Predicting Molecular Properties with Graph Attention Networks

https://youtu.be/IyWacHGrOQo

Rohan Mehrotra
rohanm2@stanford.edu
Kevin Guo
kyguo@stanford.edu

Introduction

MOTIVATION:

- Design of therapeutics is dependent on tedious assays to screen molecules
- AI has become useful to predict molecular properties
- 2014: NIH releases **Tox21**, a dataset of 7830 molecules with 12 unique tasks.

APPROACH:

- **Graph representations** of molecules have worked well to train networks
- We utilize **graph convolutional neural networks**.
- Incorporate **attention mechanisms** to condition outputs with the most relevant information possible.

RESULTS:

- With an attention mechanism, our model outperforms logistic regression, SVM, and traditional graph convolutional neural networks (GCN).

Dataset

SOURCE:

- Sourced from NIH's National Toxicology program

PROCESSING:

- Raw: 7830 chemical compounds with corresponding 12 binary labels; molecules are represented as SMILES strings
- Processed: DGL library converts SMILES strings into graph objects of *n*-dimensional node feature vectors and lists of edge pairs

SPLIT:

- 80/10/10: training/validation/test

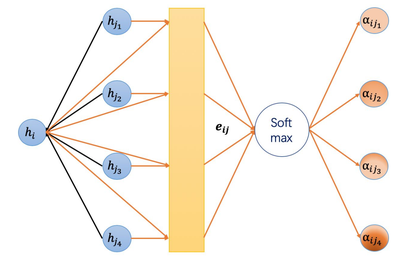
Method

Baseline Models

- Implemented a Logistic Regression model, SVM model, and GCN model via the DeepChem package as benchmark

Graph Attention Network (GAT) Layer

(1)Linear transformation of lower layer embedding and its learnable weight matrix
(2)Leaky ReLU of pair-wise unnormalized attention score between 2 neighbors
(3)Softmax to normalize attention scores
(4)Neighbor embeddings are aggregated and then scaled by attention scores



$$z_i^{(l)} = W^{(l)} h_i^{(l)}, \quad (1)$$

$$e_{ij}^{(l)} = \mathrm{LeakyReLU}(\vec{a}^{(l)^T}(z_i^{(l)} || z_j^{(l)})), \quad (2)$$

$$\alpha_{ij}^{(l)} = \frac{\exp(e_{ij}^{(l)})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik}^{(l)})}, \quad (3)$$

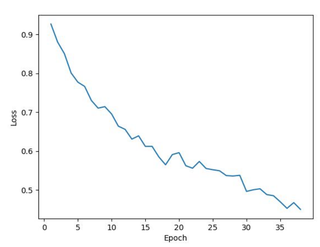
$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} z_j^{(l)}\right), \quad (4)$$

Model Architecture

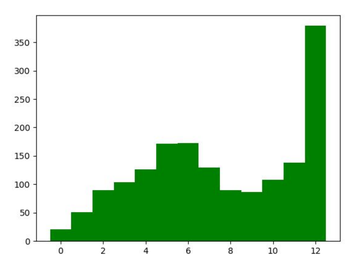
Input: molecular graph object
4-layer architecture: GAT layer (74 -> 32), GAT layer (32 -> 32), GAT later (32 -> 64),
and output sigmoid layer (64 -> 1)x12.
Hyper Parameters: Adam optimizer, batch size = 128, learning rate = 0.001

Results

Loss/Early stopping



* Distribution of Number of Tasks Correct



*Bimodal distribution on task correctness.

Performance Comparison

Model	ROC-AUC
Logistic Regression	0.7129
SVM	0.7672
GCN	0.7810
GAT (Ours)	0.8242

Task	ROC-AUC
1	0.6957
2	0.8553
3	0.8799
4	0.8379
5	0.7311
6	0.8430
7	0.8542
8	0.7856
9	0.8289
10	0.8377
11	0.9185
12	0.8227

Discussion/Conclusion

DISCUSSION:

- Our model manages to outperform all baseline models.
- From analyzing incorrect predictions, we saw that the model is more accurate on smaller, linear molecules.
- We rationalize better performance on smaller, linear molecules due to better learned local environments and lack of aromaticity.
- Moreover, attention mechanisms learn local environments even more effectively.

CONCLUSION:

- Our GAT model managed to outperform traditional deep learning methods for molecule prediction.

FUTURE WORK:

- Adding features or edge attention weights for aromaticity and other attributes to overcome shortcomings with the current model

References

[1] Duggan, Susan, and Norma Pence. "A Tough Road: Cost To Develop One New Drug Is $2.6 Billion; Approval Rate for Drugs Entering Clinical Development Is Less Than 12%." Policy & Medicine, Dec. 2014, www.policymed.com/2014/12/a-tough-road-cost-to-develop-one-new-drug-is-26-billion-approval-rate-for-drugs-entering-clinical-de.html.
[2] National Toxicology Program. Annual Report for Fiscal Year 2016. Research Triangle Park, NC: National Toxicology Program; 2017. Available from https://ntp.niehs.nih.gov/annualreport/2016
[3] Mayr, Andreas, et al. "DeepTox: Toxicity Prediction Using Deep Learning." Frontiers in Environmental Science, vol. 3, 2016, doi:10.3389/fenvs.2015.00080.
[4] Huang, R. (2019). Predictive Modeling of Tox21 Data. In Challenges and Advances in Computational Chemistry and Physics (pp. 279–297). Springer International Publishing. https://doi.org/10.1007/978-3-030-16443-0-14
[5] Thomas Unterthiner, Andreas Mayr, Günter Klambauer: "Toxicity Prediction using Deep Learning", 2015; [http://arxiv.org/abs/1503.01445 arXiv:1503.01445].
[6] Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing: "MoleculeNet: A Benchmark for Molecular Machine Learning", 2017; [http://arxiv.org/abs/1703.00564 arXiv:1703.00564].
[7] Michaël Defferrard, Xavier Bresson: "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering", 2016, Advances in Neural Information Processing Systems 29 (2016); [http://arxiv.org/abs/1606.09375 arXiv:1606.09375].
[8] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò: "Graph Attention Networks", 2017; [http://arxiv.org/abs/1710.10903 arXiv:1710.10903].
[9] Minkin, V. I., Mikhail N. Glukhovtsev, and B I. Simkin. Aromaticity and antiaromaticity : electronic and structural aspects. New York: Wiley, 1994. Print.