

# 메타 데이터를 이용한 채널 선택

최만석, 최형석, 오상욱, 설상훈  
고려대학교 전자공학과  
서울시 성북구 안암동 5가 1번지

## Intelligent Channel Selection based on User History Data

Manseok Choi, Hyungseok Choi, Sangwook Oh and Sanghoon Sull  
Department of Electronics Engineering, Korea University  
mschoi@mpeg.korea.ac.kr

### Abstract

In this paper, we propose a method of learning user preference for set-top box scenario. Our proposed method analyzes user history data to learn user preference and then automatically suggest the list of TV programs to the user under the assumption that the TV programs are regularly repeated on time and daily basis. We used MPEG-7 MDS to describe user history data. The experiment results show the proposed method yielded a good performance.

### 1. 서론

최근에 와서 TV시청자들은 위성방송의 보급과 더불어 보다 많아진 채널을 시청할 수 있게 되었다. 시청자들은 그들의 취향에 맞추어 방송 프로그램을 찾으며 선택한다. 그러나 채널 수의 증가에 비례하여 시청할 수 있는 프로그램이 다양해졌으며, 많은 프로그램 중에서 사용자의 취향에 적합한 프로그램을 전부 검색하는 것이 힘들게 됐다. 따라서 사용자가 원하는 시간대에 그들의 취향에 유사한 프로그램 리스트를 알 수 있

다면 보다 효과적으로 채널을 선택할 수 있다. 채널을 선택하는 방법에 있어서 사용자가 원하는 프로그램의 제목이나 좋아하는 배우를 통해 프로그램들을 일일이 사용자의 요구에 따라서 검색하는 것은 사용자의 취향에 보다 가까운 검색이지만, 실제적인 STB(Set Top Box)의 사용에 비추어 볼 때 활용률이 낮다. 따라서 사용자의 관심에 적합한 프로그램을 선별해 주는 기능이 필요하다.

TVAnytime Forum는 서로 다른 미디어 방식을 효과적으로 이용하기 위해 새로운 멀티미디어 서비스 도입을 만들기 위해 시작되었다[1]. 저장 매체와 TVAnytime/Anywhere를 이용한 새로운 멀티미디어 서비스 표준화가 작년부턴 이루어지고 있다. Conceptual Matching을 이용한 TV 프로그램 선택에서는 사용자가 이전에 시청하였던 프로그램들로부터 학습 데이터를 구성하고 일치되는 단어를 포함하는 EPG(Electronic Program Guide) 정보를 갖는 프로그램을 추천한다[2]. 혹은 화면내의 얼굴과 문자 처리(text processing)를 기반으로 TV 프로그램을 분류하거나 오디오 정보를 이용하여 프로그램의 장르를 분류하는 방법도 제시되었다[3][4].

그리고 메타데이터를 이용하기 위한 방법으로 MPEG-7 MDS를 이용한다. 기존의 텍스트 형식의 검색에 비해, 멀티 미디어료의 검색은 정보의 정보의 특

※ 본 논문은 1998년 한국학술진흥재단의 학술 연구비에 의해 지원되었음.

성상 찾고자 하는 것을 알아내기가 쉽지 않다. 이에 따라 MPEG-7 표준안이 정해 졌고, 시청각 자료에 대해 보다 효과적이고 빠른 찾기 및 필터를 가능하게 한다. 이러한 노력으로 MPEG-7 descriptors(D) 및 description schemes(DS)가 보장 되었다. D는 구별되는 데이터의 신택스와 시맨틱을 정의하였고, DS는 서로 다른 DS사이의 관계를 구조로 정의하였다. 우리는 MPEG-7 MDS중에서 사용자의 A/V자료에 대한 기호를 나타내는 UserPreference DS를 이용하여 user history data를 작성하였다.

본 논문에서는 프로그램 장르가 대체로 주일, 시간 단위로 정해진다는 것과 사용자가 시청하는 프로그램이 시간대별로 대체로 유사하다는 점에 착안하여, 시간대별 사용자의 취향과 유사한 프로그램 스케줄을 만들어 줌으로써 사용자에게 과도한 채널 변경 없이 보다 빠르고 효과적으로 채널을 선별 선택할 수 있는 방법을 제시한다.

