

네트워크 임베딩을 사용한 Top-N 추천 정확도 개선에 관한 연구

서창원, 정경중, 신원용
연세대학교

changwoni@yonsei.ac.kr, jeongkj@yonsei.ac.kr, wy.shin@yonsei.ac.kr

Improving Accuracy of Top-N Recommendation Using Network Embedding

Changwon Seo, Kyeong-Joong Jeong, Won-Yong Shin
Yonsei University

요 약

최근 네트워크 임베딩 기술을 사용하여 추천 시스템을 개발하는 연구가 활발히 이루어지고 있다. 하지만 대부분의 네트워크 임베딩 기술을 이용한 연구는 사용자가 특정 아이템을 소비했는지 여부 (즉, implicit feedback)에만 초점을 맞추어 수행되었다. 본 논문에서는 주어진 사용자-아이템 평점 데이터를 바탕으로 평점에 offset 을 둔 signed/weighted graph 를 생성함으로써 보다 현실적인 사용자 선호도를 반영하여 네트워크 임베딩을 수행하고, 이를 통해 top-N 정확도 개선하는 새로운 협업 필터링 기반 추천 시스템을 제안한다.

I. 서 론

네트워크 임베딩을 이용한 추천 시스템 설계 연구가 최근 관심을 끌고 있다 [1]. 그러나 사용자의 선호도와 관계없이 소비 여부 (즉, implicit feedback)에만 초점을 맞춰 추천을 해주는 모델이라는 점에 한계가 있다. 본 논문에서는 평점 offset 을 두어 signed/weighted graph 를 생성함으로써 보다 현실적인 사용자의 선호도를 반영한 네트워크 임베딩 기술을 제안한다.

II. 본론

이론 그래프 $G = (U, V, E)$ 에서 U 를 사용자의 집합, V 를 아이템의 집합, E 를 사용자-아이템 간의 간선, 가중치 $w_{ij} \in W$ 를 사용자 u_i 가 아이템 v_j 에 남긴 평점이라고 하자. 선호/비선호 아이템을 구분하기 위한 offset w_o 을 두어 그림 1과 같이 signed/weighted graph 를 형성한다. 새로 형성된 그래프에서 가중치가 양수인 경우 선호 아이템, 음수인 경우 비선호 아이템이라고 가정한다.

x_i, y_j 를 각각 $u_i \in U, v_j \in V$ 의 latent representation, $w_{ij}^{off} = w_{ij} - w_o$, $\sigma(x) = \frac{1}{1+e^{-x}}$ 라고 할 때, negative sampling 을 활용한 손실 함수는 다음과 같다.

$$L = - \sum_{(i,j) \in E} \left\{ w_{ij}^{off} \log \sigma(\text{sign}(w_{ij}^{off}) x_i \cdot y_j) \right. \\ \left. + \sum_{n=1}^K \mathbb{E}_{v_n \sim P_n(v)} [\log \sigma(-x_i \cdot y_n)] \right\}$$

위 손실 함수를 이용하여 그림 1과 같이 임베딩 공간에서 사용자 노드 기준으로 선호 아이템은 가까이, 비선호 아이템은 멀리 위치할 수 있도록 하여 사용자 노드와 아이템 노드의 내적 값이 큰 순서대로 top-N 추천 리스트를 뽑는다.

데이터셋으로는 MovieLens 100k 를 사용하고, 정확도 측정을 위한 metric 으로는 precision, recall, NDCG 를 사용한다. 사용자 선호도를 잘 반영하는지 확인하기 위해 ground-truth set 은 평점 5 점을 받은 아이템만으로 구성한다. 표 1에서 확인할 수 있듯이, 제안한 기술은 두 가지 네트워크 임베딩에 기반을 둔 방법 (즉, BiNE [1]과 LightGCN [2]) 대비 높은 성능을 보임을 알 수 있다.

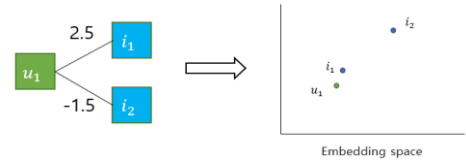


그림 1. 고려하는 signed graph 기반 노드 임베딩

Algorithm	P@25(%)	R@25(%)	NDCG@25(%)
BiNE [1]	4.6	21	11
LightGCN [2]	10.5	45	31
Our model	10.7	45	31.8

표 1. MovieLens 100k 데이터셋을 이용한 실험 결과

III. 결론

본 논문에서는 평점 offset 으로 signed/weighted graph 를 생성하여 사용자가 선호할만한 아이템을 추천하는 모델을 제안하였다. 제안된 기술은 기존 방법 대비 추천 정확도 성능을 개선하였음을 검증하였다.

ACKNOWLEDGMENT

This work was supported by the Republic of Korea's MSIT (Ministry of Science and ICT), under the High-Potential Individuals Global Training Program (No. 2020-0-01463) supervised by the IITP (Institute of Information and Communications Technology Planning Evaluation) and by the Yonsei University Research Fund of 2020 (2020-22-0101).

참 고 문 헌

- [1] M. Gao, X. He, L. Chen, and A. Zhou, "BiNE: Bipartite network embedding," in Proceedings of the 41st ACM SIGIR Conference on Research and Development in Information Retrieval, Ann Arbor, MI, July 2018, pp.715-724.
- [2] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and powering graph convolution network for recommendation," in Proceedings of the 43rd ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, July 2020, pp.639-648.