

개인화된 뉴스 추천 시스템에서 콜드 스타트 유저 문제 해결을 위한 그래프 인공 신경망 기반의 인공지능 모델

고윤용
중앙대학교

yyko@cau.ac.kr

Graph Neural Network-based Personalized News Recommendation For the Cold-Start User Problem

Yunyong Ko
Chung-Ang University

요약

본 논문에서는 개인화된 추천 시스템의 고질적 문제인 콜드 스타트 유저 문제 (cold-start user problem)를 완화하기 위한 그래프 인공 신경망 (Graph Neural Network) 모델 기반의 개인화된 뉴스 추천 기술을 제안한다. 그를 위해, 먼저 유저와 뉴스 기사간 관계를 이분 그래프(bipartite graph)로 표현하고, 해당 그래프에 세 가지의 최신 GNN 모델을 적용하여 유저와 뉴스 기사를 상호 보완적으로 업데이트한다. 이를 통해, 뉴스 플랫폼 사용 이력이 거의 없는 유저들도 상호 보완적으로 학습할 수 있으며, 나아가 콜드 스타트 유저 문제를 완화하는데 기여할 수 있다. 실험을 통해, GNN 모델이 콜드 스타트 유저 문제를 완화하는데 효과적이라는 것을 확인하였다. 또한, 세 가지 GNN 모델이 개인화된 뉴스 추천 정확도에 미치는 영향도 함께 분석하였다.

I. 서론

최근 MS 뉴스, Google 뉴스, 네이버 뉴스 등 온라인 뉴스 플랫폼의 발전 덕분에, 플랫폼 유저들이 양질의 뉴스 정보를 손쉽게 접근할 수 있게 되었다. 그러나, 매일 수많은 양의 뉴스 기사가 생성됨에 따라, 뉴스 기사들 중에서 각 유저가 진정으로 선호하는 뉴스를 찾는 것은 많은 시간과 노력을 필요로 하게 되었다. 이러한 유저들의 불편을 해소하기 위해 개인화된 뉴스 추천 시스템 (personalized news recommendation) 기술들이 활발하게 연구되고 있다 [1,2,3,4].

이러한 개인화된 뉴스 추천 기술 덕분에 유저들이 선호하는 뉴스 기사들을 손쉽게 접할 수 있게 되었지만, 여전히 플랫폼 사용이력이 전혀 혹은 거의 없는 유저들의 경우(즉, 콜드 스타트 유저), 효과적으로 뉴스 기사를 추천해주기 어렵다는 한계를 가지고 있다. 이는 뉴스 추천 시스템의 고질적인 문제로 콜드 스타트 유저 문제 (cold-start user problem)이라 불리며, 이를 해결 및 완화하는 것은 추천 시스템의 전통적인 과제이다.

본 논문에서는 이러한 콜드 스타트 유저 문제를 완화하기 위해 그래프 인공 신경망 (Graph Neural Network) 모델 기반의 개인화된 뉴스 추천 기술을 제안하고자 한다. 그를 위해, 먼저 유저와 뉴스 기사를 이분 그래프 (bipartite graph)로 표현하고, 해당 그래프에 세 가지의 최신 GNN 모델[6-8]을 적용한다. 이를 통해, 플랫폼 사용이력이 거의 없는 유저들도 상호보완적으로 학습할 수 있다. 또한, 세 가지 GNN 모델이 개인화된 뉴스 추천 정확도에 미치는 영향도 함께 분석한다. 실험을 통해, GNN 모델이 콜드 스타트 유저 문제를 완화하는데 효과적이라는 것을 확인하였다.

II. 본론

일반적으로 개인화된 뉴스 추천은 유저가 과거에 읽었던 뉴스 기사의 내용을 기반으로 수행된다. 즉, 뉴스 추천 모델이 유저가 과거에 읽었던 뉴스 기사들의 내용을 기반으로 해당 유저의 뉴스 선호도를 파악하고, 그를 토대로 유저가 아직 읽지 않은 뉴스들 중 해당 내용과 유사한 새로운 뉴스를 추천한다[5]. 이러한 내용 기반 추천 기술은 유저의 뉴스 선호도를 정확하게 파악할 수 있다는 장점을 가지고 있지만, 한편으로는 읽었던 뉴스 기사의 수가 적은 콜드 스타트 유저의 선호도를 정확하게 파악하기 어렵다는 한계도 가지고 있다.

