

데이터 스트림 마이닝 기법을 적용한 개인/커뮤니티 맞춤형 Digital TV 시스템

신 세 정[†] · 이 원 석^{††}

요 약

국내의 TV방송의 디지털 전환 프로젝트가 본격적으로 진행되고 있다. 디지털 방송 서비스는 다매체, 다채널을 통한 방송 프로그램의 증가와 양방향 TV방송 서비스로 인해 사용자에게 다양한 방송 프로그램의 선택과 개인/커뮤니티별 맞춤형 시청 기회를 제공함으로써 새로운 방송 서비스 환경을 필요로 한다. 본 논문에서는 TV-Anytime 영상 메타데이터에 대한 데이터 스트림 마이닝 기법을 이용하여 사용자의 시청 상황을 포함한 시청 패턴을 분석함으로써 개인/커뮤니티 시청 패턴 프로파일 및 시청 선호도 연관규칙 생성 기법을 적용한 개인/커뮤니티 맞춤형 Digital TV 시스템을 제안한다. 또한, 임베디드 시스템 기반의 사용자 인터페이스를 구현하여 개인/커뮤니티 사용자들에게 적절한 추천 프로그램을 제공하고, 시청 프로그램 정보에 따른 시청 상황을 자동으로 제어하는 기능을 포함한다. 또한, 스마트폰 기반의 채널 추천 시스템을 구현하여 프로파일의 활용도를 증가시켰으며, 실험을 통하여 본 논문에서 제안하는 방법의 효율성을 검증한다.

키워드 : 데이터 스트림 마이닝, DTV, 맞춤형 서비스

Customized Digital TV System for Individuals/Communities based on Data Stream Mining

Shin Se Jung[†] · Lee Won Suk^{††}

ABSTRACT

The switch from analog to digital broadcast television is extended rapidly. The DTV can offer multiple programming choices, interactive capabilities and so on. Moreover, with the spread of Internet, the information exchange between the communities is increasing, too. These facts lead to the new TV service environment which can offer customized TV programs to personal/community users. This paper proposes a 'Customized Digital TV System for Individuals/Communities based on Data Stream Mining' which can analyze user's pattern of TV watching behavior. Due to the characteristics of TV program data stream and EPG(electronic program guide), the data stream mining methods are employed in the proposed system. When a user is watching DTV, the proposed system can control the surrounding circumstances as using the user behavior profiles. Furthermore, the channel recommendation system on the smart phone environment is proposed to utilize the profiles widely.

Keywords : Data Stream Mining, DTV, Personalization System

1. 서 론

국내의 TV방송의 디지털 전환 프로젝트가 본격적으로

진행되고 있다. 최근 시작된 국내 디지털 지상파 방송으로 이제 본격적인 디지털 방송 시대가 열리게 되었다. 방송 제작부터 송수신 기술 및 모든 단말기가 디지털화 되어, 부가적인 멀티미디어 데이터 방송이 제공된다. 기존의 TV방송 서비스에서는 방송 프로그램을 단순히 수동적으로 시청하는 형태가 주를 이루지만, 디지털 방송 서비스는 양방향 TV방송 서비스를 가능하게 함으로써 정보 공급자와 사용자 간의 수평적 커뮤니케이션이 가능하게 되었다.

다매체, 다채널을 통한 방송 프로그램의 증가와 양방향

* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국과학재단의 국가지정연구실사업(No.2010-0008007) 및 특정기초연구(No.2008-0052335)지원으로 수행된 연구임.

† 정 회 원 : 연세대학교 컴퓨터과학과 석·박사 통합과정

†† 종 신 회 원 : 연세대학교 컴퓨터과학과 교수

논문접수 : 2010년 4월 13일

수정일 : 1차 2010년 7월 14일, 2차 2010년 8월 17일

심사완료 : 2010년 8월 23일

TV방송 서비스는 사용자에게 다양한 방송 프로그램의 선택과 개인별 맞춤형 시청 기회뿐만 아니라, 동호회, 지역 등 커뮤니티간의 정보교류가 활성화됨에 따라 사용자에게 다양한 방송 프로그램 선택과 개인/커뮤니티별 맞춤형 시청 기회를 제공하는 방송 서비스 환경을 필요로 하게 되었다.

본 논문에서는 맞춤형 DTV(Digital TV) 방송 서비스를 제공하기 위하여 각 사용자의 시청 패턴을 분석함으로써 개별 사용자의 취향에 맞는 방송 정보를 선별적으로 제공할 수 있도록 하는 시청 선호도 프로파일 생성 기법을 제안한다. 이러한 맞춤형 DTV 방송 서비스의 구현을 위해서는 DTV방송 스트림, 다채널 방송에 대한 EPG(Electronic Program Guide) 데이터와 시청 상황 센서 카메라 및 홈 스마트 센서로부터 전송되는 실시간 정보 등 매시간 끊임없이 생성되는 실시간 데이터들의 분석을 필요로 한다. 기존의 대용량 데이터베이스에 대한 데이터 분석기법들로 이러한 정보들을 저장/분석하는 경우 과도한 컴퓨팅 자원이 요구될 뿐만 아니라 분석 과정 동안 새로 생성된 정보들이 축적되어 실시간으로 그 결과를 활용할 수 없다. 따라서 이러한 실시간 정보들을 한정된 컴퓨팅 자원 내에서 실시간 분석을 할 수 있는 데이터 스트림 마이닝 기법을 적용한다.

또한 스마트폰이나 PMP와 같은 다양한 형태의 모바일 기기의 빠른 보급과 함께 모바일TV의 수요가 증가하면서 DMB 방송 채널수의 급증에 따른 DMB 프로그램 추천을 위해, 이동TV 사용자의 채널 시청 패턴에 대해 클러스터링 기반의 위치/시간별 분석을 통해 현재 위치와 시간에서 사용자에게 취향에 보다 적합한 선호 채널 리스트를 제공할 수 있는 시스템을 제안한다.

