

시청자 프로파일 추론과 TV Anytime 메타데이터를 이용한 표적 광고

(Target Advertisement based on a TV Viewer's Profile
Inference and TV Anytime Metadata)

김문조[†] 이범식^{**} 임정연^{**}
(Mun Jo Kim) (Bum Sik Lee) (Jeong Yon Lim)

김문철^{***} 이희경^{****} 이한규^{****}
(Munchurl Kim) (Hee Kyung Lee) (Han Gyu Lee)

요약 지상파, 위성파, 케이블 방송 같은 기존의 방송환경은 시청자 취향에 상관없이 일방적인 단방향 방송 서비스를 제공해 왔다. 하지만, 최근에는 광대역 통신망을 통한 다양한 미디어 전송이 가능하게 되었다. 또한, 방송 환경에서 양방향 통신이 가능하게 됨으로써 장르, 시청 시간대, 배우 등 시청자의 선호도를 반영한 방송 서비스가 증용한 응용으로 대두되고 있다. 따라서, 기존의 방송환경에서 시청자의 선호도를 반영한 맞춤형 방송 서비스가 중요한 방송 서비스의 하나가 될 수 있다.

본 논문에서는 표적광고를 위한 새로운 시도로서 맞춤형 방송 서비스 응용 중 하나인 시청자 프로파일 추론을 통한 표적 광고 방법을 소개한다. 제안된 시청자 프로파일 추론 알고리즘은 시청자의 TV 시청 데이터(TV Viewing history data) 분석을 통해 시청자의 성별 및 연령대를 예측한다. 예측된 시청자의 성별 및 연령대를 바탕으로 TV Anytime 메타데이터를 이용한 표적 광고 선별 방법을 통하여 광고를 선택하게 된다. 제안된 표적 광고 시스템은 시청자 프로파일 추론 알고리즘과 표적 광고 선별 방법을 이용하여 구성되어 있으며, 실제 TV 시청 데이터를 이용하여 제안된 표적 광고 시스템의 실험 결과를 제시한다.

키워드 : 표적 광고, 다단계 분류기, TV Anytime, 프로파일 추론

Abstract The traditional broadcasting services over terrestrial, satellite and cable media have been unidirectional mass media regardless of TV viewer's preferences. Recently rich media streaming has become possible via the broadband networks. Furthermore, since bidirectional communication is possible, personalcasting such as personalized streaming services has been emerging by taking into account the user's preference on content genres, viewing times and actors/actresses etc. Accordingly, personal media becomes an important means for content provision service in addition to the traditional broadcasting service as mass media. In this paper, we introduce a user profile reasoning method for target advertisement which is considered an important application in personalcasting service. The proposed user profile reasoning method predicts an unknown TV viewer's gender and ages by analyzing TV Viewing history data. Based on the estimated user's gender and ages, a target advertisement is provided with TV Anytime metadata. A proposed target advertisement system is developed based on the user profile reasoning and the target advertisement selection method. To show the effectiveness of our proposed methods, we present a plenty of experimental results by using realistic TV Viewing history data.

Key words : target advertisement, multi-stage classifier, TV Anytime, profile inference

[†] 학생회원 : 한국정보통신대학교 공학부
kimmj@icu.ac.kr

^{**} 비회원 : 한국정보통신대학교 공학부
bslee@icu.ac.kr
jylim@icu.ac.kr

^{***} 비회원 : 한국정보통신대학교 공학부 교수
mkim@icu.ac.kr

^{****} 비회원 : 한국전자통신연구원 방송미디어연구그룹 디지털방송연구단 연구원
lhk95@etri.re.kr

^{****} 비회원 : 한국전자통신연구원 방송미디어연구그룹 디지털방송연구단 팀장
hkl@etri.re.kr

논문접수 : 2005년 12월 21일

심사완료 : 2006년 7월 16일

1. 서론

인터넷의 폭발적 성장은 인터넷 사용인구의 증가와 인터넷을 통한 다양한 정보의 생산 및 소비를 가능하게 하였다. 또한, 정보 트래픽의 급속한 증가와 더불어 원활한 정보의 전송을 위해 인터넷 대역폭도 함께 증가되어 왔다. 인터넷을 통한 많은 데이터 중 동영상 전송은 매우 중요한 응용 중의 하나로 등장하였고, 인터넷 상에서의 동영상 서비스 형태인 인터넷 방송 또는 웹 캐스팅(web casting) 서비스가 매우 보편화 되었고 향후 광대역 인터넷 망을 통한 IPTV 서비스가 활성화될 것으로 기대된다[1].

인터넷은 양방향 통신을 이용하여 대화형 방송 서비스 및 개인 맞춤형 방송 서비스 분야에 매우 매력적인 수단이다. 미래의 방송 환경은 인터넷을 이용하여 개인 선호도 중심의 맞춤형 방송으로 진화될 것으로 예상된다. 이러한 움직임에 맞추어 맞춤형 방송에 사용될 방송 기술에 관한 연구가 필수적이라고 할 수 있다. 예를 들어, 현재와 같은 단 방향 방송 환경에서의 광고 방송은 시청자의 관심에 상관없이 방송사에 의해 프로그램의 인기도 혹은 시청률, 시청 연령, 시간대에 따라서 획일적으로 광고를 전송하고 있기 때문에 효과적인 광고 방송이 이루어지지 못하고 있다. 시청자의 광고 콘텐츠에 대한 선호도에 따라 표적 광고 서비스(target advertisement service)를 제공할 수 있다면 맞춤형 방송의 중요한 응용 분야로 자리잡을 것으로 예상된다.

현재와 같은 부분별한 광고 제공 문제를 해소하기 위한 기존의 표적 광고에 대한 연구로서, 유사한 소비 성향을 가지고 있는 TV 시청자의 그룹을 분석하고 그에 맞는 표적 광고 서비스를 가능하게 하는 표적 광고 서비스와 디지털 TV 협력 필터링(collaborative filtering)을 이용하여 미리 주어진 사용자의 정보를 바탕으로 실제 사용자의 정보를 파악하여 표적 광고를 가능하게 연구 등이 제시되었다[2,4]. 이러한 연구들은 디지털 TV에 연결되어 있는 셋탑 박스(Set-Top Box)나 컴퓨터에 시청자가 성별, 직업, 연령대 같은 시청자의 프로파일을 명시적으로 입력하고 이러한 입력된 정보는 인터넷을 통하여 TV 콘텐츠 서버로 전송된다. 이러한 명시적인 정보를 바탕으로 광고 콘텐츠 제공자들이 맞춤형 광고를 제공하였다. 그러나 인터넷을 통한 개인 정보 유출의 피해가 있기 때문에 시청자들은 개인 정보를 입력하는 것을 꺼려하며, 이는 올바른 표적 광고 서비스를 제공하기 힘들게 한다.

본 논문에서는 명시적인(explicit) 시청자의 입력 없이 프로그램 시청일자, 시간, 장르와 같은 묵시적인(implicit) 시청 정보인 TV 시청 데이터(TV Viewing history data)

를 이용하여 TV 시청자의 프로파일(성별/연령대)를 추론할 수 있는 방법과 추론된 시청자의 프로파일과 TV Anytime 메타데이터를 이용한 표적 광고 선별 방법에 기반을 둔 표적 광고 시스템을 제안한다. 사용자 프로파일의 추론 및 그를 이용한 표적 광고 방법은 기존의 연구와 다른 새로운 시도이며 본 논문에서는 실제 TV 시청 데이터를 이용한 프로파일 추론 알고리즘 제안 및 성능 평가와 TV Anytime을 기반으로 한 표적 광고 선별 방법을 제공하는 프로토타입 표적 광고 시스템을 구현하였다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 표적 광고 시스템 구조에 대하여 설명하고, 3장에서는 시청자 프로파일 추론 알고리즘에 관하여 설명한다. 그리고 4장에서는 TV Anytime 메타데이터를 이용한 표적 광고 선별 방법에 대하여 알아보고, 5장에서는 실험 결과를 제시하고, 마지막으로 결론 및 향후 연구계획을 6장에 제시하였다.

2. 표적 광고 시스템 구조

본 논문에서 제안된 표적 광고 시스템은 방송국, 광고주 그리고 시청자를 고려하며, 시청자의 TV 시청 데이터를 분석하여 시청자의 프로파일을 추론하는 시청자 프로파일 추론 기능, 추론된 결과를 바탕으로 추천 서비스를 제공하는 방송 프로그램 전송 기능 그리고 시청자의 프로파일을 보호하고, 사용자의 시청 내역에 따른 TV 시청 데이터(TV Viewing history data)를 수집하여 제한적인 내용만을 전송하며, 추천 서비스를 제공할 수 있는 사용자 인터페이스 기능을 요구한다. 그림 1은 표적 광고 서비스를 위한 표적 광고 시스템의 구조를 보여주고 있다. 표적 광고 시스템 구조는 사용자 프로파일(성별/연령대) 추론 기능을 제공하는 프로파일 추론 에이전트, 추론 결과를 바탕으로 광고 콘텐츠를 선별하고 전송 기능을 제공하는 콘텐츠 제공 에이전트 그리고 사용자 인터페이스 에이전트로 구성되어 있다.

