

# GCN Manual

Solubility Prediction using Graph Convolutional Network

```
# regre_run.sh
```

```
python regression.py -D './data/data01.xlsx' -X1 'Solute SMILES' -X2 'Solvent SMILES' -Y 'LogS' -O  
 './results/regression/myResult.json' -M './results/regression/myModel.pt'
```

- D './data/data01.xlsx' : 데이터셋 파일명
- X1 'Solute SMILES' : 파일에서 Solute의 SMILES 값이 있는 컬럼명
- X2 'Solvent SMILES' : 파일에서 Solvent의 SMILES 값이 있는 컬럼명
- Y 'LogS' : 파일에서 logS 값이 있는 컬럼명
- O './results/regression/myResult.json' : 결과 파일의 경로와 파일명
- M './results/regression/myModel.pt' : 결과 모델의 경로와 파일명

위 파일 실행

```
$ bash regre_run.sh
```

```
Graph Convolutional Network for logS Regression  
Soongsil University, Seoul, South Korea  
Computational Science and Artificial Intelligence Lab
```

```
[Preparing Data]
```

```
- Device : cuda
```

```
[Converting to Graph]
```

```
- Train Data : 14076
```

```
- Test Data : 3520
```

```
[Train]
```

```
- Epoch : 1
```

```
- Loss : 2.5445
```

# smiles.txt

Solute SMILES	Solvent SMILES
<chem>C=CC(=O)N</chem>	<chem>O</chem>
<chem>C=CC(=O)N</chem>	<chem>CC(=O)C</chem>
<chem>CCCC1=CC=CC=C1</chem>	<chem>O</chem>
<chem>CCCC1=CC=CC=C1</chem>	<chem>CC(=O)C</chem>

# regre\_test.sh

```
python regre_test.py -M './results/regression/myModel.pt' -I './smiles.txt' -O './results/regression/myTest.txt'
```

-M './results/regression/myModel.pt' : 모델 선택

-I './smiles.txt' : Input으로 위 파일 경로와 이름

-O './results/regression/myTest.txt' : Output으로 파일 경로와 이름

위 파일 실행

\$ bash regre\_test.sh

# myTest.txt

Solute SMILES	Solvent SMILES	logS
<chem>C=CC(=O)N</chem>	<chem>O</chem>	0.17527944
<chem>C=CC(=O)N</chem>	<chem>CC(=O)C</chem>	0.36640382
<chem>CCCC1=CC=CC=C1</chem>	<chem>O</chem>	-3.079277
<chem>CCCC1=CC=CC=C1</chem>	<chem>CC(=O)C</chem>	0.63403666

```
# class_run.sh
```

```
python classification.py -D './data/data01.xlsx' -X1 'Solute SMILES' -X2 'Solvent SMILES' -Y 'LogS' -O  
 './results/classification/myResult.json' -M './results/classification/myModel.pt' -ROC
```

- D './data/data01.xlsx' : 데이터셋 파일명
- X1 'Solute SMILES' : 파일에서 Solute의 SMILES 값이 있는 컬럼명
- X2 'Solvent SMILES' : 파일에서 Solvent의 SMILES 값이 있는 컬럼명
- Y 'LogS' : 파일에서 logS 값이 있는 컬럼명
- O './results/classification/myResult.json' : 결과 파일의 경로와 파일명
- M './results/classification/myModel.pt' : 결과 모델의 경로와 파일명
- ROC : ROC AUC 결과 저장 여부

위 파일 실행

```
$ bash class_run.sh
```

```
Graph Convolutional Network for logS Classification  
Soongsil University, Seoul, South Korea  
Computational Science and Artificial Intelligence Lab
```

```
[Preparing Data]  
- Device : cuda
```

```
[Converting to Graph]  
- Train Data : 14076  
- Test Data : 3520
```

```
[Train]  
- Epoch : 1  
- Loss : 0.7508  
- Accuracy : 65.7480
```

# smiles.txt

Solute SMILES	Solvent SMILES
<chem>C=CC(=O)N</chem>	<chem>O</chem>
<chem>C=CC(=O)N</chem>	<chem>CC(=O)C</chem>
<chem>CCCC1=CC=CC=C1</chem>	<chem>O</chem>
<chem>CCCC1=CC=CC=C1</chem>	<chem>CC(=O)C</chem>

# class\_test.sh

```
python class_test.py -M './results/classification/myModel.pt' -I './smiles.txt' -O './results/classification/myTest.txt'
```

-M './results/classification/myModel.pt' : 모델 선택

-I './smiles.txt' : Input으로 위 파일 경로와 이름

-O './results/classification/myTest.txt' : Output으로 파일 경로와 이름

위 파일 실행

\$ bash class\_test.sh

# myTest.txt

Solute SMILES	Solvent SMILES	Class
<chem>C=CC(=O)N</chem>	<chem>O</chem>	High
<chem>C=CC(=O)N</chem>	<chem>CC(=O)C</chem>	High
<chem>CCCC1=CC=CC=C1</chem>	<chem>O</chem>	Low
<chem>CCCC1=CC=CC=C1</chem>	<chem>CC(=O)C</chem>	High