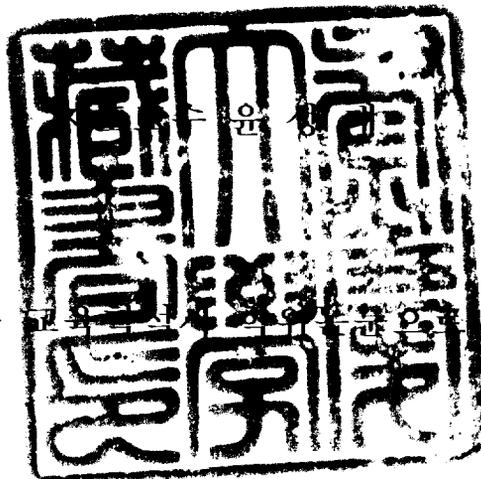


교육학석사학위논문

인터랙티브 TV에서의 pEPG를
위한 혼합 추천 기법



이 논문을 제출함

2005년 8월

부경대학교 교육대학원

전산교육전공

서 정 애

서정애의 교육학석사 학위논문을 인준함

2005년 6월 17일

주 심 공학박사 김 창 수



위 원 이학박사 신 상 욱



위 원 이학박사 윤 성 대



[차례]

표 차례	ii
그림 차례	iii
Abstract	iv
I. 서 론	1
II. 관련연구	4
2.1 인터랙티브 방송 서비스	4
2.2 추천 기법	6
2.3 MPEG-7의 사용자 기술구조	8
III. 혼합 추천 기법	13
3.1 사용자의 프로그램 시청 정보 모델링	14
3.2 데이터 필터링	16
3.3 혼합 추천 알고리즘	18
IV. 실험 및 평가	23
4.1 실험 환경	23
4.2 혼합 추천 기법의 적용	24
4.3 성능 평가	31
V. 결 론	34
참고문헌	35

[표 차례]

<표 1> 실험 환경	23
<표 2> 프로그램 장르 구분	26
<표 3> user1의 내용기반 추천 결과	27
<표 4> 유사 사용자 그룹	28
<표 5> user1을 위한 프로그램 추천 결과	29
<표 6> user1을 위한 혼합 추천 기법 적용	30
<표 7> 순수 내용기반 추천의 coverage	31
<표 8> 혼합 추천 기법의 coverage	32

[그림 차례]

(그림 1) 인터랙티브 방송 서비스	5
(그림 2) MDS의 구조	9
(그림 3) User Interaction 구조	10
(그림 4) User Preference 기술 구조	11
(그림 5) User History 기술 구조	12
(그림 6) 혼합 추천 기법의 개념도	13
(그림 7) 장르의 트리 구조의 예	19
(그림 8) EPG 화면	24
(그림 9) 프로그램 정보 데이터베이스	25
(그림 10) 사용자 시청 정보 데이터베이스	26
(그림 11) 추천 기법별 coverage	32

A Hybrid Recommendation Method for pEPG on the Interactive TV

Jung Ae Seo

*Graduate School of Education
Pukyong National University*

Abstract

According to an evolution of digital broadcasting, it is possible that the terrestrial and satellite broadcasting media providing multi-channel service. In a broadcasting terminal, we can obtain a broadcasting program and its information using the metadata describing a broadcasting program content. However it is difficult that they select the favorite program among hundreds of channels.

To solve that problem, pEPG(personalized Electronic Program Guide) is that recommends a program reflecting an inclination and preference of each user.

pEPG should be adapted properly in a preference change of user according to the broadcasting program watching tendency which is obtained by monitoring and analyzing a user's habit watching broadcasting.

In this paper, we suggest a Hybrid Recommendation Method that improves the current content-based recommendation and collaborative recommendation for pEPG.

In Simulation, we have known that the proposed method is more efficient to select the user preference program than existing methods.

I. 서 론

최근 디지털 방송의 보편화에 따라 수많은 채널의 수용이 가능하게 되었다. 우리나라도 2001년 하반기부터 디지털 지상파 방송의 본 방송을 시작함으로써 지상파, 위성, 케이블 매체를 통한 다매체·다채널 방송 시청 환경으로 변화하고 있다. 지금까지 수동적인 매체로 머물러 있던 TV는 방송 제작, 전송 및 수신체계 모두가 디지털 체계로 전환되어 컴퓨터와 같은 지능을 갖는 종합정보매체로 진화하고 있다. 이러한 환경 하에서 디지털 TV 방송 서비스는 인터랙티브 TV(Interactive TV)라는 개념으로 나타난다. 즉, 방송국이 서비스를 주관하는 서버 역할을 하고 디지털 TV는 서비스를 수신하는 클라이언트 역할을 함으로써 기존의 TV로는 할 수 없었던 다양한 양방향 서비스가 가능하게 되었다. 이러한 인터랙티브 TV 환경에서의 시청자는 시청하고자 하는 내용이나 다양한 정보 등에 대해 양방향으로 능동적인 참여를 할 수 있고, 지금까지와 같이 일방적으로 정보를 받는 시청자가 아니라 수신 단말기를 통해서 방송미디어에 대한 조정력과 통제력을 갖고 자발적으로 참여하는 이용자가 된다[16].

방송 서비스 제공자는 방송 프로그램 내용에 대한 부가데이터를 방송 프로그램과 함께 전송하기 때문에 수신 단말기에서는 전송된 방송 메타데이터를 이용하여 프로그램 정보를 얻을 수 있다. 그러나 방대한 프로그램 정보 속에서 사용자가 자신의 취향에 맞는 프로그램을 선택한다는 것은 매우 힘든 일이다. 따라서 시청자가 원하는 방송 프로그램이나 프로그램 관련 정보를 미리 제시함으로써 보다 빠르고 효율적인 채널 선택을 가능하게 하는 서비스가 필요하게 되었다.

이렇게 등장한 서비스가 전자 프로그램 가이드(Electronic Program Guide : EPG)이다. EPG는 현재 진행되고 있는 프로그램이나 앞으로 방영될 프로그램들에 대한 정보를 셋톱박스(Set-Top Box) 화면상에서 보여주는 프로그램 가이드를 말한다. 모든 셋톱박스는 이러한 서비스 정보를 추출하여 향후 일주일간의 방송 프로그램 시간표를 만들어 낼 수 있다[3].

그러나 단순히 프로그램 정보를 채널별·시간대별로 정렬하여 제시하는 EPG를 하나하나 살펴보면 사용자가 원하는 프로그램을 찾는 것은 효과적이지 못하다. 예를 들어, 100개의 채널이 있고 각 채널당 하루에 24개의 프로그램을 방송한다면 하루 분량의 EPG 데이터는 2400개가 된다. 현재 서비스되고 있는 EPG에서는 사용자가 직접 검색할 수 있도록 GUI(Graphical User Interface)를 제공하는 경우도 있으나 TV 화면에서 사용자가 원하는 키워드(Keyword)를 직접 입력하여 정보를 검색하는 일은 매우 번거로운 일이다.

이러한 문제점을 해결하기 위해 시청자 개개인의 취향 및 선호도를 반영하여 프로그램을 추천해주는 EPG가 절실히 요구되었으며, 이러한 TV 추천의 한 분야로 개인 전자프로그램 가이드(personalized Electronic Program Guide : pEPG)가 대두되었다[13].

pEPG는 시청자의 방송 프로그램 시청 성향을 통해 시청자의 선호도에 높게 부합하는 프로그램을 자동적으로 선별하여 이를 사용자에게 제시해주며, 시청자의 방송 프로그램 시청 습관 등을 모니터링하고 분석하여 시청자의 선호도 변화에 적절히 적용할 수 있어야 한다.

여기서 사용자의 프로그램 선호도에 대한 기술(Description)은 MPEG-7 MDS(Multimedia Description Schemes) 규격에서 지원하고 있다[12].

기존 pEPG 관련 연구[9, 17]들은 내용기반 추천 기법을 이용하여 사용자에게 프로그램을 추천하고 있지만 인터랙티브 TV 환경에서는 리턴 채널을

통해서 협업기반 추천도 가능하다.

본 논문에서는 기존의 pEPG를 위한 추천 기법을 개선하여 내용 기반 추천과 협업 기반 추천을 결합한 혼합 추천 기법을 제안한다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구로 본 논문에서 기반으로 하고 있는 양방향 방송 서비스에 대해 알아보고, pEPG를 위한 추천기법에 대해 살펴본다. 또한, 사용자 정보 모델링을 위한 MPEG-7의 사용자 기술 구조에 대해 알아본다. 3장에서는 본 논문에서 제안하는 혼합추천 기법에 대해 기술하며, 4장에서는 제안한 기법에 대한 실험 및 평가를 하고, 마지막으로 5장에서는 결론 및 향후 연구 과제를 제시한다.

