

Message Passing Neural Network 모델을 통한 독성 예측 연구

정홍기, 권현*

육군사관학교 AI·데이터과학과

jeonghonggi587@gmail.com, *hkwon.cs@gmail.com(교신저자)

Research on Predicting toxicity through Message Passing Neural Network

Honggi Jeong, Hyun Kwon

Department of Artificial Intelligence and Data Science

요약

본 논문은 분자 구조 정보를 활용하여 혈뇌장벽(Blood-Brain Barrier, BBB) 투과성을 예측하는 Graph Neural Network(GNN) 기반 방법을 제안한다. 본 연구의 핵심은 Simplified Molecular Input Line Entry System(SMILES) 형식으로 표현된 분자 구조를 바탕으로, 미지의 화학물질이 BBB를 투과할 수 있는지를 예측하는 것이다. 이를 위해 메시지 전달 메커니즘에 기반한 Message Passing Neural Network(MPNN)을 적용하였으며, 다양한 화학적 특성과 구조적 정보를 효과적으로 반영할 수 있도록 하였다. 또한, 제안한 방법의 실효성을 국방 분야에 확장 적용 가능성 측면에서 분석하였다.

I. 서론

화학물질의 독성 및 생물학적 특성을 예측하는 것은 신약 개발, 독성 평가, 국방 분야 등에서 중요한 과제이다. 특히, 미지의 화학물질에 대한 실험적 분석은 시간과 비용이 많이 들며, 경우에 따라 높은 위험을 수반한다. 이에 따라 분자 구조 기반의 예측 기술이 대안으로 주목받고 있다. 분자 구조가 유사하거나 부분적으로 동일한 화합물의 특성을 활용하여 새로운 화학물질의 특성을 예측할 수 있으며, 이를 위해 효과적인 기계학습 기반 방법론이 요구된다.

본 연구에서는 혈뇌장벽(Blood-Brain Barrier, BBB) 투과성 예측을 사례로 하여, Simplified Molecular Input Line Entry System(SMILES) 형식으로 표현된 분자 구조 데이터를 입력으로 하는 그래프 신경망(Graph Neural Network, GNN) 기반 예측 모델을 제안한다. 특히, GNN의 일종인 메시지 전달 신경망(Message Passing Neural Network, MPNN)을 적용하여, 분자 내 원자 간의 상호작용 정보를 학습하고 이를 바탕으로 BBB 투과성 여부를 효과적으로 예측하고자 한다.

제안한 방법은 단순한 분자 유사성 기반 접근을 넘어서, 구조적 특성과 연결 관계를 정교하게 반영함으로써 예측 성능을 향상시킬 수 있다. 또한 본 연구는 해당 기술을 국방 분야에 응용할 수 있는 가능성에 대해서도 논의한다.

II. 혈뇌장벽 투과성 예측 방법

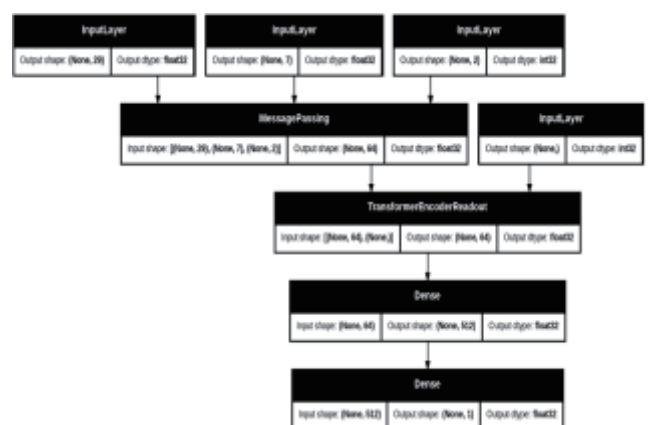
본 연구에서는 미지의 화학물질에 대한 혈뇌장벽(Blood-Brain Barrier, BBB) 투과성 여부를 예측하기 위해 Graph Neural Network(GNN) 계열의 모델인 MPNN(Message Passing Neural Network)을 적용하였다. 이 모델은 분자의 구조적 정보를 그래프 형태로 표현하여, 원자 간의 관계를 기반으로 예측을 수행할 수 있는 장점을 가진다.

모델의 입력은 SMILES 형식으로 표현된 분자 구조이며, 이를 바탕으로 그래프를 구성하였다. 이때 각 원자는 그래프의 노드(Node)로, 원자 간의 화학 결합은 엣지(Edge)로 표현된다. 각 엣지에서 파생된 특성들은 집계되어 Dense Layer로 전달되며, ReLU 활성화 함수를 통해 비선형성을 부여한다. 이후 메시지 집계(Aggregation) 과정에서는 Gated Recurrent

Unit(GRU)를 활용하여 노드 간 정보를 통합하며, 최종적으로 각 노드의 상태를 기반으로 전체 그래프 수준의 특성을 추출한다.

본 연구에서는 기존 MPNN 구조에서 두 가지 주요 변형을 적용하였다. 첫째, Readout 단계에서 보통 사용되는 MaxPooling 대신 AveragePooling을 사용하였다. 이는 객체 탐지보다는 평균적인 구조와 미세한 정보를 추출하는 데 적합하기 때문이다. 둘째, Readout 단계에서 ReLU 활성화를 거친 후 AveragePooling을 수행한다. 결과값을 Dense Layer로 전달하여 한 번 더 ReLU를 적용한 후, Sigmoid 함수를 통해 0과 1 사이의 확률 값을 출력하였다.

모델 학습 과정에서는 Binary Cross Entropy(BCE) 손실 함수를 사용하였으며, 최적화 알고리즘으로는 Adam 옵티마이저를 적용하여 모델의 성능을 향상시켰다. 이러한 구조를 통해 분자 구조 정보만으로도 혈뇌장벽 투과 여부를 효과적으로 예측할 수 있다.



(그림 1) MPNN 모델의 구조 및 흐름도

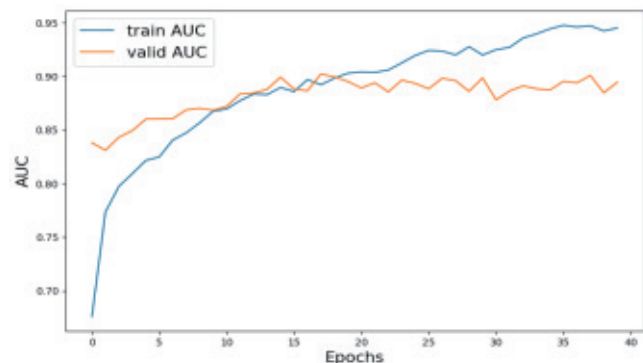
III. 실험 환경 및 실험 결과

본 연구의 성능을 검증하기 위해서, 데이터셋은 Molecular.org에 있는 오픈소스인 BBBP.csv파일 사용하였다. 데이터셋은 이름, 독성의 여부, SMILES 특징을 가지고 있으며 학습 데이터의 수는 1642개이고 테스트 데이터의 수는 21

개이다.

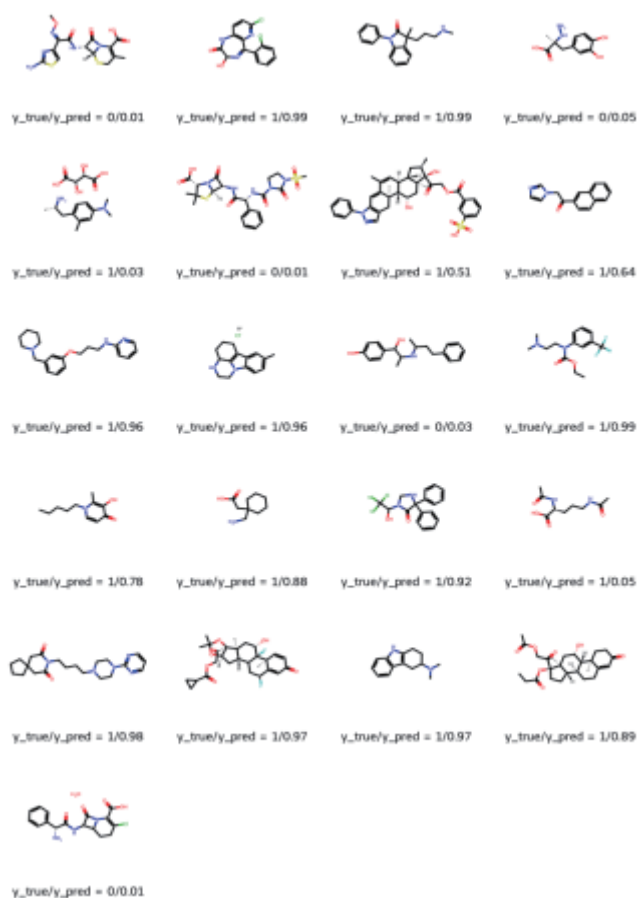
제안 모델은 MPNN 모델을 사용하였으며 MPNN 모델의 구조는 그림 1을 참고하면 된다. 하이퍼파라미터로 학습률은 5×10^{-4} 로 정하였고 epoch은 40으로 설정하였다.

