Optimizing Molecular Graphs in the Wasserstein Space
Thesis Proposal

Octav Dragoi

May 6, 2020

Abstract

The aim of this project is to create a computational framework for optimization of molecular graphs with respect to certain desirable properties, in the space of Wasserstein point clouds.

1   Motivation

Automating molecular generative processes is a key step in improving the research process for novel molecules with desirable properties. There is a wide range of applications for automated molecule discovery, for example finding new pharmaceutical drugs or composite materials. Current research processes are mostly done by hand, relying on the intuition and experience of experts. Given the lengthy and expensive testing period until a new drug receives definitive approval, there is a real need for an automated, data-driven process for coming up with novel molecular structures.

This project approaches the problem of graph generation from the perspective of **graph optimization**, that is, starting with a given graph and modifying it to improve its desirable properties (antiviral action, bioactivity, etc) and reduce its undesirable properties (toxicity, etc).

2   Current State and Limitations

Current state of the art models for molecule generation are based on Graph Neural Networks (GNNs) or Message Passing Neural Networks (MPNNs) specially designed for molecule applications, such as JTVAE [1]. These models learn embeddings of molecules in metric spaces where optimization can be performed, then decode the optimized results back into corresponding molecular structures. Essentially, this pipeline is split in three different parts: encoding, optimization, and decoding.

Existing encoding approaches seek to learn embeddings for every node in the graph, and then create a graph embedding by applying an aggregation function, such as summation or average. This approach is lossy, meaning that at this step it can potentially lose structural information, such as pairwise node interactions. Even state of the art models for molecule property prediction, such as ChemProp [2] or DimeNet [3], proceed in this way.

For decoding, existing literature focuses on sequential, or autoregressive, molecular generation. In this process, atoms are generated one by one, together with the connections to the previous

1

atoms. In [1, 4, 5, 6], tree and graph decoders pass through the nodes one by one in a depth-first fashion, predicting whether a node has children or not, and what type of connections do they have. Other graph generative models described in [7] use auto-regressive models (e.g GraphRNN) or deconvolutional models (e.g. GraphVAE). All these models have the significant drawback of being dependent on the order in which the atoms are generated, as well as yielding invalid intermediate results that might not satisfy validity constraints, even if the final result does.

3   Project Overview

In order to overcome the limitations of the existing pipeline for graph optimization detailed above, we are proposing some improvements that should expand in the critical areas mentioned in the previous section.

To reduce structural and semantic information loss and preserve graph structure, we study two different graph encoding approaches, relying on the theory of optimal transport [8]. One consists in embedding the graph nodes and transforming the graph into a point cloud, where we can measure distances between these clouds using the Wasserstein distance. Alternatively, we can embed the graph nodes together with the edges into an adjacency tensor, and use the Gromov-Wasserstein distance between such tensors.

As a start, we will begin by studying the simpler method, graph embeddings as point clouds. For our optimization step, we plan to use the Wasserstein distance between point clouds to define the so-called "Wasserstein space", a metric space of point clouds. Each molecular graph is embedded as a point cloud in this space, and then optimized by moving in certain "directions", or "velocities". In a manner similar with tools from Riemannian geometry, these directions contain the information on how should we modify our graph in order to improve its properties.

This optimization process can be formalized within the Wasserstein space and geometry, which are suitable for dealing with sets of points. To do this in an efficient manner, we construct a Riemannian-like framework in which we can average datapoints, construct geodesics, compute barycenters, gradients, tangent spaces, velocities, exponential and logarithmic maps, all necessary tools to manipulate and optimize functions over the Wasserstein space. We draw inspiration from existing frameworks such as [9], and adapt them to our use cases. We will also explore different options to optimize computational efficiency, like in [10, 11, 12].

This framework also allows us to construct molecular graphs directly and not sequentially. The latent representation as a set of node embeddings is decoded back into a molecular graph all at once, without the need to establish some (naturally biased) order in which to decode the molecule structure. This yields a permutation invariant model, useful to eliminate inductive biases that appear in sequential models.

Finally, another advantage of embedding the graph as a point cloud or adjacency tensor is the preservation of its inter-node structure. Using point cloud embeddings retains a representation for each individual node, and adjacency tensor embeddings additionally encodes edges between nodes. These inter-node relations are more explicitly captured this way, compared to compressing graphs into low-dimensional representations. However, higher complexity comes at a computational cost, both point cloud and adjacency tensor models having harder training stages and more intense distance computations.

References

[1] W. Jin, R. Barzilay, and T. S. Jaakkola, "Junction tree variational autoencoder for molecular graph generation," *CoRR*, vol. abs/1802.04364, 2018.

[2] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen, and R. Barzilay, "Analyzing learned molecular representations for property prediction," 2019.

[3] J. Klicpera, J. Groß, and S. Günnemann, "Directional message passing for molecular graphs," in *International Conference on Learning Representations*, 2020.

[4] W. Jin, K. Yang, R. Barzilay, and T. S. Jaakkola, "Learning multimodal graph-to-graph translation for molecular optimization," *CoRR*, vol. abs/1812.01070, 2018.

[5] W. Jin, R. Barzilay, and T. Jaakkola, "Hierarchical generation of molecular graphs using structural motifs," 2020.

[6] W. Jin, R. Barzilay, and T. Jaakkola, "Hierarchical graph-to-graph translation for molecules," 2019.

[7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *CoRR*, vol. abs/1901.00596, 2019.

[8] T. Vayer, L. Chapel, R. Flamary, R. Tavenard, and N. Courty, "Optimal transport for structured data with application on graphs," 2018.

[9] S. Chowdhury and T. Needham, "Gromov-wasserstein averaging in a riemannian framework," 2019.

[10] V. Seguy and M. Cuturi, "Principal geodesic analysis for probability measures under the optimal transport metric," 2015.

[11] B. Heeren, C. Zhang, M. Rumpf, and W. Smith, "Principal geodesic analysis in the space of discrete shells," *Computer graphics forum*, vol. 37, 06 2018.

[12] M. Cuturi and A. Doucet, "Fast computation of wasserstein barycenters," 2013.