

9/21

1. GNN에 대해서

Graph Neural Network

$G=(V,E)$

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

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

1. GNN - CNN과 비교

이미지 데이터 분석에 많이 사용되는 convolutional neural network (CNN)는 이미지를 작은 구역으로 나누어 지역적인 정보를 계속 취합하는 식으로 작동한다.

일반적으로 CNN을 사용하는 네트워크에서는 convolution 연산을 반복적으로 수행하는데, convolution은 각각의 구역으로부터 정보를 추출하는 작업을 한다.

GCN에서도 CNN과 마찬가지로 태스크 수행에 있어 가장 중요한 정보를 그래프에서 뽑아내는 것이 목표이다. GCN은 필터를 그래프에 대해 통과시켜 그래프 내에서 중요한 노드와 엣지로부터 정보를 취합한다.

1. GNN - CNN과 비교

둘의 차이점은

- 2D convolution은 특정 픽셀(혹은 위치)과 인접해있는 지역의 정보를 모은다
- 그래프 convolution은 노드와 연결되어 있는 노드들의 정보를 모은다

이미지에서 convolution의 아이디어는 학습 가능한 필터를 통해 중심 픽셀의 주변 픽셀을 합치는 것이다. Spatial Convolution Network의 핵심 아이디어는 이 아이디어에서 주변 픽셀 대신 **연결된 점의 특징을 적용한 것**이다.

2. 학습이 가능한지 살펴보기

5. fold값 맞춰주기

```
fold_choice = params['fold']
```

#loading the model

#fold에 따라 바꾸기

```
if fold_choice != -1:
```

6. 보통은 이쪽 best model을 사용한다.

```
model_path = os.path.join(os.getcwd(), "best_model")
```

```
model_path = os.path.join(model_path, "fold" + str(fold_choice))
```

```
model_path=os.path.join(model_path,"checkpoint.pth.tar")|
```

```
model,device=init_model(model_path,params)
```

학습을 한다거나 그런 것은 없고,
코드를 살펴보면 그냥 바로 fold에 따라
best_model로 초기화를 진행합니다.

라이브리를 사용하지 않고, 독자적으로 만든
코드로 계산을 하고 있습니다.

3. 결과가 재현이 가능한지

DATA AVAILABILITY STATEMENT

The Dockground docking dataset was downloaded from the Dockground database (<http://dockground.compbio.ku.edu>) at the link <http://dockground.compbio.ku.edu/downloads/unbound/decoy/decoys1.0.zip>.

The ZDOCK dataset was downloaded from the ZDOCK decoy sets (<https://zlab.umassmed.edu/zdock/decoys.shtml>) at the link https://zlab.umassmed.edu/zdock/decoys_bm4_zd3.0.2_6deg.tar.gz.

The CAPRI score set was downloaded from http://cb.iri.univ-lille1.fr/Users/lensink/Score_set.

해당 사이트로 가서 데이터 셋을 받고 돌려볼려고 했습니다.

3. 결과가 재현이 가능한지

일단 CAPRI 데이터는 너무 커서 그러는지 하다가 컴퓨터가 멈췄습니다.

한 pdb 파일에 너무 많은 데이터가 있어서 그런 것 같습니다.

```
1966072  ATOM  3592  CD  GLN A  84      13.255 -35.423  25.881  1.0
1966073  ATOM  3593  NE2 GLN A  84      13.729 -34.429  26.614  1.0
1966074  ATOM  3594  OE1  GLN A  84      13.669 -35.554  24.734  1.0
1966075  TER   2403594      GLN A  84
1966076  ENDMDL
```

3. 결과가 재현이 가능한지

1EER.pdb 파일의 Score은 0.9776
가 나왔습니다.

그런데, Score을 가지고 어떻게
Hit rate를 계산을 하는지, 어떻게
계산을 하는지 나와있지 않아서
어떤걸 의미하는지 잘
모르겠습니다.

TABLE 3 | Performance on the Dockground+ZDOCK testing dataset.