본 논문의 구성은 다음과 같다. 먼저, 2장에서 MPEG-7 MDS의 UserPreference DS포맷을 이용한 메타데이터에 대해 기술하고, 3장에서는 사용자의 시청 패턴을 분석 방법에 대해 살펴본다. 4장에서는 스케줄링에 대해 기술하고, 마지막으로 5장에서 실험 결과와 6장에서 결론에 대해서 기술한다.

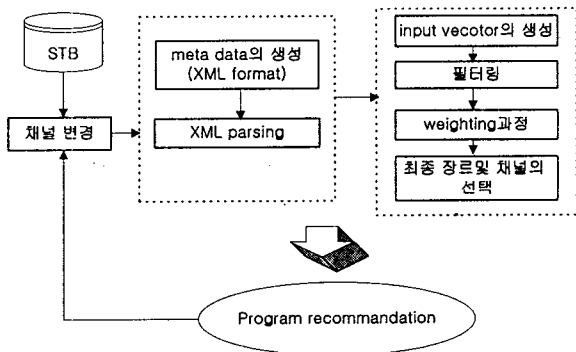


그림. 1 채널 검색 시스템

## 2. STB기반의 메타 데이터의 생성

STB를 사용하는 사용자의 취향을 알기 위한 방법으로 우리는 user history data를 만들어서 사용한다. STB를 통해 사용자가 취하는 행동과 프로그램의 정보를 채널 변경에 따라 계속 저장함으로써 user history data를 만들게 된다. user history data에 EPG 서비스를 통해 얻게 되는 프로그램 정보를 선택적으로 데이

터 구성요소로 정하여 기록한다.

User history data의 요소로는 해당 프로그램의 채널, 장르, 시청 요일, 시청 시간, 시청 상태를 포함한다. STB를 통해 사용자의 동작과 프로그램 정보가 들어있는 user history data가 자동적으로 만들어져야 하나 여건상의 어려움 때문에, 다음과 같은 방법으로 history data를 작성하였다.

TV프로그램의 방송 중에 EPG정보가 함께 온다는 가정 하에 특정 시간대(20~22시)에서 프로그램을 시청하였다. 시청하는 중에 채널 변환을 하게 되면, 시청했던 채널의 시간과 프로그램 정보, 시청상태(녹화, 시청 등)에 대해 history data가 생성된다. History data는 수작업에 의한 방법으로 기록하였다.

실제로 user history data는 STB내에서 자동적으로 생성되고 이용되기 위해서는 쉽게 parsing되고 확장성이 용이한 포맷을 가져야 한다. 이를 위해서 우리는 XML(Extensible Markup Language)를 이용하였다. User History data는 MPEG-7 MDS의 UserPreference DS 포맷을 이용하여 만들었다[5].

```
<?xml version="1.0" encoding="euc-kr" ?>
- <UsageHistory id="STB1">
- <UserIdentifier protected="ture">
- <UserName xml:lang="en">mschol</UserName>
- <UserIdentifier>
- <UserActionHistory id="uah1" protected="ture">
+ <ObservationPeriod>
- <UserActionList id="ual1" numInstances="2" totalDuration="PT01H59M">
- <ActionMode>
- <Label>Play</Label>
- <ActionMode>
- <Action>
- <ActionTime>
+ <ActionMediaTime>
- <ActionGeneralTime>
- <TimePoint>2001-09-20T20:02</TimePoint>
- <Day>THU</Day>
- <Duration>PT11M10S</Duration>
- <ActionGeneralTime>
- <ActionTime>
- <ProgramIdentifier>KBS - Baseball</ProgramIdentifier>
- </Action>
+ <Action>
- </UserActionList>
- </UserActionHistory>
- <UserChoiceHistory id="uch1" protected="ture">
+ <ObservationPeriod>
- <ClassificationHistory id="uch1" protected="ture">
+ <Language>
- <Genre term="1.4" schema="EBUEscort4.3ContentCB">
- <Label xml:lang="en">Sports</Label>
- </ClassificationHistory>
+ <CreationHistory>
- <SourceHistory>
- <Publication term="1" schema="Publication_type_mpeg7.xml">
- <Label>Terrestrial</Label>
- </Publication>
- <PublicationSource>KBS 1</PublicationSource>
- <PublicationDate>
- <TimePoint>2001-09-22T18:00</TimePoint>
- <Duration>PT2H50M</Duration>
- </PublicationDate>
- </SourceHistory>
+ <MediaFormat>
- </UserChoiceHistory>
- </UsageHistory>
```

