

정교한 이웃 노드 선택법을 활용한 그래프 합성곱 네트워크 (Graph Convolutional Networks with Elaborate Neighborhood Selection)

정연성[†] 황지영^{**}
(Yeonsung Jung) (Joyce Jiyoun Whang)

요약 그래프 합성곱 네트워크(GCNs)는 합성곱 구조를 활용하여 주변 노드들의 정보를 종합하는 방식으로 대상 노드의 표현력을 높인다. 높은 성능을 보이기 위해서는 우선적으로 대상 노드에게 필요한 정보를 전달할 수 있는 주변 노드를 선별하고, 이후 학습시 적절한 필터(filter) 값을 습득하는 과정이 수반되어야 한다. 최근 GCNs 알고리즘들은 1-hop 거리의 노드들을 선택하는 등의 비교적 간단한 이웃 노드 정의를 활용하고 있다. 이러한 경우 불필요한 정보가 대상 노드에 전파되어 성능을 저하하는 문제가 발생한다. 본 논문에서는 대상 노드와 주변 노드간의 유사도 계산을 통해 유효한 이웃 노드를 선별하여 활용하는 GCN 알고리즘을 제안한다.

키워드: 노드 분류, 그래프 합성곱 네트워크, 그래프 신경망, 그래프 마이닝

Abstract Graph Convolutional Networks (GCNs) utilize the convolutional structure to obtain an effective insight on representation by aggregating the information from neighborhoods. In order to demonstrate high performance, it is necessary to select neighborhoods that can propagate important information to target nodes, and acquire appropriate filter values during training. Recent GCNs algorithms adopt simple neighborhood selection methods, such as taking all 1-hop nodes. In the present case, unnecessary information was propagated to the target node, resulting in degradation of the performance of the model. In this paper, we propose a GCN algorithm that utilizes valid neighborhoods by calculating the similarity between the target node and neighborhoods.

Keywords: node classification, graph convolutional network, graph neural network, graph mining

· 본 연구는 정부(교육부, 과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 사업이며(NRF-2016R1D1A1B03934766, NRF-2019R1C1C1008956). 또한, 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과임(2015-0-00914)

† 학생회원 : 성균관대학교 전자전기컴퓨터공학과
ys.jung@skku.edu

** 종신회원 : 성균관대학교 소프트웨어학과 교수
(Sungkyunkwan Univ.)
jjwhang@skku.edu
(Corresponding author임)

논문접수 : 2019년 8월 12일

(Received 12 August 2019)

논문수정 : 2019년 9월 5일

(Revised 5 September 2019)

심사완료 : 2019년 9월 9일

(Accepted 9 September 2019)

Copyright©2019 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 제46권 제11호(2019. 11)

1. 서론

Graph Convolution Networks[12]는 최근 연구되는 그래프 신경망분야의 대표적인 모델 중 하나로, 이미지 처리 분야에서 높은 성능을 보이는 합성곱의 개념을 활용하는 모델이다. 합성곱[15]은 여러 정보를 합성하는 것을 의미하며 이미지에서는 주로 특정 범위 내의 픽셀 값들을 합성하여 패턴을 인식하는 목적으로 사용된다. 이미지 데이터의 경우 그리드(grid)형태로 표현이 되기 때문에, 픽셀들의 위치와 순서가 고정되어있어 문맥(context)을 정의하기 쉽다. 하지만 그래프 데이터의 경우 노드들의 위치가 고정되어있지 않고 서로간의 순서가 정해져있지 않기 때문에 해당 개념을 그대로 대입하여 사용할 수 없다.

이를 보완하기 위한 여러 연구들이 진행이 되었고 크게 스펙트럴 기반 모델과 비 스펙트럴 기반 모델로 나눌 수 있다. 먼저, 스펙트럴 기반 모델은 그래프 스펙트럴의 고유값 분해(eigen decomposition)를 이용하거나, 고유값 기저(eigen basis)에 기반을 두고 필터 학습을 진행하기 때문에 많은 계산량을 필요로 한다. 비 스펙트럴 기반 모델의 경우 군집화 연구[6]와 같이 그래프 상의 구조적인 거리를 활용하여 합성곱 개념을 보다 직관적으로 적용한다. 다시 말해, 그래프 상에서 가까운 거리에 위치한 노드들을 문맥으로 정의하고 이를 대상으로 합성곱을 진행한다. 해당 방법론의 대표적인 이슈는 노드별로 서로 다른 개수의 이웃 노드를 가지고 있는 상황에서 특정 개수의 가중치 파라미터를 공유하는 모델을 학습하는 것이다.

