

PyTorch Geometric을 이용한 그래프 기반 협업 필터링 성능 비교 연구

김경태¹, 진희국², 안진현³, 임동혁⁴

¹광운대학교 인공지능응용학과

²(주)핀다

³제주대학교 경영정보학과

⁴광운대학교 정보융합학부

kkt9601@kw.ac.kr, heegook@finda.co.kr, jha@jejunu.ac.kr, dhim@kw.ac.kr

A Comparative Study on the Performance of Graph Based Collaborative Filtering Using PyTorch Geometric

Gyoung-Tae Kim¹, Hee-Gook Jun², JinHyun Ahn³, Dong-Hyuk IM⁴

¹Dept. of Artificial Intelligence Applications, Kwangwoon University

²Finda, Seoul, Korea

³Dept. of Management Information Systems, Jeju National University

⁴School of Information Coverage, Kwangwoon University

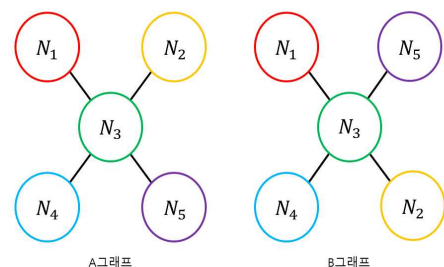
요 약

그래프 데이터는 데이터간의 관계를 효율적으로 분석할 수 있으며, 뛰어난 확장성, 다양한 종류의 데이터들을 쉽게 표현할 수 있어 화학, 의학, 추천시스템등 다양한 분야에 적용하려는 사례가 늘고 있다. 이러한 그래프 데이터를 머신러닝기법에 쉽게 사용할 수 있도록 적용된 것이 GNN모델이다. 그 중 Convolution기법을 적용한 ConvGNNs 모델이 추천 시스템 등 다양한 분야에서 많이 연구 되고 있다. 본 논문은 실험을 통해 상이한 데이터셋 환경에서 Convolution 그래프 기반 모델들의 성능을 비교하였다.

1. 서론

그래프란 관계된 객체들의 집합이며 노드(Node)와 엣지(Edge)로 이루어진 구조를 의미한다. 그래프 형태의 데이터는 데이터 간의 복잡하게 얽혀있는 관계성을 효율적으로 분석할 수 있다. 또한 하나의 노드가 가질 수 있는 엣지는 제한이 없기 때문에 확장성도 뛰어나며 다양한 분야의 데이터들을 트리플 구조(노드-엣지-노드)를 통해 쉽게 표현 할 수 있다 [1]. 그러나 그래프 데이터는 노드간의 관계가 매우 복잡하기 때문에 기존의 머신러닝 기법들로 사용하기에는 어려움이 있다[2]. GNN(Graph Neural Network)은 이러한 어려움을 해결하고 그래프 데이터에 머신러닝을 직접 사용하여 학습하는 모델이며 기존의 기법들과 비교하여 더 나은 성능을 보인다 [3]. GNN 모델 중 Convolution기법이 제약 분야에서의 약효분석, 그래프 데이터를 사용한 경로탐색, 화학, 의학에서의 분자, DNA 구조 분석, 사용자와 아이템 간의 상호작용을 통한 추천시스템등 다양한 분야에서 사용 되고 있다[4,5]. 이러한 ConvGNNs

을 쉽게 사용할 수 있도록 라이브러리들 또한 개발이 되고 있다. 그 중 PyTorch Geometric 은 PyTorch에서 지원하는 라이브러리이다. 본 연구에서는 PyTorch Geometric을 이용한 ConvGNNs의 여러 종류의 모델들을 준 지도학습에서의 성능과 협업 필터링 추천 시스템에서의 성능을 비교한다.



(그림 1) 그래프 예시

2. 배경지식

ConvGNNs은 그래프 데이터를 처리하기 위하여 CNN(Convolution Neural Network)과 같이 컨

블루션을 적용하였다[6]. 하지만 그래프의 경우 CNN에서 사용하는 이미지와 달리 고정된 형태가 아니기 때문에 (그림 1)의 A그래프, B그래프와 같이 각 노드의 위치는 다르지만 서로 같은 그래프임을 보여주고 있다. 이러한 이유 때문에 고정된 크기의 데이터를 사용하는 CNN에서의 Convolution은 사용할 수 없고 그래프에 최적화된 Convolution을 사용해야 한다[3]. ConvGNNs은 그래프를 보다 쉽게 학습할 수 있기 때문에 GCN(Graph Convolution Network), MFConv, GraphSage, GAT(Graph Attention)등 파생모델들이 존재한다[3,7,8,9].

3. 방법론

PyTorch Geometric 라이브러리를 사용하여 ConvGNNs의 여러 모델인 GCN, MFConv, GraphSage, GAT의 레이어를 Node Classification과 Rating Prediction 협업 필터링 모델에 적용하여 성능을 측정한다. PyTorch Gemetric에서 지원하는 라이브러리에서 GCN의 경우 이기종 그래프는 사용할 수 없어 Node Classification만을 실험하였다.

4. 실험

<표 1> Node Classification 실험 결과

Models	Accuracy
GCN	0.77
MFConv	0.361
GraphSage	0.785
GAT	0.797

<표 2> Rating Prediction 실험 결과

Models	RMSE
MFConv	1.646
GraphSage	1.1529
GAT	1.1947

Cora Dataset과 MovieLens Latest Small Dataset을 사용하여 실험을 진행하였다. 사용한 모델의 경우 3절에서 언급한 모델들을 사용하였다. Cora dataset은 노드가 논문을 나타내고 엣지의 경우 인용을 나타내는 인용 그래프이다. 실험은 각각의 논문을 7가지의 카테고리로 분류하는 실험을 진행하였다. MovieLens Latest Small Dataset는 사용자와 영화가 노드이며 사용자가 영화에 부여한 평점을 갖는 이기종 그래프이다. 실험은 사용자별 영화에 대한 평점 예측을 진행하였다. Cora Dataset의 평

가지표는 Accuracy를 사용하였고 Accuracy의 경우 1에 가까울수록 논문의 분류를 정확하게 수행함을 의미한다. MovieLens Latest Small Dataset의 경우 추천시스템의 평가 지표 중 하나인 RMSE를 사용하였다. RMSE의 경우는 Accuracy와 다르게 낮을수록 평점의 예측을 더 잘한 것을 의미한다. 실험 결과 GAT모델은 Node Classification(표 1), GraphSage모델은 Rating Prediction 실험(표 2)에서 더 높은 성능을 나타내는 것을 확인할 수 있다.

5. 결론

본 연구는 Convolution 그래프 기반의 머신러닝 모델인 GCN, MFConv, GraphSage, GAT를 Node Classification과 Rating Prediction에 적용하여 보다 추천시스템에 적합한 모델을 찾는 것을 목표로 하였다. 실험결과 GAT는 Node Classification, GraphSage는 Rating Prediction 성능이 더 좋은 것을 확인하였다. 향후 계획으로는 다양한 그래프 기반 추천시스템에서의 실험 후 모델 개선 방법을 연구할 예정이다.

Acknowledgement

이 논문은 한국국토정보공사 공간정보연구원의 재원과 지원을 받아 수행된 지역상생 발전 R&D 사업인 '디지털 행정민원 플랫폼 고도화를 위한 메타버스 서비스 개발'과 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2023-2018-0-01417).

참고문헌

[1] Zhu, Yanqiao, et al. "Deep graph structure learning for robust representations: A survey." arXiv preprint arXiv:2103.03036 14 (2021).
 [2] Bhatti, Uzair Aslam, et al. "Deep Learning with Graph Convolutional Networks: An Overview and Latest Applications in Computational Intelligence." International Journal of Intelligent Systems 2023.
 [3] Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." arXiv preprint arXiv:1609.02907 (2016).
 [4] Jin, Wei, et al. "Graph structure learning for robust graph neural networks." Proceedings

of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020.

[5] 김정태, et al. "추천 분야에서의 지식 그래프 기반 어텐션 네트워크 모델 성능 향상 기법 연구." 한국정보처리학회 학술대회논문집 29.2 (2022): 603-605.

[6] Wu, Zonghan, et al. "A comprehensive survey on graph neural networks." IEEE transactions on neural networks and learning systems 32.1 (2020): 4-24.

[7] Duvenaud, David K., et al. "Convolutional networks on graphs for learning molecular fingerprints." Advances in neural information processing systems 28 (2015).

[8] Hamilton, Will, Zhitao Ying, and Jure Leskovec. "Inductive representation learning on large graphs." Advances in neural information processing systems 30 (2017).

[9] Veličković, Petar, et al. "Graph attention networks." arXiv preprint arXiv:1710.10903 (2017).