II. 관련 연구

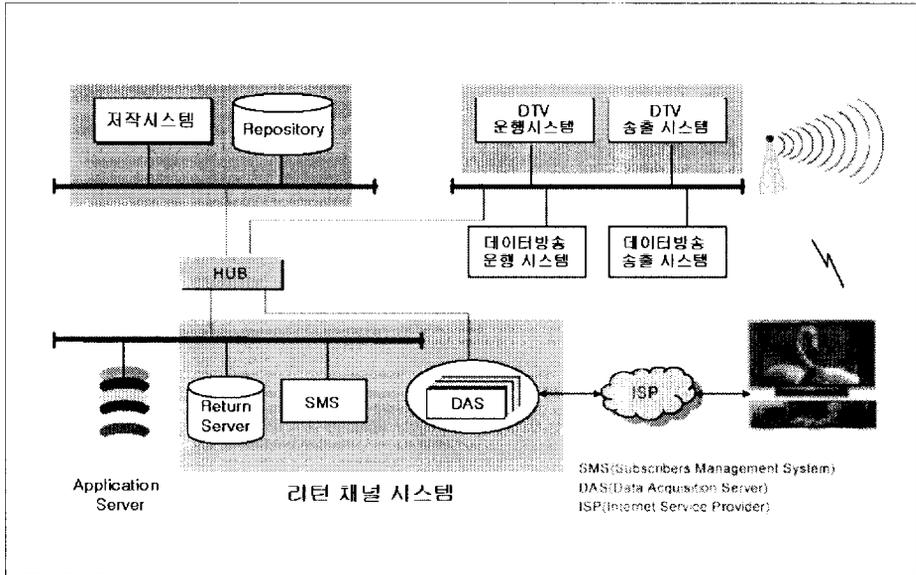
이 장에서는 본 논문에서 기반 환경으로 하고 있는 인터랙티브 방송 서비스에 대해 살펴보고, 대표적인 추천 기법과 사용자 정보 모델링을 위한 MPEG-7 의 사용자 기술 구조에 대해 알아본다.

2.1 인터랙티브 방송 서비스

인터랙티브 방송 서비스는 방송의 디지털화에 의해 가능해진 새로운 형태의 서비스로 텍스트·그래픽·문서·소프트웨어 등의 멀티미디어 데이터를 방송매체를 이용하여 전송하고, 전용 셋톱박스를 통하여 시청자가 그 정보를 이용할 수 있는 서비스를 말한다. 인터랙티브 방송 서비스는 방송과 통신의 융합이라는 시대적 흐름에 가장 잘 부합되는 서비스로서, 사용자와의 상호작용을 통해 방송정보의 개인화를 가능하게 하는 것으로 지금까지 체험해 오던 방송이라는 서비스의 속성을 획기적으로 바꾸는 새로운 서비스라 볼 수 있다.

시청자는 이 서비스를 통해 TV를 보면서 간편한 방법으로 프로그램 관련 부가정보를 얻거나 상품을 구입할 수도 있고, 날씨, 증권, 뉴스 등의 관심 정보를 검색할 수도 있으며, 집안에서 은행 업무도 처리할 수 있다. 또한 생방송 퀴즈 프로그램에 직접 참여하여 획득 점수에 따라 상품을 받을 수도 있고, 시청자 자신이 뉴스 기사를 제공하거나 의견을 방송 프로그램에 반영시키는 등의 적극적인 참여도 가능하다[18].

그림 1은 인터랙티브 방송 서비스를 위한 시스템 개념 모델을 나타낸 것으로, 광대역의 방송채널과 양방향 통신을 위한 리턴 채널을 활용하여 기존의 방송 서비스와 차별화된 사용자 중심의 인터랙티브 방송 서비스가 제공됨을 보여준다[19].



(그림 1) 인터랙티브 방송 서비스

인터랙티브 방송 시스템의 핵심은 사용자 인터랙션 정보를 인터넷과 같은 리턴채널을 통해 방송국 서버로 전달하는 것과 역으로 개별 사용자 응답을 셋톱박스로 전달하는 것으로 이를 담당하는 것이 바로 리턴채널 시스템이다.

리턴채널 시스템은 셋톱박스과 인터넷을 통해 직접 연결되며 필요한 데이터를 주고받는 리턴데이터 수집부 그리고 리턴데이터 수집부를 통한 입출력 데이터를 가공·처리하는 리턴서버, 사용자 정보를 보호하기 위한 보안·인증 시스템으로 구성되어 있다.

리턴채널 시스템은 특성상 양방향 애플리케이션이 저작될 당시에 리턴 서버에서 어떠한 리턴 데이터를 가공하여 전송할 것인지에 대한 정보도 함께 작성되어야 한다. 이때 작성되는 정보는 XML(eXtensible Markup Language) 형태이며 이러한 XML 파일은 송출 콘텐츠 DB에 저장되어 있다가 송출 매니저가 실제 방송 시간에 맞춰 IDP(Interactive Data Processor) 서버로 전달한다. IDP 서버의 양방향 애플리케이션 처리기가 XML 데이터를 분석하여 리턴서버로부터 필요한 정보를 질의(query)하며, 리턴서버에서는 애플리케이션에서 필요한 정보를 DB에 저장하고 가공하는 역할을 담당한다[5].

2.2 추천 기법

대표적인 추천 기법으로는 내용기반 추천(CBR : Content - Based Recommendation)과 협업기반 추천(CR : Collaborative Recommendation)이 있다. 이 절에서는 내용기반 추천과 협업기반 추천의 각각의 장단점에 대해 살펴본다.

2.2.1 내용기반 추천(Content-Based Recommendation)

내용기반 추천은 사용자가 평가했던 아이템에 대한 특징 정보와 다른 아이템에 포함된 특징 정보의 유사도를 이용하여 추천하는 기법이다[13]. 사용자의 과거 이력(History) 중 선호한 아이템들과 유사한 성격을 갖는 아이템과의 관계에 의해 추천하므로 'item-to-item' 추천이라고도 부른다.

내용기반 추천은 사용자의 과거 프로파일을 통해 추천이 이루어지므로 추천 결과를 쉽게 반영할 수 있어 추천 속도가 빠르다. 그러나 이 기법은

효과적인 추천을 위해 사용자의 충분한 과거 프로파일 정보를 축적해야 하며, 만약 추천의 대상이 되는 아이템에 관한 정보가 부족하다면 추천의 질이 떨어지게 된다. 또한 자신의 프로파일과 비교하여 높은 점수를 얻은 아이템을 추천함으로써 이미 평가한 콘텐츠와 유사한 아이템들만 과도하게 특수화되는 경향이 있으므로 사용자는 자신이 가진 취향의 테두리 안에 갇힐 우려가 있다[4].

2.2.2 협업기반 추천(Collaborative Recommendation)

협업기반 추천은 사용자들의 아이템에 대한 평가를 기초로 유사한 성향을 갖는 사용자들을 분류하고, 이렇게 분류된 사용자들의 아이템에 대한 평가를 이용하여 추천하는 기법이다[11]. ‘people-to-people’ 추천이라고도 부르는데 이 기법은 사용자 사이의 상관관계에 의해 추천함으로써, 대상 아이템이 상세한 설명 정보를 갖고 있지 않거나 사용자로부터의 상세한 선호 정보가 없어도 잘 동작한다. 또한, 사용자의 선호 정보와 아이템 사이의 유사성을 계산하기 위한 복잡한 과정을 전혀 수행하지 않고도 양질의 추천을 제공할 수 있다. 이러한 이유로 협업 기반 추천은 여러 추천 시스템에서 단독으로도 빈번히 이용되어 왔으며, 또 많은 경우에 있어서 충분히 훌륭한 성과를 거두기도 했다[10]. 또한, 협업기반 추천은 사용자의 직접적인 흥미에 포함되지 않는 아이템도 유사한 성향의 다른 사용자들에게 좋은 평가를 받게 되면 추천이 이루어질 수 있는 장점이 있다.

그러나 협업기반 추천 역시 만만치 않은 문제점을 안고 있다. 협업 기반 추천이 제대로 동작하기 위해서는 대상 아이템에 대한 각 사용자의 평가가 어느 정도 풍부하게 축적되어 있어야 한다. 따라서 추천의 대상이 새로 주어지는 경우, 누군가가 평가를 하기 전에는 추천이 이루어지지 않게 되며

자신이 가진 취향과 비슷한 사람이 적을 경우 추천의 질이 떨어질 수 있다. 자신과 비슷한 선호도를 가진 누군가가 평가를 해준 아이템을 추천 받아야 하기 때문에 비슷한 사람의 수가 적으면 상대적으로 손해를 볼 수밖에 없다. 따라서 다른 사용자와 어떤 뚜렷한 유사성이 나타나지 않는 사용자 층은 만족스러운 추천을 제공받지 못한다.

