

11/2

1. GNN - 학습데이터

학습시키는데, 데이터를 어떻게 해야할 것 같습니다.

labeled 데이터가 아니라서 결과값을 어디다가 맞춰야 할지 모르겠다. 그래서 일단 다음과 같이 학습 데이터를 주어진 코드로 한 결과 값들을 정답이라고 놓고 돌려했습니다.

Y =

```
np.array([0.9969,0.9969,0.9969,0.9972,0.9984,0.9969,0.9969,0.9969,0.9969,0.9974,0.9974,0.997  
7,0.9974,0.9986,0.9973,0.9995,0.9971,0.9990,0.9994,0.9994,0.9994,0.9974,0.9972,0.9978,0.997  
4,0.9974])
```

train 데이터로는 교수님이 주신 26개의 **pdb**파일을 이용했습니다.

1. GNN - 결과

그랬더니 **test data**로는 모델에서 주어진 맞는 **pdb**파일 **correct.pdb**와 틀린 **incorrect.pdb**를 둘 다 돌려보았는데,
두개 다 **1.0**이 나오고 다른 데이터로 해도 모두 **1.0**이 나오는 기적을 보았습니다.

데이터를 구하는 방법이나 매개변수에 대한 것을 물어봐야할 것 같습니다.

2. atom feature

다음과 같이 atom의 feature를 rdkit으로 하고있습니다.

```
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
```

2. atom feature - residue

residue의 info를 얻는다는 게 있긴 한데 논문에서는 interface region를 residue의 집합이라고 이미 구하고 있어서 하고있는 코드 자체가 residue들만 가지고 하고 있습니다.

For an input protein docking decoy, the interface region is identified as a set of residues located within 10.0 Å of any residues of the other protein. A residue-residue distance is defined as the shortest distance among any heavy atom pairs across the two residues.



RDKit

Open-source cheminformatics and machine learning.

[Main Page](#)[Related Pages](#)[Namespaces ▾](#)[Classes ▾](#)[Files ▾](#)

- ▶ RDKitFP
- ▶ RDTypeTag
- ▶ ScaffoldNetwork
- ▶ SLNParse
- ▶ StructureCheck
- ▶ TopologicalTorsion
- ▶ UFF
- ▶ Utils
- ▶ AdditionalOutput
- ▶ AromaticAtomIterator_
- ▶ Atom
- ▶ AtomEnvironment

RDKit::AtomPDBResidueInfo Class Reference

Captures atom-level information about peptide residues. [More...](#)

```
#include <MonomerInfo.h>
```

Inheritance diagram for RDKit::AtomPDBResidueInfo: