

컨텐츠 선호도 정보를 이용한 딥러닝 기반의 하이브리드 추천 시스템

채동규*, 김상욱ⁱ⁾*

*한양대학교 컴퓨터소프트웨어학과

e-mail:kyu899@agape.hanyang.ac.kr

A Hybrid Recommender System based on Deep Learning using Contents Preference

Dong-Kyu Chae*, Sang-Wook Kim*

*Dept of Computer and Software, Hanyang University

요 약

본 논문에서는 사용자의 상품에 대한 평점 정보와 상품의 컨텐츠 정보를 모두 이용하는 하이브리드 추천 모델에 대해서 논의한다. 기존 논문들과는 다르게, 본 논문은 추천의 정확도를 높이기 위해 사용자가 상품의 컨텐츠 (예를 들면, 영화의 장르 또는 상품의 카테고리 등) 에 가질 수 있는 선호도를 예측하고, 이를 추가적으로 활용할 수 있는 딥러닝 기반의 추천 모델을 제안한다. 실세계의 데이터를 이용해서 제안하는 방법의 우수성을 보인다.

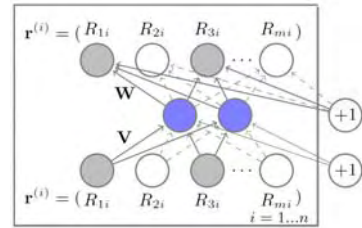
1. 서론

최근 딥러닝이 컴퓨터비전, 자연어처리, 음성인식 등 여러 분야에서 큰 성공을 거둠에 따라, 추천시스템에도 딥러닝을 적용하고자 하는 시도가 증가하고 있다. 딥러닝 기반의 여러 모델들 중, 추천시스템에 가장 활발하게 이용되는 모델로는 오토인코더 (AutoEncoder) 가 있다. 오토인코더는 입력 데이터를 그대로 재현하는 것을 목적으로 하는 모델이다. 즉, 입력 데이터와 모델의 출력 간의 차이를 최소화하는 목적 함수를 통해 학습된다. 추천시스템에서의 오토인코더는 사용자의 평점 벡터를 입력으로 받아서 이 벡터를 그대로 재현하도록 모델을 학습시키는데, 이 때 모델은 사용자의 평점 벡터의 비어있는 차원에 값을 예측할 수 있게 된다. 뿐만 아니라, 오토인코더는 사용자나 상품의 컨텐츠 정보를 표현하는 벡터 등 추가적인 정보가 있을 때 이를 함께 활용해서 모델을 학습하기에 용이하다. 대표적인 오토인코더 기반 추천 모델로는 I-AutoRec, U-AutoRec [1] 등이 있다.

사용자가 상품을 구매할 때, 그 상품의 컨텐츠 정보 (예를 들면, 영화의 장르나 상품의 카테고리 등) 는 구매를 결정하는 핵심적인 요소라고 할 수 있다. 따라서, 만약 컨텐츠 정보가 사용가능하다면, 이를 이용해서 추천 모델에 반영하는 것이 효과적이다. 이에 따라, 본 논문에서는 사용자의 상품에 대한 평점 정보와 상품의 컨텐츠 정보를 모두 활용하는 하이브리드 추천 모델에 대해서 논의한다. 구체적으로, 본 논문에서 제안하는 모델은 오토인코더를

기반으로 하며, 기존의 하이브리드 추천 모델과는 달리 사용자가 평가하지 않은 컨텐츠 정보에 대해서 사용자의 선호도를 예측하여 이를 추가적으로 모델에 반영한다.

2. 관련 연구



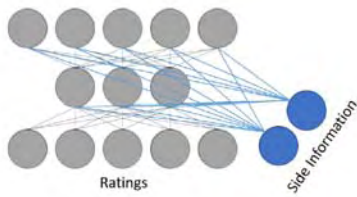
(그림 1) AutoRec [1]

본 연구는 AutoRec [1] 모델을 기반으로 한다. AutoRec은 하나의 (혹은, 다수의) 은닉 층을 가지는 오토인코더 기반의 추천 모델이다. 그림 1과 같이, AutoRec은 먼저 사용자의 평점 벡터를 은닉 층으로 인코딩한 후, 인코딩된 정보를 다시 디코딩한다. 이 때 디코딩된 결과가 최초로 입력되었던 벡터와 같도록 인코딩 모델 파라미터와 디코딩 모델 파라미터들이 학습된다. 몇 번의 반복 학습을 통해 모델 파라미터들의 학습이 완료되었다면, AutoRec은 모델 파라미터를 이용해서 각 사용자의 평점 벡터를 인코딩한 후 다시 디코딩함으로써 각 사용자의 평점벡터의 비어있는 차원의 값을 (즉, 구매하지 않은 상품에 부여할 평점)을 계산 (예측) 한다.

또 다른 관련 연구로는 Hybrid AutoEncoder [2]가 있다. 이 모델은 AutoRec 모델을 확장한 것으로, 전체적인 학습 과정은 거의 유사하면서 상품의 컨텐츠 정보 (그림 2의 side information 벡터가 이에 해당됨) 가 추가적으로

i) 교신저자

사용자의 선호도 예측에 이용된다는 점이 다르다. 영화의 장르를 예로 들면, 어떤 영화의 장르가 “액션”, “판타지” 라면 해당 장르에 해당하는 벡터의 차원이 1이고 나머지는 0인 벡터의 형태로 콘텐츠 정보가 표현된다.



(그림 2) Hybrid AutoEncoder [2]

3. 제안하는 방법

제안하는 방법은 두 개의 오토인코더를 이용하는 추천 모델이다. 본 논문은 각각을 (1) 평점 예측 네트워크, (2) 콘텐츠 선호도 예측 네트워크로 명명한다. 제안하는 방법은 다음과 같은 절차로 각 오토인코더를 학습한다.

콘텐츠 선호도 예측 네트워크: 이 오토인코더는 각 사용자가 각 콘텐츠에 가질 수 있는 선호도를 예측한다. 이를 위해서는 초기 학습 데이터가 필요하다. 이 학습 데이터는 각 사용자가 각 콘텐츠에 평균적으로 어느 정도의 선호도를 갖고 있는지를 나타내는 데이터인데, 이는 사용자가 상품에 부여한 평점들을 통해서 가공될 수 있다. 예를 들어, 어떤 사용자가 “해리포터” 영화에 5점, “스타워즈” 영화에 2점을 부여했고, “해리포터”의 장르는 “판타지, 드라마”, “스타워즈”의 장르는 “판타지, 액션”이라고 가정하자. 그렇다면 해당 사용자의 각 장르에 대한 선호도는 (평균으로 계산했을 때) “판타지”가 3.5점, “액션”이 2점, “드라마”에 5점으로 파악할 수 있다. 위는 평균을 예로 든 것이며, 평균 대신 빈도수를 사용할 수도 있다. 본 연구에서는 평균과 빈도수를 곱한 수치를 사용자의 해당 콘텐츠(즉, 장르)에 대한 선호도로 이용한다.

이렇게 사용자의 콘텐츠 선호도 예측을 위한 초기 데이터를 구성하면, 이제 사용자가 선호도를 명시적으로 표시하지 않은 장르에 대한 선호도를 예측할 수 있다. 예를 들면, 위의 예시에서 사용자는 “코메디”, 혹은 “다큐멘터리” 등의 장르의 영화는 시청하지 않았으므로 해당 장르에 대한 선호도는 아직 모르는 상황이며, 이를 예측하고자 하는 것이 본 논문의 목표이다. 예측을 위해 위에서 설명한 콘텐츠 선호도 데이터를 이용해서 오토인코더를 학습시킨다. 학습 방법은 AutoRec [1] 논문에서 이용된 방법과 동일하다. 학습이 완료되고 나면 해당 오토인코더로부터 사용자의 존재하는 모든 콘텐츠에 대한 선호도가 예측된 벡터를 얻을 수 있다.