최근 많은 연구에서 비 스펙트럴 기반 모델이 큰 스케일의 벤치마크를 대상으로 스펙트럴 기반 모델에 비해 노드 분류에서 높은 성능을 보이고 있다[11, 13]. 사용되는 벤치마크 데이터는 주로 인용 네트워크와 단백질-단백질 상호작용 네트워크와 같이 관련성이 높은 노드들 사이에 간선이 존재할 확률이 매우 높고, 대상 노드에 대한 설명력이 2-hop 혹은 그 이상의 노드에 비해 높은 네트워크다. 하지만 소셜 네트워크나 웹 네트워크 등 다른 도메인에서는 이와 같은 특성이 존재하지 않을 가능성이 더 높기 때문에 이웃 노드를 그래프 특성에 맞게 정의하는 것이 중요하다[8]. 본 논문에서는 정교한 이웃 노드 선택법을 활용하여 각 노드의 문맥을 형성하고, 이를 합성곱을 활용하여 종합하는 알고리즘을 제안한다.

2. 관련 연구

스펙트럴 기반 GCNs 모델의 경우 그래프 라플라시안의 고유값 분해를 계산하여 합성곱의 개념을 푸리에 도

메인에서 정의하는 연구[5,9]들이 진행되었는데, 많은 계산량과 위치를 활용한 지역적 필터(localized filter)를 사용하지 않는 한계가 있다. 이러한 한계점들을 보완하고 필터의 범위를 각 노드별 1-hop 노드로 제한하는 연구[12]가 진행되었지만, 스펙트럴 기반 모델의 경우 필터 학습을 위해 라플라시안 고유값 기저를 사용하기 때문에 다른 그래프에 바로 적용할 수가 없다는 공통적인 단점이 존재한다.

비 스펙트럴 기반 모델의 경우, 스펙트럴 기반 모델보다 직관적으로 합성곱 개념을 적용하였다. 해당 모델의 주요 이슈는 고정된 크기의 필터에 맞추어 이웃 노드를 특정 개수만큼 선택하는 것이다. 1-hop 노드를 모두 이웃 노드로 정의하거나, 그 중 특정 개수를 선택해서 사용하는 등의 방법을 채택한 연구[4, 11]가 진행되었다.

하지만 이러한 매우 가까운 노드를 선택하는 접근법은 관련이 높은 노드들끼리의 간선이 존재할 확률이 큰 그래프를 대상으로 한다는 전제가 존재한다. 웹과 소셜 네트워크 등 노이즈가 많은 도메인에서는, 이미지의 대상 픽셀과 주변 픽셀이 비슷한 값을 가지고 있는 것에 기반을 둔 기법을 그대로 적용하기는 어렵다. 따라서 대상 노드의 문맥이 되는 이웃 노드들을 어떻게 정의하고, 어떠한 방식으로 적절한 가중치를 부여하여 많은 정보를 종합(aggregation)하는지가 중요하다. 최근 이를 해결하기 위한 대표적인 알고리즘으로 GAT 모델[11]이 있다. 이 모델의 경우, Multi-head attention 기법을 적용하여 1-hop노드들과 특성 벡터간의 유사도를 비교하여 적절한 가중치를 부여한다. 노이즈에 해당하는 노드에게는 적은 가중치를, 유효한 노드에게는 높은 가중치를 부여함으로써 보다 높은 성능을 보인다.