그림 1에서와 같이, 프로파일 추론 에이전트는 사용자 인터페이스 에이전트로부터 TV 시청 데이터 즉, 사용자가 시청한 TV 프로그램 제목, 장르, 채널, 시청 요일 등의 정보를 얻는다. 이러한 정보를 이용하여 프로파일 추론 에이전트는 시청자의 프로파일(성별/연령대)를 사용자 프로파일 추론 모듈을 통하여 추론하며, 추론 결과를 콘텐츠 제공 에이전트로 보내어 준다. 콘텐츠 제공 에이전트는 광고 콘텐츠 선별 모듈을 통하여 표적 광고 콘텐츠를 선별한다. 선별된 광고 콘텐츠는 TV 프로그램 콘텐츠와 주문형 비디오 서비스(VOD: Video On Demand)와 같은 형태로 방송국에 의해 전송될 수 있다. 사용자 인터페이스 에이전트는 TV 단말 상에서 TV 시

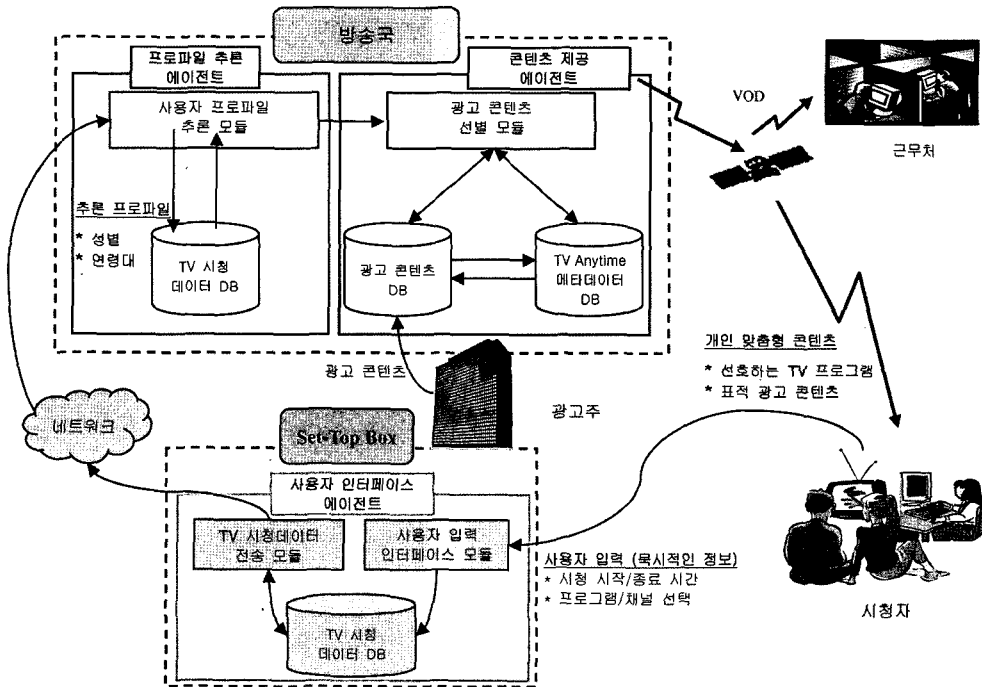


그림 1 표적 광고 시스템 구조

청자가 콘텐츠 소비 및 관련 데이터 소비를 가능하게 할 수 있는 그래픽 사용자 인터페이스를 제공한다. 이 에이전트는 보통 가정 내에서 설치되어 있는 셋탑 박스(Set-Top Box)에서 실행되며 콘텐츠 제공 에이전트로 부터 전달된 추천 서비스를 받아 사용자에게 추천 정보를 제공하고 소비를 가능케 한다. 사용자가 TV를 시청하는 동안 사용자가 시청한 TV 프로그램과 시청 시간 등의 정보는 사용자 입력 인터페이스 모듈을 통하여 셋탑 박스에 저장된다. 이렇게 저장된 시청자 정보의 개인 정보 유출되는 것을 방지하기 위해 정보 제공의 수위에 따라 제공 가능한 시청 정보와 사용자의 개인적 정보로 나뉘며, 제한된 사용자의 정보만이 TV 시청 데이터 전송 모듈을 통하여 프로파일 추론 에이전트로 전달되어 사용자의 프로파일(성별/연령대)을 추론하게 된다.

3. 사용자 프로파일 추론 알고리즘

표적 광고 시스템의 가장 중요한 핵심 기술은 사용자 프로파일을 추론하는 알고리즘이며, 프로파일 추론 알고리즘인 다단계 분류기(MSC: Multi-Stage Classifier)와 특징 벡터(Feature Vector) 추출 방법을 특징으로 한다.

3.1 특징 벡터 추출

사용자 프로파일 추론 알고리즘에 사용되는 특징 벡터는 사용자의 TV 시청 데이터로부터 획득할 수 있다. TV 시청 데이터는 전자 프로그램 가이드(EPG) 정보로

부터 시청자가 소비한 TV 프로그램의 제목, 날짜, 장르, 채널 등 TV 시청 데이터를 가지고 있는 데이터베이스를 의미하며, 이런 TV 시청 데이터의 데이터베이스에는 표 1과 같은 필드들이 존재한다.

표 1 TV 시청 데이터 데이터베이스의 필드 및 설명

필드	설명
id	시청자 아이디
profile	시청자의 성별 및 연령
date	프로그램 방송 날짜
dayofweek	시청한 요일
subscstart_t	시청자의 프로그램 시청 시작시간
subscend_t	시청자의 프로그램 시청 종료시간
programstart_t	프로그램의 정규 시작시간
programend_t	프로그램의 정규 종료시간
title	프로그램의 제목
channel	프로그램의 채널(6개 채널)
genre	프로그램의 장르(8개 장르)

또한, 시청자의 프로파일(성별/연령대)을 추론하기 위한 특징 값(Feature Value)는 다음 조건을 만족해야 한다.

$$\frac{(subscend_t - subscstart_t)}{(programend_t - programstart_t)} \geq T_{th}$$

$$\sum PGs \geq C_{th}$$

$(\text{subscend}_t - \text{subscstart}_t) / (\text{programend}_t - \text{programstart}_t)$ 는 어떤 특정 프로그램의 정규방송시간에 비례하여 얼마나 시청했는가를 의미하고, T_{Th} 는 이에 대한 문턱 값(threshold value)를 의미한다. $\sum PGs$ 는 특정기간 동안 T_{Th} 를 만족하는 프로그램의 총 개수이고, C_{Th} 는 이에 대한 문턱 값을 의미한다. 위의 두 조건을 만족하면서 데이터베이스에서 개인별로 추출할 수 있는 특징 값들과 개수는 표 2와 같다.

표 2 특징 값의 종류와 개수

특징 값 종류 및 계산식	개수
<ul style="list-style-type: none"> 시청한 장르 횟수에 대한 확률 (GPRC: Genre Probability based on the number of counts) $GPRC_{i,k,a} = GC_{i,k,a} / \sum_{i=1}^I GC_{i,k,a}$	8
<ul style="list-style-type: none"> 시청한 장르 시청시간의 장르 확률 (GPRT: Genre probability based on the amount of consumption time) $GPRT_{i,k,a} = GT_{i,k,a} / \sum_{i=1}^I GT_{i,k,a}$	8
<ul style="list-style-type: none"> 특정 기간 동안의 평균 시청 시간 (AVT: Average viewing time) $AVT_{k,a} = VT_{k,a} / TotTime$	1
<ul style="list-style-type: none"> 시청한 채널 시청시간의 채널 확률 (CPR: Channel probability based on the amount of consumption time) $CPR_{j,k,a} = CT_{j,k,a} / \sum_{j=1}^J CT_{j,k,a}$	6

표 2에서, $GC_{i,k,a}$ 는 미리 정의된 특정기간 동안 성별 및 연령대 그룹 a 에 속하는 시청자 k 가 소비한 장르 i 에 대한 빈도수를 의미하며, $GT_{i,k,a}$ 는 그룹 a 의 시청자 k 가 특정기간 동안에 소비한 장르 i 에 대한 시청시간을 말한다. 또한, $VT_{k,a}$ 는 그룹 a 에 속한 시청자 k 가 특정기간 동안의 시청시간이고, $TotTime$ 은 특정기간의 전체 시간이다. $GT_{j,k,a}$ 는 그룹 a 의 시청자 k 가 소비한 채널 j 에 대한 특정기간 동안 소비한 시청시간을 의미한다. I 는 전체 장르 수(8개) 이고, J 는 전체 채널 수(6개)를 의미한다.

표 2를 이용하여 개인별 특정기간 동안의 특정 요일에 해당하는 특징 벡터를 만들면 표 3과 같다.