2.3 MPEG-7의 사용자 기술구조

2.3.1 MPEG-7의 표준화

MPEG(Moving Picture Experts Group)에서는 멀티미디어 정보 검색을 효율적으로 지원하기 위한 기술을 개발하는 새로운 아이템을 시작하여 이를 MPEG-7이라 불렀다. 이 MPEG-7은 멀티미디어의 내용 그 자체 또는 사용자가 관심을 보이는 자료를 빠르고 효율적으로 찾기 위한 여러 가지 멀티미디어 정보 기술의 표준화로써 데이터 그 자체가 아닌 데이터의 내용에 대한 표현 방법을 다루고 있다[12].

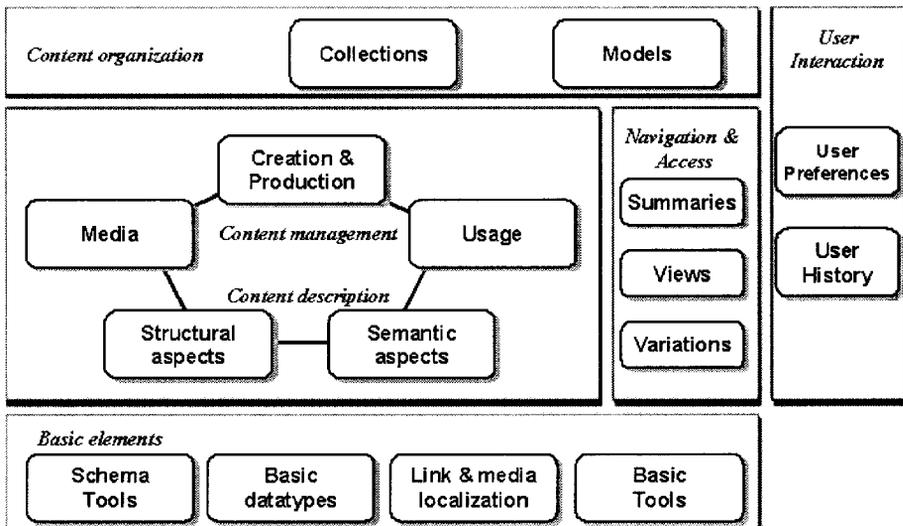
MPEG-7에는 다른 MPEG 표준들과 같은 시스템(Systems), 비디오(Video), 오디오(Audio), 참조 소프트웨어(Reference Software), 부합성(Conformance) 등의 다섯 부분 외에, 기술 정의 언어(Description Definition Language : DDL)와 멀티미디어 기술 구조(Multimedia Description Schemes : MDS)의 두 개 부분이 추가되어 있다.

각 부분에 대해서 살펴보면 시스템 부분은 전송(Transport), 저장(Storage), 동기화(Synchronization) 등이 정의되며 비디오, 오디오는 각각 비디오와 오디오에 직접 연관된 내용 기술방법을 정의하고, 참조 소프트웨어는

MPEG-7 표준과 관련된 소프트웨어의 구현, 기술 정의 언어는 새로운 기술 구조의 정의나 확장 등을 가능하게 하여주는 XML 기반의 언어이다. 또한, MDS는 일반적인 내용 기술 부분들과 기술자(Descriptor)와 기술 구조(Description Scheme)들 간의 구조적 연관관계 등을 정의한다.

인터랙티브 방송에서는 사용자 정보 기술을 위하여 MDS의 기술구조를 바탕으로 하고 있다. MDS의 구조는 그림 2와 같으며 멀티미디어 콘텐츠의 내용에 관한 정보들을 기술하고 그 기술한 내용들 사이의 연관 관계 등을 정의하는 기본 요소(Basic Elements)와 내용 정리(Content Organization), 내용 관리(Content Management), 내용 기술(Content Description), 이동과 접근(Navigation and Access), 사용자 인터랙션(User Interaction)의 6개 부분으로 나뉘어 멀티미디어 콘텐츠의 제작과 사용관계, 내용, 구조, 사용자 선호도 등을 기술하게 된다[12].

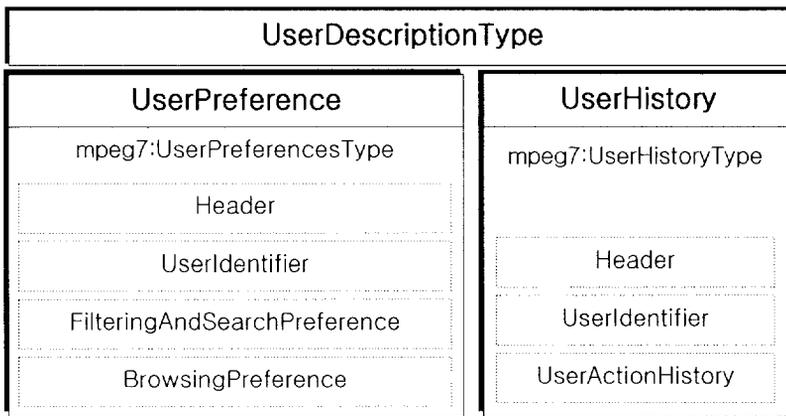
MDS는 XML을 이용하여 메타데이터를 기술하고 있는데 XML은 메타데이터를 자동적으로 생성하고 이용하는데 적합한 확장성을 가진 언어이다.



(그림 2) MDS의 구조

2.3.2 MPEG-7의 사용자 기술구조

사용자의 취향에 따른 pEPG를 생성하기 위해서는 사용자의 프로그램에 대한 선호도(User Preference)와 이용 내역(Usage History)을 알아야 한다. 이 부분은 MPEG-7의 사용자 인터랙션(User Interaction)에 기술되어 있으며 그 구조는 그림 3과 같다.



(그림 3) User Interaction 구조

(1) 사용자 선호도(User Preference)

사용자 선호도(User Preference)는 사용자의 콘텐츠의 소비와 관련된 사용자의 취향 정보를 기술하는데 이용되며 특정 사용자를 식별할 수 있는 사용자 식별자(User Identifier)와 필터링 및 검색 선호도(Filtering and Search Preference), 오디오 비주얼 콘텐츠의 요약 방식에 대한 선호도인 브라우징 선호도(Browsing Preference)로 구성되어 있다. MPEG-7 DDL로 표현한 사용자 선호도 기술 구조는 그림 4와 같다[12].

```

<complexType name="UserPreferencesType">
  <complexContent>
    <extension base="mpeg7:DSType">
      <sequence>
        <element name="UserIdentifier" type="mpeg7:UserIdentifierType"
          minOccurs="0"/>
        <element name="FilteringAndSearchPreferences"
          type="mpeg7:FilteringAndSearchPreferencesType"
          minOccurs="0" maxOccurs="unbounded"/>
        <element name="BrowsingPreferences"
          type="mpeg7:BrowsingPreferencesType"
          minOccurs="0" maxOccurs="unbounded"/>
      </sequence>
      <attribute name="allowAutomaticUpdate"
        type="mpeg7:userChoiceType" use="default" value="false"/>
    </extension>
  </complexContent>
</complexType>

```

(그림 4) User Preference 기술 구조

(2) 사용자 이용내역(Usage History)

사용자 이용내역(Usage History)은 관찰시간(Observation Period) 동안의 사용자 동작(User Action)에 대한 리스트(User Action List)를 기술하기 위해 사용되며 사용자 식별자(UserIdentifier)와 사용자 액션 히스토리(UserActionHistory)를 포함하고 있어 사용자별 프로그램 시청 리스트를 얻을 수 있다. 이를 응용하여 개인의 시청정보를 파악하여 프로그램의 제작에 반영한다거나 개인화된 맞춤형 서비스를 제공할 수 있다. 그림 5는 MPEG-7의 DDL로 정의된 사용자 히스토리 기술 구조를 보여준다[12].

```

<complexType name="UserActionHistoryType">
  <complexContent>
    <extension base="mpeg7:DSType">
      <sequence minOccurs="0">
        <element name="ObservationPeriod" type="mpeg7:TimerType"
          minOccurs="1" maxOccurs="unbounded"/>
        <element name="UserActionList" type="mpeg7:UserActionListType"
          minOccurs="1" maxOccurs="unbounded"/>
      </sequence>
      <attribute name="protected" type="mpeg7:userChoiceType"
        use="default" value="true"/>
    </extension>
  </complexContent>
</complexType>

<complexType name="UserActionListType">
  <complexContent>
    <extension base="mpeg7:DSType">
      <sequence minOccurs="0">
        <element name="ActionType" type="mpeg7:TermUseType"/>
        <element name="UserAction" type="mpeg7:UserActionType"
          minOccurs="0" maxOccurs="unbounded"/>
      </sequence>
      <attribute name="numInstances" type="nonNegativeInteger" use="optional"/>
      <attribute name="totalDuration" type="mpeg7:durationType" use="optional"/>
    </extension>
  </complexContent>
</complexType>

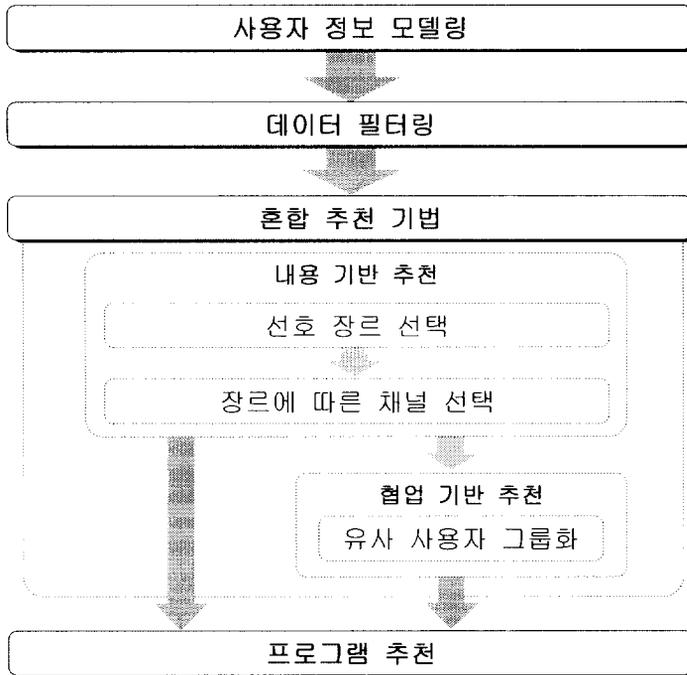
```

