

실세계 동적 네트워크 분석을 위한 그래프 및 하이퍼그래프 학습 기술 성능 평가

유승경¹, 이다은¹, 고윤용², 김상욱^{3*}

¹한양대학교 AI 응용학과 석사과정

²중앙대학교 소프트웨어학부 교수

³한양대학교 컴퓨터소프트웨어학과 교수

ssong915@hanyang.ac.kr, ddanable05@hanyang.ac.kr, yyko@cau.ac.kr, wook@hanyang.ac.kr

Performance Evaluation of Graph and Hypergraph Learning for Real-world Dynamic Networks

Songkyung Yu¹, Daeun Lee¹, Yunyong Ko², Sang-Wook Kim^{3*}

¹Dept. of Artificial Intelligence Application, Hanyang University

²School of Computer Science and Engineering, Chung-Ang University

³Dept. of Computer Science, Hanyang University

요 약

그래프 학습 기술은 실세계의 네트워크를 그래프로 모델링하여 분석함으로써, 네트워크에 내재된 유용한 정보를 도출하는 핵심적인 역할을 한다. 그러나 기존의 그래프 학습 기술에는 두 가지 한계점이 존재한다: (1) 그룹 정보 표현의 한계 및 (2) 동적 관계 학습의 한계. 각 한계를 극복하기 위해 하이퍼그래프 학습 기술과 동적 그래프 학습 기술이 활발하게 연구되었지만, 두 가지 한계를 동시에 극복하기 위한 연구들은 아직까지 충분히 수행되지 못했다. 이러한 동기로부터, 본 논문은 실세계 네트워크를 동적인 하이퍼그래프로 모델링하여, 동적 그래프 학습 기술, 정적 하이퍼그래프 학습 기술, 그리고 동적 하이퍼그래프 학습 기술들의 성능에 대해 비교 분석하고자 한다. 실험 결과, 시간에 따라 변화하는 실세계의 복잡한 관계를 정확하게 포착하는 데는 동적 하이퍼그래프 학습 기술이 가장 효과적이라는 것을 확인하였다.

1. 서론

실세계에는 소셜 네트워크, 교통 네트워크, 단백질 상호작용 네트워크 등 다양한 네트워크가 존재한다. 이러한 실세계의 네트워크를 그래프를 통하여 분석함으로써 네트워크에 내재하여 있는 유용한 정보들을 도출할 수 있으며, 해당 정보들은 추천 시스템[1, 2, 3], 소셜 네트워크 분석[4], 생명정보학[5] 등 다양한 분야에서 활용될 수 있다.

그래프(graph)란 노드(node)와 엣지(edge)로 이루어진 데이터 구조로, 실세계 네트워크의 객체들과 그들의 관계를 각각 노드와 엣지로 모델링한다. 그래프 표현 학습(graph representation learning) 기술은 노드들의 관계(즉, 이웃 관계)를 기반으로 각 노드를 임베딩 벡터로 표현한다[6]. 그러나 다음 실세계 네트워크의 특징에 의해 기존의 그래프 표현 학습 기술들은 두 가지

한계점을 가지고 있다.

첫째, 실세계에는 세 개 이상의 객체들이 함께 형성하는 그룹 관계들이 흔하게 존재한다. 따라서 기존의 일반 그래프 구조로 해당 관계를 모델링할 경우에 그룹 정보의 손실이 불가피하게 발생하며, 이는 그래프 표현 학습 기술의 성능 저하시키는 근본적인 원인이다[7,8,9]. 둘째, 실세계의 네트워크는 시간의 경과에 따라 변화한다. 이러한 동적인 변화를 고려하지 않고, 네트워크를 정적으로만 학습할 경우, 객체들 간의 관계를 정확하게 파악하기 어렵다.

최근 첫 번째 한계를 극복하기 위하여 하이퍼그래프(hypergraph)라는 새로운 데이터 구조가 활발하게 연구되고 있다[7,8]. 하이퍼그래프란 일반화된 그래프 데이터 구조로, 임의의 수의 노드를 하나의 하이퍼엣지(hyperedge)로 모델링한다. 따라서 일반 그래프와 달

* 교신 저자

리, 여러 노드들이 함께 형성하는 그룹 관계를 하나의 하이퍼엣지로 모델링함으로써, 복잡한 그룹 관계도 정보 손실 없이 정확하게 표현할 수 있다. 결과적으로 네트워크를 하이퍼그래프로 모델링하여 학습함으로써 일반 그래프 기반의 학습 기술의 성능을 개선할 수 있다[7,8].

또한, 두 번째 한계를 극복하기 위해 다양한 동적 그래프 학습(dynamic graph learning) 기술[10,11]들이 연구되고 있다. 해당 기술들은 노드들의 관계 파악 시 해당 관계가 발생한 시간 정보도 함께 고려함으로써, 동적으로 변화하는 그래프 내 노드들의 관계를 파악하고자 했다.

앞서 설명한 바와 같이, 기존 그래프 표현 학습 기술의 한계를 극복하기 위하여 하이퍼그래프 학습 기술과 동적 그래프 학습 기술이 연구되었지만, 두 한계점을 동시에 극복하기 위한 동적 하이퍼그래프 학습 기술에 대한 연구 및 분석은 아직 충분히 수행되지 않았다. 이러한 동기로부터, 본 논문에서는 동적으로 변화하는 실세계 네트워크를 동적 하이퍼그래프로 모델링하여, 기존의 그래프 및 하이퍼그래프 학습 기술들의 성능을 평가 및 분석하고자 한다.

2. 관련 연구

그래프 학습 기술은 그래프의 엣지 정보를 바탕으로 노드들의 관계를 학습하여 노드를 저차원의 임베딩 벡터로 표현한다. 그래프 학습 기술에 의해서 표현된 노드 임베딩은 노드 분류[12,13], 링크 예측[14,15] 등 다양한 다운스트림 작업(downstream task)에 적용될 수 있다.

일반적인 그래프 학습 기술들은 그래프에서 엣지로 표현된 노드들 간의 쌍상호관계를 기반으로 노드들의 관계를 학습한다[16,17]. 그러나 일반 그래프 학습 기술은 실세계 네트워크에 존재하는 다수 노드들의 복잡한 그룹 관계를 학습하는 데에 어려움이 있다.

이러한 문제를 극복하기 위해 하이퍼엣지로 표현되는 노드들의 그룹관계를 학습하기 위한 하이퍼그래프 신경망 모델(hypergraph neural network)이 제안되었다 [7,8].

다른 한편으로는, 네트워크가 시간의 흐름에 따라 진화하는 동적 특징을 학습하기 위한 동적 그래프 학습 기술도 연구되었다[10].

그러나 아직까지 두 관점을 모두 반영하여 네트워크를 학습하기 위한 기술에 대해서는 충분히 연구되지 못했다. 따라서 본 논문은 두 관점을 모두 고려하여 (1) 동적 그래프 학습 기술과 (2) 정적 하이퍼그래프 학습 기술의 성능을 평가하고 그 결과를 분석하고자 한다.

3. 실험

우리는 실험을 통하여 동적인 환경에서의 1. 일반 그래프와 하이퍼그래프 학습 기술의 성능 비교와 2. 정적 하이퍼그래프와 동적 하이퍼그래프 학습 기술의 성능 비교를 통하여 동적인 환경에서의 효과적인 그래프 학습 기술이 무엇인지 분석하고자 한다.

3.1 실험 환경

우리는 시간 정보를 포함한 동적 하이퍼그래프 데이터셋을 사용하여 실험을 진행하였다: email-Enron, email-Eu, contact-primary-school, contact-high-school. 일반 그래프 모델과 하이퍼그래프 모델에서 동일한 데이터셋을 활용한 실험이 가능하도록 하이퍼그래프 데이터셋에 클릭확장 기법(clique expansion)을 적용하여 실험을 진행하였다. 데이터셋에 대한 통계는 <표 1>에 나타내었다.

<표 1> 데이터셋에 대한 통계

Dataset	노드 수	하이퍼엣지 수
email-Enron	143	10,883
email-Eu	979	234,760
contact-high-school	327	172,035
contact-primary-school	242	106,879

동적 그래프 학습 모델 ROLAND[10]에 일반 그래프 학습 모델 GCN[17]과 하이퍼그래프 학습 모델 HGNN[7]을 적용하여 성능을 비교하였다. 공정한 비교를 위해, 우리는 모든 방법들의 임베딩 벡터 차원을 64로 설정하였다.

성능 평가는 엣지 예측 실험을 통하여 진행하였다. 엣지 예측이란 미래 엣지의 존재 여부를 얼마나 정확하게 예측하는지를 평가하는 것이다. 평가 지표로는 Average Precision(AP), Area Under the Receiver Operating Characteristic Curve(AUROC)를 사용하였다.

3.2 실험 결과

3.2.1 단순/하이퍼그래프 학습 기술 성능 비교

<표 2: Dynamic GCN/HGNN>는 동적인 환경에서의 일반 그래프와 하이퍼그래프 학습 기술의 엣지 예측 정확도를 나타낸 것이다. 우리는 대부분의 결과에서 HGNN이 GCN보다 예측 정확도 성능이 높음을 확인할 수 있다. 이러한 결과는 하이퍼그래프가 일반 그래프보다 실세계에서의 고차원 관계를 잘 반영한다는 것을 의미한다. 이를 통해 그래프 구조가 광범위하게 활용되기 위해서는 일반 그래프보다 하이퍼그래프가 유용하다는 것을 확인하였다.