3. 알고리즘

3.1 기존 연구의 문제점

기존 연구들의 경우 대상 노드를 설명하기 위해 1-hop 노드들만을 사용하거나, 1-hop 노드들을 대상으로 각 feature별 가장 큰 k개의 값을 추출하여 가상의 노드를 생성하여 이웃 노드로 사용하였다. 이러한 방법론은 여러 벤치마크 데이터를 대상으로 뛰어난 성능을 보였다. 하지만 이는 사용된 벤치마크 데이터의 도메인이 인용 네트워크, 단백질-단백질 상호작용 네트워크와 같이 관련성이 높은 노드사이에 간선이 생성되어 있을 확률이 매우 높은 분야이기 때문이다. 또한 이러한 도메인은 1-hop 노드들의 설명력이 2-hop 혹은 그 이상의 노드들에 비해 매우 강한 특성을 보인다. 하지만 소셜 네트워크와 웹 네트워크와 같이 많은 노이즈가 존재하는 도메인들이 존재한다. 이러한 도메인은 1-hop 노드들보다 설명력이 높은 노드들이 2-hop 혹은 그 이상의

거리에 존재할 가능성이 적지 않다. 따라서 다양한 도메인에서 좋은 성능을 보이기 위해서는 관련성이 높은 이웃 노드를 적절히 정의 하는 것이 매우 중요하다.

3.2 제안 알고리즘

3.2.1 유사도를 활용한 이웃 노드 선택

대상 노드를 기준으로 1-hop 노드들과 2-hop 노드들을 후보로 특성 벡터간의 유사도를 계산한 후 순위를 측정한다. 노드별 차수(degree)를 기준으로 일정 비율만큼 유사도 점수에서 상위에 위치한 노드들을 이웃 노드로 정의한다. 대상 노드와 이웃 노드들의 특성 벡터를 합성곱 구조를 활용하여 가중합함으로써 대상 노드의 표현력을 높인다. 그림 1과 같이 유사도 함수를 활용하여 보다 정교한 방법으로 이웃 노드를 정의함으로써 1-hop에 존재하는 이상치(outlier)를 제거함과 동시에 직접적인 연결이 없어 경시되기 쉬운 2-hop에 존재하는 중요 정보에 높은 가중치를 부여하는 효과를 볼 수 있다.

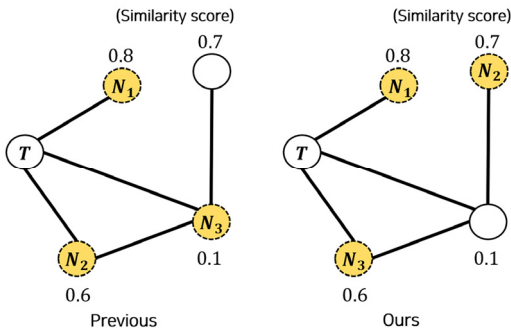


그림 1 정교한 이웃 노드 선택
Fig. 1 Elaborate selection of neighborhoods

3.2.2 Damping Factor

그래프 구조적인 측면에서 대상 노드를 기준으로 거리가 가까운 노드일수록 관련성이 높을 확률이 크고 멀수록 관련성이 낮다는 해석이 가능하다. 앞서 제안한 이웃 노드 선택법에서는 1-hop 노드와 2-hop 노드 간의 그래프 구조적인 차이를 반영하지 않고 동등하게 처리하였다. 보다 높은 성능을 위해 구조적인 측면을 반영하여 대상 노드로부터 거리가 멀수록 유사도 점수가 일정 비율로 감소할 수 있도록 그림 2와 같이 damping factor를 적용한다. 본 실험에서는 0.5의 값을 사용하였으나, 각 도메인에 따라 거리와 관련성의 상관 정도는 다르기 때문에 학습을 통해 적절한 값을 선택한다.

3.2.3 그래프 합성곱 네트워크

대상 노드를 기준으로 선택된 이웃 노드의 가중치를 Multi-head Attention 기법[2]을 활용하여 계산한다. 각 이웃 노드별 가중치를 부여하여 합하는 방식으로 그림 3과

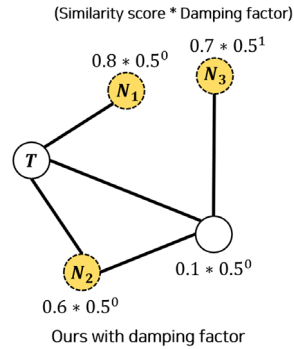


그림 2 damping factor를 활용한 이웃 노드 선택
Fig. 2 Selection of neighborhoods with a damping factor

