DHP Flexible Fairness Constraints

May 16, 2021

1 Important links

- Colab Notebook for this code
- Github page with code, instructions to run and download datasets

2 Introduction

In this notebook we provide a demo run of the algorithm suggested by Bose et al.. The code can be found on our github repository linked above. We show results on: - MovieLens-1M (Dataset used by the authors in the paper) - Freebase15k-237 (Synthetic dataset that we tried on our own)

3 Setup

- Library Imports
- Cloning our github repository
- Cloning the Knowledge graph dataset repository inside our repository

[ ]:
```
%tensorflow_version 1.x
from IPython.display import Image
! pip install torch==1.0.0 scikit-learn tqdm pickle-mixin joblib networkx␣
↪comet-ml ipdb
! git clone https://github.com/ShashwatNigam99/
↪Flexible-Fairness-Constraints-for-Graph-Embeddings.git
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings && git clone https://
↪github.com/cmoon2/knowledge_graph.git
```

4 Download and Unzip datasets

- MovieLens-1M: This is a standard recommender system benchmark, where the goal is to predict the rating that users assign movies. We treat the user features (age, gender, and occupation) as sensitive attributes. This recommendation task is an edge prediction problem between users and movies, viewing the different possible ratings as different edge relations. It contains 1 million ratings from 6000 users on 4000 movies.
- Freebase: The Freebase 15k - 237 dataset contains knowledge base relation triples and textual mentions of Freebase entity pairs. It has a total of 592,213 triplets with 14,951 entities and 1,345 relationships. Freebase 15k - 237 is a variant of the original dataset where inverse

relations are removed, since it was found that a large number of test triplets could be obtained by inverting triplets in the training set.

[ ]:
```
# Download Datasets
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings && wget https://
↪download.microsoft.com/download/8/7/0/8700516A-AB3D-4850-B4BB-805C515AECE1/
↪FB15K-237.2.zip
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings && wget␣
↪--no-check-certificate https://files.grouplens.org/datasets/movielens/ml-1m.
↪zip
# Unzip Datasets
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings && unzip -qq ./ml-1m.
↪zip
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings && unzip -qq ./
↪FB15K-237.2.zip
# Renaming to match path specified in code
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings && mv Release fb15k
```

```
--2021-05-15 17:02:07--  https://download.microsoft.com/download/8/7/0/8700516A-
AB3D-4850-B4BB-805C515AECE1/FB15K-237.2.zip
Resolving download.microsoft.com (download.microsoft.com)… 23.78.216.154,
2600:140e:6:b96::e59, 2600:140e:6:ba1::e59
Connecting to download.microsoft.com
(download.microsoft.com)|23.78.216.154|:443… connected.
HTTP request sent, awaiting response… 200 OK
Length: 146221215 (139M) [application/octet-stream]
Saving to: ‘FB15K-237.2.zip’

FB15K-237.2.zip     100%[===================>] 139.45M  96.9MB/s    in 1.4s

2021-05-15 17:02:09 (96.9 MB/s) - ‘FB15K-237.2.zip’ saved [146221215/146221215]

--2021-05-15 17:02:09--
https://files.grouplens.org/datasets/movielens/ml-1m.zip
Resolving files.grouplens.org (files.grouplens.org)… 128.101.65.152
Connecting to files.grouplens.org (files.grouplens.org)|128.101.65.152|:443…
connected.
WARNING: cannot verify files.grouplens.org's certificate, issued by ‘CN=InCommon
RSA Server CA,OU=InCommon,O=Internet2,L=Ann Arbor,ST=MI,C=US’:
  Unable to locally verify the issuer's authority.
HTTP request sent, awaiting response… 200 OK
Length: 5917549 (5.6M) [application/zip]
Saving to: ‘ml-1m.zip’

ml-1m.zip           100%[===================>]   5.64M  3.59MB/s    in 1.6s

2021-05-15 17:02:12 (3.59 MB/s) - ‘ml-1m.zip’ saved [5917549/5917549]
```

5 Model Structure

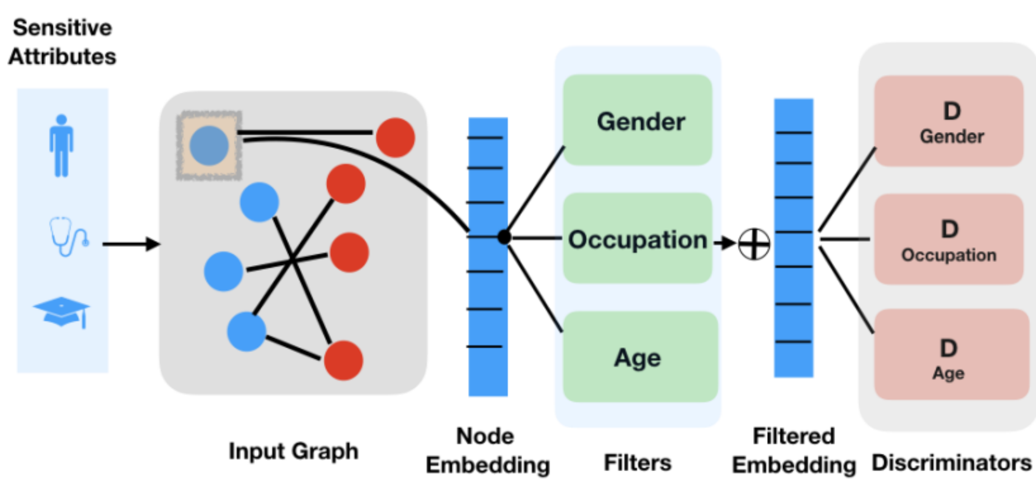
In the model figure, the general case of embedding a heterogeneous or multi-relational (social) graph is considered, which is a set of directed edge triples consisting of starting node, ending node and a relation type or edge between the two. For example, the blue nodes can be considered users, red nodes are movies, and the edges represent movie ratings. Now, consider the task is to predict these edges or relations between the nodes, but we do not want age, gender and occupation to not bias the results. The edge prediction task can essentially be defined as learning a scoring function on edges which should ideally score any true edge higher than any negative edge.

