

에이전트 기반의 개인화된 TV 프로그램 추천 시스템

홍종규⁰, 박원익^{*}, 김룡^{**}, 김영국^{**}

충남대학교 컴퓨터공학과^{*}, 충남대학교 컴퓨터공학과^{**}

{jkhong⁰, wonik78, ryong, ykim}@cnu.ac.kr

Agent-based Personalized TV Program Recommendation System

Hong Jong-Kyu⁰, Park Won-Ik^{*}, Kim Ryong^{**}, Young-Kuk Kim^{**}

Dept. of Computer Engineering, Chungnam National University^{*}

Dept. of Computer Science, Chungnam National University^{**}

요약

디지털 방송이 시작되면서 시청자가 선택할 수 있는 채널은 200여 개로 늘어났다. 지금처럼 리모컨으로 채널을 돌리거나 보거나 원하는 TV 프로그램을 찾기란 거의 불가능해진 것이다. 이러한 다채널 다매체 시대에 원하는 프로그램 시청을 도와줄 수 있는 프로그램 가이드 시스템의 필요성이 증가하게 되었고, 더 나아가 TV를 시청하는 각 개인의 선호도를 반영하는 것이 요구되었다.

본 논문에서는 r-order Markov Model을 이용한 개인화된 전자 TV 프로그램 추천 시스템을 제안한다. Markov Model은 시간이 지남에 따라 시청하는 프로그램의 변화를 모델링하기 위한 방법으로 사용하였다. 이 시스템은 시청자의 선호 프로그램을 예측하기 위해서 r-order Markov Model을 제안하는 것뿐만 아니라 TV 시청자의 프로그램 선호를 예측하기 위한 모델들을 적용하였다. 실험 결과는 Markov Model이 추천에 대한 높은 정확성을 제공할 수 있다는 것을 보여준다.

1. 서론

디지털 TV 채널 정보의 양이 크게 증가하면서 시청자는 때때로 자신이 원하는 프로그램을 찾는데 어려움을 겪게 되며, 많은 시간을 소비하게 된다. 디지털 방송은 기존의 방송보다 채널수가 비교할 수 없을 정도로 많기 때문에 TV 시청자들은 때로 자신이 선호하는 TV 프로그램을 찾으려고 검색하는데 많은 시간을 허비한다.

불특정 다수에게 TV 프로그램을 제공하는 기존의 방송 서비스 개념을 시청자가 선호하는 TV 프로그램 중심으로 시청할 수 있도록 하는 개인화된 방송 서비스가 디지털 방송의 주요 요소가 되고 있다. 개인화 방송은 고객이 선호하는 프로그램을 찾는데 걸리는 시간을 줄여줄 수 있으며, 선호하는 프로그램을 찾는 동안 이미 방송이 진행되어 원하는 방송을 놓쳐버리는 경우를 줄여줌으로써 보다 편리한 환경을 제공할 수 있다.

지금까지 연구되고 있는 다양한 추천 기술들은 크게 3가지 범주, 즉, 규칙기반 필터링, 협업 필터링, 추론 방법으로 분류할 수 있다. 규칙 기반 필터링 기술은 사전에 이미 결정된 규칙에 의해 규칙에 맞는 콘텐츠를 추천하게 된다[1]. 협업 필터링 기술은 콘텐츠 소비 성향이 고객의 성향과 비슷한 그룹이 선호하는 콘텐츠를 목표 고객에게 추천하는 기술이다[2][3]. 추론 방식은 고객 개인의 콘텐츠 소비 행위 기록을 기반으로 고객의 콘텐츠 소비 행위를 예측할 수 있는 기술이다[4]. 이와 같은 기술들의 대부분은 고객이 시스템에 직접 입력하는 명시적인 고객 프로파일 또는 고객의 과거 소비 행위를 이용한다. 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 정보는 이용되지 않고 있다.

* 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 육성·지원 사업의 연구 결과로 수행되었음.

시간 변화에 따른 고객의 콘텐츠 선호도 전이 정보문제를 해결하기 위한 많은 연구[5]에서 Markov 모델을 이용하였다. Markov 모델은 선호도 전이 정보를 이용하여 고객의 가까운 장래의 선호도를 예측하는데 이용되었다.

본 논문의 주요 연구 내용은 시간 변화에 따른 고객의 선호도 전이를 반영한 프로그램 추천 방법을 제시한다. 시간 변화에 따른 고객의 선호도 전이를 반영하기 위한 방법으로 Markov 모델 제안한다. 제안된 Markov 모델은 비교적 적은 양의 시청자의 시청 정보만을 이용하여 실시간으로 개인화 추천 서비스가 가능하

다. 기존에 연구된 대부분의 개인화 추천 방법들은 시간 변화에 따른 고객의 선호도 전이를 반영하지 않고 있다. 따라서 본 논문에서는 시간 변화에 따른 고객의 선호도 전이를 반영하여 TV 시청 중 사용자 원하는 프로그램을 실시간으로 추천하기 위한 알고리즘을 제안한다.

