

LDA 클러스터링을 이용한 TV 프로그램 추천 기법

박창용*, 정연오**, 김누리*, 이지형*

*성균관대학교 전자전기컴퓨터공학과

**성균관대학교 정보 및 지능 시스템 연구실

e-mail : {codep,pd99j,john}@skku.edu,yeounohster@gmail.com

TV Program Recommendation Method Using LDA Clustering

Chang-yong Park*, Yeounoh Chung**, Noo-ri Kim*, Jee-hyoung Lee*

*Dept. of Electrical and Computer Engineering,

Sungkyunkwan University

**Information and Intelligence Laboratory, Sungkyunkwan University

요 약

최근 TV 시청자들의 콘텐츠 소비량이 증가함에 따라 방송사에서 제공하는 TV 프로그램들의 수량이 방대해지고 장르 또한 다양해지고 있기 때문에 시청자가 TV 프로그램을 선택하는 것이 점점 더 어려워지고 있다. 이러한 문제를 해결하기 위해 TV 프로그램 추천이라는 연구가 활발하게 이루어지고 있다. 기존의 연구에서는 시청자를 기반으로 하는 협업 필터링 추천 방법과 아이템을 기반으로 하는 협업 필터링 추천 방법이 제안되었지만 시청자의 시청 의도를 고려하는 연구는 사례는 적다. 이에 본 논문에서는 LDA 모델링을 이용하여 사용자의 시청 의도를 고려한 TV 프로그램 추천 기법을 제안한다. 실험을 통해 시청자의 시청 의도가 반영된 TV 프로그램 추천이 가능하다는 것을 검증했다.

1. 서론

최근 TV프로그램의 소비량이 증가함에 따라 TV프로그램들의 수량이 방대해지고 장르 또한 다양해지고 있다. 이러한 이유로 시청자가 프로그램을 선택하는 것이 점점 더 어려워지고 있다. 기존의 TV프로그램의 구조는 공중파, 케이블 방송 정도의 소분류였다면 현재는 VOD서비스의 추가와 기존보다 방대해진 TV프로그램으로 구성되어있다.[1][2][3] TV프로그램의 증가는 시청자의 다양한 취향을 충족시킬 수 있는 장점이 있지만, 반대로 방대한 TV 프로그램으로 인해 시청자의 취향이 반영된 프로그램을 찾는 것이 어려운 단점이 있다. 이를 해결하기 위해 다양한 연구 분야에서 TV 프로그램 추천에 대한 연구가 활발하게 진행되고 있다. TV 프로그램 추천에 대한 대표적인 연구로는 시청자의 프로그램 시청기록을 비교하여 유사한 사용자의 시청기록을 이용하여 추천하는 협업 필터링 기반 추천 방법(Collaborative Filtering based Recommendation)과, 사용자가 시청한 TV프로그램과 전체 프로그램간의 유사도를 이용한 아이템 기반(Item based Recommendation) 협업 필터링 추천 방법이 있다. 협업 필터링 기반 추천 방법은 유사한 사용자의 시청기록만을 이용하여 추천하기 때문에 사용자의 개인적인 취향이 고려되지 않을 수 있는 단점이 있다. 그리고 아이템 기반 추천 기법은 아이템의 장르, 아이템 별 시청자 선호도 평균, 런타임 등의 메타데이터 정보만을 사용하므로 시청 의도가 고려되지 않을 수 있다. 이러한 단점들을 극복하기 위해 본 논문에서는 LDA(Latent Dirichlet Allocation)를 이용한 확률 모델링을 이용하여 사용자의 시청 의도를 고

려할 수 있는 TV 프로그램 추천 방법을 제안한다.