(그림 5) Usage History 기술 구조

Ⅲ. 혼합 추천 기법

이번 장에서는 TV 환경에 맞게 사용자 정보를 모델링하고, 기존 pEPG를 위한 추천 기법을 개선한 혼합 추천 기법(HR : Hybrid Recommendation)을 제안한다.

제안하는 혼합 추천 기법은 사용자 개인의 선호 프로그램과 유사 사용자 그룹의 선호 프로그램을 같이 추천해줌으로써 내용기반 추천과 협업 기반 추천의 단점을 상호보완 한다. 혼합 추천 기법의 개념도는 그림 6과 같다.



(그림 6) 혼합 추천 기법의 개념도

3.1 사용자의 프로그램 시청 정보 모델링

TV는 다른 환경과 달리 TV 프로그램을 기술하는 정보가 몇 가지 요소로 대표되어 있다. 웹 페이지를 대상으로 하는 추천 기법의 경우 해당 웹 페이지가 가진 정보는 일반적으로 단어들의 집합이며 사용자 정보도 이에 맞추어 단어들의 집합으로 나타난다. TV 환경에서도 이러한 단어들이 나타난다는 점은 같지만 각 단어가 속한 요소들에 의해 채널, 장르, 시청 시간 등 여러 가지 상위 요소로 나누어진다는 점이 다르다. 그러므로 사용자의 프로그램 시청 정보 모델링도 이에 맞추어져야 한다.

사용자의 프로그램 시청 정보는 MPEG 7 MDS의 사용자 인터랙션의 표준을 따르는 XML 데이터를 파싱(Parsing)하여 DB테이블의 형태로 관리된다. 사용자 이용내역을 기반으로 작성되며 시청자의 프로그램 선택이 장르, 채널, 시간, 요일, 날짜에 의해 좌우되므로 이 요소들로서 모델링 하였다.

식 (1)은 방송국으로부터 제공되는 프로그램 방송 정보를 나타내는 상태 벡터이다.

$$B_j = (N_j, G_j, L_j, C_j, S_j, E_j, D_j, DA_j) \quad (1)$$

where, B_j : 프로그램 j 의 방송 정보

N_j : 프로그램 j 의 이름

G_j : 프로그램 j 의 소속 장르

L_j : 프로그램 j 의 소속 서브장르

C_j : 프로그램 j 의 소속 채널

S_j : 프로그램 j 의 방송 시작 시간

E_j : 프로그램 j 의 방송 마침 시간

D_j : 프로그램 j 의 방송 요일

DA_j : 프로그램 j 의 방송 날짜

식 (2)는 사용자 이용내역에서 필요한 정보 필드만을 선택하여 사용자의 프로그램 시청 정보를 상태 벡터로 나타낸다.

$$H_{ij} = (ID_i, UG_{ij}, UL_{ij}, UC_{ij}, US_{ij}, UE_{ij}, UD_{ij}, UDA_i) \quad (2)$$

where, H_{ij} : 사용자 i 의 프로그램 j 의 시청 정보

ID_i : 사용자 i 의 아이디

UG_{ij} : 사용자 i 의 시청 프로그램 j 의 소속 장르

UL_{ij} : 사용자 i 의 시청 프로그램 j 의 소속 서브장르

UC_{ij} : 사용자 i 의 시청 프로그램 j 의 소속 채널

US_{ij} : 사용자 i 의 프로그램 j 의 시청 시작 시간

UE_{ij} : 사용자 i 의 프로그램 j 의 시청 마침 시간

UD_{ij} : 사용자 i 의 프로그램 j 의 시청 요일

UDA_i : 사용자 i 의 데이터 작성 날짜

식 (3)은 사용자가 TV를 켜고 있을 때 필요한 사용자의 프로그램 선호도를 나타내는 벡터이다.

$$F_i = (PG_i, PL_i, PC_i, T_i, D_i, DA_i) \quad (3)$$

where, F_i : 사용자 i 의 선호도

PG_i : 사용자 i 의 선호 장르

PL_i : 사용자 i 의 선호 서브장르

PC_i : 사용자 i 의 선호 채널

T_i : 사용자 i 의 선택 시간

D_i : 사용자 i 의 선택 요일

DA_i : 사용자 i 의 선택 날짜

3.2 데이터 필터링

3.2.1 프로그램 시청 시간에 따른 필터링

사용자의 프로그램 시청 정보에서 정해진 프로그램 방송 시간 중 일정 시간 이하를 시청한 데이터는 사용자의 선호도를 나타내기에 부적합하므로 필터링 되어야 한다. 예를 들어 정해진 방송 시간의 20%이하를 시청한 데이터를 부적합 데이터로 간주한다면 방송 시간이 60분인 프로그램의 경우 12분 이하를 시청한 프로그램, 방송 시간이 30분인 프로그램의 경우 6분 이하를 시청한 프로그램은 필터링 되어야 한다. 식 (4)를 통해 프로그램 시청 시간에 따른 가중치 T_{ij} 를 구하고 최소 프로그램 시청 비율(min_time)에 미치지 못하는 데이터를 필터링한다.

$$T_{ij} = \frac{WT_{ij}}{PT_j} \quad (4)$$

$$\text{if } T_{ij} < min_time, \quad T_{ij} = 0$$

where, T_{ij} : 프로그램 방송 시간에 따른 사용자 i 의
프로그램 j 의 시청 시간 비율

$$WT_{ij} = UE_{ij} - US_{ij}$$

$$PT_j = E_j - S_j$$

min_time : 최소 프로그램 시청 비율

3.2.2 데이터 저장 기간에 따른 필터링

TV 환경은 동적이기 때문에 영화나 드라마와 같은 경우가 아니면 다시 제공하는 경우도 드물고, 프로그램의 가치 자체가 떨어지기 때문에 오랜 기간이 지난 데이터는 사용자의 취향의 변화나 제공되는 프로그램의 변화를 고려할 때 신뢰할 수 없다. 따라서 프로그램 시청 정보에서 오래된 시청 정보는 필터링 되어야 한다. 식 (5)를 통해 현재로부터 얼마나 떨어져 있는지를 나타내는 가중치 R_i 를 구하고 최대 데이터 저장 기한(max_date)에 미치지 못하는 데이터를 필터링한다. 예를 들어 $DA_i = '2004-12-01'$ 이고 $UD_{ij} = '2004-11-01'$, $max_date = '15'$ 라면 2004-11-01부터 2004-11-15까지의 데이터는 필터링 된다.

