

# 모바일 환경에서의 효율적인 멀티미디어 콘텐츠 추천시스템 구조

홍종규<sup>o</sup>, 박성준<sup>\*</sup>, 김영국<sup>\*\*</sup>

(주)이노플러스<sup>o</sup>, 공주영상정보대학 멀티미디어콘텐츠전공<sup>\*</sup>, 충남대학교 컴퓨터학과<sup>\*\*</sup>  
jkhong@innoplus.com<sup>o</sup>, sjpark@cac.ac.kr<sup>\*</sup>, ykim@cnu.ac.kr<sup>\*\*</sup>

## Efficient Multimedia Contents Recommendation System in Mobile Environment

Jong-Kyu Hong<sup>o</sup>, Sung-Joon Park<sup>\*</sup>, Young-Kuk Kim<sup>\*\*</sup>  
Innoplus Co., Ltd.<sup>o</sup>

Major of Multimedia, Kongju Communication and Arts Collage<sup>\*</sup>  
Dept. of Computer Science, Chungnam National University<sup>\*\*</sup>

### 요 약

모바일 환경이 언제, 어디서든지 원하는 서비스나 콘텐츠에 접근할 수 있다는 장점에도 불구하고, 모바일 단말기는 여러 가지 취약한 단점들을 가지고 있다. DMB, 인터넷, 모바일 등에 대한 콘텐츠 또는 정보의 양이 거대하게 증가하면서 사용자는 때때로 자신이 원하는 콘텐츠를 찾는 데 어려움을 겪게 되며, 많은 시간을 소비하게 된다. 발전하는 모바일 환경 및 단말기의 장점을 최대한 이용할 뿐만 아니라 모바일 단말기가 가지는 제약 사항들의 한계를 극복하여 사용자가 원하는 정보 및 콘텐츠를 언제 어디서나 빠른 시간에 이용할 수 있는 모바일 환경에 적합한 추천시스템의 필요성은 증가하고 있다.

본 논문에서는 기존의 서버 중심의 추천시스템을 개선하여 클라이언트와 서버간의 데이터 교환을 통하여 추천 정확도를 높일 수 있는 추천시스템 구조를 제안한다. 제안하는 시스템은 사용자의 히스토리를 이용하여, 클라이언트에서는 모바일 단말기 사용자만의 히스토리를 이용한 추천 알고리즘을 적용하였고, 서버에서는 협업 필터링을 통해 다른 사용자의 히스토리를 이용한 추천 알고리즘을 적용하였다. 실험 결과는 서버 중심의 추천시스템 보다 더 높은 정확도를 제공할 수 있다는 것을 보여준다.

### 1. 서 론

무선 인터넷 환경과 모바일 단말기의 발전은 사용자의 증가와 함께 서비스의 종류 및 소비되는 정보의 양이 급격히 증가하게 하는 원인이 되었다. 현재 모바일 서비스에서는 교육, 게임, 정보, 운세, 뉴스 그리고 영화, DMB방송 같은 모바일 동영상 등의 콘텐츠를 제공할 뿐만 아니라 다양한 서비스가 이루어지고 있다. 특히 계속해서 서비스의 발전이 이루어지고 있는 멀티미디어 콘텐츠는 모바일 환경이 성장함에 따라 훨씬 많은 서비스 종류와 양을 가지게 될 것이다[1].

많은 양의 콘텐츠가 모바일 환경에서 서비스 되어지더라도 모바일 환경은 유선 환경과 비교하여 여러 가지 제약 조건을 가지고 있음에 틀림이 없다. 모바일 단말기는 작은 화면과 사용하기 불편한 인터페이스, 서버나 PC에 비해 메모리의 용량과 프로세스의 낮은 성능, 네트워크의 불안정성과 높은 비용, 배터리의 수명이 짧은 단점을 갖는 것이 바로 그것이다.

이와 같이 발전하는 무선 인터넷 환경 및 모바일 단말기의 장점을 최대한 이용할 뿐만 아니라 모바일 단말기가 가지는 제약 사항들의 한계를 극복하여 시간 변화에 따른 멀티미디어 콘텐츠 소비 성향이 변하는 환경에서 사용자가 원하는 정보를 언제 어디서나 빠른 시간에 이용할 수 있는 방법이 절실히 요구되어지고 있다.

그로 인해 불특정 다수에게 프로그램을 제공하는 기존의 방송 서비스 개념을 사용자가 선호하는 프로그램 중심으로 시청할 수 있도록 하는 개인화된 방송 서비스가 DMB 방송 시대의 주요 요소가 되고 있다. 개인화 방송은 고객이 선호하는 프로그램을 찾는 데 걸리는 시간을 줄여줄 수 있으며, 선호하는 프로그램을 찾는 동안 이미 방송이 진행되어 원하는 방송을 놓쳐버리는 경우를 줄여줌으로써 보다 편리한 서비스를 제공할 수 있다.