표 3의 특징 벡터는 23개의 특징 값을 가지고 있으며, 2가지의 특징 벡터가 존재한다. 시청자 개인별 특징 벡터와 성별/연령대 그룹을 대표하기 위한 그룹 특징 벡터가 있다. 그림 1의 프로파일 추론 에이전트는 성별/연

표 3 특징 벡터

Index	1~8	9~16	17	18~23
특징 값	GPRC	GPRT	AVT	CPR

령대 그룹을 기준으로 시청자 별 특징 벡터들로 이루어져 있는 룩업 테이블(Look Up Table)을 만들 수 있다. MSC는 특징 벡터를 가지고 있는 룩업 테이블을 이용하여 특정 시청자의 프로파일(성별/연령대)을 추론하게 된다. 또한, 특징 벡터는 주중의 데이터, 즉 월요일부터 금요일까지의 TV 시청 데이터를 이용하여 특징 벡터를 추출한다. 이는 주말에는 주중에 비해서 TV를 시청하는 패턴이 성별/연령대 그룹 별로 유사하기 때문에 특징 값들로 고려되지 않는다.

3.2 제 1 단계 분류기(새로운 벡터거리 척도)

제 1단계 분류기는 특정 요일에 해당하는 특징 벡터 간의 거리를 이용하여 구한다. 특징 벡터간의 거리 측정 은 벡터 상관법(vector correlation)과 정규화된 유클리드 거리(normalized Euclidean distance) 방법을 이용하여 특징 벡터간의 유사성을 측정한다. 벡터 상관법은 벡터간의 유사성 측정을 위하여 식 (1)을 이용한다 [5].

$$VC(x, y) = \cos \theta = \frac{x \cdot y}{\|x\| \cdot \|y\|} = \frac{\sum_{i=1}^m x_i y_i}{\sqrt{\sum_{i=1}^m x_i^2} \sqrt{\sum_{i=1}^m y_i^2}} \quad (1)$$

하지만, 벡터 상관법은 벡터간의 유사성을 벡터간 이루고 있는 각을 이용하여 구하였기 때문에 실제 벡터간의 거리를 고려하지 않고 방향성만 고려한 단점이 있다.

정규화된 유클리드 거리는 특징 벡터를 이루고 있는 특징 값들의 분산을 이용하고, 이는 식 (2)에서 보여지고 있다.

$$ED(x, y) = \sqrt{\sum_{i=1}^m \frac{(x_i - y_i)^2}{\sigma_{i,g}^2}} \quad (2)$$

식 (2)에서 g 는 특정 성별 및 연령대 그룹을 나타낸다. 정규화된 유클리드 거리는 벡터간의 거리만 고려하지 방향성을 고려하지 않은 방법이다. 그래서, 본 논문에서는 벡터간의 방향성과 거리를 같이 고려한 새로운 벡터간의 거리 척도 방법을 제안한다. 제안된 벡터간의 거리 척도 방법은 벡터 상관법과 유클리드 거리에 가중치를 적용하는데 두 벡터 V_i 와 V_l 사이의 새로운 벡터 거리 척도 방법은 식 (3)과 같이 나타낼 수 있다.

$$\begin{aligned} Dist(V_i, V_l) &= GVC(V_i, V_l) + GED(V_i, V_l) \\ GVC(V_i, V_l) &= (1 - W_{i,v}) \times (1 - VC(V_i, V_l)) \\ GED(V_i, V_l) &= W_{i,e} \times ED(V_i, V_l) \end{aligned} \quad (3)$$

식 (3)에서 $i \in I$ 이고, I 는 그룹의 인덱스이다. 또한,

$W_{I,v} = VC(G_i, V_i)$ 이며 $W_{I,E} = ED(G_i, V_i)$ 이다. G_i 는 그룹 I 의 그룹 특징 벡터를 의미한다. 즉, $W_{I,v}$ 와 $W_{I,E}$ 는 그룹 특징 벡터 G_i 와 V_i 사이의 벡터 상관 계수치와 정규화된 유클리드 거리 값이 된다. 또한, V_i 는 록업 테이블의 그룹 I 에 속해 있는 i 번째 특징 벡터이고, V_i 는 추론하고자 하는 시청자의 특징 벡터를 의미한다.

제1단계 분류기는 새로운 벡터 거리 척도 방법인 식 (3)을 이용하여 거리 값을 측정 한 뒤, 벡터 거리표(VDT: Vector Distance Table)을 만든다. 그림 2는 제1단계 분류기의 측정 방법의 예를 도시하였다.

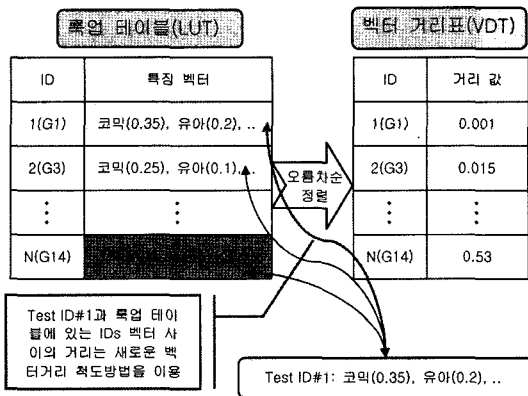


그림 2 제1단계 분류기 측정 방법 예

추론하고자 하는 시청자의 특징 벡터가 프로파일 추론 에이전트로 전달되면, 프로파일 추론 에이전트에는 이미 여러 시청자들에 대한 특징 벡터들로 이루어져 있는 록업 테이블이 있다. 이 록업 테이블 내에 있는 특징 벡터와 추론한 시청자의 특징 벡터간의 거리를 식 (3)을 이용하여 구할 수 있고, 구한 결과 값을 오름차순으로 정렬하여 벡터 거리 표(VDT)를 만들 수 있다.

3.3 제2단계 분류기(k-NN의 적용)

제2단계 분류기는 제 1단계에서 구한 VDT를 기반으로 k-NN(k-Nearest Neighbor) 방법을 이용한다. 일반적인 k-NN 방법은 상위 k개에 속하는 같은 그룹들의 개수에 의해서 결정이 된다. 즉, k-NN 방법은 상위 k개에 속한 그룹들의 개수만을 고려하며 실제 k개 속한 값들은 고려되지 않는다. 이 문제를 해결하기 위하여 제 2단계 분류기는 상위 k개에 속하는 거리 값을 이용한 가중 거리 k-NN(Weighted-Distance k-Nearest Neighbor) 방법을 이용한다[6]. 가중 거리 k-NN은 그림 2의 VDT내의 상위 k개의 개수 및 거리 값을 고려하며, 이는 식 (4)를 이용하여 구해진다.

$$WDK(I) = \frac{\sum_{i \in I, j \in k} 1/VDT(i)}{\sum_{j=1}^N \sum_{j \in k} 1/VDT(j, G_i)} \quad (4)$$

식 (4)에서 $i \in I$, I 는 k 에 속하는 그룹의 인덱스, k 는 k-NN의 k 값을 의미한다. 또한, $VDT(i)$ 는 k-NN의 k개에 속하고 그룹 I 에 속하는 벡터 거리 값을 뜻한다. N 은 그룹의 총 개수를 의미하며, $VDT(j, G_i)$ 는 k-NN의 k개에 속하는 G_i 그룹들의 벡터 거리 값이다. 식 (4)를 이용하면, VDT내의 상위 k개에 속하는 그룹들의 가중 거리 k-NN 표(WDKT: Weighted-Distance KNN Table)을 만들 수 있다.

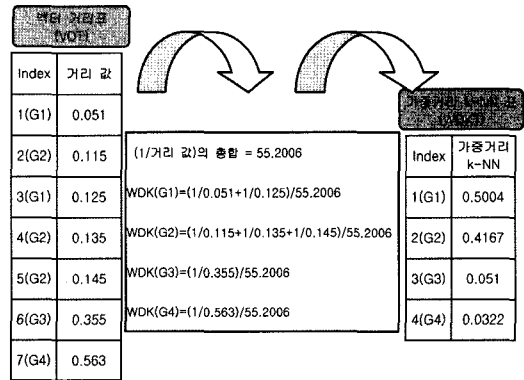


그림 3 제2단계 분류기 측정 방법 예

그림 3은 제2단계 분류기인 가중 거리 k-NN을 구하는 방법의 예를 도시하였다.

제1단계 분류기를 거치면 벡터 거리표가 만들어지는데 위의 그림 3은 k가 7인 경우의 예를 보이고 있으며, 가운데 노란색 상자는 식 (4)와 벡터 거리표를 이용하여 구하는 방법을 보이고 있다. 식 (4)를 이용하면 가중거리 k-NN 표(WDKT)를 만들 수 있다. 제2단계 분류기를 거치면 월요일부터 금요일까지의 WDKT와 WDKT내의 최대 값을 이용하여 추론 결과를 알 수 있다.

3.4 제 3 단계 분류기(다수결 원칙 적용)

사용자의 프로파일을 추론하기 위하여 제1단계와 제2 단계를 거치면, 월요일부터 금요일까지의 가중 거리 k-NN표(WDKT)를 가질 수 있다. 제3단계 분류기는 제1, 2단계 분류기를 거친 요일 별 WDKT의 최대 값과 추론 결과 값을 이용하여 다수결 원칙 표(MRT: Majority Rule Table)을 작성한다. 제3단계 분류기는 다수결 원칙(Majority Rule)과 요일 별 WDKT내의 최대 값을 이용한 정규화된 다수결 원칙(Normalized Majority Rule) 식을 이용하여 최종 추론 결과를 낸다. 제3단계 분류기를 위한 다단계 분류기의 전체 추론 구조를 그림 4에서 보여주고 있다.