그림2. user history data의 포맷

## 3. 사용자 선호도에 의한 장르와 채널의 선택

우리는 특정 시간과 요일에 대해 user history 정보를 토대로 가장 사용자의 취향에 근접한 확률을 가지는 프로그램의 장르와 채널정보를 얻기 위해 history data에 필터처리와 가중치 작업을 하였다.

$$P = (p_c, p_g, p_d, p_{date}, p_t)$$

P = 사용자의 선호도 상태 벡터

p<sub>c</sub> = 가장 높은 가중치를 가지는 채널

p<sub>g</sub> = 가장 높은 가중치를 가지는 장르

p<sub>t</sub> = 선택 시간, p<sub>d</sub> = 선택 요일, p<sub>date</sub> = 선택 날짜

즉 특정 시간, 날짜 그리고 요일에 대한 정보를 통해 p<sub>c</sub>와 p<sub>g</sub>의 상태를 구한다.

### 3.1 데이터의 필터처리

$$V_i = (v_c, v_g, v_d, v_{st}, v_{et}, v_{date}, v_{du}, v_{stu})$$

v<sub>c</sub> : 채널, v<sub>g</sub> : 장르, v<sub>d</sub> : 요일, v<sub>st</sub> : 시작 시간,

v<sub>et</sub> : 마침 시간, v<sub>date</sub> : 데이터 작성 시간,

v<sub>du</sub> : 전체 프로그램 시간, v<sub>stu</sub> : 시청 상태

p<sub>g</sub>의 값을 얻기 위하여 user history data를 위의 입력 벡터형태로 변환하였다. 그런데 STB를 통한 프로그램의 시청 시에 많은 채널 변화가 일어나므로 각각의 채널 변경 때마다 생성되는 user history data를 V<sub>i</sub>로 변환하는 것은 비효율적이다. 따라서 전체 시청 프로그램의 시간과 전체 프로그램의 시간 비율을 정의하여 임계치를 넘지 못하는 데이터는 필터링하였다.

$$w_{TR_i} = \frac{v_{et} - v_{st}}{v_{du}}, \text{ (where } w_{TR_i} < Thr_w, w_{TR_i} = 0)$$

v<sub>et</sub> - v<sub>st</sub> : 프로그램 시청 시간

그리고 사용자의 취향은 시간에 따라서 변할 수 있다. 따라서 이와 같은 데이터는 입력시간에 대한 의존도가 높다고 할 수 있다. 그러므로 현재 시간과 비교하여 최근에 생성된 데이터만을 사용자 선호도 작업에 이용하기 위해서 입력 벡터 V<sub>i</sub>는 circular linked list를 이용하여 구조화하였다.

### 3.2 데이터의 필터처리

위의 입력벡터 V<sub>i</sub>를 이용하여 신뢰도 있는 p<sub>g</sub>와 p<sub>c</sub>를 구하기 위하여 다음과 같이 가중치 함수를 정의한다. 가중치 함수 F<sub>G</sub>는 현재 시간, 요일, 날짜를 입력값으로 하여 특정 장르 G에 대한 선호도 값을 결정한다.

