메모리뱅크와 모멘텀 학습 도입을 통한 그래프 대조 학습 성능 향상

조우성, 이재구*
국민대학교

*jaekoo@kookmin.ac.kr

Performance Improvement of Graph Contrastive Learning via Introduce Memory Bank and Momentum Learning

Wooseong Cho, Jaekoo Lee*
College of Computer Science, Kookmin University

요 약

자기 지도 학습 (Self-Supervised Learning)은 최근 기계 학습 분야에서 매우 효과적인 접근 방법으로 입증되었다. 특히 대조 학습 (Contrastive Learning)을 활용하는 방법들이 부상하고 있으며, 그래프 기반의 기계 학습에서도 효과적임이 입증되었다. 본 논문에서는 메모리 뱅크 (Memory Bank)에 부정적인 쌍을 사전 할당하고, 모멘텀 (Momentum)을 이용하여 모델을 갱신하는 방법을 도입하여 그래프 기반 대조 학습의 성능을 더욱 향상시켰다. 자기 지도 학습을 통해 얻은 임베딩 (Embedding)을 활용하여 다양한 후속 과업을 수행할 때, 기존 자기 지도 학습 모델을 포함한 비교 모델 (Baseline Model)들보다 우수한 성능을 보였다. 이 결과는 이미 사진 과업 분야에서 메모리 뱅크와 모멘텀을 보완한 다양한 모델들 또한 그래프 기반 모델에 활용하여 더욱 발전시킬 수 있는 가능성을 보여준다.

Ⅰ. 서론

사진, 그래프 관련 과업 모두 지도 학습 (Supervised Learning) 기반의 모델이 뛰어난 성능을 보여 주고 있다. 하지만 지도 학습은 데이터 라벨링 (Labeling)에 대한 비용과 시간적 제약이 커 실제 적용 범위가 제한된다. 이러한 한계로 인해 최근에는 자기 지도 학습 (Self-Supervised Learning)이 많은 관심을 받고 있다.

자기 지도 학습은 라벨이 없는 데이터로 입력 특징을 사전 학습하는 방식이다. 특히 대조 학습 (Contrastive Learning)은 자기 지도 학습의 주요한 방법으로, 긍정적인 쌍 (Positive Pair)과 부정적인 쌍 (Negative Pair)의 대조를 통해 입력의 특징을 학습한다. 이러한 자기 지도 학습 방법은 대량의 데이터를 라벨 없이 활용하여 더욱 강력한 모델을 학습시킬 수 있는 장점을 가지고 있다.

본 논문에서는 그래프 기반 자기 지도 학습에 사진 관련 분야의 대조 학습 모델인 MoCo[1]에서 사용한 메모리 뱅크 (Memory Bank)와 모멘텀 (Momentum)을 도입한 새로운 시도, GMOCL (Graph MOmentum Contrastive Learning)를 제안한다. 이는 기준 모델 (Baseline Model)들보다 여섯 가지 그래프 분류 (Graph Classification) 데이터 집합 모두에서 가장 우수한 성능을 보였다. 이런 결과는 MoCo 를 응용한 다른 모델 또한 그래프 기반 자기 지도 학습에 적용하여 성능을 높일 수 있는 가능성을 보여준다.

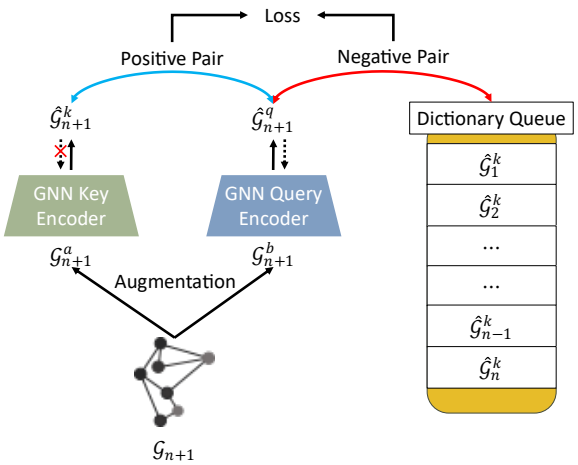

그림 1. GMOCL 모델 구조

Ⅱ. 관련 연구

메모리 뱅크 (Memory Bank)

MoCo 는 메모리 뱅크를 활용하여 부정적인 쌍을 사전 큐 (Dictionary Queue)로 미리 할당한다. 이 방법은 부정적인 쌍을 배치 (Batch) 단위로 사용하지 않고 사전 할당함으로써 더 적은 메모리로 더 많은 부정적인 쌍을 활용하여 효율적인 학습을 가능하게 한다.