그림 4에서 테스트 데이터는 프로파일을 추론 할 시청자 데이터를 이야기 하며, 테스트 데이터는 시청자 인터페이스 에이전트로부터 월요일부터 금요일까지의 데

이타를 이용하여 각 요일에 맞게 특징벡터를 추출한다. 프로파일 추론 에이전트는 월요일부터 금요일까지의 여러 시청자들에 대한 특징 벡터를 가지고 있다. 제1, 2단계 분류기를 거치면 각 요일의 추론 결과와 가중거리 k-NN 값을 가지고 있는 WDKT을 만들 수 있다. WDKT 내의 최대 값과 추론 결과를 이용하여 MRT를 작성하고 정규화된 다수결 원칙을 이용하면 최종 추론 결과를 얻을 수 있다. 식 (5)는 정규화된 다수결 원칙의 식을 보여주고 있다.

$$NMR(I) = \frac{\max\{WDKT(d) | d \in D\}}{\sum_{d=1}^D \max\{WDKT(d) | d \in D\}} \quad (5)$$

식 (5)에서 I 는 요일 별 추론 결과의 인덱스이며, D 는 월요일부터 금요일까지를 의미하며, $WDKT(d)$ 는 특정 요일의 WDKT내의 값을 의미한다. 제3단계 분류기에서 MRT와 식 (5)를 이용하면 시청자의 프로파일을 추론할 수 있고, 그림 5는 이에 대한 예를 보이고 있다.

그림 5의 다수결 원칙 표는 가중거리 k-NN 표에서 최대 값과 각 요일의 추론 결과를 이용하여 만든 표이다. 또한, 그림 5의 노란색 상자 내에는 식 (5)를 이용하여 추론하는 방법을 보이고 있다.

4. TV Anytime을 이용한 표적광고 콘텐츠 선별 방법

본 장에서는 사용자의 프로파일 추론 결과를 이용하여 실제로 광고를 선별하는 방법에 대해서 알아본다. 광고를 선별하는 방법은 TV Anytime과 한국방송광고공사에서 제공하는 성별/연령대 별 광고의 선호도 값을

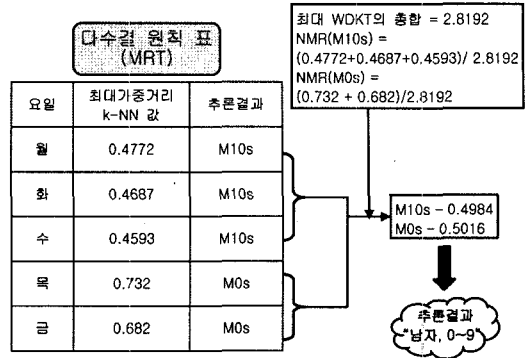


그림 5 제3단계 분류기 추정 방법 예

이용하여 선별한다.

TV Anytime은 저장매체를 갖는 사용자 단말, 즉 PDR(Personal Digital Recorder) 또는 PVR(Personal Video Recorder)를 이용하여 사용자가 원하는 콘텐츠를 원하는 시간에 선택, 소비할 수 있는 TV 프로그램 콘텐츠의 Anytime 서비스를 위한 메타데이터를 표준화하는 민간 표준기구이다. TV Anytime(TVA) 표준은 방송환경과 제공하고자 하는 기능에 따라 1단계(Phase 1)와 2단계(Phase 2)의 표준 규격을 제정하여 왔다. 1단계 표준규격에는 PDR이나 PVR을 중심으로 메타데이터를 이용한 콘텐츠 검색 및 선택을 통해 TV 프로그램 콘텐츠를 소비하는 환경에 필요한 콘텐츠 참조, 메타데이터 등에 대한 표준을 제정하였고, 2단계 표준에서는 양방향 통신 네트워크를 통한 향상된 콘텐츠 소비를 위해 홈 네트워크 환경에서 단말간의 콘텐츠 공유 및 재분배, 그리고 다양한 사용자 환경에 맞는 콘텐츠를 제공

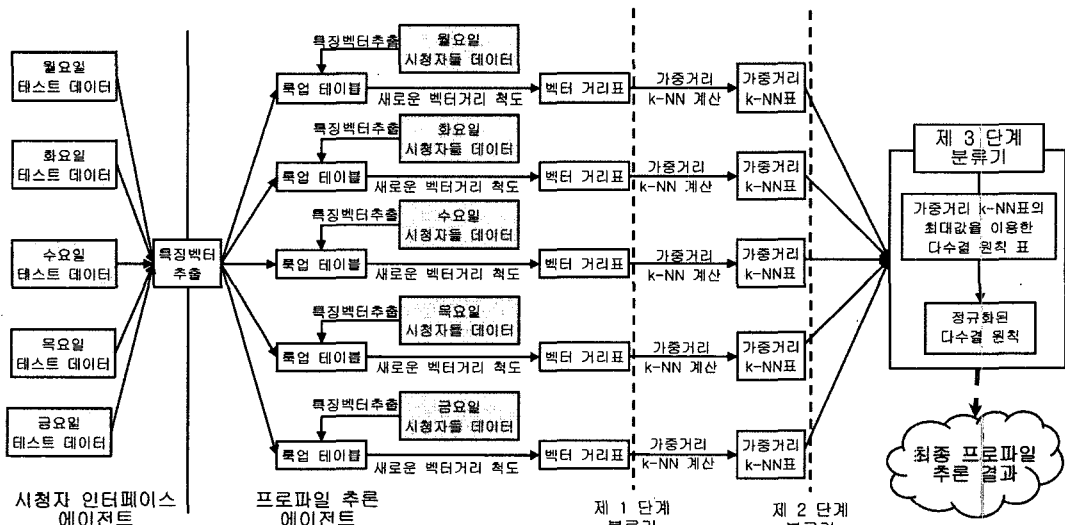


그림 4 다단계 분류기의 전체 추론 구조

하기 위한 기술 규격을 제정하였다. 본 논문에서는 TVA 1단계 제3부 규격인 메타데이터와 제4부 규격인 콘텐츠 참조(Content Referencing)규격을 이용하여 광고 콘텐츠에 대한 메타데이터를 생성하고, 이를 이용하여 실제 광고 콘텐츠를 선별하고 전송한다.

4.1 TV Anytime 메타데이터 구조

TVA 메타데이터는 사용자들로 하여금 로컬 PDR이나 네트워크에 연결된 PDR 등에 저장되어 있는 TV 프로그램 콘텐츠를 검색, 선택, 소비를 가능하게 한다[7]. 이러한 TVA 메타데이터 구조는 그림 6과 같이 구성되어 있고, “TVAMain”이라는 최상위 요소(Root Element) 내에 크게 4가지 종류의 메타데이터가 있다.

- 콘텐츠 서술 메타데이터(Content Description metadata) - 프로그램에 대한 전반적인 사항들을 정의할 수 있다. 프로그램 정보 테이블(ProgramInformationTable)에는 콘텐츠의 제목, 장르, 출연진들의 역할에 대한 정보를 정의할 수 있으며, 그룹 정보 테이블(GroupInformationTable)에서는 콘텐츠가 어떤 그룹에 속해있는지에 대한 정보를 정의할 수 있다. 크레디츠 정보 테이블(CreditsInformationTable)에는 프로그램 정보 테이블에서 정의된 출연진들의 실명을 기입할 수 있으며 프로그램 리뷰 테이블(ProgramReviewTable)에서 프로그램에 대한 시청자들이나 전문가들의 의견을 기록할 수 있다.
- 인스턴스 서술 메타데이터(Instance Description metadata) - 서비스 정보 테이블(ServiceInformationTable)에서는 콘텐츠 공급자의 이름과 소유권자에 대한 것을 정의할 수 있다. 프로그램 위치 테이블(ProgramLocationTable)에서는 프로그램의 스케줄에 관한 정보를 기입하는 것이 가능하다.
- 소비자 메타데이터(Consumer metadata) - 시청자가 어떻게 얼마나 콘텐츠를 소비하였고, 사용자의 선호도를 정의할 수 있다. 이는 MPEG-7에 있는 사용자 선호도

서술구조(User Preference Description Scheme)에 정의되어 있는 표준안을 따르고 있다[8].

- 세그먼트 메타데이터(Segmentation metadata) - 세그먼트는 특정 콘텐츠 내부의 내용을 시간단위로 접근, 소비, 검색이 가능하게 하는 단위를 이야기 한다. 세그먼트 정보 테이블(SegmentInformationTable)내에서는 특정 콘텐츠에 대한 이벤트, 내용, 주제, 배경 등 자세한 내용을 정의할 수 있다.

본 논문에서는, 표적 광고 콘텐츠를 위해 콘텐츠 서술 메타데이터내의 프로그램 정보 테이블, 크레디츠 정보 테이블과 인스턴트 메타데이터내의 서비스 정보 테이블을 이용하여 TVA 메타데이터를 작성한다. 광고 콘텐츠에 대한 메타데이터는 콘텐츠를 검색하는데 사용하기 때문에 광고에 대한 세그먼트 메타데이터나 소비자 메타데이터는 고려되지 않는다.