$$F_G(p_t, p_{day}, p_{date}) = \sum_{i=0}^{n-10} (w_i \delta(G, v_{g_i})) + \alpha \sum_{i=n-10}^n (w_i \delta(G, v_{g_i}))$$

$$\delta(G, v_{g_i}) = \begin{cases} 1, & \text{where } G = v_{g_i} \\ 0, & G \neq v_{g_i} \end{cases}$$

α = 최근 입력벡터 10개에 대한 가중치상수

$$w_i = w_{stu_i} w_{DT} w_{TR} w_{DW} w_{GR}$$

$$w_{GR} = \frac{M_C - M_G}{M_C}$$

(where, M<sub>G</sub> << M<sub>C</sub> 이면 w<sub>GR</sub> = 1)

(M<sub>C</sub> : 채널의 총 수, M<sub>G</sub> : 장르의 총 수)

$$w_{STU_i} = \begin{cases} 1, & \text{where } v_{stu_i} = record \\ 0.8, & \text{otherwise} \end{cases}$$

$$w_{D_i} = \begin{cases} 1, & \text{where } P_d = v_{d_i} \\ 0.8, & \text{otherwise} \end{cases}$$

$$w_{DT} = e^{-0.023(P_{date} - v_{date_i})}$$

w<sub>GR</sub>은 전체 채널의 수에 대한 장르의 수를 정의한다. 예를 들어 50개의 채널 중 현재 시청하는 장르와 같은 채널이 2개 정도라면, 이 채널의 장르가 보다 사용자의 취향을 잘 나타낼 수 있는 기준이 된다. 그러나 본 연구에서는 history data를 통해 얻는 M<sub>C</sub> 값이 낮았기 때문에 가중치 w<sub>GR</sub>는 의미 있는 값을 가질 수가 있었다. 실제 STB의 채널수는 많으므로 w<sub>GR</sub>의 값은 무시될 수 있다. w<sub>DT</sub>는 현재 날짜와 가까운 데이터일수록 높은 가중치를 가지는 것을 의미한다.

마찬가지로 특정채널에 대한 가중치 함수 F<sub>C</sub>는 다음과 같이 정의한다.

$$F_C(p_t, p_{day}, p_{date}) = \sum_{i=0}^{n-10} (w_i \delta(C, v_{c_i})) + \alpha \sum_{i=n-10}^n (w_i \delta(C, v_{c_i}))$$

$$\delta(C, v_{c_i}) = \begin{cases} 1, & \text{where } C = v_{c_i} \\ 0, & C \neq v_{c_i} \end{cases}$$

위에서 얻어진 장르 및 채널에 대한 함수 값을 토대로 p<sub>g</sub>를 구한다.

$$p_g = \arg \text{Max } F_{G_i}, i = 1, 2, \dots, M_G$$

$$p_c = \arg \text{Max } F_{C_i}, i = 1, 2, \dots, M_C$$

## 4. 프로그램 스케줄링의 생성

우리는 위에서 제안된 방법을 사용자 취향의 프로그램을 제안해 주는 agent에 적용하였다. 3장에서 얻은 p<sub>g</sub> 값을 이용하여 다음의 두 가지 방법으로 프로그램의 선택을 할 수 있게 하였다. 특정 시간대에 가장 가중치가 높은 프로그램을 추천하거나 가중치가 높은 프로그램들에 대한 스케줄 리스트를 작성하였다.

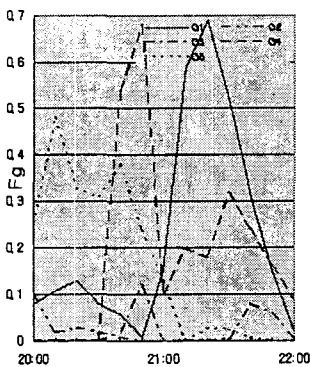
#### 4.1 프로그램의 선택

3장에서 얻은  $p_g$  값을 토대로, 특정 시간에 대해서 가장 가중치가 높은 프로그램의 장르를 구한다. 우리는 EPG정보를 통해서 현재 방송되는 프로그램들 중에서  $P_g$ 과 같은 값을 가지는 프로그램을 선택하여 제시하였다. 따라서 제시된 프로그램은 특정시각에 사용자가 가장 좋아하는 취향의 프로그램의 장르가 될 것이다.

