COMP396 Meeting 6:

Utilizing deep infomax to learn unsupervised representations of molecular graphs

Summary

1. Recall
2. Objective function
3. Implementation
4. Experiments
5. Questions

Recall

Last time we discussed about:

- the implementation (Preferably Unsupervised instead of Semi-Supervised)
- the derivation of the JSD (f-divergence generalization)

Objective function (Motivation)

Mutual information I, is a measure of mutual dependence of two random variables X and Y. It is equivalent to the Kullback Leibler divergence between the joint probability P(X,Y) and the product of marginals P(X)P(Y) .

$$I(X;Y) = \sum P(x,y) log \frac{P(x,y)}{p(x)P(y)} \quad (1)$$

The Kullback Leibler Divergence (Dkl) quantifies how much one probability distribution differs from another probability distribution. For two probability distributions P and Q, it can be simply defined as follows.

$$D_{KL}(q||p) = \int q(x) log \frac{q(x)}{p(x)} dx = \sum q(x) log \frac{q(x)}{p(x)}$$
$$1. D_{KL}(q||p) \neq D_{KL}(p||q)$$
$$2. D_{KL}(q||q) = 0$$
$$3. D_{KL}(q||p) \geq = 0 \quad (2)$$

This measure of distance is asymmetric and always non negative. The Jensen Shannon Divergence is a symmetrization of the KL divergence and it has the following properties.

$$D_{JS}(q||p) = \frac{1}{2} D_{KL}(P||\frac{1}{2}(P+Q)) + 1/2 D_{KL}(q||\frac{1}{2}(P+Q))$$
$$1. D_{JS}(q||p) = D_{JS}(p||q)$$
$$2. D_{JS}(q||q) = 0$$
$$3. D_{JS}(q||p) \geq = 0 \quad (3)$$

In this project we use the f-divergence generalization of the Jensen Shannon divergence measure taken from Infograph/ DIM https://arxiv.org/abs/1808.06670. The definition is a combination of the convex, lower-semicontinuous function and its Fenchel conjugate.

$$I_{\phi,\psi}(h^i_\phi(G); H_\phi(G)) :=$$
$$\mathbb{E}_{\mathbb{P}}[-\text{sp}(-T_{\phi,\psi}(\vec{h}^i_\phi(x), H_\phi(x)))] - \mathbb{E}_{\mathbb{P}\times\tilde{\mathbb{P}}}[\text{sp}(T_{\phi,\psi}(\vec{h}^i_\phi(x'), H_\phi(x)))]$$

Implementation

Dataset: QM9 (11000 samples)

GNN: MPNN (enn-s2s)

SMILES encoder: LSTM

Regression parameters: Laplacian kernel, GridSearch (alpha {0.01,0.1,1.0 ,10,100}), cv=5, Adam optimizer

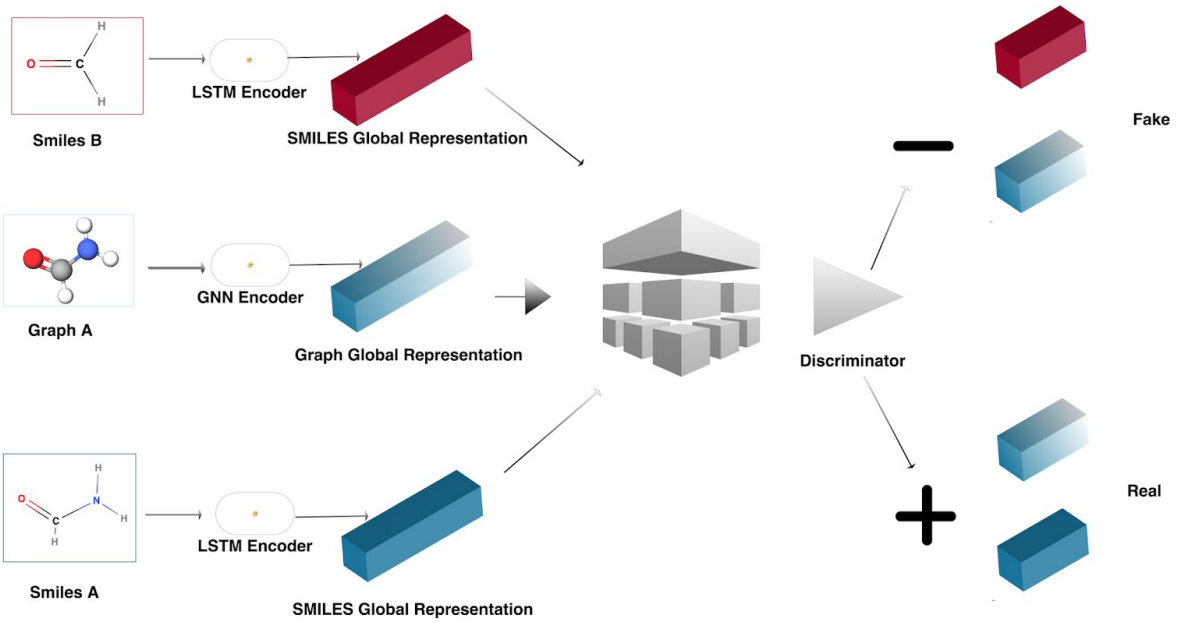
The fake samples are generated by shuffling the feature map at each location across a batch.