본 논문에서는 이러한 서비스를 보다 효율적으로 이용할 수 있도록 데이터 스트림 마이닝 기법을 적용한 개인/커뮤니티 맞춤형 Digital TV 시스템을 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 알아보고 3장에서는 제안한 시스템의 구조와 TV-Anytime 영상 메타데이터 스트림을 실시간으로 분석하여 시청 패턴 프로파일을 생성하는 알고리즘과 시청상황을 포함하는 시청정보를 활용하는 시청 선호도 연관규칙 알고리즘에 대하여 설명한다. 4장에서는 커뮤니티별 시청 패턴/상황 프로파일을 생성하는 기법을 설명하고 5장에서 스마트폰 사용자 맞춤형 채널 추천 시스템에 대하여 기술한다. 6장에서는 사용자 인터페이스와 실험결과 및 성능에 대해서 논의 하고, 마지막으로 7장에서 결론을 기술한다

2. 관련연구

사용자 맞춤형 방송을 위해서는 방송 콘텐츠의 내용과 사용 환경을 기술하기 위한 메타데이터가 제공되어야 한다. 멀티미디어의 콘텐츠 서술을 위한 표준으로는 MPEG-7[2]

이 있으며 멀티미디어의 전송, 소비 등의 사용 환경 기술에 중점을 둔 표준으로는 MPEG-21[3]이 있다. 메타데이터 기반의 맞춤형 방송을 위한 표준은 현재 TV-Anytime Forum[1]을 중심으로 활발하게 연구되고 있다. TV-Anytime Forum은 개인 저장장치(PDR: Personal Digital Recorder)를 이용하여 사용자가 원하는 콘텐츠를 원하는 시간에 효율적으로 선택 시청할 수 있는 맞춤형 디지털 방송 제공을 목적으로 설립되었다. TV-Anytime의 메타데이터는 MPEG-7에서 정의된 내용 기반 정보와 EPG정보를 포함하며 크게 프로그램 기술 메타데이터와 사용자 기술 메타데이터로 구분된다. 프로그램 기술 메타데이터는 방송 프로그램에 대한 고유번호 부여 방식과 콘텐츠 내용 표현을 위한 메타데이터들로 이루어져 있다. 사용자 기술 메타데이터는 다양한 값들로 사용자 선호도를 기술할 수 있도록 정의하고 있다.

사용자 맞춤형 서비스를 위한 대표적인 추천 기법으로는 Content-based 추천[4], Collaborative Filtering[5]이 있으며, 두 방법을 혼용한 방법[6] 외에 추천 및 사용자 패턴 학습 데이터를 사용하는 방법들이 제안되었다.

디지털 영상 저장장치를 통한 방송 서비스를 제공하는 TiVo[5]는 협동 필터링(Collaborative Filtering)방법을 사용하여 프로그램의 선택에 유사한 성향을 보였던 사용자들은 새로운 프로그램에 대해서도 유사한 성향을 보인다는 가정을 바탕으로 하는 추천 시스템을 이용한다. 이 방법은 내용 분석의 용이성에 관계없이 어떠한 프로그램에 대해서나 분석이 가능하다는 장점이 있지만, 새로운 사용자들이 등장하면, 기존 사용자 관련 데이터를 저장하는 데 드는 비용과 계산을 수행하는 데 걸리는 시간이 급격히 증가하게 되는 단점이 있다. 또한, 사용자의 피드백을 필요로 하기 때문에 데이터에 많은 오차를 포함할 수 있다.

[7]에서는 사용자들의 시청 패턴을 분석하여 표적 광고를 제공하는 Advertising Delivery System(ADS)을 제안한다. ADS는 PVR(Personal Video Recorder)로부터 사용자의 시청 데이터를 받아서 분류한다. 이러한 정보는 사용자의 인구통계학적/심리적 정보들과 함께 프로파일로 생성되며, 다중 데이터 마이닝을 통해서 정제된 후, 사용자의 투표 시스템에 의해 결과를 조정한다. 이 시스템은 개별적인 사용자보다는 가족단위의 분류에 중점을 두고 있으며 프로파일이 끊임없이 생성되는 프로그램 시청 정보에 즉각적으로 적용하지 못하는 단점이 있다.

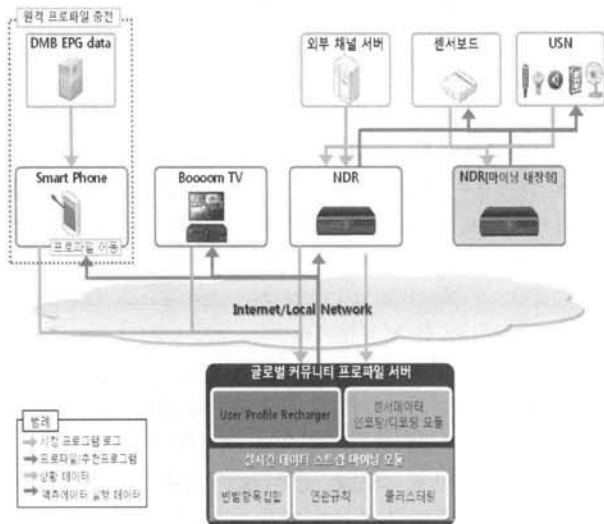
Conceptual Matching[8]은 사용자가 이전에 시청한 프로그램들로부터 학습 데이터를 구성하여 추천 프로그램을 제공하는 방법을 제안하였다. 그러나 이러한 관련 연구들에서는 실시간으로 생성되는 무한한 시청 정보를 즉각적으로 처리하여 추천 프로그램을 제공하는 기능을 수행할 수 없는 한계를 가지고 있다.

3. 시청 패턴 프로파일 및 시청 선호도 연관규칙

본 논문에서 제안하는 지 데이터 스트림 마이닝 기법을 적용한 개인/커뮤니티 맞춤형 Digital TV 시스템 의 전체 구성도는 다음 (그림 1)과 같다.

기존의 연구[12]에서는 (그림 1)의 마이닝 내장형 NDR(Network Digital Recorder)과 같이 글로벌 커뮤니티 서버와 연결하지 않고 NDR 자체에 마이닝 모듈이 포함되어 독립적으로 서비스를 제공하였으나, 이는 개인 사용자 환경에서만 적합하므로 가족과 같은 시청 단말기기의 공유환경이나, 커뮤니티별 맞춤 서비스와 모바일 기기를 사용하는 환경에서도 적용할 수 있도록 서버를 통해 상호 데이터 교환이 가능하도록 확장하였다.

또한, 사용자의 시청 시간대, 시청 장소별로 다양하게 프로파일을 생성하여 활용할 수 있는 방법을 제안하였다. 추가적으로, 다양한 센서로부터 시청 상황 정보를 입력 받아 추천 프로그램 시청과 동시에 적절하게 시청 상황을 제어할 수 있는 시스템을 제안하였다. 스마트폰에서는 DMB 방송 정보를 해당 휴대폰의API를 이용하여 추출하고, NDR 기기는 외부 채널 서버로부터 시청 채널 정보를 수신 받는다. 또한 시청공간에 설치된 센서들을 통하여 현재 시청 상황 정보를 수신 받게 된다. 각 단말에서DTV 방송 스트림, DMB 방송 스트림과 시청 상황 정보를 수신하면, 주요 키워드의 종류와 구조에 따라 필요 정보를 추출한 후, 글로벌 커뮤니티 프로파일 서버로 전송한다. 서버에서는 데이터 스트림 마이닝을 적용하여 사용자/커뮤니티별 시청 패턴/상황 프로파일 및 선호도 연관규칙을 생성한다. 각 프로파일들은 사용자 맞춤형 추천 프로그램을 생성하여 다시 각 단말에 전송한다. 또한, 선호도 연관규칙을 통하여 시청 프로그램 정보에 따른 시청 상황을 제공할 수 있도록 시청 상황 제어



