분자 구조 분석을 위한 GNN 에 설명 가능한 인공지능 적용

조우성, 이재구*
국민대학교

*jaekoo@kookmin.ac.kr

Application of Explainable Artificial Intelligence for Molecular Structure Analysis Using GNN

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

요 약

비정형 데이터들을 다루기 위해서 노드와 엣지로 정의되는 그래프를 입력으로 하는 그래프 신경망(Graph Neural Network; GNN)이 늘어나고 있다. 특히 분자 데이터를 그래프로 표현하고, GNN 을 이용하여 새로운 분자의 특성을 예측하여 신약 개발에 이용하는 등 많은 분야에서 사용되고 있다. 그러나 이때 인공지능의 블랙박스(Blackbox) 특성이 왜 이 분자가 원하는 특성을 갖게 되는지 파악하기 어렵게 만든다. 따라서 본 논문에서는 분자 그래프의 특성을 예측하는 GNN 에 설명 가능한 인공지능(Explainable Artificial Intelligence; xAI) 기술들을 도입하고, 이를 시각화한 결과가 인간지능과 비교해 합리적인지 알아보고자 한다.

Ⅰ. 서 론

현재 인공지능 기술은 급격하게 발전하여 신약, 의료 영상, 헬스케어, 금융 등 인간에게 직접적인 영향을 끼치는 분야까지 확장되었다. 특히 신약 부분에서는 분자를 그래프(Graph)로 만들어 분자의 특성을 예측하는 그래프 신경망(Graph Neural Network; GNN)이 많이 쓰이며, 이들 특성을 분류하는 데 높은 성능을 보이고 있다.

하지만 인공지능의 블랙박스(Blackbox) 특성은 왜 인공지능이 그와 같은 의사결정을 했는지 알지 못하게 만든다. 따라서 인공지능의 의사결정에 대한 이해가 매우 중요해지고 있으며, 이를 위한 기술들을 설명 가능한 인공지능(Explainable Artificial Intelligence; xAI) 기술이라 한다. 설명 가능한 인공지능 기술은 중요성 맵(Saliency Map; SM)[1], 경사기반 클래스 활성 맵(Gradient CAM; Grad-CAM)[2], 계층별 관련성 전파(Layer-wise Relevance Propagation; LRP)[3] 등 여러 가지가 존재한다. 하지만 이 기술들은 그래프가 아닌 사진 등의 정형 데이터를 기반으로 발전된 기술이기 때문에 그래프에 적용했을 때 시각화 결과를 사람이 해석하기 쉬울 것이라 보장하지 못한다.

따라서 본 논문에서는 앞서 언급한 분자 특성을 예측하는 그래프 신경망에 설명 가능한 인공지능 기술을 적용한다. 그리고 그 결과를 인간지능과 비교하여 정형 데이터에서의 설명 가능한 인공지능 기술이 분자의 특성을 알아내는 데 유의미할지 확인해보려 한다.

Ⅱ. 관련 연구

본 논문에서 사용한 설명 가능한 인공지능 기술은 계층별 관련성 전파(LRP)[3]와 경사기반 클래스 활성 맵에서 음의 값을 살린 UGrad-CAM[4] 두 가지이다.

사용한 모든 인공지능 기술은 역전파(Backpropagation)를 이용한다. UGrad-CAM 은 마지막 층에서 얻은 특징 맵(Feature Map)을 기반으로 각 데이터의 특징 맵에 분류 결과에 대한 경사를 곱해가며 얻은 값을 합산하여 계산한다[식 1]. 한편 LRP 는 모델의 가중치(w)와 활성도(a)를 사용해서 모델의 결과를 입력 층으로 역전파하는 방법이다. 이때 층별로 각 값의 기여도를 관련성(Relevance)라 하며, 연속하는 두 층 j,k에서 관련성 전파는 [식 2]와 같이 일어난다.

$$\text{UGrad-CAM} := \sum_k \alpha_k^{l,c} F_{k,n}^l(X,A)$$
$$\text{이때, } \alpha_k^{l,c} = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial y^c}{\partial F_{k,n}^l} \qquad [식\ 1]$$

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} R_k \qquad [식\ 2]$$

실험에 사용된 데이터집합은 MUTAG[6]이다. 이는 살모넬라 티파이무름(Salmonella Typhimurium)에 해당 분자가 돌연변이를 일으키는지 여부를 이진 분류하는 데이터셋으로 188 개의 데이터를 128, 20, 40 개로 나누어 각각 학습, 검증, 테스트 집합으로 사용하고 정확도를 이용해서 성능을 계산하였다.

Ⅲ. 모델 학습

분자의 원소 사이 여러 종류의 결합을 다룰 수 있는 관계성 그래프 어텐션 망(Relational Graph Attention Networks; RGAT)[6]을 사용하여 그래프 분류 모델을 구성하였다. RGAT 에서 노드를 갱신 하는 방식은 [그림 1]에 나와있다. 그래프로 만든 분자에서 빨간색으로 표시한 하나의 노드를 타겟 노드로 정했을 때, 세 이웃 노드 중 하나는 단일 결합(주황색), 나머지는 방향 결합(초록색)이다. 같은 결합으로 연결된 이웃 간에 로짓 $n_{i,j}^r$을 구하고, 소프트맥스(Softmax)를 통해 어텐션 $a_{i,j}^r$을 구한다. 층은 총 세 개를 사용하였다. 첫 번째 층은 4 개의 다중 어텐

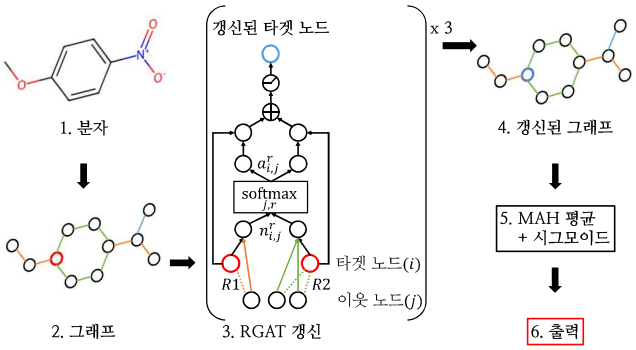

그림 1. RGAT 를 이용한 그래프 분류 모델 모식도

션 헤드(Multi-Attention Head; MAH)를 사용하였고, 각각은 32 개의 특징을 계산하여 총 128 개의 차원으로 매핑된다. 두 번째 층도 마찬가지로 4 개의 MAH 를 사용하여 32 개 씩 128 개 차원에서 계산한 뒤 마지막 층에서는 각각 1 개 씩의 특징을 계산하는 6 개의 MAH 를 사용하여 6 차원이 만들어진다. 이들을 평균 내어 최종 출력을 정하게 된다. MUTAG 은 과업이 이진 분류이기 때문에 시그모이드(Sigmoid)를 적용하였다. 학습 결과 83.42%의 정확도를 보였다.

IV. 설명 가능한 인공지능 결과

[그림 2]는 MUTAG 에 있는 분자 중 4-니트로아니졸(4-Nitroanisole)이란 분자에 대한 설명 가능한 인공지능 기술 적용 결과이다. LRP 와 UGrad-CAM 모두 -1 에서 1 의 범위를 가지며 파란색이 낮은 값, 빨간색이 높은 값을 지칭한다. [그림 2]에서 LRP 의 결과는 O 에 파란색, N 에 빨간색이 표시되어 있음을 볼 수 있다. Debnath 등은 살모넬라 티파이무름에 대한 돌연변이성은 소수성(Hydrophobicity)과 니트로기(Nitro Compound; NO2) 등이 결정적이라고 말하고 있는데[5], 이는 LRP 의 결과와 일치함을 볼 수 있다. 파란색으로 표시된 산소 부분은 소수성을 방해하기 때문에 음의 영향 값을 갖고, 빨간색으로 표시된 N 부분은 니트로기를 지칭하기 때문에 양의 영향 값을 가짐을 알 수 있다.

따라서 분자 그래프에 대해서는 LRP 가 좀 더 인간지능과 일치함을 알 수 있고, 따라서 LRP 를 쓰는 것이 좀 더 결과를 잘 볼 수 있음을 알 수 있다.

V. 결론

본 논문에서는 MUTAG 데이터 집합에서 설명 가능한 인공지능 기술을 적용했을 때의 결과를 인간지능과 비교하여 해당 기술이 적절한지를 알아보았다. LRP 의 결과에서 실제 화학적 지식과 같은 부분을 확인할 수 있었으며, 따라서 다른 분자 그래프에 대한 결과에서도 LRP 를 적용했을 때 사람이 해석가능한 결과를 얻을 것이라 기대할 수 있을 것으로 보인다.

이 가능성은 특히 신약 개발 분야에서 특히 중요한 부분으로 적용될 수 있을 것이다.

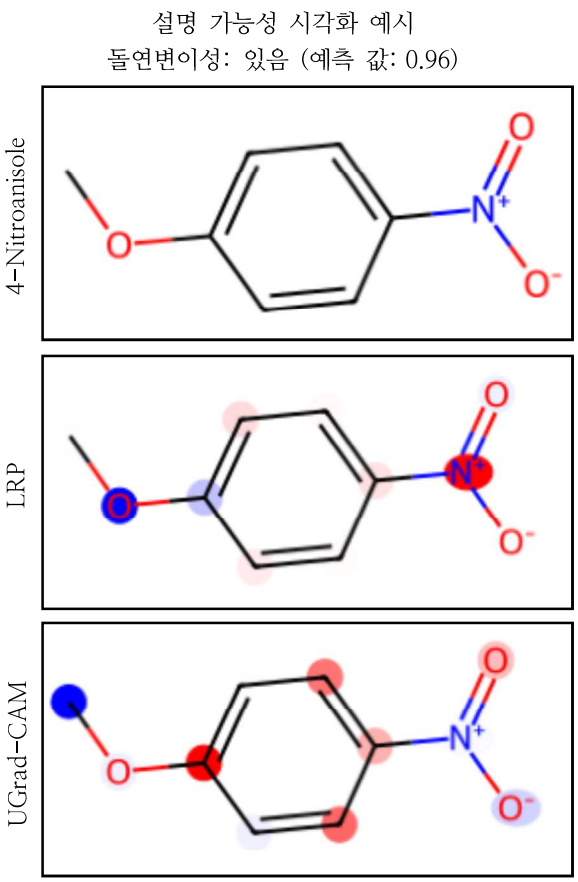

그림 2. MUTAG 데이터 중 4-니트로아니졸에 대한 설명 가능한 인공지능 기술 적용 시각화 예시

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] K. Simonyan, et al. (2013) Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034

[2] R. R. Selvaraju, et al. (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In ICCV, pages 618-626.

[3] Binder, A., et al. (2016, 9). Layer-wise relevance propagation for neural networks with local renormalization layers. In International Conference on Artificial Neural Networks (pp. 63-71). Springer, Cham.

[4] Pope, P. E., et al. (2019). Explainability methods for graph convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10772-10781).

[5] Debnath, A.K., et al. (1991). Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity. J. Med. Chem. 34(2):786-797.

[6] Busbridge, D., et al. (2019). Relational graph attention networks. arXiv preprint arXiv:1904.05811.