모멘텀 (Momentum)

일반적으로 긍정적인 쌍과 부정적인 쌍을 만들 때 같은 모델을 공유하여 사용하는 반면[2], MoCo 에서는 서로 다른 두 모델, 쿼리 인코더 (Query Encoder)와 키 인코더 (Key Encoder)를 사용한다. 이때 쿼리 인코더만 역전파를 통해 갱신되며 키 인코더는 역전파가 흐르지 않는다. 대신, 키 인코더의 가중치(θ_k)는 쿼리 인코더의 가중치(θ_q)를 이용하여 식과 같이 업데이트 된다.

$$\theta_k \leftarrow m\theta_k + (1-m)\theta_q \quad (1)$$

안정성을 위해 0.999와 같은 큰 m을 이용하여 키 인코더를 천천히 갱신하며, 이와 같은 방법을 모멘텀이라 한다.

그래프 대조 학습 (Graph Contrastive Learning)

본 논문에서는 GraphCL[3]을 그래프 대조 학습의 기준 모델로 사용하였다. GraphCL 은 SimCLR[2]과 마찬가지로 같은 데이터에 대한 두 변형 (Augmentation)에 대

표 1. 실험 결과

학습 방법	데이터 / 모델	COLLAB[5]	DD[5]	MUTAG[5]	NCI1[5]	PROTEINS[5]	PTC_MR[5]
지도 학습	GAT[6]	67.4±2.9	64.3±1.4	89.4±6.1	66.6±2.2	71.7±3.6	66.7±5.1
	GIN[4]	80.1±1.9	75.1±5.1	89.4±5.6	82.7±1.7	78.5±5.3	64.6±7.0
그래프 커널	WL[7]	78.9±1.9	72.3±3.4	80.7±3.0	80.0±0.5	72.9±0.6	58.0±0.5
	DGK[8]	73.1±0.3	67.0±0.6	87.4±2.7	80.3±0.5	73.3±0.8	60.1±2.6
비지도 학습	InfoGraph[9]	70.7±1.1	72.9+ 1.8	89.0±1.1	76.2±1.4	74.4±0.3	61.7±1.4
	GraphCL[3]	71.4±1.2	78.6±0.4	86.8±1.3	77.9±0.4	74.4±0.5	61.3±2.1
	GMOCL (ours)	**84.8±4.1**	**84.7±2.2**	**90.0±0.0**	**82.9±8.2**	**83.2±7.4**	**70.9±1.1**

한 두 특징 벡터를 긍정적인 쌍으로, 나머지를 부정적인 쌍으로 정의하여 대조 손실 함수 (Contrastive Loss)로 학습한다. GraphCL 은 그래프 신경망 (Graph Neural Network, GNN)을 통해 특징 벡터를 생성하며, 변형은 노드 제거 (Node Dropping), 엣지 제거 (Edge Removing), 특징 마스킹 (Attribute Masking)과 부분 그래프 추출 (Subgraph Extraction)의 네 가지를 사용한다.

Ⅲ. 실험

본 논문에서 제안한 GMOCL 은 GraphCL 모델에 메모리 뱅크와 모멘텀을 도입한 것으로, 구조는 [그림 1]과 같다. [그림 1]은 크기 n의 사전 큐를 사용하고, $n+1$번째 반복에서의 상황을 나타낸다. 입력의 두 변형이 각각 쿼리와 키 인코더로 들어가서 입력 쿼리, 키 특징 벡터 $(\hat{g}^q, \hat{g}^k)$를 생성한다. 이 둘은 긍정적인 쌍을, 사전 큐의 모든 벡터는 부정적인 쌍을 구성한다. 사전 큐는 큐의 FIFO (First In First Out) 특징에 따라 가장 오래된 벡터가 제거되고 새로 생성된 키 특징 벡터가 들어가서 갱신된다. 한편, 쿼리 인코더의 가중치 θ_q 만 역전파를 통해 학습되며 키 인코더의 가중치 θ_k는 [수식 1]의 모멘텀 방식으로 학습된다. GNN 인코더는 GIN[4]을 사용하였다.