이러한 동기로부터, 본 논문에서는 뉴스 기사의 내용과 함께 이웃 유저들의 협업 신호 정보(collaborative signal)를 상호 보완적으로 활용하기 위한 GNN 모델 기반의 뉴스 추천 기술을 제안하고자 한다. 이를 위해, 먼저 (1) 유저와 뉴스 기사 간 관계를 이분 그래프로 표현한다. 그림 1에서 볼 수 있듯이, 이분 그래프의 각 노드는 유저(위 노드) 또는 뉴스 기사(아래 노드)를 나타내고, 유저가 뉴스 기사를 읽었을 경우 두 노드는 간선으로 연결된다. 즉, 유저-노드 간선만 존재할 뿐, 유저-유저 혹은 뉴스-뉴스 간선은 존재하지 않는다. 우리는 해당 이분 그래프를 $A \in \mathbb{R}^{|U| \times |N|}$ 로 표현한다

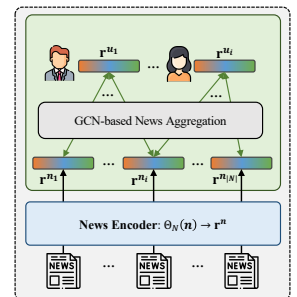


그림 1 그래프 인공 신경망 기반의 개인화된 뉴스 추천 과정

(수식에서 $|U|$ 는 유저의 수를 나타내고, $|N|$ 은 뉴스 기사의 수를 나타낸다. 그리고 유저와 뉴스 기사에 대한 임베딩 벡터를 $R^u \in \mathbb{R}^{|U| \times d}$ 와 $R^n \in \mathbb{R}^{|N| \times d}$ 로 각각 나타낸다.

다음 단계로, 구축된 이분 그래프에 GNN 모델을 적용하여 유저와 뉴스를 상호 보완적으로 업데이트한다. 구체적인 GNN 모델 과정은 수식 1 과 같다.

$$R^u = \sigma(AR^nW_g^l + b_g^u), R^n = \sigma(AR^uW_g^l + b_g^n) \quad (1)$$

수식 1 에서 W_g^l 와 b_g^u 는 GNN 모델의 학습 가능한 파라미터를 나타낸다. 본 방법론은 특정 GNN 모델에 종속된 방법이 아니기 때문에 다양한 GNN 모델이 적용될 수 있다. 본 논문에서는 GCN[6], LightGCN[7], 그리고 GraphSAGE[8] 세 가지 GNN 모델을 고려했으며, 최종적으로 GraphSAGE 모델을 사용했다. GNN 모델에 따른 개인화된 추천 성능은 실험 섹션에서 자세히 보인다.

이러한 과정을 통해 이력이 적은 유저도 주변의 이웃 유저들과 관련 뉴스 기사의 정보를 상호보완적으로 수렴할 수 있으며, 이는 콜드 스타트 문제를 완화하는데 핵심적으로 작용할 수 있다.

III. 실험

본 섹션에서는 본 논문에서 제안한 GNN 모델 기반의 개인화된 뉴스 추천 기법을 검증하기 위해, 실험을 수행하여 다음 두 가지 연구 질문에 답하고자 한다.

- RQ1. 그래프 인공 신경망 (GNN) 모델이 정말로 콜드 스타트 유저 문제 완화에 도움이 되는가?
- RQ2. 각 그래프 인공 신경망 모델에 따라 개인화된 뉴스 추천 정확도를 어떻게 변화하는가?

본 실험에서는 대표적 뉴스 데이터인 MIND-small [9] 데이터를 사용한다. 해당 뉴스 데이터의 정보는 다음과 같다 (표 1 참조).

