

9/28

1. GNN_DOVE

Graph Neural Network

$G=(V,E)$

Node V 는 원자, E edge는 if atom i and atom j are connected by a covalent bond

인접 행렬 A is the adjacency matrix, which numerically represents the connectivity of the graph

1. GNN_DOVE - Feature 논문

Features	Representation
Atom type	C, N, O, S, H (one hot)
The degree (connections) of atom	0, 1, 2, 3, 4, 5 (one hot)
The number of connected hydrogen atoms	0, 1, 2, 3, 4 (one hot)
The number of implicit valence electrons	0, 1, 2, 3, 4, 5 (one hot)
Aromatic	0 or 1

원핫 인코딩을 통해서

feature의 벡터의 길이는 23개(원자 타입 5 + 원자 degree 6 + 연결된 수소의 개수 5 + 내포된 원자가 전자의 수 6 + 향기가 나는 방향성 2)

1. GNN_DOVE - Feature 실제 코드

```
def atom_feature(m, atom_i):  
    atom = m.GetAtomWithIdx(atom_i)  
    return np.array(one_of_k_encoding_unk(atom.GetSymbol(),  
                                           ['C', 'N', 'O', 'S', 'F', 'P', 'Cl', 'Br', 'B', 'H']) +  
                    one_of_k_encoding_unk(atom.GetDegree(), [0, 1, 2, 3, 4, 5]) +  
                    one_of_k_encoding_unk(atom.GetTotalNumHs(), [0, 1, 2, 3, 4]) +  
                    one_of_k_encoding_unk(atom.GetImplicitValence(), [0, 1, 2, 3, 4, 5]) +  
                    [atom.GetIsAromatic()])  # (10, 6, 5, 6, 1) --> total 28 Feature
```

2. GNN_DOVE - hit?

Dockground Unbound Docking Decoys Set 1에서

Hit은 데이터 셋 자체의 설명으로는 hit이 각각의 decoy set에서 near-native 해답의 수라고 하는데
(The number of near-native solution kept in each decoy set.)

near-native는 L-RMSD가 5.0Å보다 적은 걸 near-native라고 한다.

(A solution with L_rmsd (RMSD of the backbone atoms of the ligand) less than 5.0 Å are defined as near-native one.)

2. GNN_DOVE - hit?

방향을 잘못 생각하고 있었던 것 같습니다

이 코드가 protein docking을 만드는 것이 아니라

이미 도킹이 된 pdb들이 가능한지 평가를 하는 거라서 웬만한 데이터 셋의 결과가 다 90% 이상의 값이 나온다는 것

이해를 한게 mRNA-miRNA 하듯이 protein을 바꿔가면서 뭐가 맞네 뭐가 틀리네 이런 형식이라고 생각을 했었습니다.

2. GNN_DOVE - hit?

그래서 data를 돌려도 Predict result가 좋게 나오고,

큰 데이터를 나눠서 돌려도 좋게 나오는데, 그래프와 같이 hit rate를 표시하는 방법이나 이 모델을 통해서 hit을 했다는 기준을 모르겠습니다.