시간에 따라 제공되는 콘텐츠가 다양하게 변하는 환경에서 콘텐츠에 대한 고객의 향후 선호도를 예측하는 주요 실마리가 될 수 있는 시간 변화에 따른 고객의 콘텐츠 선호도 전이 정보문제를 해결하기 위한 많은 연구[2]에서 Markov 모델을 이용하였다. 전통적인 Markov 모델은 선호도 전이 정보만 이용하여 고객의 가까운 장래의 선호도를 예측하는데 이용되었다. 일반적으로 방송 콘텐츠를 이용하는 고객은 요일에 따라서 최근에 소비한 선호 콘텐츠를 다시 소비하는 경향이 있다. 이런 특성을 고려하여 Markov Model을 확장한 Weighted Markov Model은 사용자의 방송 콘텐츠에 대한 선호도를 예측하는데 높은 성능을 나타낸다.

현재 추천 기술을 이용한 모바일 환경의 멀티미디어 추천 시스템의 구조는 대부분 서버 중심적이다[3]. 클라이언트에서는 필요한 서비스를 선택하여 서버에 요청을 하게 된다. 서버에서는 클라이언트에서 요청한 추천 서비스를 해결하기 위해 추천 기술들을 이용하여 사용자의 선호도를 판단한다. 사용자의 선호도에 따라 콘텐츠의 우선순위를 결정하게 된다. 서버는 선호도 순위에 맞추어 콘텐츠의 리스트를 클라이언

\* 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 육성 지원 사업(IITA-2005-C1090-0502-0016)의 연구결과로 수행되었음.

언트에게 응답한다. 서버 중심적인 구조는 클라이언트의 수가 증가하게 되면 서버의 비용이 증가하게 되는 확장성(Scalability)의 문제를 갖게 된다. 또한 추천 서비스를 사용하기 위해서는 서비스 시마다 서버에 요청하게 되므로 통신 비용의 증가 및 서비스 지연의 발생 가능성이 존재하게 된다.

클라이언트에서 추천 알고리즘을 실행하고, 개인 사용자 데이터만을 이용해서 높은 정확도를 나타내는 Weighted Markov Model을 이용한 추천 시스템은 모바일 단말기의 특성상 데이터 희박성(Sparsity)의 문제가 발생할 수 있다. 모바일 단말기는 배터리의 용량도 적을 뿐만 아니라 연속적인 콘텐츠 소비가 어려운 특성을 가지고 있다. 그러므로 사용자의 콘텐츠 소비에 대한 기록인 히스토리가 부족하여도 높은 정확도를 얻을 수 있는 방법의 모색이 필요하다.

본 논문의 목적은 디지털 방송과 같이 시간 변화에 따라 고객의 방송 콘텐츠 선호도가 전이되는 환경에서 다음과 같은 특성을 지원하는 추천시스템을 제공하는데 있다. 첫째, 사용자 개인 히스토리가 콘텐츠를 중심으로 많은 연수가 이루어져 왔으며, 최근에는 디지털 방송의 프로그램이 많아짐으로 인해서 개인화에 대한 중요성이 더욱 더 강조되고 있다. 시청자는 수많은 프로그램에서 자신이 원하는 프로그램을 찾을 때마다 많은 시간이 필요로 한다. 따라서 개인화는 불필요한 정보는 제외하고 시청자에게 필요한 정보만을 제공함으로써 이를 극복할 수 있다.

2. 관련 연구

2.1 개인화

개인화는 고객이 원하거나 필요로 하는 정보를 제공하여, 이를 찾는 데 걸리는 시간과 비용을 절약해 주고, 손쉽게 접근하도록 사용자 선호도에 따라 동적으로 제공하는 것이다. 개인화는 웹 사이트를 중심으로 많은 연구가 이루어져 왔으며, 최근에는 디지털 방송의 프로그램이 많아짐으로 인해서 개인화에 대한 중요성이 더욱 더 강조되고 있다. 시청자는 수많은 프로그램에서 자신이 원하는 프로그램을 찾을 때마다 많은 시간이 필요로 한다. 따라서 개인화는 불필요한 정보는 제외하고 시청자에게 필요한 정보만을 제공함으로써 이를 극복할 수 있다.

개인화를 수행하는데 필요한 정보는 크게 묵시적인(Implicit Profile) 정보와 명시적인 프로파일(Explicit Profile) 정보로 구분할 수 있다. 묵시적인 프로파일 정보는 고객의 다양한 콘텐츠 소비 행위로부터 프로파일을 획득하거나 추천하는 정보이다. 명시적인 프로파일 정보는 고객으로부터 직접 입력 받은 정보를 의미한다.

2.2. Markov 모델

Markov 모델은 일반적으로 일련의 확률변수의 통계 값을 예측하기 위해 이용된다[4]. Markov 모델로부터 유도되는 기술은 시간에 따라 변하는 상태가 존재하며, 고객이 이미 수행된 연속적인 행위가 주어져 있으면, 고객이 다음에 행할 행위를 예측하기 위해 포괄적으로 이용된다[5].

Markov 모델은 <행위(A), 상태(S), 전이(T)> 3개의 파라미터로 표현한다. 여기서, 행위(A)는 고객에 의해 수행될 수 있는 모든 가능한 행위들의 집합이다. 상태(S)는 Markov 모델이 만든 모든 가능한 상태의 집합이다. Markov 모델의 전이 확률 행렬은 다음 행위를 예측하는데 이용된 사전 행위의 수에 종속적이다. 가장 단순한 Markov 모델은 고객의 의해 수행된 마지막 행위만 보고 다음 행위를 예측하는 것으로 1-order Markov 모델이라 한다. 이와 같은 방법으로 고객에 의해서 수행된 마지막 r개의 행위를 보고 예측도를 계산하는 r-order Markov 모델로 일반화 할 수 있다.