실험결과로 학습데이터를 모델에 학습하면, 그림 2와 같은 그래프가 얻을 수 있다. 그림 2는 학습데이터와 검증데이터에 대해 epoch에 따른 제안방법의 AUC 성능을 보여준다. 40 epochs만큼 돌리게 되면 검증데이터의 기준으로 약 88%의 AUC를 갖는다.



(그림 2) 훈련데이터와 검증데이터에 해서 Epoch에 따른 제안방법의 AUC

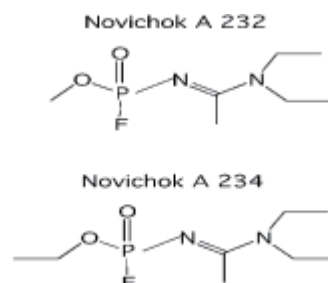
그림 3은 테스트 데이터에 대한 제안방법의 성능 결과이다. 테스트 데이터에 대한 성능 검증 결과를 보면, 대부분 실제값과 예측값이 거의 일치하는 것을 볼 수 있다.



(그림 3) 테스트 데이터에 대한 제안방법의 성능결과

IV. 국방분야에 대한 고찰

이 연구의 주요 목적은 북한 화학무기의 위협에 대한 방어이다. 특히, 공산주의 국가에서 개발 중인 노비촉(Novichok)이라는 신경작용제의 위협에 대응하는 데 중점을 두었다. 노비촉은 대체로 무색무취이기 때문에 북한이 이를 사용하더라도 화학무기 공격 사실을 인지하지 못하게 할 수 있다. 북한은 이러한 특성을 활용해 새로운 노비촉을 개발하려는 노력을 기울이고 있다. 북한의 이러한 시도를 저지하기 위한 방안으로 한 가지 접근법을 제안하고자 한다. 분자의 구조의 유사성을 활용하는 것이다. 예를 들면, 그림 4에서 A-232와 A-234의 구조를 보게 되면 굉장히 비슷한 경향을 볼 수 있다. 만일 이 모델을 사용하게 된다면 A-234가 밝혀지지 않았을 때 A-232를 통해 A-234가 신경작용제의 유독하다는 점을 알 수 있다. 더 나아가 이 독성을 파악하고 방독면과 방호복을 착용하지 못한 국군에게 해독제를 미리 나누는 역할을 하거나 독성을 바로 없앨 수 있는 제거제를 만들게 된다면, 북한 화학무기의 위협에서 벗어날 수 있다고 기대한다.



(그림 4) Novichok A-232, A-234의 분자구조

V. 결론

본 연구에서는 분자 구조 정보를 기반으로 화학물질의 혈뇌장벽 투과성을 예측하기 위한 방법으로 MPNN을 활용한 모델을 제안하였다. 제안한 모델은 SMILES 형식으로 표현된 분자 구조를 그래프 형태로 변환하여, 원자 간 상호작용 정보를 효과적으로 학습하고 예측 성능을 향상시킬 수 있음을 보였다.

향후 연구에서는 탈리도마이드와 같은 예외적인 사례를 고려하기 위해, 단순한 분자 구조 외에도 입체화학 정보, 대사 경로 등 생물학적 활성에 영향을 미칠 수 있는 다양한 요인을 통합한 확장된 모델링이 필요하다. 이를 통해 구조적으로 유사한 분자 간의 미세한 특성 차이까지 반영할 수 있는 정밀한 예측 모델을 구축할 수 있을 것이다.

참고문헌

- [1] 황순열, 이은경, 이영민. (2022-05-11). Tox21 데이터의 독성 예측을 위한 Message Passing Neural Network 모델의 초매개변수 최적화. 환경독성보건학회 심포지엄 및 학술대회, 부산.
- [2] Datasets, <https://moleculenet.org/datasets-1> (accessed May 13, 2025).
- [3] H. M. Bolt and J. G. Hengstler, "Recent research on Novichok," Archives of Toxicology, vol. 96, no. 5, pp. 1137-1140, Mar. 2022. doi:10.1007/s00204-022-03273-7
- [4] K. Team, "Keras documentation: Message-passing neural network (MPNN) for molecular property prediction," Keras, <https://keras.io/examples/graph/mpnn-molecular-graphs/> (accessed May 12, 2025).