Global-Global Infomax

Objective function:

Let I denote the mutual information between the global representation after applying READOUT of the LTSM encoder Y1 and Graph encoder Y2 of the molecular Graph.

$$max_{\phi\omega_1} I_{\phi\omega_1}(Y_{1\phi}(G); Y_{2\phi}(G)) \quad (4)$$



Colab code sample:

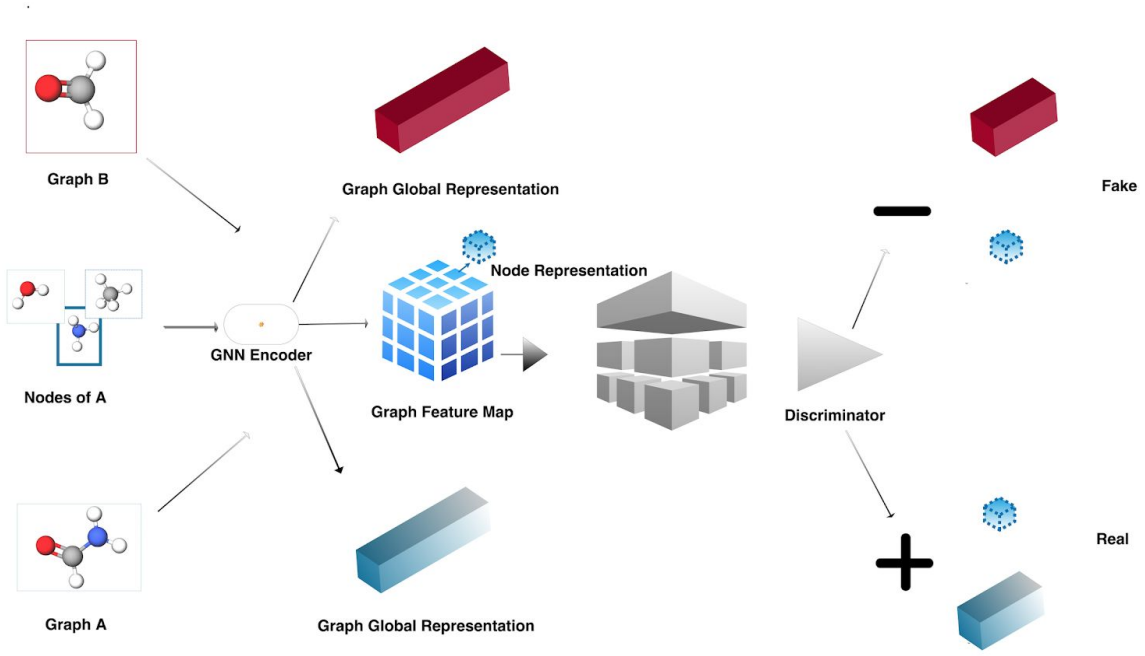
https://colab.research.google.com/drive/1YrWSaq8LJ3Fx8YNgqYROwm_t6eS0mwwo

Local-Global Infomax

Objective function:

Let I denote the mutual information between the summarized patch representation centered at a node of G and the global representation after applying READOUT Y.

$$max_{\phi\omega} \sum_{G\epsilon G} \frac{1}{|G|} I_{\phi\omega}(h_\phi(G); Y_\phi(G))$$