3. 제안 시스템

3.1 시스템 구조

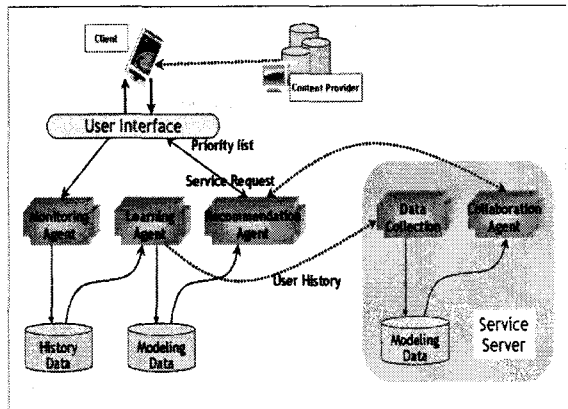
최근의 모바일 단말기의 저장 공간과 프로세싱 성능은 모바일 단말기가 서비스를 요청하고 보여주는 역할을 넘어서게 하고 있다. 멀티미디어 콘텐츠를 요청하여 모바일 단말기에 저장하고, 실행시킬 수 있는 환경이 되고 있다. 모바일 단말기의 성능 향상이 가져오는 변화는 멀티미디어 콘텐츠의 서비스가 가능하다는 측면뿐만 아니라 시스템 구조에 대한 변화를 야기하는 것이다.

기존의 멀티미디어 콘텐츠 추천 시스템에서 사용자의 선호도를 파악하기 위해서는 사용자의 히스토리를 서버에 저장하고, 저장된 히스토리를 기반으로 사용자에게 가장 적합한 멀티미디어 콘텐츠를 추천하였다. 하지만 모바일 단말기의 성능 향상은 모바일 단말기가 사용자의 콘텐츠 소비에 대한 히스토리를 저장할 수 있게 하였고, 추천 시스템도 모바일 단말기에 탑재할 수 있게 하였다. 추천에 필요한 선호도의 계산이 서버와 클라이언트 양쪽에서 모두 가능하게 된 것이다.

클라이언트에서 히스토리를 저장할 수 있고, 추천을 위한 선호도 계산이 가능하다면 모든 컴퓨팅을 클라이언트에서 하는 것이 가장 효율적이게 된다. 계산된 선호도를 통해 사용자에게 가장 적합한 멀티미디어 콘텐츠를 추천해주고, 사용자는 추천된 콘텐츠를 비교하여 선택하면, 클라이언트는 콘텐츠 제공자에게 원하는 콘텐츠를 요청하여 서비스 받을 수 있게 되는 것이다.

하지만 이러한 구조는 사용자의 히스토리가 적은 경우에는 사용자의 선호도를 예측하는데 어려움을 겪을 수 있다. 협업 필터링에서의 새로운 사용자에 대한 문제는 히스토리가 부족하여 선호도를 예측하기 어려운 상황과 비슷한 경우가 발생할 수 있는 것이다.

본 연구에서 제안하는 모바일 환경에서의 멀티미디어 콘텐츠 추천 시스템 구조는 [그림 1]과 같다.



[그림 1] 멀티미디어 콘텐츠 추천 시스템 구조

사용자의 히스토리가 부족하여 선호도를 예측하기 어려운 사용자일 경우와 선호도를 예측하기 충분한 히스토리를 갖는 사용자를 구분하였다.

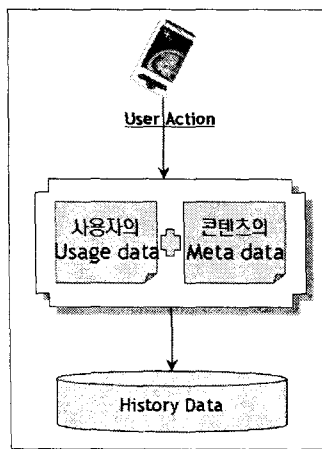
첫째, 사용자의 히스토리가 충분한 경우에는 클라이언트에서 사용자 개인의 히스토리만을 사용하여 Weighted Markov 모델을 이용하여 선호도를 예측하였다. 둘째, 사용자의 히스토리가 부족한 경우는 서버에 사용자의 프로파일과 현재 소

비하고 있는 히스토리를 보내고, 서버는 협업 필터링을 통해 얻어진 히스토리를 얻는다. 협업 필터링을 통해 얻은 히스토리는 Weighted Markov 모델을 이용하여 모델링한다. 최종적으로 모델링된 데이터를 클라이언트에 보내게 되고, 클라이언트는 모델링된 데이터를 수신하여 선호도를 계산함으로써 예측 정확도를 향상시킬 수 있게 한다.