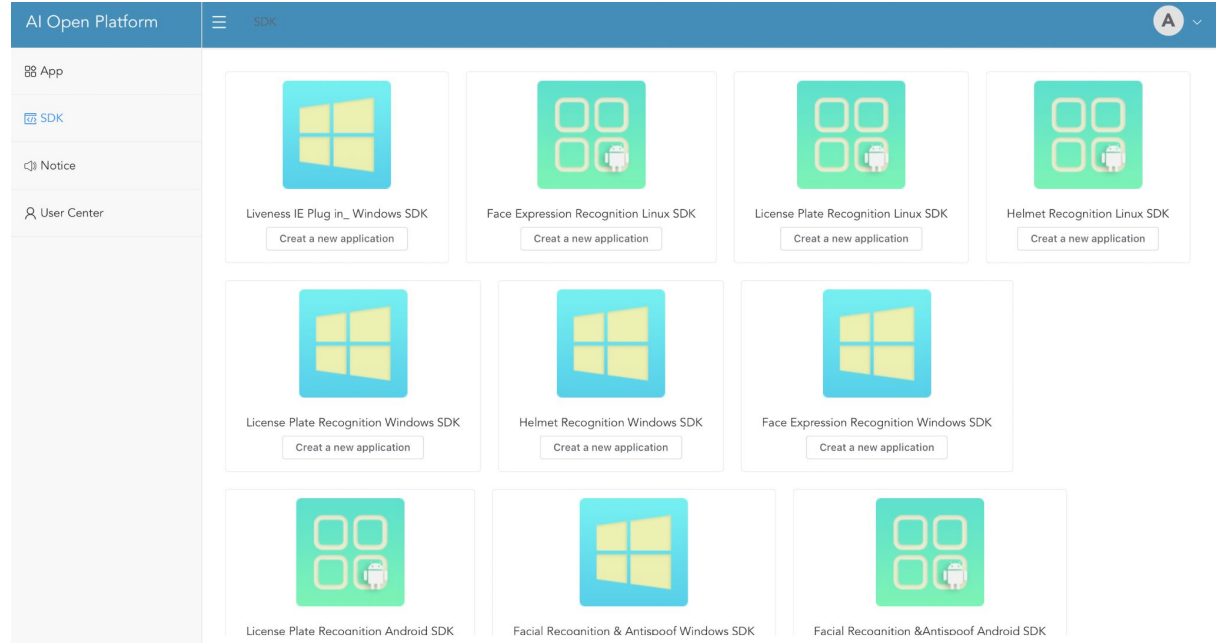
ID	GNN-DOVE	GOAP	ITScore	ZRANK	ZRANK2	IRAD	Total
1AK4	1	10	1	1	7	0	179
1AY7	8	0	3	9	8	8	176
1EER	0	0	0	0	3	0	41
1GLA	5	1	0	8	4	8	165
1HCF	9	0	8	3	3	7	183
1JIW	3	0	2	0	1	2	106
1JTG	8	0	10	10	0	10	177
1KAC	7	0	5	8	2	6	183
1KTZ	0	1	0	1	3	0	77
1MAH	9	0	8	9	0	9	179
2MTA	7	0	4	9	0	9	186
2VDB	9	1	9	7	2	6	173
3D5S	7	0	10	6	1	5	156
1BUH (1)	3	8	9	6	4	9	183
1FQ1 (1)	0	0	0	0	0	0	20
1JWH (1)	6	6	7	6	2	8	171
2OZA (1)	1	0	1	0	0	0	19
1EFN (2)	1	0	0	4	3	4	130
1GCQ (2)	2	9	0	1	8	4	142
Hit rate	0.842	0.368	0.684	0.789	0.737	0.737	—
Group HR	0.867	0.333	0.717	0.833	0.767	0.767	

In the ID column, the number in a parentheses indicates which group the target belongs to. Thus, four complexes belong to the same similarity group, and the other two belong to another group. The rest of the complexes are single entry groups. Group HR indicates the group hit rate. In Group HR, the fraction of complexes within each group that have at least one hit (acceptable model) within the top 10 ranks was first computed, and then averaged across all the groups. The total column indicates the total number of acceptable docking models for a given target.

4. Deep cam

<https://ai.deepcam.com/#/web/sdk>

라는사이트가 나오는데 이걸
application을 만들고 활용하는
느낌의 것이고,



4. Deep cam

<https://www.nowyouknow.com.au/>

논문을 찾아봐도 말씀하신 라이브러리같은 코드가 찾지 못했습니다.

<https://www.deepcam.com/>

DeepCAM사이트는 중국어로 되어있어서 번역기 돌리면서 보는데 코드같은게 나오지는 않았습니다.

github에도 나오지 않아서 어떤걸 말씀하셨는지 잘 모르겠습니다.



Home

Solutions

Pricing

Case Studies

About

Contact

Do you know?
if you can identify VIPs
coming into your shop or
attending your venue