#### 4.2 스케줄링의 작성

시간에 따라 사용자의 시청 취향이 변하므로 특정시간 동안 여러 장르의 프로그램이 선택될 수 있다. 우리는 다이너믹 프로그래밍 기법 중 flow shop scheduling 알고리즘을 이용하여, 그 시간동안 사용자가 시청 가능한 프로그램의 스케줄을 최적으로 배치함으로써 사용자에게는 별도의 시간 편성표 없이 그들의 취향에 유사한 스케줄을 알 수 있는 시스템을 구성하였다. 예를 들어 20:00~23:00시간대의 스케줄을 보고 싶다면, user history data중에서 해당 시간대에 대한 데이터를 분류하고 사용자의 취향을 토대로 현재 프로그램의 정보를 검색하여 사용자가 시청 할 수 있는 최적의 프로그램 스케줄러를 만들어 준다.

### 5. 실험결과



	장르
G1	뉴스
G2	코믹
G3	다큐멘터리
G4	드라마
G5	오락

그림.3 시간에 대한  $F_g$ 의 변화 표.1 장르 정보

우리는 특정 시간대(20~22시)에서 프로그램의 프로그램 시청을 통해 한달 동안 약 350개의 history data를 생성하였다. 특정 시간대(20~22시)에서 공중파 프로그램 시청을 통해 한달 반 동안 약 350개의 history data를 생성하였다. 이 데이터를 이용하여 가중치 함수값  $F_g$ 를 구하였다. 그림 3은 10월23일 화요일의 입력값에

대해서 시간대별(20시~22시)사이의  $F_g$ 값을 나타낸다. 결과를 통해 9시 10분경은 장르 '뉴스'가 가장 많이 선호됨을 알 수 있다.

### 6. 결론

본 논문에서는 user history data를 이용하여 효과적으로 사용자의 취향에 유사한 채널을 선택하는 방법을 제시하였다. 우리는 STB를 통해 사용자가 시청한 프로그램 정보를 메타 데이터로 기록하였다. 기록된 데이터의 필터 처리 및 시청 시간과 빈도 등을 토대로 데이터에 가중치를 부여함으로써 보다 시간대별 사용자의 취향에 가장 가까운 장르 및 채널을 구하였다. 현재 STB를 통해 시청할 수 있는 채널을 검색하여, 위에서 구한 장르 및 채널을 비교함으로써 사용자 중심의 프로그램을 선택하고 추천해 주었다. 우리는 제시된 방법을 이용하여 어플리케이션 시스템을 디자인하여 시뮬레이션하였다. 실험 결과에서 보여지듯이 본 논문에서 제안된 기법이 사용자의 채널 선택에 매우 효과적인 도움이 되었다. 프로그램의 장르에 대해 좀 더 세밀한 분류가 된다면 보다 개선된 사용자 취향의 분류가 될 것이다.

### 참고문헌

- [1] TVAnytime Forum <http://www.tv-anytime.org/>
- [2] Takagi. T, Kasuya. S, Mukaidono. M and Yamaguchi. T, "Conceptual matching and its applications to selection of TV programs and BGMs", Systems, Man, and Cybernetic, IEEE SMC.99, Vol. 3, 1999, pp. 269-273.
- [3] Wei. G, Agnihotri. L and Dimitrova. N, "TV program classification based on face and text processing", Multimedia and Expo. 2000. ICME 2000, Vol. 3, 2000, pp. 1345-1348.
- [4] Jasinschi. R.S and Louie. J, "Automatic TV program genre classification based on audio patterns", Euromicro Conference 2001. Proceedings 27th, 2001, pp. 370-375.
- [5] MPEG-7 Document, ISO/IEC 15938-5/CD, "Information Technology-Multimedia Content Description Interface - Part5 Multimedia Description Schemes" 2000.ISO/IEC JTC 1/SC 29 N 3705. User History DS.