제안하는 시스템에서 클라이언트에는 3개의 에이전트로 구성된다. 클라이언트에 위치하는 에이전트는 모니터링(Monitoring) 에이전트, 러닝(Learning) 에이전트, 추천(Recommendation) 에이전트가 있다. 또한 히스토리 데이터와 모델링 데이터의 2가지 데이터를 저장하게 된다. 모니터링 에이전트는 사용자의 행위에 의한 히스토리와 멀티미디어 콘텐츠를 제공하는 제공자에 의해 콘텐츠의 메타 데이터를 얻어 히스토리 데이터를 생성하고 저장한다. 사용자의 히스토리는 여러 가지 이유로 쓰레기 데이터를 가지고 있을 수 있기 때문에 수집된 데이터를 정제할 필요가 있다. 러닝 에이전트는 히스토리 데이터를 기반으로 히스토리를 스테이트(state)로 모델링한다. 사용자의 히스토리를 Weighted Markov 모델에 적용하기 위해서는 각각의 히스토리가 스테이트(State)가 되어야 한다. 그러므로 사용자의 히스토리를 스테이트화 하여 시간에 따른 순차적인 스테이트의 전이를 나타내어야 한다. 이 스테이트의 전이는 사용자의 멀티미디어 콘텐츠 소비 성향을 나타내는 데이터가 되는 것이다. 추천 에이전트는 시간에 따라 변하는 멀티미디어 콘텐츠의 소비 성향의 특징을 반영하기 위하여 모델링된 데이터를 이용한 Weighted Markov 모델을 이용하여 분석함으로써 사용자가 현재 소비중인 멀티미디어 콘텐츠 바로 다음에 선호하게 될 콘텐츠를 예측하여 추천한다.

3.2 모니터링 에이전트

모니터링 에이전트의 구조는 [그림 2]와 같다.

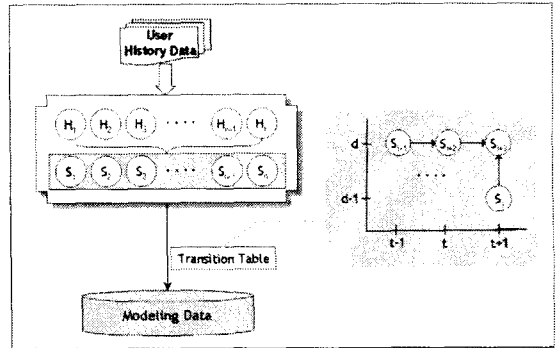


[그림 2] 모니터링 에이전트의 구조

사용자의 행위에 의해 발생하는 Usage 데이터와 멀티미디어 콘텐츠 제공자가 보내주는 콘텐츠의 메타 데이터를 기반으로 사용자의 히스토리 데이터를 구성한다. Weighted Markov 모델은 시간의 변화에 따른 사용자의 선호도 변이를 예측하기 위한 알고리즘이므로 사용자의 Usage 데이터는 콘텐츠를 소비하고 있는 시간에 대한 데이터를 갖게 된다. 콘텐츠의 메타데이터는 제목과 장르와 같은 콘텐츠의 특징을 설명해주

는 데이터를 갖고 있다.

3.3 러닝 에이전트

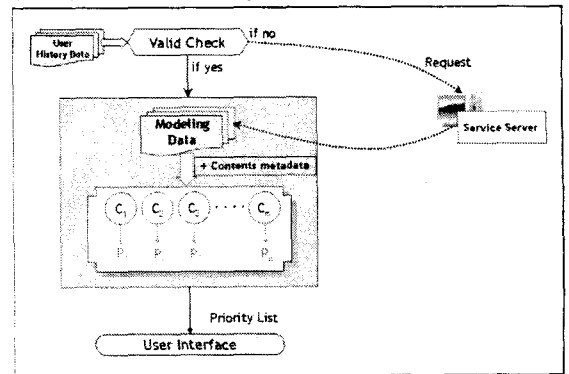


[그림 3] 러닝 에이전트의 구조

멀티미디어 추천 시스템의 러닝 에이전트 구조는 [그림 3]과 같다. 정제된 히스토리 데이터(H)는 Weighted Markov 모델 분석을 위한 스테이트(S)가 된다. 스테이트에서 스테이트로의 전이는 시간의 흐름에 따라 이루어지게 된다. 시간의 흐름에 따른 스테이트 전이를 테이블로 구성할 수 있고, 이 테이블을 저장한 것이 모델링 데이터이다. Weighted Markov 모델에서 가장 추천 정확도가 높은 모델은 2-order, 1-weight Markov 모델이므로 이 모델을 사용하여 전이 테이블을 구성하였다[6]. 구성된 전이테이블을 저장한 것이 모델링 데이터이다.

3.4 추천 에이전트

추천 에이전트의 구조는 [그림 4]와 같다.



[그림 4] 추천 에이전트의 구조

추천 에이전트는 아주 가까운 미래, 즉 바로 다음 시간에 사용자가 선호하는 콘텐츠를 추천하기 위한 분석 또는 훈련을 위하여 러닝 에이전트에서 사전에 정의된 기간 동안 수집되어 정제된 모델링 데이터와 사용자의 히스토리 데이터를 이용한다. 과거 히스토리 데이터에서 현재 시간 슬롯에서의 선호도로부터 바로 다음 시간 슬롯에서의 선호도로의 콘텐츠에 대한 사용자의 선호도 전이 정보는 가까운 미래에 선호하는 콘텐츠를 결정하기 위한 주요 요인이 될 수 있다.