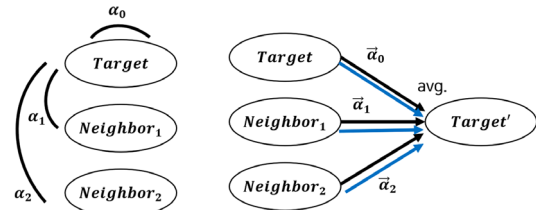


그림 3 Attention기법을 활용한 합성곱
Fig. 3 Convolution using Multi-head Attention

같이 합성곱을 진행한다. 선별된 이웃 노드들을 활용하여 필터별로 새로운 대상 노드의 표현을 학습한 뒤, 학습된 표현들을 연결(concatenate)하거나 평균(average)을 계산하여 최종적인 특성 벡터를 도출한다.

4. 실험

4.1 실험 데이터

다양한 도메인에서의 노드 분류 실험을 진행하기 위해 인용 네트워크 데이터 셋 Cora[10], Citeseer[10], 블로그 소셜 네트워크 데이터 셋 BlogCatalog[14] 그리고 Co-purchasing 네트워크인 Amazon Photo[7]를 사용하였다. 데이터 별 노드와 간선은 각각 논문과 인용관계, 블로그와 친구관계, 상품과 공동구매를 의미하고, 레이블은 논문, 블로그, 상품이 속한 카테고리를 의미한다. 기본적인 데이터 정보는 표 1과 같다.

표 1 데이터 요약
Table 1 Data analysis

Dataset	Blog Catalog	Amazon Photo	Cora	Citeseer
Classes	6	8	7	6
Nodes	5,196	7,487	2,708	3,312
Edges	171,743	126,530	5,429	4,715
Features	8,189	745	1,433	3,703

4.2 실험 세팅

각 노드별 1-hop과 2-hop 노드들을 대상으로 유사도를 계산한다. 이때 유사도 계산은 코사인 유사도(cosine similarity), 내적(dot product), L1-norm, L2-norm을 사용하여 진행되었다. 계산된 유사도 점수를 기준으로 일정 개수의 상위 노드를 선택한다. CNN의 필터에 해당하는 attention heads의 개수와 아웃풋의 차원, 활성화 함수는 [11]과 같이 각각 8, 8, Exponential Linear Unit(ELU)를 사용하였다. 이후 완전 연결 계층(Fully-

connected layer)을 활용하여 아웃풋의 차원이 카테고리(label) 개수가 되도록 한다. 이후 softmax cross-entropy 함수를 활용하여 노드 분류 실험을 진행한다. 기존 연구들과 같은 조건으로, 학습시 클래스별 20개 데이터와 검증 데이터 500개 그리고 테스트 데이터 1,000개를 사용하였다. 자세한 정보는 표 2와 같다.

4.3 평가 방법

이웃 노드 선택시 노드별 차수의 10%부터 100%까지 10%단위로 개수를 세분하여 실험하였고, 유사도 함수는 앞서 말한 4가지에 대한 결과를 보였다. 또한 데이터 별로 가장 좋은 성능을 보이는 모델에 damping factor 0.5를 적용하여 그림 2에 'Ours(damping)'으로 제시하였다. 기존 연구와의 성능 비교를 위해 GAT[11], LGCN[4], Deepwalk[3], node2vec[1] 모델을 베이스라인으로 사용하였고, 성능 평가는 모델별 총 30회 수행의 평균 정확도(accuracy)를 기준으로 사용하였다.