$$R_i = \begin{cases} \frac{1}{distance} & , \text{ if } (distance < max_date) \\ 0 & , \text{ otherwise} \end{cases} \quad (5)$$

where, R_i : 사용자 i 의 데이터 저장 기간에 따른 가중치

$distance = DA_i - UDA_i$

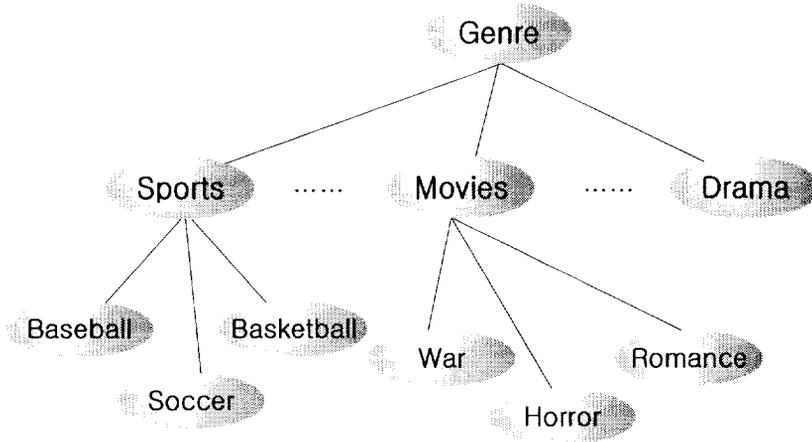
max_date : 최대 데이터 저장 기한

3.3 혼합 추천 알고리즘

TV 환경이 아닌 다른 추천 시스템의 경우에는 일반적으로 각 아이템에 대한 사용자들의 평가 점수를 바탕으로 추천이 이루어진다. 그러나 TV 환경의 경우는 프로그램을 보거나 보지 않는 두 가지의 경우 밖에 존재하지 않으며 이런 프로그램 시청 여부만을 평가 정보로 이용하여 프로그램을 추천하는 것에는 한계가 있다. 따라서 제공되는 프로그램의 장르와 채널을 이용하여 추천을 하는 것이 보다 타당하다.

3.3.1 선호 장르의 선택

TV 프로그램의 장르에 대한 선호도는 프로그램 선호도를 결정하는 중요한 요소이다. 프로그램의 장르 구성은 그림 7과 같이 장르와 서브장르의 트리 구조 형태를 이루고 있다. 장르와 서브장르라는 계층적인 구조로 표현되는 이유는 방송 프로그램의 특징 때문이다. 영화나 음악 추천 시스템의 경우에는 영화나 음악이라는 하나의 형식이 정해져 있고 그 형식 내에서 장르로 분류되는 것이 일반적이다. 그러나 방송 프로그램의 경우에는 여러 형식의 정보가 함께 제공되기 때문에 이를 추천하기 위해서는 먼저 형식에 따른 분류(ex : 영화, 드라마, 스포츠 등)가 필요하고 그 하위에 내용에 따른 분류(ex : 영화인 경우 내용에 따라 공포, 로맨스, 코미디 등)가 이루어진다.



(그림 7) 장르의 트리 구조의 예

필터링된 데이터들을 이용하여 시청 프로그램에 대한 장르의 선호도 V_G 를 구한다. V_G 는 식 (6)과 같이 전체 프로그램 시청 횟수에 대한 각 장르의 시청 횟수의 비로 계산되며 현재의 날짜에 영향을 받는다. 즉, 과거의 선호 장르에 비해 최근의 선호 장르에 더 비중을 두는 것이다. 이렇게 얻은 값 중 가장 큰 V_G 를 갖는 장르를 선호 장르 PG_i 로 선택한다.

$$V_G(G_{ij}, UG_{ij}) = R_i \times \frac{\sum_{j=0}^n \text{count}(G_{ij}, UG_{ij})}{\sum_{j=0}^n P_{ij}} \quad (6)$$

$$\text{count}(G_{ij}, UG_{ij}) \begin{cases} 1, & \text{if } (G_{ij} = UG_{ij}) \\ 0, & \text{otherwise} \end{cases}$$

$$PG_i = \text{Max}(V_G)$$

where, V_G : 사용자의 장르 선호도

P_{ij} : 사용자 i 의 시청 프로그램 j

G_{ij} : T_i, D_i, DA_i 에서의 사용자 i 의 프로그램 j 의 소속 장르

n : 최대 저장 데이터 수

식 (6)을 통해서 구해진 선호 장르에서 선호 서브장르 PL_i 를 구한다. V_L 는 식 (7)과 같이 선호 장르의 전체 시청 프로그램 수에서 특정 서브장르의 시청 횟수로 구한다. 이렇게 얻은 값 중 가장 큰 V_L 을 갖는 장르를 선호 서브장르 PL_i 로 선택한다.

$$V_L(L_{ij}, UL_{ij}) = \frac{\sum_{j=0}^n \text{count}(L_{ij}, UL_{ij})}{\sum_{j=0}^n P_{ij}} \quad (\text{단, } P_{ij} \subset PG_i) \quad (7)$$

$$\text{count}(L_{ij}, UL_{ij}) \begin{cases} 1, & \text{if } (L_{ij} = UL_{ij}) \\ 0, & \text{otherwise} \end{cases}$$

$$PL_i = \text{Max}(V_L)$$

where, V_L : 사용자의 서브 장르 선호도

L_{ij} : PG_i 에서의 사용자 i 의 프로그램 j 의 소속 서브장르

3.3.2 장르에 따른 채널의 선택

디지털 방송에 의해 수백 채널의 수용이 가능해짐으로써 사용자의 장르 선호도만으로 프로그램을 추천한다면 같은 시간에 같은 장르를 방송하는 채널이 상당수 존재할 수 있으므로, 추천된 여러 채널 중에서 원하는 채널을 선택해야 하는 불편함이 생기게 된다. 예를 들어 24시간 영화를 내보내는 채널 A와 B가 있다면 장르는 똑같은 영화로 구분이 되지 않지만, 채널이 무료인지 유료인지 어떤 서브장르의 영화를 많이 내보내는지에 따라 사람들의 채널 선호도는 달라질 수 있다. 이러한 문제점을 해결하기 위해 식 (8)을 이용하여 선호 장르가 여러 채널에서 동시에 방송 중일 경우 채널의 선호도 V_L 를 고려한다.

$$V_C(C_{ij}, UC_{ij}) = \frac{\sum_{j=0}^n \text{count}(C_{ij}, UC_{ij})}{\sum_{j=0}^n P_{ij}} \quad (\text{단, } P_{ij} \subset PG_i) \quad (8)$$

$$\text{count}(C_{ij}, UC_{ij}) \begin{cases} 1, & \text{if } (C_{ij} = UC_{ij}) \\ 0, & \text{otherwise} \end{cases}$$

$$PG_i = \text{Max}(V_C)$$

where, V_C : 사용자의 채널 선호도

C_{ij} : PG_i 에서의 사용자 i 의 프로그램 j 의 소속 채널

3.3.3 유사 사용자 그룹화

TV 프로그램의 선택은 다른 사람들의 선호도보다는 개별 사용자의 선호도에 크게 영향을 받기 때문에 개별 사용자의 프로그램 시청 내역을 이용하여 프로그램 추천이 이루어진다. 그러나 개인 사용자의 선호도만으로 프로그램을 추천할 경우 편협한 프로그램 추천이 이루어지는 문제가 발생한다. 이런 문제를 해결하기 위하여 식 (9)를 이용해서 선호 장르와 채널이 유사한 사용자들을 하나로 묶어 그룹을 생성하고 그 그룹 내의 사람들이 즐겨 보는 프로그램을 추천에 함께 이용한다. 식 (9)에서 이용되는 선호 장르와 채널은 개별 사용자들의 일주일 동안 축적된 선호 장르와 채널의 정보이다.

$$\text{sim}(U_i, U_j) = \frac{|U_i(PG_i, PL_i, PC_i) \cap U_j(PG_j, PL_j, PC_j)|}{|U_i(PG_i, PL_i, PC_i) \cup U_j(PG_j, PL_j, PC_j)|} \quad (9)$$

where, $\text{sim}(U_i, U_j)$: 사용자 i 와 j 의 유사도

3.3.4 선호 프로그램 선택

(1) 개인 사용자의 선호 프로그램 선택

사용자의 프로그램 시청 정보에서 특정 시간에서의 프로그램 선호도를 추출하여 가장 높은 선호도를 갖는 프로그램을 선택한다. 식 (6), (7), (8)을 특정 시간으로 한정지어서 계산해내고 추출된 선호 장르, 선호 서브장르, 선호 채널과 일치하는 프로그램을 선호 프로그램으로 선택한다.

(2) 유사 사용자 그룹의 선호 프로그램 선택

유사 사용자 그룹에 속한 사람들의 프로그램 선호도를 추출하여 가장 높은 선호도를 갖는 프로그램을 선택한다. 식 (6), (7), (8)을 그룹 내의 사용자들의 선호도로 확대 적용하여 그룹의 선호 장르, 선호 서브장르, 선호 채널을 추출해 내고 이와 일치하는 프로그램을 선호 프로그램으로 선택한다.

(3) 선호 프로그램 추천

개인 사용자의 프로그램 선호값과 유사 사용자 그룹의 프로그램 선호값을 정렬하여 높은 값을 가지는 프로그램 순으로 추천한다.