과거 일정 기간 동안의 선호도 전이 누적 테이블, 즉 모델링 데이터로부터 우리는 [식 1]에서 정의된 것처럼 히스토리 데이터를 이용하여 시간 슬롯 t에서 Contents가 소비되었다

는 조건에서 모든 가능한 전이로 시간 슬롯  $t+1$ 에서 소비된 Contents의 빈도수에 대한 비율을 계산함으로써 시간 슬롯  $t$ 에서 시간 슬롯  $t+1$ 로 콘텐츠의 통계적 선호도 전이를 계산할 수 있다.

$$P(\text{contents}_{i,t}, \text{contents}_{j,t+1}) = \frac{T(i,j)}{\sum_{j=1}^n T(i,j)}$$

[식 1] 콘텐츠의 통계적 선호도 전이 계산

여기서  $\text{contents}_{i,t}$ 은 시간 슬롯  $t$ 에서 소비된 콘텐츠이며,  $\text{contents}_{j,t+1}$ 은 시간 슬롯  $t+1$ 에서 소비된 콘텐츠이다. 또한  $T(i,j)$ 는 시간 슬롯  $t$ 에서 소비된  $\text{contents}_i$ 로부터 시간 슬롯  $t+1$ 에서 소비된  $\text{contents}_j$ 로의 전이 빈도수이다.

일반적으로 사용자는 최근에 소비한 콘텐츠를 다시 소비하는 경향을 가지고 있다. 이와 같은 상황을 고려하기 위해서 본 논문에서는 [식 2]와 같이 통계적 전이 행렬을 계산하는데 2일간 계속해서 연이어 소비된 콘텐츠의 빈도수에 가중치를 부여한다. 모델링 데이터에 가중치를 부여한 값을 계산하는 것이다.

$$P_d(\text{contents}_{i,t}, \text{contents}_{j,t+1}) = \frac{\alpha_{d-1} \delta(j,i) T(i,j)}{\sum_{j=1}^n \alpha_{d-1} \delta(j,i) T(i,j)}$$

[식 2] 가중치를 적용한 선호도 전이 계산

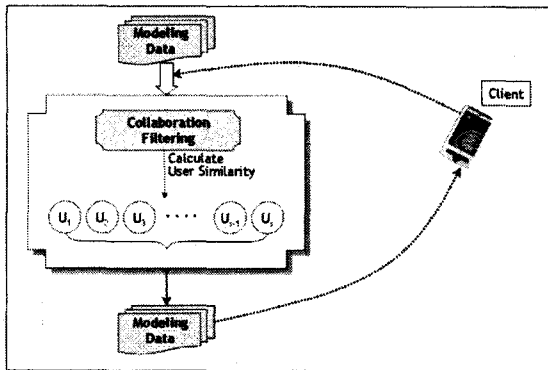
여기서,

$$\alpha_{d-1} \delta(j,i) = \left(1 + \frac{n_w(t+1)}{N_w(t+1)}\right)$$

여기서  $N_w(t+1)$ 은 히스토리 데이터에 속하는 전체 주(weeks)를 의미한다.  $n_w(t+1)$ 은  $\text{contents}_j$ 가 히스토리 데이터에서 연속해서 2일간 시간 슬롯  $t+1$ 에서 연속해서 소비된 횟수이다.

### 3.5 협업 에이전트

[그림 5]는 협업 에이전트의 구조를 나타낸다.



[그림 5] 협업 에이전트의 구조

각 클라이언트에서 Weighted Markov 모델을 세우기 위해

필요한 데이터가 부족할 경우 클라이언트는 서버에 협업 필터링을 요청하게 된다. 서버는 클라이언트들이 콘텐츠를 소비하는 히스토리가 생길 때마다 클라이언트로부터 사용자의 히스토리를 받아서 저장하게 된다. 만약 클라이언트로부터 요청이 생기게 되면 클라이언트 사용자와 다른 사용자들간의 유사도를 계산하여 가장 유사한 사용자들을 클러스터링한다. 클러스터링한 사용자들의 히스토리 데이터를 이용하여 모델링 데이터를 생성하고, 생성된 모델링 데이터를 클라이언트에게 전송한다.

유사한 사용자를 클러스터링 하는 방법은 2가지를 제시하였다. 첫째, 사용자의 프로파일의 성, 나이를 기준으로 사용자들을 클러스터링 하였다. 이 방법은 사용자의 프로파일 이 이미 정해져 있으므로 정적으로 사용자들을 클러스터링 할 수 있다. 프로파일을 기준으로 미리 유사한 사용자들의 데이터를 저장함으로써 클라이언트에서 요청이 생겼을 때 즉각적인 데이터를 전송해 줄 수 있다. 프로파일을 기반으로 하는 클러스터링은 클라이언트에서 오는 히스토리를 즉각적으로 저장할 있고, 요청된 데이터들도 데이터를 생성하기 위해 다른 컴퓨팅을 할 필요 없이 클라이언트에 전송할 수 있다. 그로 인해 프로파일 기반의 클러스터링은 서버의 컴퓨팅을 줄일 수 있는 장점이 있다.