표 2 실험 세팅
Table 2 Experimental setting

Dataset	Blog Catalog	Amazon Photo	Cora	Citeseer
Training	120	160	140	120
Validation	500	500	500	500
Test	1,000	1,000	1,000	1,000

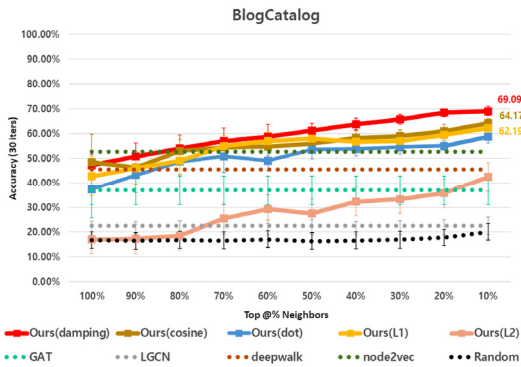


그림 2(a) BlogCatalog 실험 결과
Fig. 2(a) BlogCatalog result

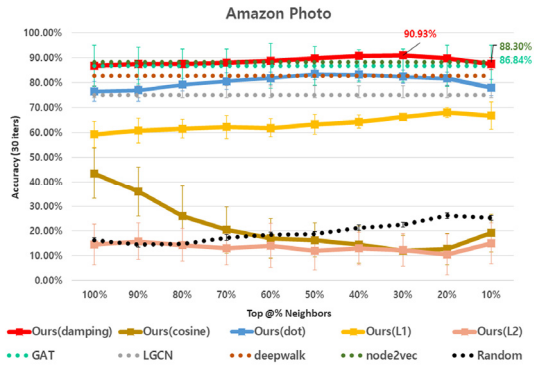


그림 2(b) Amazon Photo 실험 결과
Fig. 2(b) Amazon Photo result

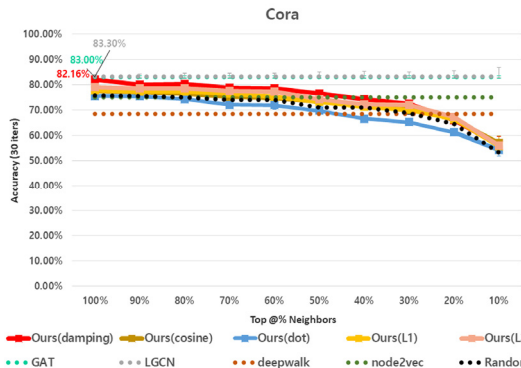


그림 2(c) Cora 실험 결과
Fig. 2(c) Cora result

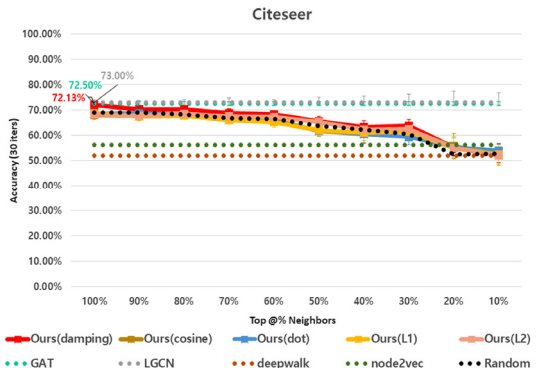


그림 2(d) Citeseer 실험 결과
Fig. 2(d) Citeseer result

4.4 실험 결과

BlogCatalog는 웹과 소셜 네트워크의 특성을 함께 가지고 있는 데이터로, 대상 블로그와 친구관계인 블로그들이 다른 카테고리에 속하는 경우가 많다. 또한 2-hop 거리에 대상 블로그와 같은 카테고리에 속한 유사 블로그가 존재할 확률도 높다. Amazon Photo는 아마존의 공동구매 네트워크로 인용 네트워크만큼은 아니지만 함께 구매한 제품은 같은 카테고리에 속할 확률이 높은 특성을 보인다.

BlogCatalog 실험 결과, 4가지 유사도 함수를 적용하였을 때 기존 연구들에 비해 성능이 향상된 것을 그림 2(a)에서 확인할 수 있다. Amazon Photo 실험 결과, damping factor를 적용했을 때 기존 연구들보다 더 나은 성능을 보이는 것을 그림 2(b)에서 확인할 수 있다.

