommend: A new issue on TV show recommendation," Information Sciences, pp. 261-274, 2014.
- [2] Jinoh Oh, et al., "Time-dependent user profiling for TV recommendation," Cloud and Green Computing (CGC), 2012 Second International Conference on. IEEE, pp. 783-787, 2012.

- [3] Roberto Turrin, Andrea Condorelli, Paolo Cremonesi, and Roberto Pagano, "Time-based TV programs prediction," 1st Workshop on Recommender Systems for Television and Online Video, 2014.
- [4] Hye-Woo Lee, Heera Kim, Jaedong Lee, and Jee-Hyong Lee, "Channel recommendation using TV viewing time log," Proceedings of Korean Institute of Intelligent Systems Spring Conference, vol. 24, no. 1, pp. 81-82, 2014.
- [5] Sungtak Oh, Noo-ri Kim, Jaedong Lee, and Jee-Hyong Lee, "Comparison of techniques for time aware TV channel recommendation," Joint 7th International Conference on Soft Computing and Intelligent Systems and 15th International Symposium on Advanced Intelligent Systems, pp. 989-992, 2014.