둘째, 모델링 데이터를 요청한 사용자가 현재 소비하고 있는 콘텐츠와 유사한 콘텐츠를 소비하는 사용자들을 클러스터링한다. 현재 소비하고 있는 콘텐츠와 유사한 콘텐츠를 가지고 클러스터링하는 것은 시간이 흐름에 따라 변하게 되므로 첫째 방법과 다르게 동적으로 클러스터링 되어진다. 현재 소비하는 콘텐츠를 사용하여 사용자들 간의 유사도를 계산하게 되고, 유사한 사용자들의 히스토리 데이터를 사용하여 모델링 데이터를 생성한다. 생성된 모델링 데이터는 첫째 방법에서와 마찬가지로 클라이언트에게 전송하게 된다. 동적 클러스터링 방법은 현재 소비하고 있는 콘텐츠의 유사도를 이용하기 때문에 정적인 방법보다 더 정확한 유사도를 계산할 수 있을 것이다. 하지만 요청시마다 동적으로 클러스터링하게 되므로 클라이언트로부터의 요청이 동시에 많아지게 되면 서버의 컴퓨팅이 상당히 늘어나게 되는 단점을 가질 수 있다.

### 4. 실험 및 평가

본 논문에서 제안한 Markov 모델을 구현하고 시험하기 위한 시험 데이터로는 한국의 대표적인 시장 조사 기관 중 하나인 AC Nielsen Korea에 의해 2002년 12월 1일부터 2003년 5월 31일까지 2,518명의 TV 시청자로부터 수집된 3,199,990건의 TV 시청 데이터를 이용하였다. 시험 데이터가 가지는 각 프로그램은 8개의 장르, 즉 News, Entertainment, Drama & Movie, Information, Sports, Education, Children, Others로 구분한다.

실험 데이터로 이용된 전체 6개월 데이터 중 처음 5개월 데이터는 훈련 데이터로 이용하고, 나머지 1개월 데이터는 테스트 데이터로 이용하였다. 실험 데이터는 오후 7시부터 12시까지 시청한 데이터를 1시간 간격으로 이용하였다. 이 시간 외의 데이터는 시청 빈도수가 매우 적기 때문에 제외하였다. 시험 데이터에 대한 시청자의 선호 장르를 예측하거나 추천하기 위한 초적의 파라미터를 찾기 위한 훈련 데이터는 길이의 일관성을 위해 시청자의 TV 시청 데이터 중 가장 오래된 날의 데이터는 제거하고 대신에 최근에 실험된 날의 데이터가 훈련 데이터에 포함된다.

성능 평가 모델은 다음의 3가지로 구성하였다. 첫째, 서버 중심의 추천 시스템 구조이다. 이 구조는 클라이언트에서 받

생하는 모든 히스토리 데이터를 서버에 전송한다. 서버는 모든 사용자들의 히스토리 데이터를 저장하고 있으면서, 클라이언트에서 다음 멀티미디어 콘텐츠에 대한 추천을 요구할 때마다 협업 필터링을 통해 우선순위가 높은 순으로 클라이언트에게 콘텐츠 리스트를 클라이언트에게 전송한다. 이 구조는 클라이언트의 콘텐츠 추천 요구에 대해서 항상 협업 필터링을 사용하여 선호도를 예측하게 된다. 둘째, 클라이언트 중심의 추천 시스템 구조는 클라이언트에서 발생하는 개인 사용자의 히스토리 데이터를 개별 저장하여 저장된 히스토리 데이터만을 사용한다. 개인 사용자의 데이터를 사용하므로 서버에 히스토리 데이터를 전송하거나, 콘텐츠 추천을 요구할 필요가 없다. 따라서 서버의 컴퓨팅과 저장 공간이 필요하지 않으므로 클라이언트의 수가 증가하더라도 쉽게 확장하여 서비스할 수 있는 시스템 구조이다. 셋째, 본 논문에서 제안하는 추천 시스템 구조는 클라이언트 중심의 추천 시스템 구조를 기반으로 한다. 기본적으로 사용자의 히스토리 데이터를 클라이언트에 저장하며, 저장된 데이터를 사용하여 콘텐츠를 추천한다. 만약 사용자의 히스토리 데이터가 추천 정확도를 현저하게 저하시킬 수 있을 정도로 적은 데이터를 갖게 되면 클라이언트는 콘텐츠 추천을 서버에게 요청하게 된다. 서버 중심의 추천 시스템 구조에서와 마찬가지로 서버는 협업 필터링을 통해 클라이언트에게 선호도가 높은 순으로 클라이언트에게 콘텐츠 리스트를 전송한다.