사용한 데이터 집합은 TUDataset[5]의 그래프 분류 데이터 집합 중 6 가지, COLLAB, DD, MUTAG, NCI1, PROTEIN, PTC_MR 을 사용하였다. 기준 모델은 지도 학습 기반의 GAT[6], GIN, 그래프 커널 (Graph Kernel) 기반의 WL[7], DGK[8], 비지도 학습 기반의 InfoGraph[9], GraphCL 의 총 6 가지이다.

사전 학습 후의 성능 평가는 GraphCL 과 같은 방법으로 진행하였다. 10-겹 교차 검증 (10-Fold Cross Validation)을 서로 다른 랜덤 시드 (Random Seed)에서 5 번 반복하고, 선형 서포트 벡터 머신 (Linear Support Vector Machine, LinearSVM)으로 그래프 분류를 수행하였다. LinearSVM 은 사전 학습과 같은 교차 검증 설정으로 학습하여 그중 가장 높은 정확도를 사용하였다. 실험 결과는 서로 다른 랜덤 시드에서 얻은 다섯 개의 정확도에 대한 평균±표준편차를 [표 1]에 나타냈다. 굵은 글씨는 데이터 집합 별로 가장 높은 성능을 나타낸다.

실험 결과 GMOCL 은 모든 데이터 집합에서 가장 높은 성능을 보였다. 이는 부정적인 쌍을 더 많이 사용하여 입력 특징을 잘 추출하였고, 서로 다른 인코더가 모멘텀 방식을 통해 더 안정적으로 학습되었기 때문이다.

Ⅳ. 결론

본 논문에서는 그래프 기반 자기 지도 학습 중 대조 학습에 메모리 뱅크와 모멘텀을 활용한 방법인 GMOCL 을 제안하였다. GMOCL 은 그래프 데이터의 특성을 다른 모델보다 효율적으로 추출하여 실험 결과 다른 자기 지도 학습 모델 뿐 아니라 지도 학습, 그래프 커널 모델보다도 우수한 성능을 보여주었다.

즉, 본 논문은 사진 과업 분야와 마찬가지로 MoCo 의 메모리 뱅크와 모멘텀 방법이 그래프 분야에도 탁월한 성과를 보임을 확인했다. 이는 MoCO 를 응용하여 발전시킨 다른 대조 학습 모델 구조 또한 그래프 기반 대조 학습에 효과적일 것이란 가능성을 보여준다.

이러한 결과들은 특히 라벨링이 어려운 그래프 기반 자기 지도 학습의 다양한 응용분야에서 그래프 데이터의 효과적인 학습에 기여할 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2022-00167194,미션 크리티컬 시스템을 위한 신뢰 가능한 인공지능)

참 고 문 헌

[1] He, K., et al. (2020). Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9729-9738).

[2] Chen, T., et al. (2020). Simclr: A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning (pp. 1597-1607).

[3] You, Y., et al. (2020). Graph contrastive learning with augmentations. Advances in neural information processing systems, 33, 5812-5823.

[4] Xu, K., et al. (2018). How powerful are graph neural networks?. arXiv preprint arXiv:1810.00826.

[5] Morris, C., et al. (2020). Tudataset: A collection of benchmark datasets for learning with graphs. arXiv preprint arXiv:2007.08663.

[6] Veličković, P., et al. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.

[7] Shervashidze, N., et al. (2011). Weisfeiler-lehman graph kernels. Journal of Machine Learning Research, 12(9).

[8] Yanardag, P., & Vishwanathan, S. V. N. (2015, August). Deep graph kernels. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1365-1374).

[9] Sun, F. Y., et al. (2019). Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. arXiv preprint arXiv:1908.01000.