개인 사용자와 유사 사용자 그룹에서의 선호 프로그램을 함께 추천해줌으로써 자신의 선호 프로그램뿐만 아니라 자신과 성향이 유사한 사람들의 선호 프로그램이 같이 추천되어 내용기반 추천에서 볼 수 있는 편협한 프로그램 추천에서 벗어날 수 있게 된다.

또한, 새로운 프로그램이 처음으로 시작할 때 전송된 프로그램 메타데이터를 기준으로 개인 사용자의 선호 프로그램을 추천함으로써 협업기반 추천에서의 초기 평가자 문제를 극복할 수 있다.

IV. 실험 및 평가

이 장에서는 본 논문에서 제안하고 있는 pEPG를 위한 혼합 추천 기법을 기존의 순수 내용기반 추천과 비교하여 제안하는 혼합 추천 기법의 우수함을 보인다.

4.1 실험 환경

프로그램 시청 시 셋톱박스를 통해 사용자 시청 정보가 자동적으로 만들어져야 하나, 현재 셋톱박스에서 개인 사용자 시청 정보를 얻기 어려운 관계로 인터넷에서 서비스 중인 TV 편성표에서 프로그램 정보를 얻어 사용자 시청 정보를 생성하였다.

실험을 위한 시스템으로는 Windows 2000 Professional에서 ASP로 작성된 웹 페이지에서 사용자들의 TV 시청 행위들을 입력하였으며 이를 MS Access 2000으로 데이터베이스를 구축하였다. 본 논문에서의 실험 환경은 표 1과 같다.

<표 1> 실험 환경

구 분	구축 환경
OS	Windows 2000 Professional
DBMS	MS-Access 2000
Language	ASP(Active Server Page)
Web Server	MS IIS 5.0
Web Browser	MS Internet Explorer 6.0
Web Editor	Namo Web Editor FX

4.2 혼합 추천 기법의 적용

4.2.1 사용자 정보 모델링

TV를 시청한 사용자의 실제 시청 데이터를 얻는 것이 불가능하여 EPG에 프로그램과 관련된 정보를 링크로 처리하여 사용자들의 시청 데이터를 수집하였다. 실제 TV환경에서는 시청 기록 중 일정 시간 이상의 것만을 선택해야 하지만 실험에서는 사용자가 끝까지 프로그램을 시청했다고 가정한다. 또한, 실험의 편의를 위해서 실험 대상 방송을 장르에 특성화되어 있는 케이블 채널을 배제하고 종합 방송을 하는 지상파 5개 채널(KBS1, KBS2, MBC, SBS, EBS)만 이용하였다.



(그림 8) EPG 화면

또한 대상 프로그램을 특정 시간대(19:00~24:00)의 편성 프로그램으로 한정하였고, 같은 시간대의 다른 프로그램을 동시에 선택하는 것은 배제하였다. 그림 8은 ASP로 작성된 EPG이며 인터넷에서 제공하는 TV 편성표를 바탕으로 작성하였다.

그림 8의 EPG를 이용하여 총 30명의 사용자로부터 2005년 2월 20일부터 2005년 3월 5일까지의 2주간의 19:00시~24:00시 사이에 방영된 데이터를 획득하였다. 2주간 방영된 프로그램은 총 485개 (KBS1 : 101개, KBS2 : 90개, MBC : 91개, SBS : 90개, EBS : 113개)이며 30명의 사용자가 선택한 프로그램은 총 2268개이다.

사용자가 프로그램을 일정 시간 시청하였다고 가정하였고 2주간의 짧은 기간에 대한 실험이므로 식 (4)와 (5)의 T_{ij} 와 R_i 는 고려하지 않고 사용자 시청 정보를 그대로 이용한다. 그림 9와 10은 실험에 이용된 프로그램 정보와 사용자의 프로그램 시청 정보이다.

tv_channel	tv_date	tv_time	tv_program	tv_genre
KBS1	2005-02-20	오후 7:00:00	KBS 뉴스	뉴스
KBS1	2005-02-20	오후 7:10:00	도전 골든벨	오락
KBS1	2005-02-20	오후 8:00:00	KBS 스페셜	교양/사회
KBS1	2005-02-20	오후 9:00:00	KBS 뉴스 9	뉴스
KBS1	2005-02-20	오후 9:30:00	불멸의 미순신	드라마
KBS1	2005-02-20	오후 10:20:00	취재파일 4321	시사/정치
KBS1	2005-02-20	오후 11:00:00	KBS 영상실록	교양/사회
KBS1	2005-02-20	오후 11:30:00	영상기록 병원 2	다큐
KBS2	2005-02-20	오후 7:00:00	즐거운 일요일	오락
KBS2	2005-02-20	오후 7:55:00	부모님 전상서	드라마
KBS2	2005-02-20	오후 8:55:00	개그 콘서트	오락
KBS2	2005-02-20	오후 10:05:00	비타민	의료/건강
KBS2	2005-02-20	오후 11:15:00	드라마시티	드라마
SBS	2005-02-20	오후 7:00:00	일요일이 좋다	오락
SBS	2005-02-20	오후 8:00:00	SBS 8 뉴스	뉴스
SBS	2005-02-20	오후 8:35:00	오늘의 스포츠	스포츠

(그림 9) 프로그램 정보 데이터베이스

user_id	channel	date	time	program	genre
user1	kbs1	2005-02-20	오후 7:00:00	KBS 뉴스	뉴스
user1	kbs1	2005-02-20	오후 7:10:00	도전 골든벨	오락
user1	kbs2	2005-02-20	오후 7:55:00	부모님 전상서	드라마
user1	kbs2	2005-02-20	오후 8:55:00	개그콘서트	오락
user1	kbs2	2005-02-20	오후 10:05:00	비타민	의료/건강
user1	sbs	2005-02-20	오후 10:55:00	세븐 데이즈	드라마
user1	sbs	2005-02-20	오후 11:45:00	인정사형 목걸이	영화
user2	MBC	2005-02-20	오후 7:00:00	일요일 일요일 밤에	오락
user2	kbs2	2005-02-20	오후 7:55:00	부모님 전상서	드라마
user2	sbs	2005-02-20	오후 8:45:00	토지	드라마
user2	sbs	2005-02-20	오후 9:45:00	몸날	드라마
user3	kbs2	2005-02-20	오후 11:15:00	드라마시티	드라마
user4	kbs2	2005-02-20	오후 8:55:00	개그 콘서트	오락
user4	sbs	2005-02-20	오후 9:45:00	몸날	드라마
user4	kbs1	2005-02-20	오후 11:00:00	KBS 영상실록	교양/사회
user4	kbs1	2005-02-20	오후 11:30:00	영상기록 병원 24시	다큐
user5	kbs2	2005-02-20	오후 7:00:00	즐거운 일요일 해피선데이	오락
user5	kbs1	2005-02-20	오후 9:00:00	KBS 뉴스	뉴스

(그림 10) 사용자 시청 정보 데이터베이스

4.2.2 선호 장르 및 채널 추출

지상파 채널의 장르 구분은 표 2와 같이 17개로 분류하였으며 아직 서브 장르까지 제공되지 않고 있어 이는 배제하였다.

<표 2> 프로그램 장르 구분

장르 구분	
뉴스	문화/생활
오락	만화
교양/사회	어린이
드라마	경제/부동산
시사/정치	스포츠
다큐	난씨
음악	교육
영화	환경
의료/건강	

장르와 채널에 대한 일주일간의 user1의 선호도는 표 3과 같다. 3장에서 제안한 식 (6)과 (8)을 이용하여 계산되었으며 서브장르까지 제공되고 있지 않아 서브장르에 대한 선호도는 배제하였다.

<표 3> user1의 내용기반 선호도

날짜	선호장르(V_G)	선호 채널(V_C)
1일	드라마(0.29)	KBS2(0.43)
2일	드라마(0.49)	MBC2(0.33)
3일	드라마(0.58)	KBS2(0.36)
4일	드라마(0.63)	KBS2(0.32)
5일	드라마(0.62)	KBS2(0.30)
6일	드라마(0.63)	KBS2(0.32)
7일	드라마(0.64)	KBS2(0.31)

4.2.3 유사 사용자 그룹 생성

3장에서 제안한 식 (9)를 이용하여 선호 장르와 채널이 유사한 사용자들을 그룹화 하였다. 각 사용자들의 일주일간의 시청 정보를 이용하여 사용자 그룹을 생성하였고 각 그룹은 선호 장르에 따라 교양 선호형, 드라마 선호형, 오락 선호형의 3종류로 나타났으며 소속된 그룹에서 선호 채널에 의해 다시 분류 되어진다. 각 사용자들의 소속 그룹은 표 4와 같다.