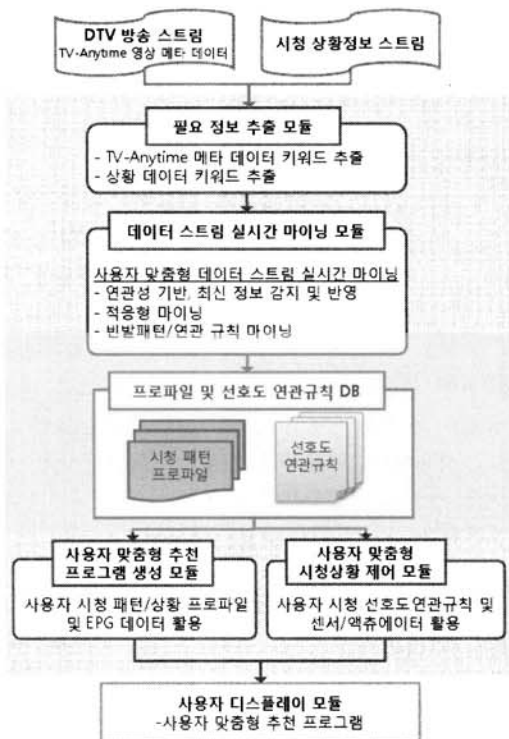
(그림 1) 전체 시스템 구성도

모듈에 결과를 전송한다. 또한 사용자가 스마트폰과 같은 모바일 기기를 통해 시청하는 경우에도 네트워크를 통하여 프로파일을 충전하고 이동할 수 있도록 제어한다.

3.1 영상 메타데이터 스트림 마이닝

영상 메타데이터의 실시간 마이닝을 위해서는 (그림 2)와 같은 단계를 거친다. DTV 방송 스트림과 시청 상황 정보 스트림이 실시간으로 생성되어 전송되면, 필요정보 추출모듈에서는 TV-Anytime의 프로그램 기술 메타데이터에 대한 전처리 과정을 거쳐 주요 키워드들이 추출된다. 즉, 사용자의 시청 선호도 프로파일 구성에 이용되는 프로그램의 제목, 방송 시간, 장르, 출연진, 제작진 등과 같은 데이터들이 추출되며, 영상 키워드 해쉬 테이블을 이용해 추출된 영상 키워드와 해당 키워드의 출현빈도수가 함께 관리된다. 데이터 스트림 형태로 제공되는 영상 메타데이터는 사용자가 시청한 한 프로그램의 기술 메타 데이터가 하나의 트랜잭션을 이루며 각 트랜잭션에 출현하는 키워드들을 각각 개별적인 항목으로 간주한다.

각 트랜잭션들을 입력 받는 영상 메타데이터 및 시청 상황 정보의 실시간 마이닝을 위해 이전 연구에서 제안된 온라인 데이터 스트림 환경에서 빈발항목집합을 탐색하는 *estDec*방법[9]을 사용하였다. *estDec*방법은 지연추가와 전지작업을 통하여 빈발항목집합이 될 가능성이 있는 항목집합들만을 관리한다. 데이터 스트림에서 출현한 모든 항목들은 해당 항목이 가까운 미래에 빈발항목이 될 수 있을 때까지 전



(그림 2) 영상 메타데이터/시청 상황 정보 실시간 마이닝

위트리에 대한 추가작업이 지연된다. 즉, 잠재적으로 중요한 항목이 충분히 큰 지지도를 갖는 적절한 시점에 전위트리 구조에 추가된다. 또한 효율적인 항목 전지 방법을 통해 이전 데이터 집합에서 중요한 항목이었다라도 해당 항목의 현재 지지도가 사전에 정의된 최소 지지도 보다 매우 작다면 해당 항목은 전위트리 구조로부터 전지된다. 이러한 방법을 통해 메모리에 관리되는 전위트리의 크기를 감소시킬 수 있다. DTV 방송 스트림에서 추출된 트랜잭션을 구성하는 키워드들은 전위트리의 각 노드에 관리된다. 사용자가 프로그램을 시청할 때마다, 이 노드들은 실시간으로 업데이트 과정을 거치며 사용자가 원하는 시점에 사용자의 빈발키워드집합을 생성하여 제공할 수 있는 장점을 갖는다. 즉, 이러한 데이터 스트림환경에서 빈발항목집합을 찾는 방법인 *estDec* 방법을 적용함으로써 누적되어 있는 시청 로그를 재탐색하는 과정을 거치지 않고 지속적으로 전위트리를 갱신하여 최신의 데이터가 반영된 결과를 즉각적으로 제공할 수 있다.

DTV 시청 로그 데이터 스트림이 실시간 마이닝 모듈을 통해 전위트리 구조에 관리되면, 빈발 프로그램/상황 키워드로부터 생성되는 빈발 프로그램/상황 키워드 집합을 탐색하여 관리한다. 이는 전위트리를 깊이 우선 탐색(*DFS: Depth First Search*) 방법으로 순회하며 사전에 정의된 사용자 임계값인 최소 지지도 이상의 값을 갖는 빈발 프로그램/상황 키워드를 추출하고 이로부터 생성되는 빈발 프로그램/상황 키워드 집합을 생성한다. 즉 (그림 2)와 같이 데이터 스트림 실시간 마이닝 모듈에서 각 마이닝 알고리즘 수행 후 결과 집합이 시청 패턴 프로파일과 선호도 연관규칙의 형태로 저장된다. 각 시청 패턴 프로파일과 선호도 연관규칙은 사용자 별/시간별/장소별로 구분되어 생성되며 한 트랜잭션이 입력 될 때마다 갱신 과정을 거쳐 가장 최신의 데이터가 반영된다.

3.2 사용자 시청 패턴/상황 프로파일

사용자 시청 선호도 프로파일은 사용자별 시청 선호도 특성 분석을 위해 영상 메타 데이터 및 시청 상황 정보 스트림 마이닝 구조에서 관리되는 빈발 프로그램/상황 키워드 집합으로부터 시청 패턴을 분석하여 생성된다. 또한, 그룹 사용자 그룹을 구성하는 단일 시청 선호도 프로파일로부터 공통된 빈발 영상 키워드 집합을 추출하여 그룹 사용자의 시청 선호도 특성을 반영하는 커뮤니티 시청 프로파일을 구축하는데 사용된다. 사용자간 선호도 거리란 단일 사용자 시청 선호 키워드 집합들간의 거리 계산 지표로서 [식 1]과 같이 계산된다.

$$Distance(A\ to\ B) = \frac{\sum_{e_i \in (A-B)} S_A(e_i) + \sum_{e_i \in (A \cap B)} |S_A(e_i) - S_B(e_i)|}{\sum_{e_i \in A} S_A(e_i)}$$

(0 ≤ Distance ≤ 1)

A : 사용자 A의 선호 키워드 항목집합

B: 사용자 B의 선호 키워드 항목집합

S(e_i): 선호 키워드 항목 e_i의 지지도

[식 1] 사용자간 선호도 거리 계산

생성된 프로파일은 EPG 데이터와 매칭하여 추천 프로그램을 생성한다. 즉 프로파일의 각 선호 키워드항목집합 지지도를 이용하여 현재 EPG 데이터와 얼마나 부합하는가를 나타내는 매칭 포인트(matching point)를 계산한다. 즉, 전체 N개의 EPG 데이터들 중 한 프로그램 데이터에 속하는 키워드 수를 *keyword(epg_i)* (1 ≤ i ≤ N) 라 하고, 프로파일을 구성하는 M개의 키워드 항목집합에 속한 키워드의 수를 *keyword(profile_j)* (1 ≤ j ≤ M) 라고 할 때 매칭 포인트는 다음과 같이 정의한다.

$$매칭포인트 = \frac{keyword(epg_i)}{keyword(profile_j)} * 100 * S(profile_j)$$

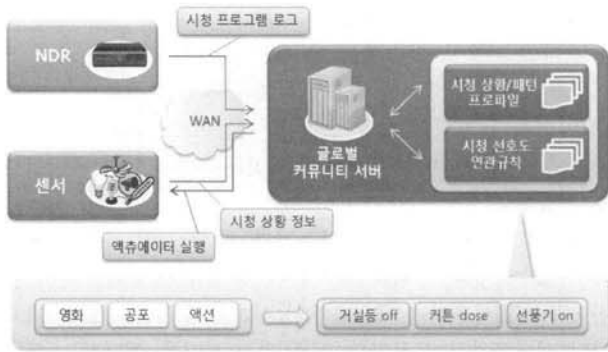
[식 2] 매칭 포인트

즉, 매칭 포인트가 높은 프로그램일수록 사용자의 프로파일들 중 가장 높은 지지도를 가지며 많은 수의 선호 키워드가 일치함을 나타내며, 매칭 포인트가 높은 순서대로 정렬하여 사용자에게 제공하여 맞춤형 서비스를 제공할 수 있다.

3.3 시청 선호도 연관규칙

시청 선호도 연관 규칙이란 시청 정보 안에 존재하는 키워드들 간의 종속 관계를 나타낸다. 시청 선호도의 지지도는 시청 프로그램 및 시청 상황 정보들 중 특정 키워드가 전체에서 차지하는 비율을 나타내며, 신뢰도란 시청 상황의 특정 입력 값들에 대해 시청 패턴 키워드가 발생할 확률을 나타낸다. 시청 선호도 연관규칙은 [10]에서 제안한 데이터 스트림 환경에서 연관규칙을 생성하는 기법을 활용하였다. 기존의 연관규칙을 생성하는 방법들은 다음의 두 가지 단계를 거친다; 1. 빈발항목집합 탐색, 2. 연관규칙 생성. 그러나 이러한 방법들은 모든 데이터를 저장하여 빈발항목집합을 생성한 후 저장하고 다시 이 데이터 집합에 접근하여 연관규칙을 생성하므로 데이터 스트림 환경에 적합하지 않다. [10]에서 제안한 방법은 별도로 빈발항목집합을 저장하지 않고 실시간으로 연관규칙을 생성할 수 있는 방법으로써, 지속적으로 추가 생성되는DTV 방송 스트림에 적합하다. 또한 주로 관련 연구들이 빈발항목집합을 효율적으로 찾는 것을 목적으로 하고 있어 비교알고리즘을 찾기 어렵다.

(그림 3)과 같이 DTV 단말과 센서로부터 전송된 키워드를 동기화하여 하나의 트랜잭션으로 구성한다. 모든 키워드들간의 연관규칙이 생성되면, (그림 3)의 예제와 같이 연관규칙의 조건부는 영화, 공포, 액션과 같은 시청 프로그램 키워드, 결과부는 거실등, 커튼, 선풍기와 같은 시청 상황으로 구성된 규칙만 남도록 필터링한다. 시청 선호도 연관규칙은 사용자가



(그림 3) 시청 선호도 연관규칙의 생성

시청하는 프로그램에 따라 액츄에이터들을 제어할 수 있도록 사용자 맞춤형 시청 상황 제어 모듈에 전달된다. 즉, 시청 프로그램과 시청시의 주변상황(온도, 조도, 소리 등)을 유기적으로 연결하여 보다 효과적인 프로그램 시청이 가능하다.

4. 글로벌 커뮤니티 실시간 마이닝

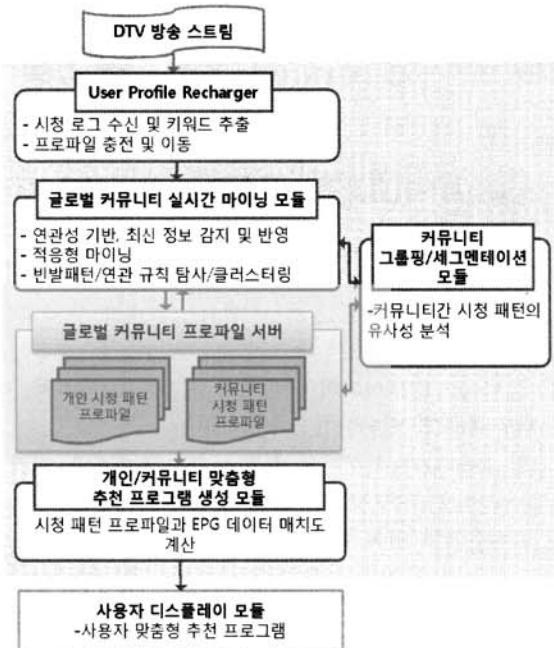
본 시스템은 3장에서 언급한 개인의 선호도가 반영된 개별 시청 상황/패턴 프로파일을 확장하여 개인이 속한 커뮤니티별 DTV 서비스를 제공한다. 즉, 시청 패턴 프로파일과 시청 선호도 연관규칙을 커뮤니티 기반으로 확대하기 위해 (그림 4)와 같이 글로벌 커뮤니티 서버를 고안하였다. 글로벌 커뮤니티 서버는 각 DTV 단말로부터 시청 로그와 시청 상황 정보를 전송 받아 실시간 데이터 스트림 마이닝을 거쳐 프로파일을 생성하고, 이들을 인덱싱 하여 관리한다.