4.2 TV Anytime 콘텐츠 참조 구조

TVA에서 콘텐츠는 CRID(Content Reference Identifier)를 이용하여 참조한다[9]. CRID는 TVA 메타데이터와 함께 콘텐츠 정보를 표현하는 기본 도구이며, 특정 콘텐츠를 가리키기 위한 도구이다. CRID의 형식은 “CRID://<authority>/<data>”이며, <authority>는 CRID의 중심을 이루고 있는 부분이다. <authority>는 <DNS name> <name_extension>으로 구성되어 있다. DNS (Domain Name Server)는 콘텐츠 공급자간에 콘텐츠를 구별하는데 효율적이기 때문에, DNS를 CRID의 <authority> 부분을 구성하기 위해서 사용된다. “www.kbs.co.kr;drama”는 <authority>의 예가 된다. <data>는 어떠한 문자열을 이용하여 구성이 가능하다. 다음은 CRID 문법을 이용한 CRID의 예들을 보이고 있다.

- CRID://www.kbs.co.kr/LarrayKingShow
- crid://www.kbs.co.kr;comedy/friends
- CrId://www.kbs.co.kr/3.1.3.2_friends

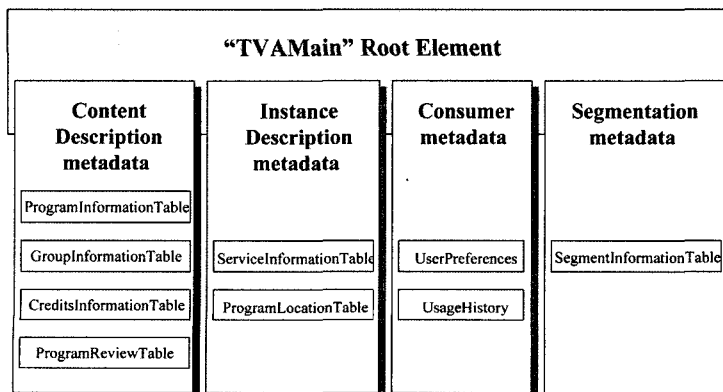


그림 6 TV Anytime 메타데이터 구조[7]

이렇게 생성한 CRID는 TVA 메타데이터의 프로그램 정보 테이블, 그룹 정보 테이블 및 세그먼트 정보 테이블에서 사용된다. CRID는 홈 서버의 CRID 테이블을 이용하여 CRID와 물리적인 콘텐츠 위치를 가리킬 수 있다. 그림 7은 이러한 CRID의 구조를 나타낸다.

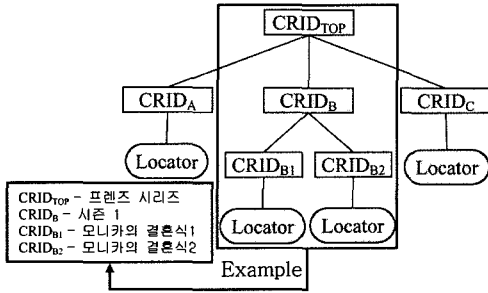


그림 7 CRID 구조[9]

CRID는 다른 CRID를 참조할 수 있는데, 예를 들어 “프렌즈” TV 사이트 시리즈는 콘텐츠가 여러 시즌으로 나누어져 있다. CRID를 바탕으로 다른 시즌에 속해 있는 콘텐츠 하나를 참조하거나 시즌 전체에 속해있는 콘텐츠 전체를 참조할 수 있다. 그림 7에서 네모 박스가 그려져 있는 부분을 보면, CRID_{TOP}은 프렌즈 시리즈 전체를 나타내는 CRID가 되고, CRID_B는 프렌즈 시리즈 중 시즌 1을 가리킨다. CRID_{B1}은 시즌 1에서의 ‘모니카의 결혼식1’이 될 수 있으며, CRID_{B2}는 시즌 1에서의 ‘모니카의 결혼식2’가 된다.

그림 8에서는 그림 7과 같은 구조를 가지고 CRID를 이용해서 콘텐츠의 위치를 얻어오는 방법에 대해서 설명하고 있다. 사용자의 콘텐츠 검색이나 선택을 통하여 특정 콘텐츠를 선택하게 된다. 선택된 콘텐츠의 CRID는 실제 물리적인 위치를 가지고 CRID 테이블을 통하여 사용자가 선택한 콘텐츠의 실제 위치를 얻게 된다. 이때, 사용자가 선택한 콘텐츠가 시리즈 중의 하나라면, 서버는 시리즈에 속해있는 다른 콘텐츠의 CRID를 사용

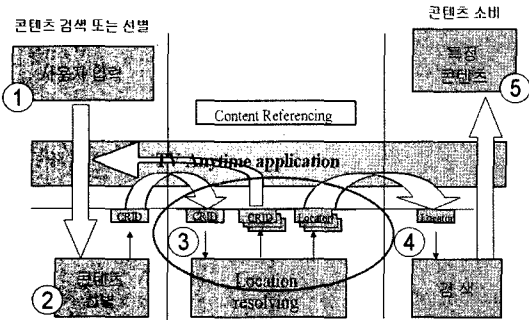


그림 8 CRID를 이용한 콘텐츠 참조 방법[9]

자에게 전해준다. 사용자는 서버에서 전달받은 콘텐츠 목록 중 자신이 원하는 콘텐츠를 선택하게 되고, 서버는 사용자가 선택한 콘텐츠의 CRID를 넘겨받아서 실제 물리적 위치의 콘텐츠 위치를 얻어서 사용자가 선택한 콘텐츠를 소비가 가능하게 한다.

4.3 표적 광고 콘텐츠 선별 방법

TVA 메타데이터를 이용하여 광고에 대한 규격화 작업이 완성이 되면, 이를 이용하여 시청자 프로파일(성별/연령대)에 맞게 광고 콘텐츠를 선별해야 한다. 하지만, 광고를 선별하기 위해서 시청자가 선호하는 광고 모델, 광고 유형, 광고 제품에 대한 선호도 정보를 알아야 하고, 이런 선호도 정보가 공신력이 있어야 한다. 본 논문에서는 광고 모델, 광고 유형, 광고 제품에 대한 선호도 정보를 알기 위해서 한국방송광고공사에서 제공하는 광고 모델, 유형, 제품에 대한 성별/연령대 별 광고에 대한 선호도 조사 결과를 이용한다[10]. 표 4에서 표 6은 한국방송광고공사에서 제공하는 광고 모델, 광고 유형, 광고 제품에 대한 선호도를 성별 및 연령대에 맞게 조사해 놓은 결과를 보여주고 있다.

한국방송광고공사에서 제공하는 표 4는 광고에 등장하는 연예인을 중심으로 성별/연령대 별로 광고 모델 선호도를 퍼센트 단위로 표시하였고, 표 5와 6은 광고 유형과 광고 제품 관심도를 최대 6까지 표시하여 성별/연령대별로 선호도 값을 표시하였다. 위의 표 4에서 표 6까지의 선호도 값을 바탕으로, 성별/연령대별로 광고 모델, 광고 유형, 광고 제품에 대한 표를 만들 수 있다. 표 7은 남자 10대에 대한 광고 모델, 유형, 제품에 대한 선호도를 표로 작성한 예를 보이고 있다.

성별/연령대별로 표 7과 같이 광고 선호도 표를 만들 때, 광고 모델, 유형, 제품은 각각 선호도 값을 가지고 있는데, 이런 값을 기준으로 내림차순으로 정렬을 한다. 또한, 각 항목 별 개수를 보면, 광고 모델은 총 11개, 광고 유형은 총 14개, 광고 제품은 총 16개의 목록들이 존재한다. 각 항목 별 다른 개수가 존재하기 때문에, 광고 모델, 유형, 제품을 적절히 선택해야 한다. 만약, 가장 선호하는 광고 모델, 광고 유형, 광고 제품을 기준으로 하면, 시청자는 거의 똑같은 광고를 보게 될 것이다. 이런 문제점을 해결하기 위하여 어떤 광고 모델, 유형, 제품의 광고를 선택하는 기준이 필요하다. 하루 24시간을 6시간씩 4등분을 하고, 사용자가 TV 시청 선호도 시간을 기준으로 광고 모델, 유형, 제품의 선호도를 나눈다. 즉, 그림 9와 같이 광고 모델, 유형, 제품을 TV 시청 선호도 시간을 기준으로 나눌 수 있고, 그림 9의 회색 상자는 TV 시청 선호도 시간의 순서를 나타낸다.

