Matheus_Schmitz_hw7_task_2

Matheus Schmitz

USC ID: 5039286453

In [7]:

```
# File manipulation imports for Google Colab
from google.colab import drive
drive.mount('/content/drive')
import os
os.chdir("/content/drive/My Drive/Colab Notebooks/DSCI 558")
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount
("/content/drive", force_remount=True).
```

In [8]:

```
!pip install -q ampligraph
!pip install -q tensorflow==1.15
```

In [9]:

```
import requests
from ampligraph.datasets import load_from_csv
import numpy as np
import pandas as pd
```

Train-Test Split

In [10]:

```
from ampligraph.evaluation import train_test_split_no_unseen

X = load_from_csv('', 'freebase-237-merged-and-remapped.csv', sep=',')
print(f'X.shape: {X.shape}')
X_train, X_test = train_test_split_no_unseen(X, test_size=10000)
print('Train set size: ', X_train.shape)
print('Test set size: ', X_test.shape)
```

```
X.shape: (308722, 3)
Train set size:  (298722, 3)
Test set size:  (10000, 3)
```

In [11]:

```
# ComplEx was the best performing model in task 1
from ampligraph.latent_features import ComplEx
from ampligraph.evaluation import evaluate_performance
import tensorflow as tf
tf.logging.set_verbosity(tf.logging.ERROR)
from ampligraph.evaluation import mr_score, mrr_score, hits_at_n_score

model = ComplEx(batches_count=100,
                seed=0,
                epochs=200,
                k=300,
                eta=5,
                optimizer='adam',
                optimizer_params={'lr':1e-3},
                loss='multiclass_nll',
                regularizer='LP',
                regularizer_params={'p':3, 'lambda':1e-5},
```

```
                verbose=True)

positives_filter = X

tf.logging.set_verbosity(tf.logging.ERROR)

model.fit(X_train, early_stopping = False)

ranks = evaluate_performance(X_test,
                             model=model,
                             filter_triples=positives_filter,   # Corruption strategy fi
lter defined above
                             use_default_protocol=True, # corrupt subj and obj separatel
y while evaluating
                             verbose=True)

mrr = mrr_score(ranks)
print("MRR: %.2f" % (mrr))

hits_10 = hits_at_n_score(ranks, n=10)
print("Hits@10: %.2f" % (hits_10))
hits_3 = hits_at_n_score(ranks, n=3)
print("Hits@3: %.2f" % (hits_3))
hits_1 = hits_at_n_score(ranks, n=1)
print("Hits@1: %.2f" % (hits_1))
```

```
Average Loss:   0.027143: 100%|██████████| 200/200 [07:10<00:00,  2.15s/epoch]
```

```
WARNING - DeprecationWarning: use_default_protocol will be removed in future. Please use
corrupt_side argument instead.
```

```
100%|██████████| 10000/10000 [01:52<00:00, 88.62it/s]
```

```
MRR: 0.21
Hits@10: 0.37
Hits@3: 0.24
Hits@1: 0.13
```

In [12]:

```
# Randomly select 10 different triples to run the prediction and show the predicted resul
t in your notebook using the format below
random_test_indexes = np.random.choice(X_test.shape[0], 10)
X_unseen = X_test[random_test_indexes]

unseen_filter = np.array(list({tuple(i) for i in np.vstack((positives_filter, X_unseen))
}))

ranks_unseen = evaluate_performance(
    X_unseen,
    model=model,
    filter_triples=unseen_filter,   # Corruption strategy filter defined above
    corrupt_side = 's+o',
    use_default_protocol=False, # corrupt subj and obj separately while evaluating
    verbose=True
)

scores = model.predict(X_unseen)

from scipy.special import expit
probs = expit(scores)

import pandas as pd
pd.DataFrame(list(zip([' '.join(x) for x in X_unseen],
                      ranks_unseen,
                      np.squeeze(scores),
                      np.squeeze(probs))),
             columns=['statement', 'rank', 'score', 'prob']).sort_values("score")
```

```
100%|██████████| 10/10 [00:00<00:00, 55.70it/s]
```

Out[12]:

	statement	rank	score	prob
9	bafta award for best film /award/award_categor...	1067	4.594365	0.989993
7	europe /location/location/contains czech republic	1213	5.777078	0.996912
4	don cheadle /award/award_winner/awards_won./aw...	267	8.492310	0.999795
5	kathleen quinlan /people/person/place_of_birth...	2015	12.807221	0.999997
1	harmonica /music/performance_role/regular_perf...	39	14.166807	0.999999
8	crunk /music/genre/artists lil jon	16	15.645445	1.000000
6	soprano saxophone /music/performance_role/trac...	30	15.965117	1.000000
3	university of delaware /education/educational_...	551	16.134884	1.000000
0	jeffrey jones /award/award_nominee/award_nomin...	6	17.202389	1.000000
2	ratatouille /award/award_winning_work/awards_w...	38	17.536339	1.000000