2. 관련 연구

2.1 개인화

개인화는 고객이 원하거나 필요로 하는 정보를 제공하여, 이를 찾는데 걸리는 시간과 비용을 절약해 주고, 손쉽게 접근하도록 사용자 선호도에 따라 동적으로 제공하는 것이다. 개인화는 웹사이트를 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 추론하는 정보이다. 명시적인 프로파일 정보는 고객으로부터 직접 입력 받는 정보를 의미한다.

2.2. Markov 모델

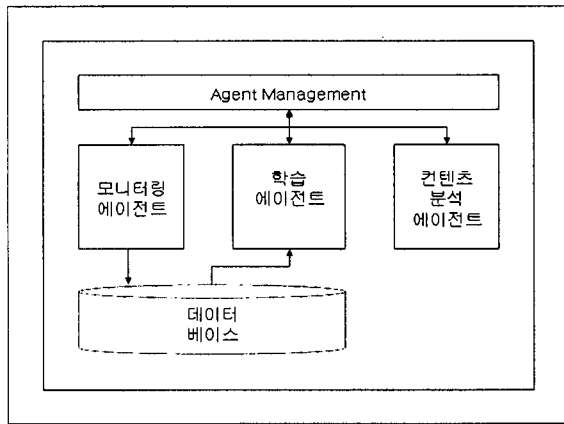
Markov 모델은 일반적으로 일련의 확률변수의 통계 값을 예측하기 위해 이용된다[6]. Markov 모델로부터 유도되는 기술은 시간에 따라 변하는 상태가 존재하며, 고객이 이미 수행된 연속적인 행위가 주어져 있으면, 고객이 다음에 행할 행위를 예측하기 위해 포괄적으로 이용된다[7].

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 일반화할 수 있다.

3. 제안 시스템

3.1 시스템 구조

개인화된 프로그램 추천 시스템을 구현하기 위해서, 제안하는 시스템은 3개의 에이전트로 구성되어 있다. 모니터링 에이전트는 사용자의 선호도를 예측하기 위하여 사용자가 시청하는 프로그램과 시간을 데이터베이스에 저장한다. 콘텐츠 분석 에이전트는 각 프로그램을 정해진 데이터베이스 구조로 변환한다. 학습 에이전트는 사용자의 선호도를 분석하고, 모델링한다. 제안하는 프로그램 추천 시스템 구조는 [그림 1]과 같다.

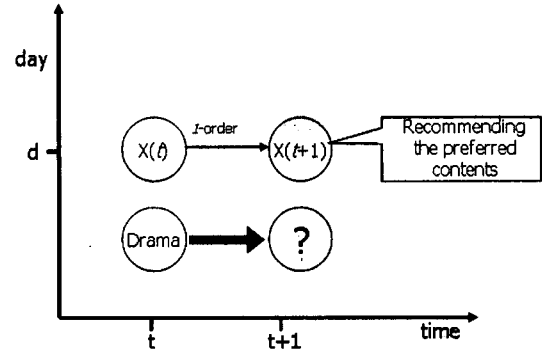


[그림 1] 프로그램 추천 시스템 구조도

3.2 추천 알고리즘

본 시스템에서는 사용자의 선호도와 프로그램의 정보를 분석하여 사용자에게 프로그램을 추천한다. 사용자의 선호도는 Markov 모델을 사용하여 계산한다. Markov 모델은 r-order의 통계적 전이 행렬을 사용하여 사용자가 현재 시청하는 프로그램으로부터 다음에 어떤 프로그램을 시청할 것인지를 예측한다.

1-order Markov 모델은 [그림 2]와 같이 현재 시간 t 시간에 소비하고 있는 프로그램의 종류를 알고 있을 때, 과거 프로그램의 시청 성향 전이 행렬을 이용하여 바로 다음 시간 t+1 시간에 어떤 종류의 프로그램을 시청하게 될지 선호 프로그램을 예측하기 위한 모델이다. 과거 프로그램 시청 성향 전이 행렬은 [표 1]과 같이 나타낼 수 있다. [그림 2]에서 원으로 표시된 것은 소비 성향의 상태를 의미한다.