3.2.2 동적/정적 하이퍼그래프 학습 기술 성능 비교

<표 2: Dynamic HGNN/Static HGNN>는 동적인 환경을

<표 2> 엣지 예측 정확도

	Dynamic				Static	
	GCN		HGNN		HGNN	
	AUROC	AP	AUROC	AP	AUROC	AP
email-Enron	0.6609	0.7360	0.6622	0.7430	0.5588	0.5413
email-Eu	0.4920	0.5306	0.5077	0.5475	0.5046	0.5001
contact-high-school	0.6146	0.9905	0.6028	0.9906	0.5156	0.5083
contact-primary-school	0.5336	0.5443	0.5167	0.5708	0.4735	0.4852

고려한 학습 기술인 동적 하이퍼그래프 학습과 고려하지 않은 정적 하이퍼그래프 학습 기술의 엣지 예측 정확도를 나타낸 것이다. 우리는 동적 구조를 활용한 동적 하이퍼그래프 학습 기술에서 엣지 예측 정확도 성능이 높음을 확인할 수 있다. 이러한 결과를 통해 정적 하이퍼그래프는 실세계의 시간 정보를 담는 데에 어려움이 있음을 확인할 수 있다. 이를 통해 동적 실세계 동적으로 변화하는 관계를 반영하기 위해서는 동적 그래프 구조가 효과적임을 확인하였다.

4. 결론 및 향후 연구

본 논문에서는 그래프 기반 학습의 두가지 주요한 측면인 하이퍼그래프와 동적 그래프를 결합한 동적 하이퍼그래프 학습 기술에 대한 효과를 탐구하였다. 4개의 데이터셋들을 이용한 실험을 통하여, 대부분의 결과에서 동적 하이퍼그래프의 결과가 우수함을 확인하였다. 이러한 결과는 실세계의 동적인 관계를 학습하는 데에 하이퍼그래프와 동적 그래프를 모두 활용해야 한다는 것을 보여준다. 이를 위하여 정적인 환경, 동적인 환경에서 모두 사용될 수 있는 일반화된 동적 하이퍼그래프 학습 기술의 연구가 필요할 것으로 보인다.

사사

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2022-0-00352)과 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS-2022-00155586, 실세계의 다양한 다운스트림 태스크를 위한 고성능 빅 하이퍼그래프 마이닝 플랫폼 개발(SW 스타랩))

참고문헌

- [1] Fan, Wenqi, et al. "Graph neural networks for social recommendation." The world wide web conference. 2019.
- [2] Yu, Chia-An, et al. "Modeling multi-way relations with hypergraph embedding." Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018.
- [3] Zhang, Muhan, et al. "Beyond link prediction: Predicting hyperlinks in adjacency space." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018.
- [4] Fan, Wenfei. "Graph pattern matching revised for social network analysis." Proceedings of the 15th international conference on database theory. 2012.
- [5] Zhang, Xiao-Meng, et al. "Graph neural networks and their current applications in bioinformatics." Frontiers in genetics 12 (2021): 690049.
- [6] Grover, Aditya, and Jure Leskovec. "node2vec: Scalable feature learning for networks." Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. 2016.
- [7] Feng, Yifan, et al. "Hypergraph neural networks." Proceedings of the AAAI conference on artificial intelligence. Vol. 33. No. 01. 2019.
- [8] Dong, Yihe, Will Sawin, and Yoshua Bengio. "Hnhn: Hypergraph networks with hyperedge neurons." arXiv preprint arXiv:2006.12278 (2020).
- [9] Yang, Chaoqi, et al. "Semi-supervised hypergraph node classification on hypergraph line expansion." Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022.
- [10] You, Jiaxuan, Tianyu Du, and Jure Leskovec. "ROLAND: graph learning framework for dynamic graphs." Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining. 2022.
- [11] Rossi, Emanuele, et al. "Temporal graph networks for deep learning on dynamic graphs." arXiv preprint arXiv:2006.10637 (2020).
- [12] Rong, Yu, et al. "Dropedge: Towards deep graph convolutional networks on node classification." arXiv preprint arXiv:1907.10903 (2019).
- [13] Wu, Qitian, et al. "Nodeformer: A scalable graph structure learning transformer for node classification." Advances in Neural Information Processing Systems 35 (2022): 27387-27401.
- [14] Yadati, Naganand, et al. "Nhp: Neural hypergraph link prediction." Proceedings of the 29th ACM international conference on information & knowledge management. 2020.
- [15] Yoon, Se-eun, et al. "How much and when do we need higher-order information in hypergraphs? a case study on hyperedge prediction." Proceedings of The Web Conference 2020. 2020.
- [16] Defferrard, Michaël, Xavier Bresson, and Pierre Vandergheynst. "Convolutional neural networks on graphs with fast localized spectral filtering." Advances in neural information processing systems 29 (2016).
- [17] Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." arXiv preprint arXiv:1609.02907 (2016).