<표 4> 유사 사용자 그룹

사용자 그룹	사용자	선호장르	장르값	선호채널	채널값	
교양 선호형	user7	뉴스	0.71	KBS1	0.58	
	user24	다큐	0.34	KBS1	0.25	
	user26	교양/사회	0.74	MBC	0.42	
드라마 선호형	A	user5	드라마	0.64	KBS1	0.27
		user15	드라마	0.65	KBS1	0.29
		user29	드라마	0.34	KBS1	0.34
	B	user1	드라마	0.64	KBS2	0.31
		user3	드라마	0.57	KBS2	0.29
		user11	드라마	0.25	KBS2	0.35
	C	user2	드라마	0.58	MBC	0.48
		user4	드라마	0.66	MBC	0.38
		user8	드라마	0.78	MBC	0.48
		user12	드라마	0.57	MBC	0.28
		user16	드라마	0.78	MBC	0.21
	D	user28	드라마	0.78	MBC	0.28
		user6	드라마	0.49	SBS	0.33
		user22	드라마	0.63	SBS	0.26
	오락 선호형	E	user30	드라마	0.54	SBS
user9			오락	0.62	KBS2	0.43
user17			오락	0.42	KBS2	0.27
user19			오락	0.39	KBS2	0.34
user25			오락	0.39	KBS2	0.31
F		user27	오락	0.66	KBS2	0.28
		user10	오락	0.43	MBC	0.32
		user18	오락	0.47	MBC	0.48
		user20	오락	0.59	MBC	0.47
G		user23	오락	0.26	MBC	0.42
		user13	오락	0.62	SBS	0.43
		user14	오락	0.41	SBS	0.26
	user21	오락	0.41	SBS	0.26	

4.2.4 선호 프로그램 추천

user1을 위한 날짜별 프로그램 추천 결과는 표 5와 같다. 선택시간은 20:30~21:00인 경우와 22:00~23:00인 경우이고 사용자의 내용기반 추천과 협업기반 추천이 함께 제공된다. 내용기반 추천과 협업기반 추천의 결과가 동일한 경우도 나타나지만 서로 다른 경우도 나타나므로 사용자의 프로그램 선택의 폭이 넓어진다. 협업기반 추천의 경우 user1은 표 4에 나타난 바와 같이 드라마 선호 그룹 B에 소속되어 user3과 user11의 시청 정보까지 고려하여 추천이 이루어진다.

<표 5> user1을 위한 프로그램 추천 결과

선택 시간	날짜	내용기반 추천		협업기반 추천	
		프로그램명	채널	프로그램명	채널
20:30 ~ 21:00	8일	부모님 전상서	KBS2	부모님 전상서	KBS2
	9일	어여쁜 당신	KBS2	어여쁜 당신	KBS2
	10일	어여쁜 당신	KBS2	어여쁜 당신	KBS2
	11일	어여쁜 당신	KBS2	어여쁜 당신	KBS2
	12일	어여쁜 당신	KBS2	어여쁜 당신	KBS2
	13일	어여쁜 당신	KBS2	어여쁜 당신	KBS2
	14일	부모님 전상서	KBS2	부모님 전상서	KBS2
22:00 ~ 23:00	8일	비타민	KBS2	봄날	SBS
	9일	쾌걸 춘향	KBS2	영웅 시대	MBC
	10일	쾌걸 춘향	KBS2	영웅 시대	MBC
	11일	해신	KBS2	슬픈 연가	MBC
	12일	해신	KBS2	슬픈 연가	MBC
	13일	VJ특공대	KBS2	사랑공감	SBS
	14일	느낌표	MBC	봄날	SBS

일주일간의 사용자 선호 정보를 누적한 결과를 이용하여 유사 사용자 그룹을 생성하기 때문에 유사 사용자 그룹이 생성되기 이전의 일주일간의 프로그램 추천은 사용자의 내용기반 추천만으로 이루어진다. 본격적인 혼합 추천은 8일째부터 이루어지며 user1의 8일째의 시청 정보는 표 6과 같다.

<표 6> user1을 위한 혼합 추천 기법 적용

선택 시간	추천 리스트		시청 리스트	
	프로그램명	채널	프로그램명	채널
19:00	도전 감동 골든벨 일요일 일요일 밤에	KBS1 MBC	일요일 일요일 밤에	MBC
20:00	부모님 전상서	KBS2	부모님 전상서	KBS2
21:00	개그콘서트 토지	KBS2 SBS	개그콘서트	KBS2
22:00	비타민 봄날	KBS2 SBS	봄날	SBS
22:20	취재파일 4321 시사매거진2580	KBS1 MBC	시청기록없음	
23:00	드라마시티	KBS2		
24:00	CSI 마이애미 영화특급	MBC SBS		

4.3 성능 평가

*coverage*는 사용자가 실제 시청한 전체 프로그램에 대해서 추천 기법을 통해 추천된 프로그램의 시청 여부를 나타내는 척도이다. 본 논문에서의 실험은 pEPG가 구현된 상황이 아니기 때문에 이미 수집된 시청자의 프로그램 시청 기록과 추천된 프로그램의 일치도를 비교하여 성능을 평가한다.

*coverage*는 식 (10) 같이 계산된다.

$$coverage = \frac{|R \cap W|}{|W|} \quad (10)$$

where, *R* : 프로그램 추천 리스트
W : 프로그램 시청 리스트

표 7은 기존의 pEPG 관련 연구에서 이루어졌던 내용기반 추천을 했을 때의 *coverage*이다. 표에 나타난 *coverage*는 30명의 사용자 *coverage*의 평균으로 나타냈다.

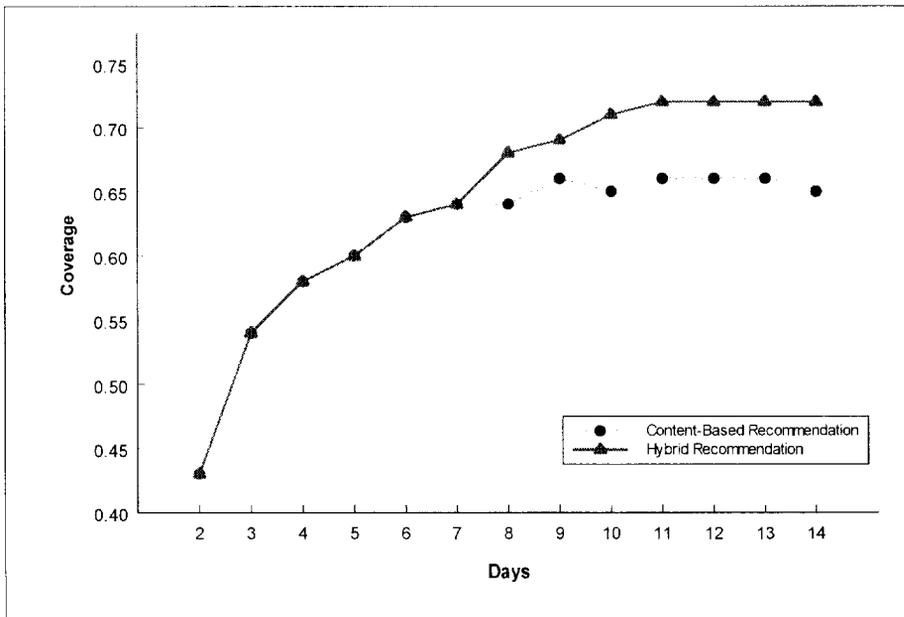
<표 7> 순수 내용기반 추천의 *coverage*

날짜	<i>coverage</i>	날짜	<i>coverage</i>
2일	0.43	8일	0.64
3일	0.54	9일	0.66
4일	0.58	10일	0.65
5일	0.60	11일	0.66
6일	0.63	12일	0.66
7일	0.64	13일	0.66
		14일	0.65

표 8은 제안하는 혼합 추천 기법에서의 *coverage* 결과이다. 협업기반 추천이 함께 제공되기 전의 일주일 동안은 내용 기반 추천이 실시되기 때문에 내용기반 추천의 *coverage* 결과와 동일하다. 표에 나타난 *coverage*값은 30명의 사용자 *coverage*의 평균으로 나타냈다.

<표 8> 혼합 추천 기법의 *coverage*

날짜	<i>coverage</i>	날짜	<i>coverage</i>
2일	0.43	8일	0.68
3일	0.54	9일	0.69
4일	0.58	10일	0.71
5일	0.60	11일	0.72
6일	0.63	12일	0.72
7일	0.64	13일	0.72
		14일	0.72



(그림 11) 추천 기법별 *coverage*

그림 11에서 나타나듯이 본 논문에서 제시하고 있는 혼합 추천 기법은 기존의 내용기반 추천을 개선한 것으로 일주일 동안은 동일한 성능을 나타내지만, 유사 사용자 그룹을 생성하여 협업기반 추천의 결과를 함께 제공하면서부터는 추천 효율이 증가한다. 제안하는 혼합 추천의 평균 *coverage*는 0.65로 내용기반 추천의 평균 *coverage* 0.61보다 0.04 높게 나타난다.