표 1 실험 데이터셋 정보

	<i>MIND-small</i>
유저 수	94,057
뉴스 기사 수	65,238
유저가 읽은 뉴스 수	347,727
뉴스 기사의 카테고리 수	18 (270)
제목에 포함된 단어 수	11.67
내용에 포함된 단어 수	41.01

본 논문에서 제안하는 방법의 우수성을 검증하기 위해, 다음 다섯 가지의 기존 개인화된 추천 기술을 비교 방법으로 선택했다. LibFM[1], DSSM[2], NPA[3], NRMS[4], NAML[5]. 각 방안들을 비교하기 위한 평가 척도로 AUC, MRR, nDCG@5, nDCG@10 을 사용한다.

● 실험 1: 콜드 스타트 유저에 대해 효과 검증

본 실험에서는 제안하는 GNN 기반의 개인화된 추천 기술의 우수성을 검증하기 위해, 뉴스 사용 이력은 적은 콜드 스타트 유저들에 대한 추천 정확도를 비교한다. 읽은 뉴스가 다섯 개 이하인 유저들을 콜드 스타트 유저로 정의하고, 해당 유저들에 대한 각 비교 방법들의 추천 정확도를 측정했다. 표 2 는 콜드 스타트 유저들에 대한 비교 방법들의 추천 정확도를 보여준다. 표 2 에서 볼 수 있듯이, 제안하는 GNN 모델 기반의 추천 기술이

기존 기술과 비교하여, 콜드 스타트 유저들에게 보다 정확한 추천을 제공할 수 있음을 확인하였다.

표 2 콜드 스타트 유저에 대한 추천 정확도

	AUC	MRR	nDCG@5
LibFM	0.5020	0.2509	0.2592
DSSM	0.5832	0.2810	0.3036
NPA	0.6022	0.3080	0.3351
NRMS	0.6021	0.3123	0.3377
NAML	0.6118	0.3139	0.3405
제안방법	0.6286	0.3254	0.3578

● 실험 2: 그래프 인공 신경망의 효과 검증

본 실험에서 세 가지 GNN 모델을 이용하여 GNN 모델이 개인화된 뉴스 추천 정확도에 미치는 효과를 검증하고자 한다. GNN 모델로 세 가지 최신 모델을 사용했으며, 실험 결과는 표 3 과 같다. 표 3 에 명시되어 있듯이, 어떤 GNN 모델을 적용하든 항상 우수한 추천 정확도를 달성했으며, 특히 GraphSAGE 모델의 효과가 가장 크다는 사실을 확인했다.

표 3 GNN 모델에 따른 추천 정확도

	AUC	MRR	nDCG@5
GCN	0.6813	0.3348	0.3691
LightGCN	0.6819	0.3350	0.3705
GraphSAGE	0.6823	0.3354	0.3697

IV. 결론

본 논문에서는 개인화된 추천 시스템의 고질적 문제인 콜드 스타트 유저 문제를 완화하기 위한 그래프 인공 신경망 (Graph Neural Network) 모델 기반의 개인화된 뉴스 추천 기술을 제안한다. 해당 기술을 통해 유저와 뉴스 기사를 상호 보완적으로 업데이트함으로써, 콜드 스타트 유저들에게도 정확한 추천을 해줄 수 있다. 실험을 통해, 제안하는 방법의 우수성 또한 검증하였다.

참고 문헌

- [1] Steffen Rendle. Factorization machines with libfm. ACM TIST, 3, 3 (2012), 1-22.
- [2] Po-Sen Huang et al., Learning deep structured semantic models for web search using clickthrough data. In Proceedings of the ACM CIKM, 2333-2338.
- [3] Chuhan Wu et al., NPA: neural news recommendation with personalized attention. In Proceedings of the 25th ACM SIGKDD, 2576-2584.
- [4] Chuhan Wu et al., Neural news recommendation with multi-head self-attention. In Proceedings of EMNLP-IJCNLP, 6389-6394.
- [5] Chuhan Wu et al., Neural news recommendation with attentive multi-view learning. In Proceedings of the IJCAI, 3863-3869.
- [6] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907 (2016).
- [7] He, X., et al. Lightgen: Simplifying and powering graph convolution network for recommendation. In Proceedings of the SIGIR 2020.
- [8] Hamilton, W., et al. Inductive representation learning on large graphs. In Proceedings of the NeurIPS 2017.
- [9] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, et al. 2020. Mind: A large-scale dataset for news recommendation. In Proceedings of the ACL, 3597-3606.