생성된 각 프로파일을 현재의 EPG 데이터와 매칭하여 추천 프로그램을 생성하고 이를 DTV 단말로 전송하는 기능을 갖는다. 또한, 시청 선호도 연관규칙 생성 단계 역시 로컬이 아닌 서버에서 수행함으로써 각 단말이 갖는 부하를 줄이고 효율적으로 관리할 수 있다.

글로벌 커뮤니티 실시간 마이닝을 통한 커뮤니티 맞춤형 DTV 프로그램 추천은 (그림 5)와 같은 단계를 거친다. TV-Anytime 메타데이터로부터 추출된 프로그램의 주요 키워드들로 시청 로그가 구성되고 이러한 시청 로그는 글로벌



(그림 4) 글로벌 커뮤니티 서버



(그림 5) 글로벌 커뮤니티 실시간 마이닝

커뮤니티 서버내의 User Profile Recharger로 전송된다. User Profile Recharger는 시청로그 수신 및 생성된 프로파일을 DTV 단말로 전송하는 역할을 한다. 이 때, 글로벌 커뮤니티 서버내의 개인 사용자가 속한 커뮤니티 정보를 이용하여 개인 사용자의 시청 로그는 사용자가 속한 커뮤니티들의 시청 로그에도 전달된다. 글로벌 커뮤니티 마이닝을 통해 각자의 특성에 맞는 커뮤니티의 시간 및 위치별 시청 패턴/상황 프로파일이 생성된다. 즉, 한 커뮤니티의 시청 로그는 커뮤니티에 속한 개인들의 시청 로그의 집합으로 구성되며, 커뮤니티 추천 프로그램 생성 모듈은 이를 바탕으로 각 커뮤니티의 기호에 맞는 추천 프로그램을 제공한다. 즉, 사용자는 사용자 개인의 추천 프로그램 이외에도 자신이 속한 커뮤니티의 선호도를 파악하고 활용할 수 있다.

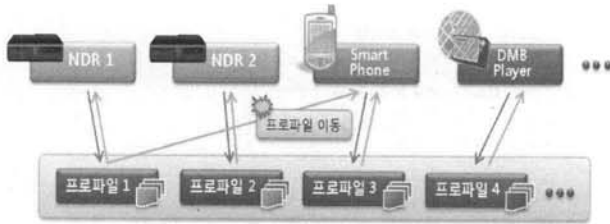
또한, 커뮤니티간 프로파일 유사성 검사를 통해서 커뮤니티를 그룹핑 하거나 하나의 커뮤니티를 나누어 보다 신뢰도가 높은 프로그램 추천 기능을 제공할 수 있는 커뮤니티 그룹핑/세그멘테이션 기법을 적용하였다.

5. 모바일 사용자 맞춤형 채널 추천 시스템

5.1 프로파일 이동 기법과 원격 프로파일 충전

모바일 사용자 맞춤형 채널 추천 시스템에서는 사용자의 이동시 시청 상황/패턴 프로파일 활용을 위해 프로파일 이동 기법으로 휴대할 수 있는 프로파일 이동 기법 및 인터넷을 통한 원격 프로파일 충전 기술을 적용하였다.

(그림 6)과 같이 프로파일 이동 기법을 통해 사용자의 현재 위치/시간이 변경되었을 경우의 변경된 위치/시간에서의



(그림 6) 프로파일 이동

프로파일이나 사용자가 원하는 위치/시간대의 프로파일을 글로벌 커뮤니티 서버로부터 이동하기로 가져와 활용할 수 있으며, 해당 기기 이외의 단말에서 생성된 사용자의 프로파일까지 활용이 가능하다.

(그림 7)은 원격 프로파일 충전 기술을 보여준다. 원격 프로파일 충전 기술은 실시간 스트림 마이닝 모듈을 통해 생성되는 사용자의 현재 위치/시간에서의 시청 로그를 글로벌 커뮤니티 서버에 전송하여 실시간 마이닝을 적용함으로써 갱신되는 시청 상황/패턴 프로파일을 주기적으로 모바일 기기로 전송하는 기술이다. 이를 통해 사용자들에게 고정된 DTV 단말기 뿐만 아니라, 모바일 기기의 시청 패턴도 분석이 가능하며, 언제 어디서나 원하는 프로그램을 서비스할 수 있다.



(그림 7) 원격 프로파일 충전

5.2 위치/시간 클러스터

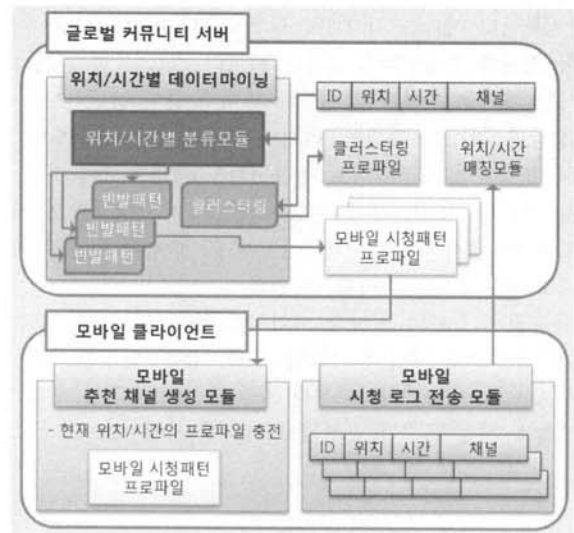
DMB 와 같은 모바일TV의 사용자 채널 시청 패턴은 시청 시간 또는 시청 위치 속성에 따라서 변화하기 때문에 시청 위치와 시간에 따른 프로파일의 생성이 필요하다. 따라서 본 시스템에서는 모바일 기기 클라이언트에 내장된 GPS의 좌표값 (x,y) 및 모바일 기기의 시스템 시간을 이용하여 위치/시간별 프로파일을 생성한다. 모바일 기기의 위치 및 시간 데이터는 글로벌 커뮤니티 서버내의 위치/시간별 데이터 마이닝 모듈에 입력 가능한 형태로 전처리 단계를 거친다. 클러스터링 모듈에서는 [11]에서 제안된 바와 같이 다차원 데이터 트랜잭션을 구성하는 각 항목들의 값이 사용자가 정의한 임계값 이하일 경우 클러스터를 생성하게 된다. 본 시스템에서는 위치와 시간의 2차원 데이터에 대하여 클러스터를 생성하고, 각 클러스터에 위치/시간 클러스터ID를 부여한다. 위치/시간 클러스터의 ID는 모바일 기기로부터 전송되는 시청 로그 데이터를 위치/시간별로 구분하여 마이닝하기 위해 분류하는 방법으로 사용된다.