[그림 2] 1차 Markov 모델

[표 1]은 t 시간과 t+1 시간 사이에 시청한 TV 프로그램 장르에 대한 전이 행렬을 표로 나타낸 것이다. 1차 Markov 모델을 이용하여 현재에서 바로 다음에 무엇을 할 것인지를 예측하기 위해서는 [표 1]과 같은 고객의 과거 통계적 전이 행렬을 이용하여 고객의 선호 프로그램을 예측할 수 있다.

[표 1] 1-order Markov 모델의 TV 시청 장르 전이 예

t \ t +1	Drama	Sports	News
Drama	6	9	7
Sports	5	6	3
News	3	5	2

예를 들어, 시청자가 특정 날 t 시간에 드라마 TV 프로그램을 시청하고 있었다면, 시청자가 현재 선호하는 장르는 드라마이다. 그러면, 만약 시청자가 같은 요일 t+1 시간에 스포츠 TV 프로그램을 시청한다면, 시청자의 상태는 드라마에서 스포츠로 소비 성향이 변화 되었다고 말할 수 있다. [그림 2]에서 X(t)는 시간 t에서 고객의 프로그램 시청 성향 변수를 의미한다. 변수는 [X1, X2, ..., Xi, ..., Xc]와 같이 c 개의 프로그램 장르를 가질 수 있다. 여기서 Xi는 ith 번째 프로그램 장르를 의미한다. 즉, 변수 X={Drama, Sports, News, ..., Others}로 표현할 수 있다.

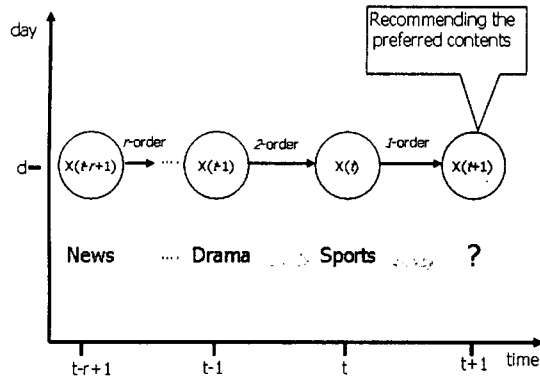
일정 기간 동안 고객의 프로그램 시청 기록으로부터 같은 요일에 특정 시간 구간 t에서 다음 시간 구간 t+1 사이에 i 번째 프로그램 장르 Xi에서 j 번째 프로그램 장르 Xj로의 추정된 통계적 성향의 변화는 [식 1]과 같이 표현될 수 있다.

$$P(X(t+1, a) = x_j | X(t, a) = x_i) = \frac{n_{ij}}{\sum_{j=1}^c n_{ij}}$$

[식 1] 프로그램 예측을 위한 1-order Markov 모델 계산식

여기서 nij는 고객의 프로그램 시청 기록으로부터 같은 요일에 시간 구간 t에서 Xi가 시청되었다는 사실이 주어졌을 때, 시간 구간 t+1에서 Xj를 시청한 빈도수를 의미한다. 그리고 j=1, 2, ..., c 이다.

1차 Markov 모델은 과거 연속적인 시청 성향의 변화를 r개까지 확장함으로써 r차 Markov 모델로 확장할 수 있다. [그림 3]은 시간에 따라 연속적으로 발생하는 시청 성향 전이 상태의 r-order Markov 모델을 나타낸 것이다.



[그림 3] r-order Markov 모델

r-order Markov 모델은 [그림 3]과 같이 과거 r개의 상태 전이 정보가 주어졌을 때, t+1 시점에서의 특정 프로그램을 선택할 조건 확률로 표현할 수 있다.

예를 들어, 2-order Markov 모델을 생각해 보자. 바로 전 시간(t-1 시간)에 드라마를 시청하고, 현재 시간(t 시간)에 스포츠를 시청하고 있는 경우, 바로 다음 시간(t+1 시간)에는 어떤 장르의 프로그램을 시청할 것인지 예측하기 위한 2차 Markov 모델은 [표 2]와 같이 표현할 수 있다.

[표 2] 2-order Markov 모델의 TV 시청 장르 전이 예

				Drama	Sports	News
		t-1	t			
Drama	Drama			0	1	1
	Sports			1	1	2
	News			2	1	0
Sports	Drama			2	1	0
	Sports			1	1	0
	News			0	0	1
News	Drama			0	2	1
	Sports			1	1	2
	News			1	0	0

t-1 시간에 드라마를 시청하고, t 시간에 스포츠를 시청한 경우, t+1 시간에 드라마를 시청할 확률은 1/4=0.25이다. t+1 시간에 스포츠를 시청할 확률 또한 1/4=0.25이다. 그리고 t+1 시간에 뉴스를 시청할 확률은 2/4=0.5이다. 따라서 확률이 가장 높은 뉴스를 추천하게 된다.