표 4 광고 모델 선호도에 대한 한국방송광고공사 제공 자료

	남 10대	남 20대	남 30대	남 40대	남 50대이상	여 10대	여 20대	여 30대	여 40대	여 40대이상
1	전지현 22.4	전지현 24.4	이효리 12.5	이효리 11.3	이영애 9.5	전지현 16.8	전지현 15.1	권상우 11.7	이영애 11.2	이영애 9.7
2	권상우 12.4	이효리 11.2	이영애 12.2	이영애 8.9	이효리 8.7	권상우 12.0	권상우 13.8	이영애 11.6	권상우 7.7	권상우 7.1
3	이효리 8.0	이영애 6.6	전지현 11.4	전지현 7.7	안성기 3.2	강동원 9.8	이영애 6.8	이효리 4.8	안성기 5.3	이효리 4.1
4	김C 4.4	송혜교 4.6	송혜교 5.2	송혜교 4.0	김혜자 3.0	원빈 5.9	이효리 4.5	전지현 4.2	채시라 4.5	채시라 3.7
5	이영애 3.6	권상우 3.4	안성기 3.4	안성기 3.3	최불암 2.8	비 5.6	강동원 3.8	비 4.0	송혜교 4.4	김정은 3.4
6	송혜교 3.6	김C 2.5	권상우 2.9	권상우 2.9	김정은 2.5	이나영 4.2	송혜교 3.2	송혜교 3.9	김정은 4.2	안성기 3.4
7	비 2.6	김정은 2.1	한석규 2.6	김정은 2.3	고두심 2.3	이영애 3.1	장동건 2.9	장동건 3.9	전지현 3.9	김혜자 3.4
8	한예슬 2.3	정우성 2.1	김정은 2.5	김남주 2.0	전지현 1.7	이효리 3.1	이나영 2.9	김정은 3.5	장동건 3.9	김희애 2.6
9	이나영 2.1	한예슬 1.8	김남주 2.2	전인화 1.9	채시라 1.7	송혜교 2.8	비 2.7	안성기 3.2	이효리 3.8	송혜교 2.4
10	보아 2.1	이나영 1.6	송윤아 1.7	최민식 1.9	송혜교 1.5	김C 2.8	원빈 2.7	이미연 2.6	전인화 2.9	고두심 2.2

표 5 광고 유형 선호도에 대한 한국방송광고공사 제공 자료

	유머	전통/인간적 유대감	유아/어린이 등장	소비자 등장	동물 등장	애니메이션/만화	사회저명 인사등장	연예인 등장	외국스타 등장	성적인 느낌강조	비교 광고	이미지 강조광고	제품내용 강조광고	호기심 유발
남10대	4.8	3.8	3.8	3.6	3.9	4.0	2.9	4.4	3.9	2.8	2.8	2.8	2.8	3.2
남20대	4.8	4.2	3.9	3.8	3.7	3.7	2.9	4.0	3.7	3.3	3.0	3.0	3.0	3.2
남30대	4.6	4.3	4.1	3.9	3.6	3.6	2.9	3.7	3.3	3.0	3.0	3.1	3.1	2.9
남40대	4.3	4.3	4.0	3.9	3.6	3.4	3.1	3.6	3.2	2.9	2.9	3.0	3.1	2.8
남50대	4.3	4.4	3.9	3.8	3.6	3.1	3.2	3.6	3.1	2.6	2.9	3.1	3.1	2.7
여10대	4.8	3.9	4.2	3.7	3.9	4.1	2.9	4.5	3.8	2.5	2.5	2.7	2.7	3.1
여20대	4.8	4.5	4.4	4.0	4.0	3.8	3.0	4.1	3.4	2.6	2.6	2.9	3.0	3.1
여30대	4.7	4.4	4.4	4.0	3.8	3.9	3.1	3.9	3.2	2.5	2.7	3.0	3.1	2.8
여40대	4.5	4.5	4.4	4.1	3.8	3.9	3.2	3.8	3.1	2.3	2.7	3.1	3.3	2.7
여50대	4.3	4.4	4.2	4.0	3.7	3.3	3.2	3.7	3.0	2.3	2.8	3.0	3.1	2.6

표 6 광고 제품 선호도에 대한 한국방송광고공사 제공 자료

	음료	제과	식품	술	생활용품	화장품	자동차	의약품	가전제품	컴퓨터	휴대폰/이동전화	백화점	가구	의류	금융	학습지
남10대	4.0	4.1	3.9	2.8	2.8	2.5	3.4	2.6	3.0	4.3	4.7	3.0	2.4	3.5	2.1	2.3
남20대	3.6	3.4	3.5	3.6	3.1	3.0	4.2	3.0	3.5	4.2	4.5	3.2	2.7	3.6	2.8	2.2
남30대	3.3	3.2	3.3	3.5	3.0	2.7	4.3	3.2	3.4	3.9	4.1	3.0	2.7	3.0	3.1	2.5
남40대	3.3	3.1	3.3	3.5	3.0	2.8	4.0	3.5	3.4	3.6	3.8	3.0	2.7	3.0	3.2	2.6
남50대	3.2	3.0	3.1	3.4	3.0	2.7	3.7	3.5	3.3	3.0	3.4	2.9	2.7	2.9	3.0	2.2
여10대	4.1	4.3	3.9	2.9	3.8	3.9	3.1	2.7	3.2	4.1	5.0	3.5	2.9	4.3	2.4	2.7
여20대	3.8	3.8	3.7	3.4	4.0	4.5	3.6	3.2	3.8	3.7	4.5	3.7	3.3	4.3	3.0	2.6
여30대	3.6	3.5	3.7	3.3	3.9	4.1	3.6	3.6	4.1	3.7	4.0	3.7	3.4	3.9	3.4	3.5
여40대	3.5	3.5	3.6	3.2	3.9	4.0	3.5	3.7	4.0	3.6	3.7	3.6	3.4	3.7	3.4	3.0
여50대	3.2	3.1	3.4	2.9	3.7	3.7	3.1	3.6	3.9	2.9	3.2	3.4	3.1	3.4	3.0	2.0

표 7 남자 10대에 대한 광고 선호도 표

남자 10대		
광고 모델	광고 유형	광고 제품
전지현	유머	휴대폰/이동전화
권상우	연예인 등장	컴퓨터
⋮	⋮	⋮
이나영	이미지 강조 광고	학습지
보아	제품 내용 강조 광고	금융

다음 그림 9와 같이, 18시에서 24시까지 TV를 가장 많이 시청하는 시간이고, 06시부터 12시까지가 2번째로 시청을 많이 하며, 3번과 4번도 이와 같이 시청시간

을 기준으로 나누었다. 그러면, 아래와 같이 사용자 프로파일 추론결과 시점을 기준으로 TV 시청 선호도 시간대 별로 광고 선호도에 맞게 광고를 제공할 수 있다.

- $0 \leq$ 추론결과 시점 $< 6 \rightarrow$ “4” 광고
- $6 \leq$ 추론결과 시점 $< 12 \rightarrow$ “2” 광고
- $12 \leq$ 추론결과 시점 $< 18 \rightarrow$ “3” 광고
- $18 \leq$ 추론결과 시점 $< 24 \rightarrow$ “1” 광고

예를 들어, 표 7이 광고 모델, 유형, 제품에 대한 선호도 값을 기준으로 정렬이 되었고, 추론결과 시점이 18시에서 24시 사이이면, 광고 모델은 1위~3위 중 한 명, 광고 유형은 1위~4위 중 한 개, 광고 제품은 1위~4위

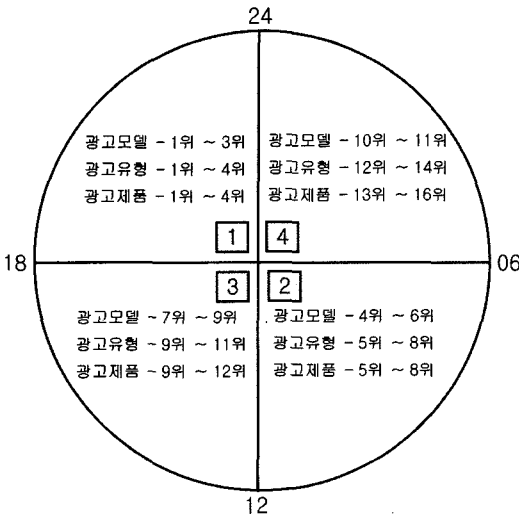


그림 9 TV 시청 선호도 시간을 기준으로 광고 모델, 광고 유형, 광고 제품 별 분류 예

중 하나씩을 임의적으로 선택한다. 만약, 임의적으로 선택된 광고모델, 유형, 제품들의 질의 조건을 “and” 조건으로 하여 광고 콘텐츠에 대한 TVA 메타데이터 데이터베이스에서 검색 결과가 존재하면 선택된 광고의 CRID를 이용하여 실제 광고 콘텐츠를 제공한다. 광고 검색 질의의 결과가 존재하지 않으면, 검색 질의를 위해 선택된 광고 모델, 유형, 제품들의 질의 조건을 “or”로 두고 TVA 메타데이터 데이터베이스에서 광고 메타데이터를 검색하게 된다. 검색 결과가 여러 개의 TVA 메타데이터가 존재하면, 그 중 하나를 임의적으로 선택하여 광고 콘텐츠를 시청자에게 제공한다.