5.3 위치/시간별 채널 추천 시스템

모바일 사용자 맞춤형 채널 추천 시스템은 (그림 8)과 같이 모바일TV가 내장된 클라이언트와 글로벌 커뮤니티 서버내의 데이터 전처리 모듈, 모바일 시청 로그 데이터 스트림 마이닝 모듈 및 마이닝 결과를 처리해주는 모듈과 유기적으로 연결된다.

모바일 클라이언트는 1)사용자가 현재 위치/시간에 시청하고 있는 채널 정보를 파악하여 글로벌 커뮤니티 서버로 전송하고 2)사용자가 채널을 이동하고자 할 때, 글로벌 커뮤니티 서버로부터 사용자의 현재 위치/시간에 해당하는 프로파일을 전송 받아 채널을 추천해주는 기능을 한다.

전처리된 모바일 시청 로그 데이터는 글로벌 커뮤니티 서버내의 클러스터링 프로파일과 매칭작업을 수행한다. 사용자가 전송하는 모바일 기기의 위치와 시간대에 따라 클러스터가 구성되며 각 클러스터별 ID를 부여한 뒤 위치/시간별 데이터 마이닝 모듈 내의 빈발패턴 생성 모듈[9]로 전달된다. 빈발패턴 생성 모듈을 통해 각 위치/시간에 맞는 빈발 시청 채널 패턴 프로파일이 생성된다. 원격 프로파일 충전 기술을 통해 이 프로파일들은 실시간으로 갱신되어 최신의 데이터를 유지한다. 또한, 프로파일 이동 기법을 사용하여 글로벌 커뮤니티 서버로부터 생성된 프로파일들 중 모바일 클라이언트의 위치/시간 조건에 맞는 프로파일을 모바일 클라이언트에 전송해준다.



(그림 8) 위치/시간별 채널 추천 시스템

6. 구현 및 성능평가

6.1 시스템 구현 및 디스플레이

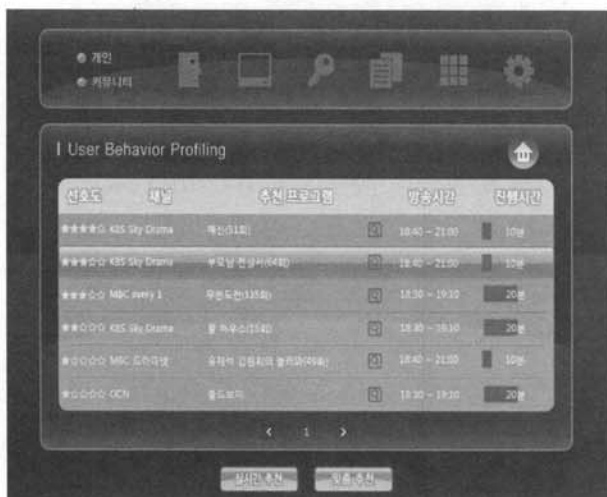
데이터 스트림 마이닝 기법을 적용한 개인/커뮤니티 맞춤형 Digital TV 시스템은 디지털 TV를 시청하는 단말에 적용되는DTV 사용자 인터페이스, 글로벌 커뮤니티 서버 관리

를 위한 User Profile Recharger와 모바일 기기를 위한 스마트폰 인터페이스로 구성된다. 본 시스템은 디지털 TV방송을 수신할 수 있는 settop box 또는 디지털 TV에 적용할 수 있도록 임베디드 시스템 기반으로 사용자 인터페이스를 구현하였다.

(그림 9)의DTV 사용자 인터페이스에서는 제목, 출연진, 개요 등과 같은 세부 프로그램 정보를 포함하는 전체 방송 프로그램 정보외에 각 사용자별 선호도순으로 정렬된 추천 프로그램 리스트가 제공된다. 사용자는 본인의 추천 프로그램 리스트외에도 커뮤니티의 추천 프로그램 리스트를 확인하고 시청할 수 있다. 사용자 인터페이스를 통해 프로그램을 시청하면, 시청한 프로그램은 다시 마이닝되어 추천 프로그램이 실시간으로 갱신된다. 또한, 글로벌 커뮤니티 서버로부터 시청 선호도 연관규칙을 전송 받아 현재 시청 프로그램과 적합한 시청 환경을 제어 한다.

DTV에서 발생한 시청로그는 (그림 10)과 같이 글로벌 커뮤니티 서버 내의 User Profile Recharger에서 수집되어 글로벌 커뮤니티 실시간 마이닝 모듈로 전달된다. User Profile Recharger는 글로벌 커뮤니티 서버를 관리하는 사용자 인터페이스로서 다음과 같은 기능을 가진다.

- 사용자 정보: 현재 서버에 등록된 사용자 와 사용자 가 소속된 커뮤니티에 대한 정보를 보여준다. 또한 프로파일 필터링을 통해 원하는 프로파일을 선택하여 보여 주는 기능을 제공한다.
- 실시간 시청 로그: 개인/커뮤니티별 시청 상황은 실시간 시청 로그 리스트를 통해 확인할 수 있다. 실시간 시청 로그 리스트는 프로그램을 시청 중인 주체 즉, 개인 또는 커뮤니티와 단말 정보, 시청자의 현재 위치 및 시간, 시청 중인 프로그램 정보만 선별하여 최신 정보순으로 표시한다. 실시간 시청 로그 리스트에 갱신된 목록은 실시간 마이닝 모듈을 통해서 개인/커뮤니티 별



(그림 9) 사용자 인터페이스(추천 프로그램)



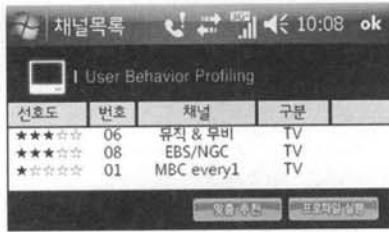
(그림 10) User Profile Recharger

로 마이닝된 후 해당하는 개인/커뮤니티의 시청 패턴 프로파일을 갱신한다. 실시간 마이닝을 통해 갱신된 개인/커뮤니티 프로파일은 프로파일 관리 모듈에 적용되며, 프로파일 갱신 목록에 등록된다.

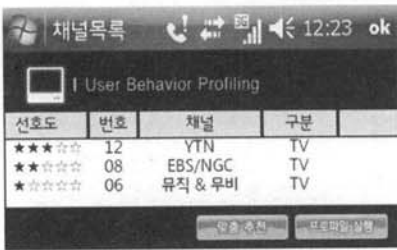
- 실시간 마이닝 옵션: 실시간으로 제공되는 개인/커뮤니티 프로그램 시청 로그를 데이터 스트림 실시간 마이닝 모듈에 전송하는 단위를 정의한다.
- 프로파일 관리: 각 개인/커뮤니티 프로파일은 시간대와 시청 위치별로 관리되며 EPG 데이터와 매칭하여 추천 프로그램을 생성한다. 개인/커뮤니티 프로파일 및 추천 프로그램 목록을 실시간으로 조회할 수 있는 기능을 제공하며, 각 단말에 프로파일을 제공하는 실시간 프로파일 이동 및 원격 프로파일 충전 기능을 제공한다.
- 선호도 연관규칙 관리: 선호도 연관 규칙 리스트는 커뮤니티 시청 상황/패턴 프로파일에서 프로그램 키워드와 시청 상황 키워드간의 상호 연관성을 지지도와 신뢰도와 함께 보여준다. 이러한 시청 선호도 연관규칙 중 시청 프로그램과 가장 부합하는 연관규칙이 각 DTV 단말에 전송되면 사용자가 선호하는 시청 상황을 제어할 수 있다.