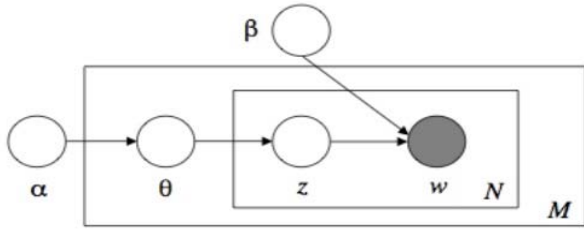
본 논문에서는 모델링 기법을 이용하여 전체시청자의 시청기록을 모델링하고 생성된 확률 모델을 이용하여 시청자들의 숨은 시청 의도를 고려하여 TV프로그램을 추천하는 방법을 제안한다. 논문의 구성은 다음과 같다. 2장에서는 TV 프로그램 추천 연구에 대한 장단점을 분석하고, 3장에서는 LDA 클러스터링을 이용한 추천 기법에 대해 소개한다. 4장에서는 제안하는 방법을 이용하여 실험을 진행한다. 마지막으로 5장에서는 실험의 결과를 분석하여 본 연구의 결론을 내리고 향후 연구 방향에 대해 제시한다.

2. 관련 연구

TV프로그램을 추천하는 연구는 일반적으로 크게 2가지로 분류된다. 첫 번째는 협업 필터링을 기반한 추천 방법이고, 두 번째는 아이템 유사도를 기반한 추천 방법이다. 본 장에서는 기존 TV 프로그램 추천 연구들을 분석하고 장단점을 분석한다.

2.1 시청자 기반 협업 필터링 추천 기법

[2]에서는 시청자를 기반으로 협업 필터링을 적용한 추천 기법을 제안했다. 먼저, 전체 시청자의 TV 프로그램 선호도 정보를 이용하여 2차원 행렬을 생성한다. 생성된 행렬의 행은 선호도 정보, 열은 시청자를 의미한다. 그리고 생성된 선호도 정보를 이용하여 시청자간의 유사도를 계산한다. 다음으로, 유사도가 가장 높은 시청자의 시청 목록을 추천하는 방법을 제안했다. 제안된 방법은 사용자간의 유사도를 계산하여 기존 연구 결과보다 좀 더 높은 정확도를 나타낸다. 하지만 장르 정보와 선호도 정보만을 이용



(그림 1) LDA model

하여 실험이 진행되었기 때문에 유사한 사용자의 시청 목록만을 이용하여 TV 프로그램이 추천되었을 수 있다. 이러한 이유로 시청자 기반 협업 필터링 추천 기법은 시청자의 시청 의도가 고려되지 않을 수 있는 단점이 있다.

2.2 아이템 기반 협업 필터링 추천 기법

[1]에서는 아이템을 기반으로 한 협업 필터링 추천 기법을 제안했다. 기존의 시청자의 유사도를 이용하여 협업 필터링 기법을 적용한 것과 유사한 형태를 띠고 있지만 단순히 시청기록이 아닌 아이템의 속성을 기반으로 유사도를 측정하기 때문에 아이템의 특성을 고려한 추천 방법이다. 하지만 과거 선호 데이터가 없으면 추천되지 않는 초기 데이터 부재(Cold start problem)의 단점이 있다.

2.3 LDA

본 논문에서 제안하는 방법은 [4]에서 제안한 확률 그래프 모델 LDA를 이용한다. LDA 모델은 그림 1과 같이 표현된다. 그림의 w 는 문서의 단어를 가리키고 z 는 문서의 주제를 나타낸다. 그림은 단어 w 가 어떤 주제 z 에 포함될 확률을 찾아내는 것을 나타낸 것이다. 그림 1의 기호에 대

표 1. LDA 기호 정리

α	주제들에 대한 Dirichlet 분포
β	단어들에 대한 Dirichlet 분포
θ	문서 레벨의 변수
z	해당 주제에 대한 단어가 관찰될 확률
w	해당 주제에 대한 실제 관찰된 단어

한 설명은 표 1과 같다.