5. 실험결과 및 프로토타입 표적 광고 시스템 구현 결과

5.1 사용자 프로파일 추론 알고리즘 실험 결과

사용자 프로파일(성별/연령대) 추론 알고리즘에 대한 실험 결과는 AC Nielsen Korea에서 제공된 실제 TV 시청 데이터(TV Viewing history data)를 이용하여 실험하였고, TV 시청 데이터는 2002년 12월부터 2003년 5월까지 2522명(남자: 1243명, 여자: 1279명)에 해당하는 데이터이다. 실험을 수행하기 위하여 TV 시청자의 성별/연령대에 대해 TV 시청 데이터를 학습 데이터(training data)와 테스트 데이터(testing data)의 두 그룹으로 나누었다. 학습 데이터는 전체 사용자 TV 시청 데이터로부터 70%(1,764명)를 무작위로 선별하였고, 나머지 30%(758명)는 테스트 데이터로 이용되었다. 학습 데이터는 1,764명에 대한 6개월 간의 TV 시청 데이터이며 테스트 데이터는 758명의 6개월간의 TV 시청 데

이터이다. 또한, 실험을 위해 서로 다른 시드 번호(seed number)를 이용하여 총 8가지의 학습 데이터와 테스트 데이터 쌍을 생성하였다.

학습 데이터로부터 특징 벡터 추출 시 사용한 조건 값은 $C_{Th} = 30$, $T_{Th} = 0.1$ 을 사용하였다. 즉, 학습 데이터의 특징 값들은 프로그램의 정규 시간에서 최소 10% 이상을 시청한 프로그램을 기준으로 특징 값을 구하였고, 이를 만족하면서 6개월 동안 시청한 횟수가 30회 이상이 되는 데이터를 학습 데이터로 사용하였다. 표 8은 다단계 분류기를 이용하여 정확도를 측정된 결과이고, 표 8의 정확도는 8개의 학습 데이터와 테스트 데이터 쌍의 정확도를 평균한 결과이다.

표 8 다단계 분류기(MSC)의 실험 결과

성별 및 연령대	정확도(%)		
	VC	ED	MSC
M0s	76.69	71.96	88.21
F0s	67.14	67.86	89.28
M10s	66.89	71.28	89.29
F10s	67.76	68.75	65.79
M20s	60.72	62.95	86.50
F20s	68.18	73.58	78.49
M30s	63.69	72.42	78.80
F30s	63.15	66.88	76.17
M40s	59.82	64.51	86.59
F40s	69.71	64.83	72.25
M50s	54.86	64.58	82.86
F50s	60.86	60.20	67.91
M60s	65.76	67.94	89.77
F60s	56.90	50.86	86.61
평균 정확도	64.54	66.81	80.17

다른 방법과의 실험 결과를 비교하기 위하여, 벡터 상관법(VC)과 유클리드 거리(ED)와의 정확도를 비교 하였다. 표 8에서 M은 남자, F는 여자를 뜻하고, M10s는 남자 10대를 뜻한다. 본 논문에서 제안한 알고리즘의 정확도가 기존의 방법보다 약 15%가 높은 것을 확인할 수 있다. 표 8에서 F10s와 F50s에서 MSC의 결과가 다른 성별 및 연령대의 결과보다 정확도가 낮는데 이는 특징 벡터의 패턴이 다른 연령대와 유사한 형태를 가지고 있기 때문에 낮은 성능 결과를 보이고 있다.

5.2 광고 콘텐츠에 대한 TVA 메타데이터 결과

광고 콘텐츠를 TVA 규격에 맞게 작성하기 위한 TVA 메타데이터 내에는 여러 가지 메타데이터와 테이블이 존재하지만, 본 논문에서는 프로그램 정보 테이블, 서비스 정보 테이블, 그리고 크레디츠 정보 테이블을 이용하여 표현하였다. 표 9는 “삼성 애니콜 v420”의 휴대폰 광고에 대한 콘텐츠를 TVA 규격에 맞게 메타데이

표 9 광고 콘텐츠에 대한 TVA 메타데이터 결과

```

<?xml version="1.0"?>
<TVAMain xmlns:lang="ko" publisher="삼성 애니콜" xmlns="urn:tva:metadata:2002"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:tva:metadata:2002 E:\WMajorStudy\WStandard\WTV-Anytime\WTV-
AnytimeSchema\Wtva_metadata_v13.xsd">
  <!-- 광고 콘텐츠 제공회사 -->
  <CopyrightNotice>Copyright (c) 삼성 애니콜 2005</CopyrightNotice>
  <ProgramDescription>
    <ProgramInformationTable>
      <!-- 광고 콘텐츠의 CRID -->
      <ProgramInformation programId="crid://www.samsung.com/AnycallV420">
        <BasicDescription>
          <Title type="main">삼성 애니콜 v420</Title> <!-- 광고 콘텐츠의 제목 -->
          <MediaTitle><!-- 광고 콘텐츠의 대표 영상 주소 -->
            <mpeg7:TitleImage> <mpeg7:MediaUri>
              http://mccb.icu.ac.kr/metadata/picture/AnycallV420.bmp
            </mpeg7:MediaUri></mpeg7:TitleImage>
          </MediaTitle>
          <Synopsis>성적인 느낌 강조</Synopsis><!-- 광고 콘텐츠의 광고 유형 기재 -->
          <Keyword>이효리</Keyword><!-- 광고 콘텐츠의 광고 모델 기재 -->
          <Keyword>성적인 느낌 강조</Keyword><!-- 광고 콘텐츠의 광고 유형 기재 -->
          <Keyword>휴대폰/이동전화</Keyword><!-- 광고 콘텐츠의 광고 제품 기재 -->
          <Keyword>삼성 애니콜</Keyword><!-- 광고 콘텐츠의 광고 공급자 기재 -->
          <!-- 광고 콘텐츠의 광고 제품 기재 -->
          <Genre href="urn:tva:metadata:cs:ContentCommercialCS:2002:3.50.5.7">
            <Name>휴대폰/이동전화</Name>
          </Genre>
          <!-- 광고 콘텐츠의 광고 모델 -->
          <CreditsList>
            <CreditsItem role="urn:tva:metadata:cs:TVARoleCS:2002:V709">
              <PersonNameIDRef ref="credit_1"/>
              <Character> <mpeg7:GivenName>광고 모델</mpeg7:GivenName> </Character>
            </CreditsItem>
          </CreditsList>
        </BasicDescription>
      </ProgramInformation>
    </ProgramInformationTable>
    <ServiceInformationTable>
      <ServiceInformation serviceId="Ch2">
        <Name>v420</Name><!-- 광고 콘텐츠의 광고 제품의 실제 이름 -->
        <Owner>삼성 애니콜</Owner><!-- 광고 콘텐츠의 광고 공급자 이름 -->
      </ServiceInformation>
    </ServiceInformationTable>
    <CreditsInformationTable>
      <PersonName personNameId="credit_1">
        <!-- 광고 콘텐츠의 광고 모델의 이름 -->
        <mpeg7:GivenName>이효리</mpeg7:GivenName>
      </PersonName>
    </CreditsInformationTable>
  </ProgramDescription>
</TVAMain>

```

타화한 결과이다.

표 9에서 <TVAMain> 밑에 있는 <CopyrightNotice> 내에는 광고 콘텐츠를 제공하는 회사의 정보를 게재한다. <ProgramInformationTable>에는 광고 콘텐츠에 대한 정보를 기술하고, <ProgramInformation>내에 있는 속성 값 중 "programId"를 이용하여 광고 콘텐츠의 CRID를 표시한다. <Synopsis>는 광고 콘텐츠의 광고 유형을 기재함으로써 광고의 유형을 <Synopsis>를 통하여 파악한다. <Genre>에는 광고 제품을 기재함으로써 어떤 광고인지를 알게 한다. 또한, <CreditsList>에는 몇 명의 광고 모델이 나오는 지를 파악한다. <Keyword>는 나중에 광고 검색에 필요한 키워드를 제공하기 위하여 광고 공급자, 광고 모델, 광고 유형, 광고 제품 등을 기재한다. 마지막으로, <ServiceInformationTable> 내에는 광고 콘텐츠의 공급자와 상품명을 기입한다. <CreditsInformationTable>내에는 <ProgramInformationTable>내에 있는 <CreditsList>에 기재된 실제 모델의

이름을 기입한다.

5.3 프로토타입 표적 광고 시스템 구현 결과

본 논문에서 제안된 시청자 프로파일(성별/연령대) 추론 알고리즘과 표적 광고 선별 방법을 이용하여 프로토타입 표적 광고 시스템 구현 결과를 제시한다. 프로토타입 표적 광고 시스템을 위한 광고 콘텐츠는 NGTV (<http://www.ngtv.net>)에서 무료로 제공되는 총 28개의 광고 콘텐츠를 이용하였고, 광고 콘텐츠에 대하여 TVA 메타데이터로 작성을 하였다. 그림 10은 표적 광고 프로토타입 시스템 구현 결과를 보이고 있다.

