COMP 396: Milestone 2

Representation learning : Graphs & LSTM's

Summary

1. Graph representational learning
2. Smiles + LSTM
3. Next steps (Expectations & report) ?

The second objective of the research project is to reproduce the baseline model Message passing neural networks paper https://arxiv.org/pdf/1704.01212.pdf (i.e Gated graph sequence neural network) and/ or the main message passing neural network that were implemented in the paper with the use of Pytorch Geometric.

MPNN papers : The following are the models listed in the paper as examples that use the neural message passing framework. We will be mainly focusing on the gated graph neural network and the final MPNN that was implemented.

Convolutional Networks for Learning Molecular Fingerprints, Duvenaud et al. (2015)
https://arxiv.org/pdf/1509.09292.pdf

Gated Graph Neural Networks (GG-NN), Li et al. (2016)

https://arxiv.org/pdf/1511.05493.pdf

Interaction Networks, Battaglia et al. (2016)
https://arxiv.org/pdf/1612.00222.pdf

Molecular Graph Convolutions: Moving beyond Fingerprints, Kearnes et al. (2016)
https://arxiv.org/pdf/1603.00856.pdf

Deep Tensor Neural Networks, Schutt et al. ¨ (2017)
https://arxiv.org/pdf/1609.08259.pdf

Laplacian Based Methods, Bruna et al. (2013); Defferrard et al. (2016); Kipf & Welling (2016) ->

Spectral Networks and Deep Locally Connected Networks on Graphshttps://arxiv.org/pdf/1312.6203.pdf

Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering https://arxiv.org/pdf/1606.09375.pdf

Graph representation learning

Variants used in experiments (Neural message passing for quantum chemistry)

Message functions: . -Matrix Multiplication

-Edge Network

-Pair Message

Virtual Graph elements : -Adding a separate “virtual” edge

-Adding master node

Readout functions: - GG-NN readout function

Set2Set readout function

Multiple Towers: ? I don’t really understand this yet

Input Variations:

Nodes: All of the information is concatenated as a vector of size 13 in the pytorch Geometric implementation

Node information	**Description**
Atom type	H, C, N, O, ,F (one-hot or null).
Atomic number	Integer electronic charge.
Partial Charge	-
Acceptor	If the atom accepts electrons

Donor	If the atom donates electrons
Aromatic	If the atom is part of an aromatic system.
Hybridization	sp, sp2 , or sp3 (one-hot or null).

Edges

NB: The chemical graph type is used in the Pytorch Geometric implementation

Types	**Description**
Chemical graph	Bond type
Distance bin	-
Raw distance	Magnitude of the vector

Google colab links

NB: SDF files were used in this experiment.

Gated graph sequence neural network:

Dataset Size: 13k

Test size -1000 Validation size (1000) Training size 10000

Message function -Matrix Multiplication:

Neighborhood aggregation function; add/sum

Readout function : Set2Set -> 1 step

Loss function: MSE

GG-NN + set2set for target mu

https://colab.research.google.com/drive/1yH6_UW00UAzjZBGQSfSyUPDIDmKScyLz

GG-NN + set2set for target alpha

https://colab.research.google.com/drive/1o7aN38ghk5Xekyhzk3JT7fMej9wROVA2

Attempt: **GraphSage with QM9 (**Need to fix feature dimensions in both input and output channels)

SAGE + set2set for target alpha

Message function -Matrix Multiplication:

Neighborhood aggregation function: mean

Readout function : Set2Set -> 1 step

Loss function: MSE

Two different implementations

https://colab.research.google.com/drive/1_Qh-j-ft1OtfXpjAFd1PHEKeqNzPRJBL

https://colab.research.google.com/drive/1V20QPxNNneJWbMjNcz3DShXGH-l2TnWO

NB: I'm not 100% confident yet about how the models are implemented . I followed the same and/or similar structure as the Pytorch Geometric examples.

SMILES + LSTM(still in progress)

-> What I am still working on..

NB: I took a break to finish the graphs

https://colab.research.google.com/drive/12VWhXR4hNd1f5y-H73fdh77yv-oNJiEH

-> What I have tried

Failed attempt: SMILES + LSTM autoencoder

- I was not able to retransform the smiles to their initial values after the training
- No finetuning of results was done

https://colab.research.google.com/drive/15cG4Ju4nPTcsq7LXGf3nwlWlUwPdqFf-

Next Steps/last steps

- Finish LSTM model
- Finish running experiments for all properties
- Validate results
- **Do an analysis of results**
- Write final report