Now, the next step is to encode the node to an embedding as shown in the model figure, where the blue node is mapped to a cell in the node embedding array. Generally, the intuition in embedding-based approaches is that the distance between two node embeddings should encode the likelihood that there is an edge between the nodes.

The invariance to sensitive attributes is enforced by introducing an adversarial loss and a technique to "filter" the embeddings generated by the encoding function. Note, that the set of sensitive attributes we want to be invariant with respect to — is not fixed across nodes; i.e., we may want to enforce invariance on different sets of sensitive attributes for different nodes. Now, to perform this filtering, there is a discriminator defined for every sensitive attribute as shown in the model figure. This attempts to predict the sensitive attribute from the node embeddings and essentially filters it out.!

[2]:
```
Image("./model.png")
```

[2]:



6 Freebase (FB15K-237 Knowledge Base Completion Dataset)

6.1 Preprocess the Dataset and Generate training files

[ ]:
```
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings && python␣
↪construct_ent_attributes.py --dataset FB15k
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings && python parse.py␣
↪--dataset FB15k
# need to enter 'c' for ipdb to continue
```

```
> /content/Flexible-Fairness-Constraints-for-Graph-
Embeddings/construct_ent_attributes.py(85)main()
     84     ''' Count attributes
'''
---> 85     train_attr_count =
count_attributes(train_data,
attr_to_idx)
     86     valid_attr_count =
count_attributes(valid_data,
attr_to_idx)

ipdb>

WARNING: your terminal
doesn't support cursor position requests (CPR).
ipdb>

ipdb>

c
Dataset: FB15k
# entities: 14940; # Attributes: 3851
Dataset: FB15k
# entities: 14541; # relations: 237
train set size: 272115; valid set size: 17535; test set size: 20466
```

6.2 Training

- Comet ml for logging

[ ]:
```
! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings &&  \
python paper_trans_e.py --namestr='FB15k Comp Gamma=1000' \
                        --do_log --num_epochs=100 --embed_dim=20␣
↪--test_new_disc\
                        --sample_mask=True --use_attr=True --gamma=1000␣
↪--valid_freq=50\
                        --api_key wLsj2vcZRKJpZaOZ2p4NpUuOr \
                        --project_name dhp-flexible-fairness-constraints \
```

```
                    --workspace shashwatnigam99
```

6.3 Results and Metrics

AUC scores to calculate the effectiveness of predicting sensitive attributes in Freebase 15K - 237 dataset.

Freebase (FB15K-237)	Baseline	With Invariance Constraints
Attribute 0	0.97	0.80
Attribute 1	0.99	0.80
Attribute 2	0.98	0.82

On the Freebase15k - 237 dataset it was not possible to completely remove the sensitive information without incurring a significant decrease in accuracy on the original edge prediction task and thus the values in the table above, reflect that as it is not close 0.5. This result is not entirely surprising, since for this dataset the "sensitive" attributes were synthetically constructed from entity type annotations, which are presumably very relevant to the main edge/relation prediction task.

Metrics from a previous run can be found at https://www.comet.ml/shashwatnigam99/dhp-flexible-fairness-constraints/f749d3b93e4345f49bfd36145611ca83

7 MovieLens-1M

```
[ ]: ! cd ./Flexible-Fairness-Constraints-for-Graph-Embeddings &&  \
python main_movielens.py --namestr='100 GCMC Comp and Dummy'␣
 ↪--use_cross_entropy \
--num_epochs=200 --test_new_disc --use_1M=True --show_tqdm=True␣
 ↪--report_bias=True \
--valid_freq=5 --use_gcmc=True --num_classifier_epochs=200 --embed_dim=30 \
--sample_mask=True --use_attr=True --gamma=10 --do_log \
--api_key wLsj2vcZRKJpZaOZ2p4NpUuOr --project_name␣
 ↪dhp-flexible-fairness-constraints \
--workspace shashwatnigam99
```

7.1 Results and Metrics

AUC score to calculate the effectiveness of predicting sensitive attribute like gender in MovieLens - 1M dataset and for age and occupation attributes the score is micro averaged F1.

Movielens-1M	Baseline	With Invariance Constraints
Gender	0.71	0.50
Age	0.41	0.33
Occupation	0.14	0.12

The MovieLens - 1M dataset was able to achieve a reasonable tradeoff, with the near-complete removal of the sensitive information leading to a roughly 10% relative error increase on the edge prediction tasks from the baselines. In other words, on the dataset the sensitive attributes were nearly impossible to predict from the filtered embeddings, while the accuracy on the main edge prediction task was roughly 10% worse than a baseline approach that does not include the invariance constraints as is evident in the table above. For, the age and occupation average F1 score is used as it is not a binary attribute. The closer the value of the F1 score to 0, the better is the filtering of the node embeddings.

Metrics from a previous run can be found at https://www.comet.ml/shashwatnigam99/dhp-flexible-fairness-constraints/609f4d6397264b2689cd8594aefb84ba