성능 평가 절차는 첫 번째, 각 사용자들의 데이터 비율을 계산한다. 데이터 비율은 콘텐츠의 소비 히스토리의 비율이다. Weighted Markov 모델에서 스테이트는 히스토리이고, 스테이트의 전이 빈도를 계산하기 위해서는 콘텐츠를 소비한 히스토리가 존재해야 한다는 것을 의미한다. 히스토리가 존재하지 않는 경우에는 전이 행렬을 계산할 수 없기 때문에 전체 히스토리를 기준으로 히스토리의 비율을 계산한 것이다. Weighted Markov 모델을 빌당하기 위한 5개월 동안의 데이터를 기준으로 했을 때, 매주의 같은 요일마다 모델을 빌당하기 위해 필요한 히스토리가 존재하는 것을 100%로 한다. 두 번째는 2가지의 협업 필터링의 정확도를 계산한다. 서버에서 이루어지는 협업 필터링을 프로파일을 이용한 성, 나이 기준으로 필터링을 했을 때와 현재 소비하고 있는 콘텐츠의 유사도를 기준으로 했을 때의 정확도를 계산한다. 2가지 협업 필터링의 차이점은 정적으로 필터링을 하는 것과 동적으로 필터링을 하는 것이다. 데이터 비율에 따른 2가지 협업 필터링의 정확도를 계산함으로써 효율적인 서버의 추천 알고리즘을 평가할 수 있을 것이다. 세 번째는 성능 평가 모델에서 제안한 3가지 추천 시스템 구조에서의 정확도를 계산한다. 데이터 비율에 따른 추천 정확도를 계산함으로써 데이터 비율에 따라 클라이언트와 서버에서 추천 알고리즘을 계산해야 할지를 결정할 수 있을 것이다.

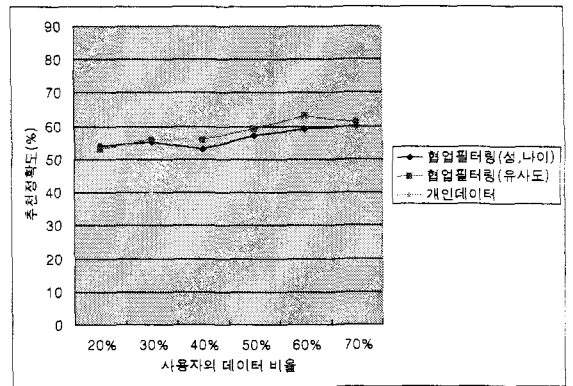
정확도는 다음과 같이 계산할 수 있다.

$$\text{정확도} = \frac{\text{실제 선호 콘텐츠 수} \cap \text{추천 콘텐츠 수}}{\text{실제 선호 콘텐츠 수}}$$

성능 분석은 성능 평가 절차에 따라 서버 중심의 추천 시스템 구조와 클라이언트 중심의 추천 구조, 그리고 본 논문에서 제안하는 추천 시스템 구조를 데이터 비율에 따라 정확도를 비교하였다.

[그림 6]은 데이터 비율에 대한 협업 필터링과 개인 데이터를 이용한 모델링의 정확도를 나타낸 것이다. [그림 6]에서 보듯이 2가지 협업 필터링의 방법에서는 현재 소비하고 있는 콘텐츠의 유사도를 이용한 방법이 약 2% 정도의 높은 정확도를 나타내었다. 2가지의 협업 필터링 모두 데이터 비율이

20%이하이면 약 53% 정도의 추천 정확도를 나타내었고, 데이터 비율이 60%~70%인 경우에는 약 60% 정도의 추천 정확도를 나타내었다. 협업 필터링을 이용한 추천 정확도에서는 약간의 차이는 있지만 대체로 비슷한 정확도를 나타내는 것을 볼 수 있다. 이는 다른 유사한 다른 사용자들이 선호하는 콘텐츠를 예측하여 추천해 주는 것은 데이터 비율에 상관없이 비슷한 추천 정확도를 나타낸다는 것을 알 수 있다. 즉, 협업 필터링을 사용하면 데이터 비율이 낮은 사용자에게 높은 정확도를 갖는 추천을 할 수 있다는 의미이다.



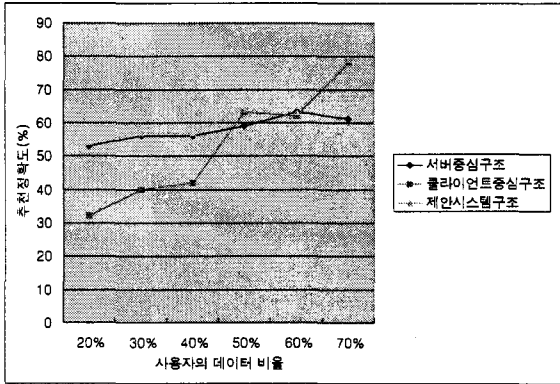
[그림 6] 협업 필터링과 개인 데이터를 이용한 추천 정확도

개인 데이터를 이용한 방법에서는 데이터 비율에 크게 의존하는 추천 정확도를 나타내었다. [그림 6]에서 보는 바와 같이 데이터 비율이 20%인 사용자는 약 30% 정도의 추천 정확도를 나타내었고, 데이터 비율이 60%~70%인 사용자는 거의 80%에 이르는 추천 정확도를 나타내고 있다. 따라서 [그림 6]은 사용자가 콘텐츠를 소비한 히스토리가 적으면 적용수록 개인 히스토리를 사용한 추천 시스템은 낮은 추천 정확도를 나타낸다는 것을 보여준다. 반대로 콘텐츠를 소비한 히스토리가 많으면 개인 히스토리를 사용한 추천 시스템은 높은 추천 정확도를 나타내는 것을 알 수 있다.