스마트폰 사용자 맞춤형 채널 추천 시스템의 모바일 인터페이스는 삼성 T*OMNIA (Windows Mobile 6.1) C++로, 서버 어플리케이션은 JAVA로 개발하였다.

(그림 11)(a)는 오전 시간대(06:00~12:00)의 연세대학교 위치(126.93, 37.56)에서의 추천 리스트이고 (그림 11)(b)는 오후 시간대(12:00~18:00)의 서울시 방학동 위치(127.04,

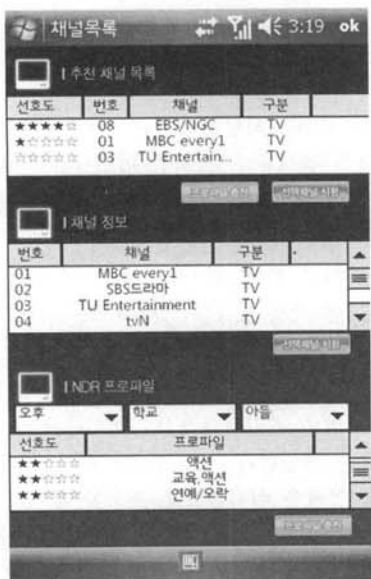


(a) 추천리스트(연세대, 오전)



(b) 추천리스트(방학동, 오후)

(그림 11) 스마트폰 추천 채널 리스트



(그림 12) 프로파일 이동

37.66)에서의 추천 리스트이다. 사용자의 위치/시간에 따라 추천 리스트가 달라지는 것을 확인할 수 있다.

(그림 12)는 스마트폰의 인터페이스를 통한 프로파일이동 기법이 적용된 모습이다. NDR 프로파일 항목은 사용자가 DTV settop box 에서 생성된 사용자의 프로파일을 이동기기로 이동한 것이다. 사용자는 이동기기를 통해서 생성된 DMB 채널 추천 서비스 외에도 다른 단말기를 통해 생성된 프로파일을 글로벌 커뮤니티 서버로부터 전송 받아 활용한다.

6.2 실험 결과 및 성능평가

디지털 방송은 TV-Anytime 영상 메타데이터가 수신되

어 EPG 데이터와 함께 제공되어야 하나, 현재 이러한 데이터를 얻기 어려우므로 인터넷에서 서비스 되는 EPG제공 사이트[16]에서 프로그램 정보를 추출하여 사용하였다. 총 105개의 채널로부터 <표 1>에 가정한 사용자들의 2개월간의 시청 로그를 생성하고, 1주일간의 EPG 데이터를 바탕으로 추천 프로그램 리스트를 생성하도록 하였다.

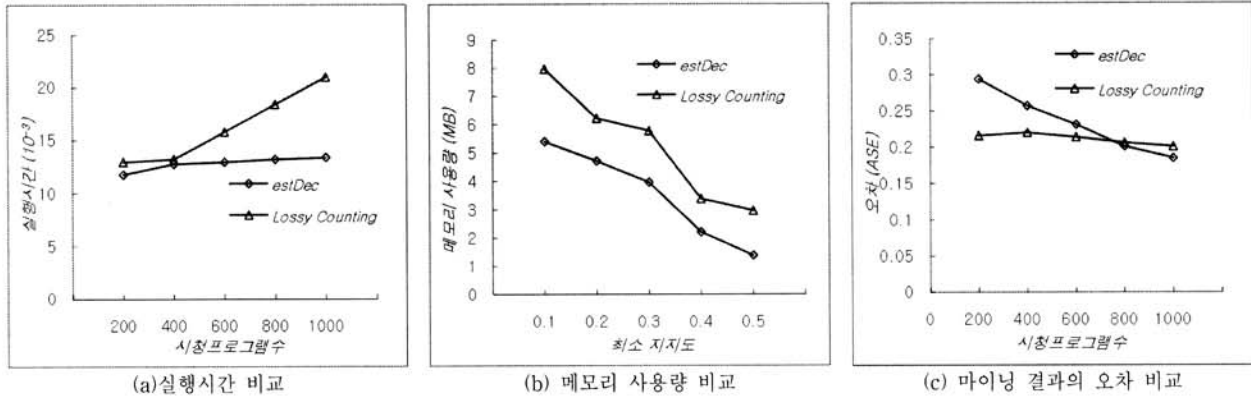
개인/커뮤니티 맞춤형 Digital TV 시스템은 각 DTV단말과 글로벌 커뮤니티 서버를 인터넷을 통해 연결하고, 사용자의 선호도에 기반하여 추천 프로그램들과, 추천 시청 환경을 제어하는 기능들을 제공한다. 이러한 기능을 제공하기 위하여 사용되는 데이터 스트림 마이닝 기법이 전체 시스템의 성능을 좌우하게 된다. 본 논문에서는 estDec방법과 이를 응용한 연관규칙 방법[10], 클러스터링 방법[11]이 사용되므로 시스템의 성능평가를 위해 데이터 스트림 환경에서 빈발항목 탐색을 위한 마이닝 방법인 Lossy Counting 방법[15]을 estDec 방법과 비교하였다.

(그림 13)(a)와 같이 실행 시간 측면에서는 두 알고리즘이 비슷한 성능 보이나 시청 로그가 증가할수록 estDec가 효율적이다. (그림 13)(b)는 최소 지지도의 변화에 따른 각 알고리즘의 메모리 사용량의 변화를 보인다. Lossy Counting 알고리즘은 버퍼의 크기와 일괄처리하는 트랜잭션의 수가 비례하므로, 버퍼의 크기가 증가함에 따라 높은 효율을 보이지만 결과적으로 빈발항목집합 탐색에 필요한 메모리 사용공간이 증가하는 단점이 있다. 이에 반해 estDec방법은 지연 추가와 전지작업을 통해 효율적으로 메모리를 관리하므로 상대적으로 적은 메모리를 사용한다.