평점 예측 네트워크: 이제 각 사용자의 콘텐츠 선호도 정보를 이용해서, 최종 목표인 상품에 대한 선호도를 예측하는 방법에 대해서 설명한다. 이 방법은 Hybrid AutoEncoder [2] 에서 사용된 방법과 유사하다. 사용자의 평점 벡터와 함께, 사용자의 콘텐츠 선호도 벡터가 또 다른 입력으로 오토인코더에 들어간다. 모델 파라미터들은

입력되었던 사용자의 평점 벡터와 출력이 최대한 같도록 학습된다. 학습이 완료되면 사용자의 상품 평점 벡터와 콘텐츠 선호도 벡터를 입력할 경우 출력 벡터에는 사용자가 평가하지 않은 상품에 대한 예측 평점(즉, 선호도)을 얻을 수 있다. 최종 추천은 해당 예측 평점이 가장 높은 상위 N개의 상품을 선택함으로써 이루어진다.

4. 실험

우리는 제안하는 방법의 정확도를 평가하기 위해 MovieLens 100K 데이터를 이용하였다. 이 데이터는 943명의 사용자와 1682편의 영화, 10만 건의 평점으로 구성되어 있다. 우리는 10만 건의 평점들 중 임의로 선택한 80%의 평점을 모델 학습을 위한 데이터로 사용하였고, 나머지 20% 데이터를 추천 정확도를 평가하기 위한 데이터로 사용하였다. 기존의 U-AutoRec, I-AutoRec, Hybrid AutoEncoder 모델과 제안하는 모델을 각각 학습 데이터를 이용해서 학습시키고, 평가 데이터를 이용해서 각 모델의 top-5 추천 정확도를 계산해서 비교하였다. 또한, baseline으로써, 단순히 상품의 인기도만으로 추천하는 방법의 정확도와도 비교하였다. 정확도 계산을 위한 척도는 기존에 널리 이용되는 precision, recall, nDCG, MRR [3]을 이용하였다. 표 1은 실험 결과를 나타낸다. 표에서 볼 수 있듯이 제안하는 방법이 기존의 다른 모델보다 항상 정확한 추천을 수행한다. 이는 제안하는 방법이 기존 방법들과는 다르게 사용자가 각 콘텐츠에 가질 수 있는 선호도를 예측하고 이를 추천 모델 학습에 활용했기 때문에 더 정확한 추천을 수행한 것으로 보인다.

<표 1> 실험 결과

	Precision	Recall	nDCG	MRR
인기도	0.081	0.102	0.117	0.204
U-AutoRec	0.141	0.170	0.202	0.308
I-AutoRec	0.127	0.141	0.172	0.271
HybridAE	0.138	0.156	0.181	0.274
제안 방안	0.153	0.177	0.209	0.311

5. 감사의 글

본 연구는 (1) 2017년도 과학기술정보통신부의 재원으로 한국연구재단-차세대정보·컴퓨팅기술개발사업 (No. NRF-2017M3C4A7069440), (2) 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업 (IITP-2017-2013-0-00881), (3) 중소기업청에서 지원하는 2017년도 산학협력기술개발사업 (No. C0511678)의 지원으로 수행되었음.

참고 문헌

- [1] S. Sedhain et al. “Autorec: Autoencoders Meet Collaborative Filtering” in Proc. of WWW, 2015
- [2] F. Strub et al. “Hybrid Recommender System based on Autoencoders” in Proc. of RecSys, 2016
- [3] P. Cremonesi et al. “Performance of Recommender Algorithms on Top-N Recommendation Tasks” in Proc. of RecSys, 2010