협업 필터링을 한 경우와 개인 데이터를 사용한 경우를 비교해 보면 데이터 비율이 40%이상일 경우에 협업 필터링을 사용한 경우와 개인 데이터를 사용한 경우가 비슷한 추천 정확도를 나타내는 결과를 얻었다. 그러므로 콘텐츠를 소비한 히스토리의 데이터 비율이 40%이하인 경우에는 서버에 추천 리스트를 요청하여 협업 필터링을 이용한 추천 알고리즘을 사용하고, 데이터 비율이 40%이상인 경우에는 클라이언트에서 개인 데이터를 이용한 추천 알고리즘을 사용하는 것이 높은 정확도를 얻을 수 있음을 알 수 있다.

특히, 개인 데이터를 이용한 추천 알고리즘에서 대체로 데이터 비율이 낮은 경우에 낮은 정확도를 나타내었으나 40%~50%의 데이터 비율을 갖는 사용자의 평균이 50%~60%의 데이터 비율을 갖는 사용자들의 평균 추천 정확도보다 높음을 보이고 있다. 이는 40%~50%의 데이터 비율을 갖는 사용자 중에서 추천 정확도가 90%인 사용자가 다른 데이터 비율 구간에서보다 많기 때문임을 데이터 분석을 통해 나타냈다.

[그림 7]은 서버 중심의 추천 시스템 구조와 클라이언트 중심의 추천 시스템 구조, 그리고 본 논문에서 제안하는 추천 시스템 구조의 데이터 비율에 대한 정확도를 나타낸 것이다.



[그림 7] 평가 모델들의 추천 정확도

본 논문에서 제안하는 추천 시스템 구조는 데이터 비율이 개인의 데이터를 사용하여도 높은 정확도를 가질 수 있는 사용자에 대해서는 클라이언트에서 추천 알고리즘을 실행하고, 사용자의 콘텐츠 소비에 대한 히스토리가 부족하여 정확도가 떨어질 때는 서버에서 협력 필터링을 이용하여 추천을 함으로써 정확도를 높일 수 있게 하였다.

데이터 비율이 0%~40%인 사용자는 서버에서 협력 필터링을 이용하여 추천하게 되므로 개인 데이터를 이용했을 때의 30%~40%이었던 정확도를 약 52% 정도로 추천 정확도를 향상시킬 수 있었고, 데이터 비율이 40%~70%인 사용자들은 협력 필터링을 이용했을 때의 약 60%의 정확도를 평균 70% 이상으로 추천 정확도를 향상시킬 수 있었다.

5. 결론 및 향후 연구 과제

본 논문에서는 모바일 환경에서 개인화된 멀티미디어 콘텐츠를 제공하기 위한 방법으로 Weighted Markov 모델을 이용한 효율적인 추천 시스템을 제공하였다. 추천 시스템은 클라이언트에 개인 사용자의 히스토리를 기반으로 추천 알고리즘을 실행하되, Weighted Markov 모델을 세우기 위해 필요한 데이터의 비율이 낮아서 정확도가 떨어지는 경우에는 서버에서 추천 알고리즘을 실행하도록 하였다. 서버에서는 다른 사용자들의 히스토리를 이용하여 협력 필터링을 하였다. 추천 시스템의 구조는 클라이언트에서 모니터링 에이전트, 러닝 에이전트, 추천 에이전트를 중심으로 실행되고, 서버에서는 클라이언트의 히스토리 데이터를 모으는 데이터 수집과 협력 필터링 에이전트가 실행되도록 하였다.

본 논문에서는 사용자의 히스토리 데이터 비율에 따라 추천 알고리즘 실행을 클라이언트와 서버로 나누었다. 이 방법은 서버 중심의 추천 시스템 구조보다 평균 10%이상의 추천 정확도를 향상시켰으며, 서버 중심의 추천 시스템 구조보다 8%이상의 추천 정확도를 향상시켰다. 또한 클라이언트에서 실행되는 것을 기본으로 하고 있으므로 히스토리 데이터가 충분한 경우 추천을 위해 필요한 시간이 아주 적으므로 서비스 지연이 거의 발생하지 않았다. 또한 실제 시청한 데이터를 이용하여 우리의 방법이 서버 중심의 추천 시스템 구조와 클라이언트의 추천 시스템의 정확도 보다 효과적임을 보임으로써 실질적으로 유용할 수 있음이 입증되었다.

하지만 모바일 특성상 사용자의 히스토리가 적어지면 서비스 요청시마다 서버에 요청하게 되므로 서버 중심의 추천 시스템 구조와 차이점이 발생하지 않을 수 있는 단점이 존재한

다.

6. 참고 문헌

- [1] 김용한, "지상파 DMB 서비스", 정보처리학회지 제 11권 제 5호, 2004. 9
- [2] Xing Dongshan and Shen Junyi, "a New Markov Model for Web Access Prediction", Computing in Science & engineering, November-December 2002
- [3] W.-P. Lee, J.-H. Wang, "A User-Centered Remote Control System for Personalized Multimedia Channel Selection", Contributed Paper, 2004. 7
- [4] Athanasios Papoulis, "Probability, Random Variables, and Stochastic Process", McGraw Hill, 1991
- [5] Mukund Deshpande, George Karypis, "Selective Markov Models for Predicting Web-page", ACM Transactions on Internet Technology(TOIT), Vol. 4, Issue 2, May 2004
- [6] Dongshan, X., Junyi, S.: Link prediction and path analysis using Markov chains. Computing in Science & Engineering (2002) Vol. 4, No. 6, 34-39