(그림 13)(c)는 시청프로그램수의 증가에 따른 각 알고리즘들의 오차를 비교한 그래프이다. 제안된 방법들의 상대적 정확도를 표현하기 위해, [5]와 같은 방법으로 두 빈발항목 결과 집합에 대해 average support error (ASE)를 사용하였다. ASE값이 작을수록 비교하는 마이닝 결과 집합이 유사함을 나타낸다. 데이터의 양이 증가할수록 estDec 방법의 마이닝 오차가 안정화 되어감을 알 수 있다.

<표 1> 사용자별 시청 패턴 가정

사용자	시청 방송의 주요 키워드
사용자1	부동산, 경제, 증권, 주식 제테크, 뉴스, 국제정세, 경제동향, 정치, 생활, 날씨, 바둑, 스포츠, 골프, 축구, 프로야구, 해외스포츠
사용자2	드라마, 연예/오락, 홈쇼핑, 패션, 뷰티, 부동산, 경제, 증권, 주식, 제테크, 뉴스, 국제정세, 경제동향, 정치, 생활, 날씨
사용자3	드라마, 연예/오락, 교육, 수능, 영화, 뉴스, 국제정세, 경제동향, 정치, 생활, 날씨
사용자4	영화, 음악, 연예/오락, 스포츠, 축구, 프로야구, 해외스포츠, 컴퓨터게임, 교육
사용자5	관광/여행, 바둑, 뉴스, 국제정세, 경제동향, 정치, 생활, 날씨, 시사/다큐, 의료/건강



(그림 13) 데이터 스트림 마이닝 알고리즘의 성능

7. 결 론

본 논문에서는 맞춤형 디지털 방송 서비스를 제공하기 위해 TV-Anytime 메타데이터와 시청 상황 정보를 포함하는 사용자의 시청 이력을 입력받아 데이터 스트림 마이닝 기법을 적용하여 개인/커뮤니티 맞춤형 Digital TV 시스템을 개발하였다. 또한, 임베디드 시스템 기반의 사용자 인터페이스를 구현하여 사용자에게 적절한 추천 프로그램을 제공하고, 시청 프로그램 정보에 따른 시청 상황을 자동으로 제어하는 기능을 포함한다. 개인 시청자의 선호도 이외에도 커뮤니티 맞춤형 DTV 프로그램 추천 서비스를 제공하기 위해 사용자의 시청 이력을 커뮤니티별로 재구성하여 실시간 마이닝 기법을 적용하였다. 시청 패턴 프로파일과 EPG 데이터를 이용하여 개인뿐 아니라 커뮤니티의 기호에 맞는 적합한 추천 프로그램을 제공할 수 있는 시스템을 개발하였다.

또한, 사용자의 시청 채널 정보를 시간 및 위치 정보와 함께 분석함으로써 이동기기를 이용한 DMB 시청 환경에서 보다 적응적인 사용자의 시청 패턴을 파악할 수 있는 시스템을 구현하였다. 프로파일 이동 및 원격 프로파일 충전 기법을 적용하여 사용자의 프로파일의 활용도를 높일 수 있으며, 제안한 방법의 효율성과 정확성을 비교 실험을 통해 입증하였다.

참 고 문 헌

[1] TV-Anytime Forum, <http://www.tv-anytime.org/>
 [2] MPEG-7 Document, ISO/IEC 15938-5, "MPEG-7 Overview v.10" ISO/IEC JTC1/SC29/WG11N6828, 2004.
 [3] MPEG-21 Document, ISO/IEC 21000, "MPEG-21 Overview v.5," ISO/IEC JTC1/SC29/WG11/N5231, 2002
 [4] Diana Weiß et al. "A user profile-based personalization system for digital multimedia content" Proc. of the 3rd Intl. conf. on DIMEA, pp.281-288, 2008.
 [5] Ali K., Stam WV. "TiVo: making show recommendations

using a distributed collaborative filtering architecture," In Proc. of the 10th ACM SIGKDD Intl. conf. on Knowledge Discovery and Data Mining, pp.394-401, 2004.
 [6] Dai, W. and Cohen, R. "Dynamic Personalized TV Recommendation System", In Proc. of the UM 2003 Workshop on Personalization in Future TV, 2003.
 [7] Spangler W., Gal-Or M., Masy J., Using data mining to profile TV viewers-ACM CACM, December 2003.
 [8] Takagi, T, Kasuya, S, Mukaidono, M and Yanaguchi, T, "Conceptual matching and its applications to selection of TV Programs and BGMs," Systems, Man, and Cyber-netic, IEEE SMC. 99. Vol.3, pp.269-273, 1999.
 [9] J.H. Chang and W.S. Lee. "Finding recent frequent itemsets adaptively over online data streams," In Proc. of the 9th ACM SIGKDD Intl. conf. on Knowledge Discovery and Data Mining, pp.487-492, 2003.
 [10] S.J Shin and W.S. Lee. "On-line generation association rules over data streams", Information and Software Technology, Vol.50, Issue. 6 pp.569-578, 2008.
 [11] J.W.Lee, N.H.Park and W.S.Lee, "Efficiently Tracing Clusters Over High-Dimensional On-line Data Streams", Data & Knowledge Engineering, Vol.68, 2009.
 [12] 신세정, 이원석, "TV-Anytime 메타데이터 연속 데이터 마이닝을 이용한 시청 선호도 프로파일 생성 기법" 한국정보처리학회 학술대회, 2006.
 [13] 신세정, 이원석, "커뮤니티 기반 맞춤형 DTV 프로그램 추천 시스템" 대한전자공학회 추계종합학술대회, 2009.
 [14] R. Agrawal and R. Srikant. "Fast Algorithms for Mining Association Rules," In Proc. of the 20th Intl. Conf. on Very Large Data Bases, pp.487-499, 1994.
 [15] G.S. Manku and R. Motwani. "Approximate Frequency Counts over Data Streams," In Proc. of the 28th Intl. Conf. on Very Large Data Bases, pp. 346-357, 2002.
 [16] <http://www.epg.co.kr>



신 세 정

e-mail : starofu@database.yonsei.ac.kr
2004년 연세대학교 컴퓨터과학과(공학사)
2004년~현 재 연세대학교 컴퓨터과학과
석·박사. 통합과정
관심분야: 데이터베이스, 데이터스트림
마이닝



이 원 석

e-mail : leewo@database.yonsei.ac.kr
1985년 미국 보스턴대학교 컴퓨터공학과
(공학사)
1987년 미국 퍼듀대학교 컴퓨터공학과
(공학석사)
1990년 미국 퍼듀대학교 컴퓨터공학과
(공학박사)
1990년~1992년 삼성전자 선임연구원
1993년~1999년 연세대학교 컴퓨터과학과 조교수
1999년~2004년 연세대학교 컴퓨터과학과 부교수
2004년~현 재 연세대학교 컴퓨터과학과 교수
관심분야: 분산데이터베이스, 미디어이터시스템, 데이터마이닝,
데이터스트림