Using Graph Convolutional Network to Predict Small‑molecule Synthesizability

# Introduction

Synthesizability…

Deep learning has achieved success in various domains, such as machine translation, speech recognition, object detection. In structural biology, Alphafold [2], a deep learning neural network, has shown its power in determining protein structure and has achieved unprecedented accuracy compared with other traditional methods in the Critical Assessment of Protein Structure Prediction (CASP13).

While deep learning effectively captures hidden patterns in Euclidean data such as image or audio, it is challenging to apply deep learning to small-molecules due to its graph data structure [3]. For one thing, the variable size of unordered atoms, and different neighbors for an atom makes some important operations (e.g., convolutions) easy to compute in the image domain, but difficult to apply to the small-molecule domain. For another, a core assumption of existing machine learning algorithms is that instances are independent of each other. This assumption no longer holds for small-molecule because each atom is related to others by bonds.

In order to apply machine learning algorithms, graph embedding or network embedding methods have been proposed [1]. Network embedding essentially represents atoms as low-dimensional vector representations, preserving both molecule topology structure and atom information, so off-the-shelf machine learning algorithms can be applied. Examples include … Meanwhile, instead of using hand-crafted features, GNN are deep learning models aiming at addressing graph-related tasks in an end-to-end manner.

Here, we propose to use graph convolutional neural network (GCN) to predict the synthesizability of a designed small-molecule.

# Method

Introduction to GCN.

Aim I: design structure

Aim II: benmark on different dataset. Intro to some databases

# Broader Impact

The project seeks to improve synthesizability prediction using a graph-based machine learning method. This opens the door to the prediction of many other small-molecule properties.

# Bibliography

[1] Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu, 2017. A Survey on Network Embedding. In *arXiv e-prints*, arXiv:1711.08752.

[2] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis, 2020. Improved protein structure prediction using potentials from deep learning. *Nature 577*, 7792 (2020/01/01), 706-710. DOI= <http://dx.doi.org/10.1038/s41586-019-1923-7>.

[3] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip Yu, 2019. *A Comprehensive Survey on Graph Neural Networks*.