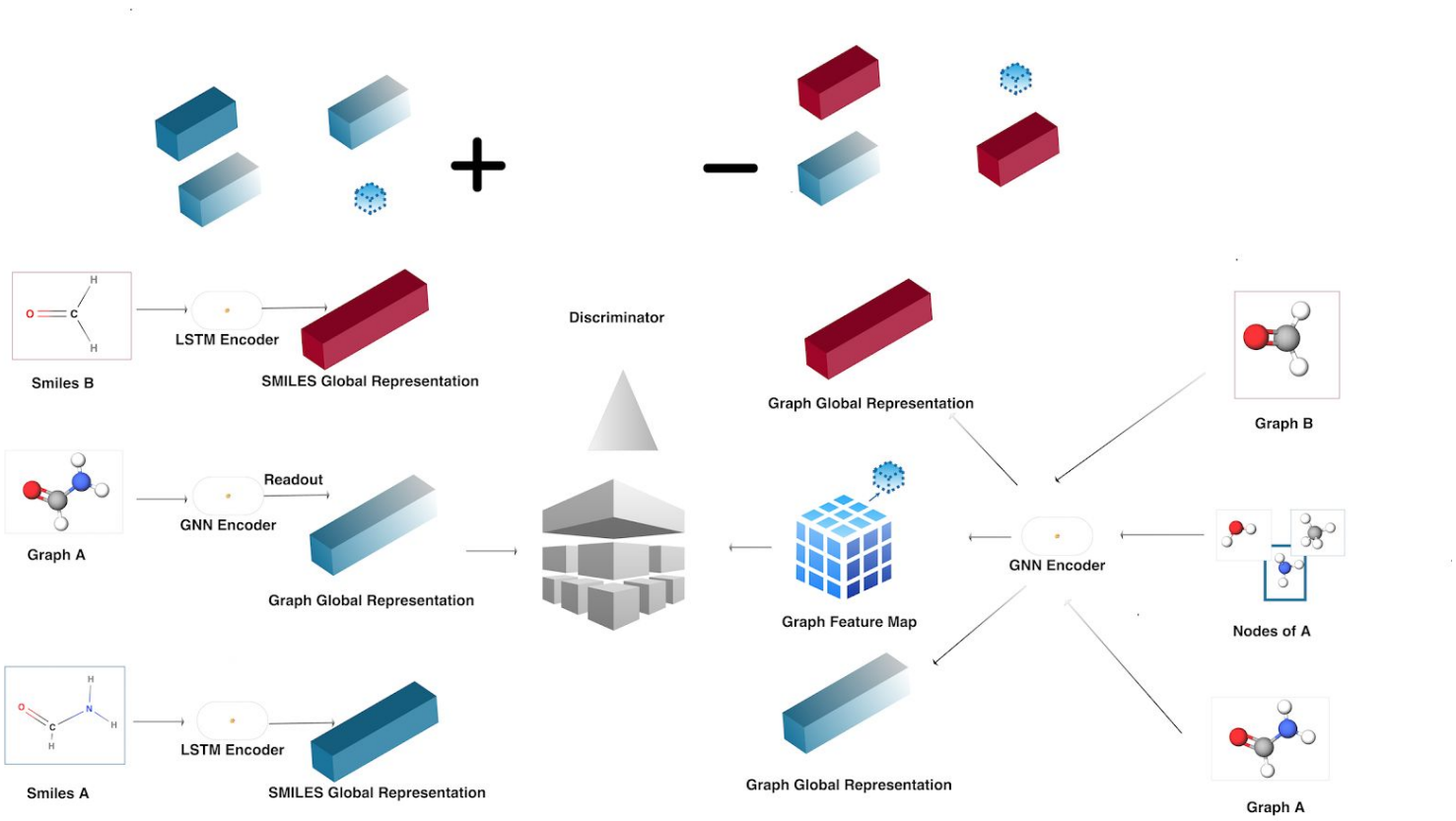
Colab code sample:

https://colab.research.google.com/drive/1yRAf_acWT_AWv8Q9YYo1yiENaMocwfqh

Deep (Molecular) Graph Infomax

The objective for the whole DMGI is defined below. where α, and β are hyperparameters for the global and local objectives, respectively.

$$\alpha \, max_{\phi\omega_1} \, I_{\phi\omega_1}(Y_{1\phi}(G); Y_{2\phi}(G)) + \beta \, max_{\phi\omega_2} \sum_{G\epsilon G} \frac{1}{|G|} I_{\phi\omega_2}(h_\phi(G); Y_\phi(G)) \quad (6)$$



Colab code sample:

https://colab.research.google.com/drive/1TNcCRgA-_ayQf-Scy-osrDHRXpbt_l3l

Experiments

NB:Still in completion

Target	**Global-global** MAE	**Local-global** MAE	**DIM** MAE	**Supervised** MAE(10epochs only)
Mu	0.73064	0.7236	0.7947	0.4091
Alpha	1.6321	1.8344	-	0.8198
Homo	0.0091	0.0079	-	0.0095
Lumo	0.01133	0.0110	-	0.0092
Gap	0.0118	0.0137	-	0.0104
R2	23.9726	35.3079		
ZVPE	0.0030	0.0037		
U0	7.1710			
U				
H				
G				
CV				

Comparisons

Past SMILES LSTM MAE:

Target	Mu	Alpha	HOMO	LUMO	gap	R2	ZPVE	U0	U	H	G	cv
LSTM	0.65684	1.202	0.0100	0.01031	0.0131	78.8518	0.0008	0.70606	0.6116	0.82404	0.6499	0.6373

Previous Baseline results:

Target	BOB	CM	GG-NN	MPNN
mu	0.6724	0.8078	0.7639	**0.1523**
alpha	0.7782	1.4809	0.9228	**0.3847**
HOMO	0.0074	0.017	0.0081	**0.0034**
LUMO	0.0106	0.0171	0.0111	**0.0037**
gap	0.0122	0.0196	0.0118	**0.0065**
R2	22.2674	37.7441	76.7278	**2.5781**
ZPVE	0.0007	0.0010	0.0007	**0.0004**
U0	0.7360	3.4899	**0.3955**	0.5545
U	0.7360	3.4895	**0.2621**	0.5218
H	0.7360	3.4898	**0.3899**	0.3991
G	0.7360	3.4901	**0.4603**	0.4632
CV	0.4156	0.6600	0.5121	**0.1516**

Questions

Should I change the encoder?

Differences in number of epochs during training between unsupervised and supervised learning models?

What is expected for the report ?

Next Steps?