그리고 공통적인 특징으로 시간이 지날수록 *coverage*값이 증가하다가 일정 수준에 도달하면 더 이상 증가하지 않고 일정하게 유지되는 현상이 나타난다. 먼저 시간에 따른 *coverage*의 증가 현상은 시간의 경과에 따라 사용자의 시청 정보가 누적되어 사용자의 선호도가 잘 반영되어 추천이 이루어지고 있음을 보여주고, *coverage*가 일정 수준에 도달하면 더 이상 증가하지 않고 일정하게 유지되는 것은 각 시간별로 프로그램의 추천은 계속 이루어지나 사용자가 항상 TV를 시청하는 것은 아니기 때문에 나타나는 현상이라 분석된다.

V. 결 론

디지털 기술의 발달과 통신 네트워크의 고도화로 인해 수백 채널의 수용이 가능해졌다. 그러나 시청 채널수의 증가는 폭발적인 프로그램의 증가를 가져오고, 이로 인해 사용자가 원하는 프로그램을 선택하는데 어려움이 나타나게 되었다. 이러한 문제점을 해결하는 방안으로 시청자가 모든 채널을 검색하지 않아도 본인의 시청 기록과 방송 메타데이터를 이용하여 자신의 선호 프로그램을 알아서 추천해주는 pEPG(personalized Electronic Program Guide)가 대두되었다.

기존의 pEPG에서의 추천 기법은 개별 사용자의 선호도를 기반으로 한 내용기반 추천으로 추천 프로그램이 장르에 특수화되어 나타나는 문제점이 있었다. 기존 TV 환경에서는 협업기반 추천으로 프로그램을 추천하는 것이 불가능하나 인터랙티브 TV 환경으로 진화하면서 협업 기반 추천도 가능해진다. 협업기반 추천은 내용기반 추천의 장르에 특수화되는 문제점은 보완할 수 있으나 새로운 프로그램 정보가 입력되었을 때 사용자들의 평가 정보가 존재하지 않아 추천 자체가 곤란한 문제가 발생 할 수 있다.

본 논문에서는 기존의 내용기반 추천과 협업기반 추천을 TV 환경에 맞게 개선하였으며 이들을 결합한 혼합 추천 기법을 제안하고 있다. 실험을 통해서 제안한 혼합 추천 기법이 평균 *coverage*가 기존의 내용기반 추천보다 0.04 높은 수치를 나타냄을 보였다.

향후의 과제로는 사용자의 선호도의 변화에 따라 유사 사용자 그룹을 재편성하는 기법에 대한 연구와 사용자마다 내용기반 추천과 협업기반의 의존비율이 다르므로 이를 이용하여 내용기반과 협업기반의 의존가중치를 따로 두어 동적인 추천을 하는 것이다.

[참고 문헌]

- [1] B. Smyth, P. Cotter and J. Ryan, “Evolving the personalized EPG – an alternative architecture for the delivery of DTV services”, In Proceedings of the AH’02 Workshop on Personalization in Future TV, pp. 161-164, 2002.
- [2] Bagic. R. and Medcini. M., “User’s requirements for electronic program guide (epg) in interactive television (itv)”, IEEE Video/Image Processing and Multimedia Communications, pp. 457-462, June, 2002.
- [3] Derry O’ Sullivan, Barry Smyth, David C. Wilson, Kieran Mc-Donald, Alan F. Smeation, “Improving the Quality of the Personalized Electronic Program Guide”, In Proceedings of the 2nd Workshop on Personalization in Future TV, pp. 5-36, May, 2002.
- [4] Douglas W. Oard, Gray Marchionini, “Conceptual framework for text filtering”, Technical Report CS TR3643, pp. 1-32, May, 1996.
- [5] Georgios Lekakos¹, Dimitris Papakiriakopoulos¹, Kostas Chorianopoulos, “An Integrated Approach to Interactive and Personalized TV Advertising”, UM 2003 Workshop on Personalization in Future TV, pp. 1 10, 2003.
- [6] James M. Barton, “Television Meta-Data Standardization Overview and Opportunities”, In Proceedings of International Conference on Consumer Electronics, pp. 28 29 , June, 2000.

- [7] Jiangshan Xu, Liang-Jie Zhang, Haiming Lu, Yanda Li, “The development and prospect of personalized TV program recommendation systems”, In Proceedings of the IEEE 4th International Symposium on Multimedia Software Engineering (MSE 02), pp. 82-89, December, 2002.
- [8] L. Ardissono, F. Portis, P. Torasso, F. Bellifemine, A. Chiarotto, and A. Difino. “Architecture of a system for the generation of personalized Electronic Program Guides”, In Proceedings of the UM2001 Workshop on Personalization in Future TV, 2001.
- [9] Munchurl Kim, Jeewoong Ryu, Beetnara Bae, Jeho Nam, Jyoungok Kang and Jinwoong Kim, “Intelligent Program Guide for Digital Broadcasting” In Proceedings of the International Workshop on Advanced Image Technology(IWAIT 2002), pp. 257-263, January, 2002.
- [10] P. Melville, R. J. Mooney, R. Nagarajan, “Content-Boosted Collaborative Filtering for Improved Recommendations”, In Proceedings of the Eighteenth National Conference on Artificial Intelligence(AAAI-2002), pp. 187-192, July, 2002.
- [11] R. Burke, “Hybrid Recommender System : Survey and Experiments”, User Modeling and User Adapted Interaction, Vol. 12, No. 4, pp. 331-370, 2002.
- [12] Salembier. P., Smith. J.R., “MPEG 7 multimedia description schemes” Circuits and Systems for Video Technology, IEEE Transactions, Vol. 11, Issue 6, pp. 748-759, June 2001.

- [13] Weizhen Dai, Robin Cohen, “Dynamic Personalized TV Recommendation System”, In Proceedings of the 3rd Workshop on Personalization in Future TV, June, 2003.
- [14] 김병만, 이경, 김시관, 임은기, 김주연, “추천시스템을 위한 내용기반 필터링과 협력필터링의 새로운 결합 기법”, 정보과학회논문지 : 소프트웨어 및 응용, 제31권 제3호, pp. 332-342, 2004.
- [15] 김상근, 김성주, “TV 프로그램 선호도에 관한 주관성 연구”, 주관성 연구 제8호, pp. 52-79, 2003.
- [16] 박종석, 김민수, “디지털TV”, 정보과학회지, 제18권 제9호, pp. 13-20, 2000.
- [17] 유상원, 이홍래, 이형동, 김형주, “TV 프로그램을 위한 내용기반 추천 시스템”, 정보과학회논문지 : 컴퓨팅의 실제, 제9권 6호, pp. 683-692, 2003.
- [18] 정충영, “디지털TV 방송 서비스 수익모델과 사업자별 T-commerce 통합 전략”, 한국해양정보통신학회논문지, 제7권 4호, pp. 589-597, 2003.
- [19] 최진수, 김진웅, “데이터방송 잠정 표준(TTALKO-07.0015)”, TTA 저널, 제77호, pp. 57-65, 2001.
- [20] 한은영, “인터랙티브 TV(Interactive TV)서비스의 전개현황 및 주요 쟁점”, 정보통신정책 ISSUE, 제12권 6호, pp. 1-63, 2000.

감사의 글

대학원을 입학하면서 멀게만 느껴지던 마지막 학기가 얼마 남지 않았습니다. 길고도 짧았던 2년 반의 대학원 생활은 아쉬움도 많이 남지만 보람되고 뜻깊은 시간이었습니다. 많은 사랑과 관심으로 부족한 제가 더욱 성장할 수 있도록 도움을 주신 많은 분들께 진심으로 감사의 말씀 드립니다.

이 논문이 완성되기까지 저의 부족한 면을 채워 주시고, 학문의 길로 인도해주신 지도교수님 윤성대 교수님께 진심으로 감사의 말씀을 드립니다. 바쁘신 와중에도 부족한 이 논문을 심사해 주신 김창수 교수님, 신상욱 교수님께도 깊은 감사드립니다. 또한 학부부터 석사 과정까지 많은 가르침을 주신 박만곤 교수님, 김영봉 교수님, 정순호 교수님, 박승섭 교수님, 이경현 교수님, 박홍복 교수님, 여정모 교수님께도 감사의 마음을 전합니다.

같이 대학원 생활을 하며 많은 도움을 줬던 귀녀 선배, 은아 선배, 미나에게도 고마움을 전하며, 순환 선배, 상일 선배, 박현호 선생님, 고석범 선생님, 윤종찬 선생님, 이경미 선생님, 소영이 언니, 영선이 언니, 그 외 연구실의 모든 식구들에게도 감사 드립니다.

그리고 대학원과 직장 때문에 소홀했던 사랑하는 우리 가족과 상윤 선배에게도 감사와 사랑한다는 말을 전합니다.

2005. 6. 21.

서 정 애 드림