Cora와 Citeseer는 기존 연구들에서 많이 다루는 인용 네트워크로 대상 논문을 기준으로 인용한 논문을 보았을 때 같은 카테고리에 속할 확률이 매우 높고, 관련도가 매우 높은 논문과의 인용관계가 존재하지 않을 확률도 매우 낮다. 다시 말해 1-hop 노드의 설명력이 매우 강한 그래프 구조를 띤다. 실험 결과 기존 연구와 유사한 경향성을 보이는 것을 그림 2(c), 그림 2(d)에서 확인할 수 있다. 후보 노드들 중에서 랜덤하게 이웃 노드를 선택한 실험의 결과가 최고 성과와 약 8%의 적은 차이가 나는 점에서, 다른 도메인과 달리 1-hop노드와 2-hop 노드들 중에 이상치가 매우 적은 수로 존재하고 대상 노드의 표현력을 높여줄 잠재성이 비슷하다고 볼 수 있다. 다른 도메인 네트워크의 경우 랜덤하게 선택한 실험의 경우 최고 성과와 50~60%의 성능차이를 보이는 것을 보면, 논문 인용 네트워크의 특수성을 확인할 수 있다.

5. 결론

본 논문에서는 정교한 이웃 노드 선택법을 활용한 GCN 모델을 제안하였다. 기존 연구들은 1-hop 노드들을 대상으로 범위를 한정하였으나, 다양한 도메인에서 좋은 성능을 보이지 못한다는 단점이 있었다. 이를 해결하기 위해 범위를 2-hop으로 늘리고 특성 벡터간의 유사도 계산을 통해 순위를 측정함으로써, 중요한 정보는 선택하고 노이즈는 제거하는 효과를 불러왔다. 여러 도메인에서의 실험을 통해 제안 방법이 다양한 도메인에서 기존 연구와 유사하거나 더 좋은 성능을 보이는 것을 확인하였다.

References

[1] A. Grover, and J. Leskovec, "node2vec: Scalable Feature Learning for Networks," *Proc. of the 22nd ACM SIGKDD international conference on Know-*

ledge discovery and data mining, 2016.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, "Attention is All You Need," *Advances in neural information processing systems*, 2017.

[3] B. Perozzi, R. Ai-Rfou, S. Skiena, "Deepwalk: Online Learning of Social Representations," *Proc. of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014.

[4] H. Gao, Z. Wang, S. Ji, "Large-Scale Learnable Graph Convolutional Networks," *Proc. of the 24th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2018.

[5] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, "Spectral Networks and Locally Connected Networks on Graphs," *International Conference on Learning Representations*, 2014.

[6] J. J. Whang, Y. Hou, D. Gleich, I. S. Dhillon, "Non-exhaustive, Overlapping Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[7] J. McAuley, C. Targett, Q. Shi, A. Van Den Hengel, "Image-based Recommendations on Styles and Substitutes," *Proc. of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015.

[8] A. Lancichinetti, M. Kivela, J. Saramaki, S. Fortunato, "Characterizing the community structure of complex networks," *PLoS one* 5.8, 2010.

[9] M. Defferrard, X. Bresson, P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," *Advances in neural information processing systems*, 2016.

[10] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Gallagher, T. Eliassi-Rad, "Collective Classification in Network Data," *AI magazine* 29.3, 2008.

[11] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, "Graph Attention Networks," *International Conference on Learning Representations*, 2018.

[12] T. N. Kipf, M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," *International Conference on Learning Representations*, 2017.

[13] W. L. Hamilton, R. Ying, J. Leskovec, "Inductive Representation Learning on Large Graphs," *Advances in Neural Information Processing Systems*, 2017.

[14] X. Huang, J. Li, X. Hu, "Label Informed Attributed Network Embedding," *Proc. of the Tenth ACM International Conference on Web Search and Data Mining*, 2018.

[15] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-Based Learning applied to document recognition," *Proc. of the IEEE* 86.11, 1998.



정 연 성

2018년 성균관대학교 통계학과, 컴퓨터공학과 졸업(학사). 2018년~현재 성균관대학교 전자전기컴퓨터공학과 석사 과정. 관심분야는 그래프마이닝, 기계학습, 데이터마이닝



황 지 영

2010년 이화여자대학교 컴퓨터공학과 졸업(학사). 2015년 텍사스 오스틴 대학교 (University of Texas at Austin) 컴퓨터 과학과 졸업(박사). 2016년~현재 성균관대학교 소프트웨어학과 조교수. 관심분야는 빅데이터, 데이터 마이닝, 그래프마이닝, 기계학습