3. LDA 클러스터링을 이용한 추천 기법

본 논문에서 제안하는 방법은 시청자의 과거 시청기록을 LDA 모델을 이용하여 각 시청자가 각 TV 프로그램을 시청할 확률을 생성한다. 다음으로 LDA 모델을 이용하여 생성된 확률을 이용하여 비슷한 확률을 가진 시청자 별로 클러스터링한다. LDA 모델링과 클러스터링 과정을 통합하여 LDA 클러스터링이라고 정의한다. LDA 클러스터링을 통해 생성된 클러스터 K 개를 시청 의도가 K 개로 나누어져서 클러스터링된다. 생성된 K 개의 클러스터는 추천할 수 있는 K 개의 집단이 된다. 이를 통해 추천 대상 사용자의 시청 데이터를 모델링하여 가장 유사한 시청 의도 집단의 시청 기록을 이용하여 추천하는 기법이다. 이 때 유사도는 코사인 유사도 측정 방법을 이용한다.



(그림 2) TV 프로그램 추천 프로그램

4. 실험

4.1 데이터 셋

실험을 위해 TVing에서 제공하는 사용자의 시청기록 데이터를 이용했으며, 총 약 45000여명의 시청 데이터를 이용했으며, 포함된 프로그램의 개수는 약 700개이다. 프로그램에 포함되는 에피소드의 개수는 약 2000개다. 수집기간은 2012년 1월부터 2012년 10월까지 수집되었다. 실제 시청자의 환경과 동일한 조건에서 실험하기 위해 Smart TV Emulator를 이용하여 실험했다.

4.2 제안 방법 실험

실험은 45,000명 중 가장 시청 횟수가 많은 시청자를 200명 제외하고 LDA로 시청 의도를 모델링하였고, 나머지 200명을 테스트 대상으로 사용하였다. 먼저 45,000명의 데이터를 이용하여 LDA 모델을 생성하고, 클러스터링을 통해 유사한 시청 의도를 가진 시청자로 클러스터링한다. 나머지 200명의 시청 기록 중 절반으로 LDA 모델링을 한다. 이를 통해 생성된 모델을 이용하여 가장 유사한 시청 의도 집단을 검색하고, 검색된 집단의 시청 기록을 이용하여 TV 프로그램을 추천한다. 논문에서 제안하는 추천 기법을 적용한 TV프로그램 추천 프로그램은 그림 2와 같다. 왼쪽의 원그래프 중 위쪽 그래프는 사용자의 시청기록에 나타나는 장르 분포를 의미하고 아래 그래프는 추천된 프로그램들의 장르 분포를 의미한다. 오른쪽의 리스트는 시청자가 과거에 시청한 시청 기록 목록이다. 실험을 통해 시청자의 시청 의도를 고려하여 TV 프로그램을 추천하면 유사한 장르와 시청 의도가 반영된 추천이 가능하다는 것을 검증할 수 있었다.

5. 결론

4장에서 실험을 통해 본 논문에서 제안하는 방법이 사용자의 시청 의도를 반영하여 추천할 수 있다는 것을 검증했다. 이전 연구들보다 정확도는 향상되지 않았지만 다양한 시청자의 의도를 고려할 수 있었다. 향후에는 시청자의 시청 의도를 고려하면서 예측 정확도도 향상될 수 있는 방법에 대한 연구를 진행할 예정이다.

참고문헌

- [1] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl, "Item-based collaborative filtering recommendation algorithms," International Conference on World Wide Web, 2001, pp. 285-295.
- [2] Hyoseop Shin, Minsoo Lee, and Eun Yi Kim, "Personalized Digital TV Content Recommendation with Integration of User Behavior Profiling and Multimodal Content Rating," Consumer Electronics, IEEE Transactions on, vol 55, 2009, pp. 1417-1423.
- [3] Hyeong-Joon Kwon, Kwang-Seok Hong, "Personalized Smart TV Program Recommender Based on Collaborative Filtering and a Novel Similarity Method," IEEE Transactions on Consumer Electronics, 2011, Vol. 57, pp1416-1423.
- [4] David Blei, Andrew Ng, Michael Jordan, "Latent Dirichlet Allocation", Journal of Machine Learning Research3, 2003, pp.993-1022.