광고 콘텐츠를 제공받기 위하여 20대 남자의 특징 벡터를 표적 광고 콘텐츠 공급자 서버로 전송하면, 서버에서는 시청자의 프로파일(성별/연령대)를 다단계 분류기를 통하여 추론한다. 추론 결과가 20대 남자로 판단을 하게 되고, 광고 모델(이효리), 광고 유형(연예인 등장) 및 광고 제품(휴대폰/이동전화)을 표 7과 같은 광고 선호도 표로부터 선택을 하고, 이에 해당하는 광고를

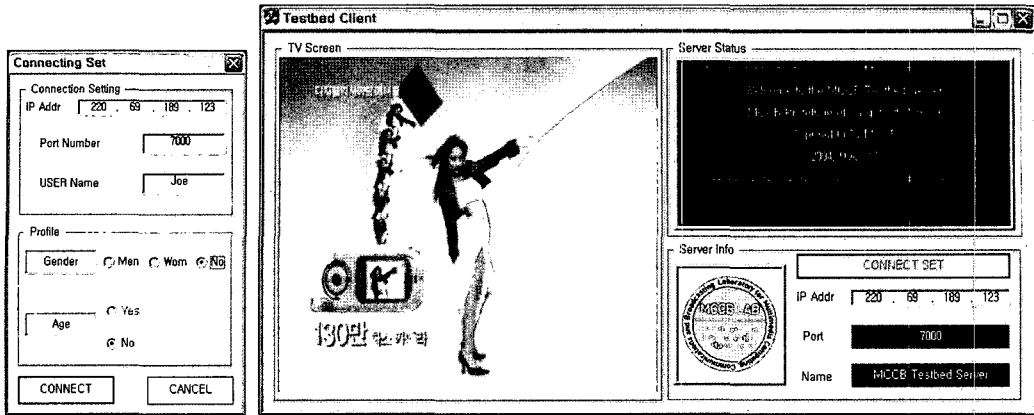


그림 10 프로토타입 표적 광고 시스템 구현 결과

TVA 메타데이터 검색을 통하여 해당하는 광고 콘텐츠에 대한 메타데이터를 찾고, 검색된 메타데이터의 CRID를 이용하여 표적 광고를 시청자에게 전송하게 된다. 그림 10에서, 남자 20대를 위한 모델은 “이효리”이고, 광고 유형은 “연예인 등장”이며 광고 제품은 “휴대폰/이동전화”인 광고 방송이 전송된 것을 프로토타입 표적 광고 시스템 구현 결과를 통해 확인할 수 있었다.

6. 결론

본 논문은 디지털 방송의 다채널 시청환경에서 시청자의 TV 시청 데이터(TV Viewing history data)를 이용한 표적 광고 서비스에 관한 연구로서, 시청자의 TV 시청 데이터로부터 특징 벡터를 추출하여 다단계 분류기(Multi-Stage Classifier)를 이용하여 시청자의 프로파일(성별/연령대)를 추론한다. 추론 알고리즘의 정확도는 약 80%로 다른 알고리즘에 비해서 높은 정확도를 보이는 것을 확인할 수 있었다. 또한, 추론한 시청자 프로파일(성별/연령대), 광고에 대한 TV Anytime 메타데이터, 그리고 한국방송광고공사에서 제공하는 성별/연령대 별 광고 선호도 값을 이용하여 시청자의 프로파일에 맞는 표적 광고 서비스를 제공할 수 있는 표적 광고 방법을 제안하였고, 이를 구현하였다. 본 논문에서 제안하는 표적 광고 방법을 통하여 시청자들은 자신의 성별과 연령대에 맞는 광고를 볼 수 있으며, 콘텐츠 공급자는 광고로 인한 수익 증대를 꾀할 수 있고, 광고주는 회사가 표적으로 삼고 있는 고객들에게 광고를 제공함으로써 보다 효과적인 광고 효과를 볼 수 있다.

향후 연구 계획으로는 본 연구의 사용자 연령대 및 성별 추론에 있어서 정확도 향상 및 장르 뿐만 아닌 다른 선호도 속성에 대해서도 다양한 추론 방법을 고안하여 보다 복잡한 사용자의 선호도를 반영한 개인 맞춤형 방송 서비스 응용에 적용하는 것이다.

참고 문헌

- [1] Katsaros D. and Manolopoulos Y., "Broadcast program generation for webcasting," *Source Data & Knowledge Engineering*, Volume 49, Issue 1, pp. 1~21, April 2004.
- [2] Bozios T., Lekakos G., Skoularidou V. and Chorianopoulos K., "Advanced Techniques for Personalised Advertising in a Digital TV Environment: The iMEDIA System," *Proceedings of the E-business and E-work Conference*, 17-19 October, Venice, Italy, pp. 1025-1031, 2001.
- [3] Miyahara K. and Pazzani M. J., "Collaborative filtering with the simple bayesian classifier," *Proceedings of the Sixth Pacific Rim International Conference on Artificial Intelligence*, PRICAI 2000.
- [4] Shahabi C., Faisal A., Kashani F. B. and Faruque J., "INSITE: A Tool for interpreting Users," *Proceeding of Interaction with a Web Space*, Volume00, pp. 635-638.
- [5] Yu Z., Zhou X., "TV3P: an adaptive assistant for personalized TV," *Journal of IEEE Transactions on Consumer Electronics*, vol. 50, Issue 1, pp. 393~399, Feb 2004.
- [6] Yuan W., Liu J., and Zhou H. B., "An Improved KNN Method and Its Application To Tumor Diagnosis," *Proceedings of the 3rd International Conference on Machine Learning and Cybernetics*, August 2004.
- [7] TV Anytime Forum, "Metadata," S-3v1.3, Part-A: Metadata Schemas, Dec. 2002.
- [8] ISO/IEC JTC1/SC29 WG11 (MPEG), "MPEG-7 Multimedia Description Schemes," ISO/IEC 15938-5 FDIS, N4242/MPEG57, Sydney, Australia, July 2001.
- [9] TV Anytime Forum, "Content Referencing," S-4v1.2, June 2002.
- [10] Korea Broadcasting Advertising Corporation, "Media & Consumer Research 2004 -소비자 행태 조사-," <http://www.kobaco.co.kr/kor/information/studydata/>

studydata_research_annual.asp



김문조

2004년 2월 동명정보대학교 메카트로닉스공학과 학사. 2006년 2월 한국정보통신대학교 공학부, 석사. 2006년 2월~현재 삼성전자 디지털 미디어 연구원. 관심분야는 멀티미디어 방송 서비스를 위한 정보추론, TV-Anytime, 디지털 멀티미디어 방송, MPEG-7/21

어 방송, MPEG-7/21



이범식

2000년 8월 고려대학교 전기전자전파공학부 학사. 2006년 8월 한국정보통신대학교 공학부, 석사. 2006년 9월~현재 한국정보통신대학교 공학부, 박사과정. 2000년 9월~2003년 9월 (주)기산텔레콤, 연구원. 관심분야는 멀티미디어 정보 처리, 멀티미디어 방송 서비스를 위한 정보 추론, 내용기반 영상 검색, 스케일러블 영상 및 비디오 압축, 내용기반 영상 검색, 멀티미디어 통신

멀티미디어 방송 서비스를 위한 정보 추론, 내용기반 영상 검색, 스케일러블 영상 및 비디오 압축, 내용기반 영상 검색, 멀티미디어 통신



임정연

1999년 3월 충남대학교 정보통신공학과 학사. 2001년 3월 충남대학교 정보통신공학과 대학원, 석사. 2001년 9월~현재 한국정보통신대학교 공학부, 박사과정. 관심분야는 멀티미디어 정보 처리, 멀티미디어 방송 서비스를 위한 정보 추론, 대화형 멀티미디어 및 영상 통신, MPEG-4/7/21/A

대화형 멀티미디어 및 영상 통신, MPEG-4/7/21/A



김문철

1989년 2월 경북대학교 전자공학과 학사 1992년 12월 University of Florida, Electrical and Computer Engineering, 석사. 1996년 8월 University of Florida, Electrical and Computer Engineering, 석사. 1997년 1월~2001년 2월 한국전자통신연구원, 선임연구원. 2001년 2월~현재 한국정보통신대학교 공학부 부교수. 관심분야는 MPEG-4/7/21/A/E, 멀티미디어 정보처리, 대화형 멀티미디어 방송, 지능형 에이전트 기반 멀티미디어 응용, 영상 및 신호처리

통신연구원, 선임연구원. 2001년 2월~현재 한국정보통신대학교 공학부 부교수. 관심분야는 MPEG-4/7/21/A/E, 멀티미디어 정보처리, 대화형 멀티미디어 방송, 지능형 에이전트 기반 멀티미디어 응용, 영상 및 신호처리



이희경

1999년 2월 영남대학교 컴퓨터공학과 학사. 2002년 2월 한국정보통신대학원대학교 공학부 석사. 2002년 2월~현재 한국전자통신연구원 연구원. 관심분야는 MPEG-7, TV-Anytime, 디지털 방송, 맞춤형 방송



이한규

1994년 2월 경북대학교 전자공학과 학사 1996년 2월 경북대학교 전자공학과 석사 2005년 2월 한국정보통신대학교 공학부 박사과정 수료. 1996년~현재 한국전자통신연구원 선임연구원. 2005년 9월~현재 한국전자통신연구원 맞춤형방송연구팀장

관심분야는 영상신호처리 및 분석, 맞춤형방송 기술 및 표준화