일반적으로 높은 차수의 Markov 모델이 낮은 차수의 Markov 모델보다 더 세부적이기 때문에 더 높은 차수의 Markov 모델이 낮은 차수의 Markov 모델보다 예측에 대한 정확도가 더 높다. 그러나 이와는 대조적으로 발생 가능한 패턴의 조합이 증가함으로써 소비 기록이 충분하지 않은 경우 low coverage가 발생할 수 있으며, 발생 가능한 상태의 수가 지수적으로 증가하는 경향이 있다. 그러므로 Markov 모델에서 차수의 선정은 고객의 콘텐츠 소비 기록의 크기에 종속적이다[7].

4. 실험 및 평가

본 논문에서 제안한 Markov 모델을 구현하고 시험하기 위한 시청 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. 시청 데이터가 가지는 각 프로그램은 8개의 장르, 즉 News, Entertainment, Drama & Movie, Information, Sports, Education, Children, Others로 구분한다.

성능 평가는 TOP-N 모델과 Markov 모델을 비교하였다. TOP-N 모델은 훈련 데이터에서 추천하고자 하는 특정 요일, 시간대에 시청한 정보를 이용하여 8개의 장르 중 시청 빈도수가 높은 순으로 추천하는 방법이다. Markov 모델은 시간에 따른 연속된 행위에서 바로 이전 행위를 기반으로 향후 행위를 예측할 수 있는 모델로 장르 간에 시간 변화에 따른 통계적 선호 전이 행렬을 이용한다. 즉 연속적으로 시청하는 TV 시청 프로그램을 기반으로 시간 변화에 따라 한 장르에서 다른 장르로 이동하는 확률을 고려한다. 1차 Markov 모델의 경우, 장르에서 장르로 전이되는 상태의 수는 64가지(8장르 × 8장르)이다. 2차 Markov 모델의 경우는 512(8×8×8)개의 상태를 가진다.

추천한 프로그램에 대한 정확도 계산은 실제 시청한 프로그램에 대한 추천된 프로그램을 시청한 비율로 계산하였다.

실험 결과 TOP-N 모델은 49%의 정확도를 나타내었고, Markov 모델은 72%의 정확도를 나타내었다.

5. 결론 및 향후 연구 과제

본 논문에서는 고객의 현재 상태의 시청 성향에서 다음 상태로의 프로그램 선호도를 예측하기 위한 가중치 적용 Markov 모델을 제안하였다. 제안된 방법의 도출 과정을 제공함으로써 우리의 방법이 유용함을 입증할 수 있었다. 또한 2000명의 TV 시청자가 시청한 데이터를 이용하여 Markov 모델의 성능이 TOP-N 모델의 정확도보다 효과적임을 보임으로써 실질적으로 유용할 수 있음이 입증 되었다.

본 연구에서는 보다 많은 시청자들의 시청 정보를 수집하는 것이 필요하며, 각각의 나이, 장르 그룹에 대한 최적의 파라미터 값을 추정하기 위한 자동 알고리즘을 개발하기 위한 더 많은 연구를 행할 필요가 있다. 또한, 고객 자신의 과거 데이터만을 이용하고 있으므로, 기존에 웹 사이트 개인화를 위해 연구되어온 기법들과의 결합을 통해 고객과 유사한 집단으로부터의 정보를 활용함으로써 정확도를 높일 수 있는 연구가 필요하다.

6. 참고 문헌

- [1] Ashrafi, M. Z., Tanizr, Smith, K., "An optimized distributed association rule mining algorithm", IEEE Distributed System Online, Vol. 3, No., 2004
- [2] Maltz, D. A., "Distributing Information for collaborative filtering on Usenet net news", SM Thesis, Massachusetts Institute of Technology, 1994
- [3] Schafer, J.B., Konstan J., Riedl, J., "Recommender systems in e-commerce", ACM Conference on Electronic Commerce, 1999
- [4] Cole, j., Gray, M., Lloyd, J., Ng, K. S., "Personalization for smart personal assistants", http://user.rsise.anu.edu.au/~jwl/crcc_paper_2004.pdf
- [5] Xing Dongshan, Shen Junyi, "A New Markov Model for Web Access Prediction", Computing in Science & engineering, November-December 2002
- [6] Athanasios Papoulis, "Probability, Random Variables, and Stochastic Process", McGraw Hill, 1991
- [7] Mukund Deshpande, George Karypis, "Selective Markov Models for Predicting Web-page", ACM Transactions on Internet Technology(TOIT), Vol. 4, Issue 2, May 2004