Appendix for "Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation"

Jiaxuan You[1]*
jiaxuan@stanford.edu

Bowen Liu[2]*
liubowen@stanford.edu

Rex Ying[1]
rexying@stanford.edu

Vijay Pande[3]
pande@stanford.edu

Jure Leskovec[1]
jure@cs.stanford.edu

[1]Department of Computer Science, [2]Department of Chemistry, [3]Department of Bioengineering
Stanford University
Stanford, CA, 94305

1 Appendix

Validity. We define a molecule as valid if it is able to pass the sanitization checks in RDKit.

Valency. This specifies the chemically allowable node degrees for an atom of a particular element. Some elements can have multiple possible valencies. At each intermediate step, the molecule environment checks that each atom in the partially completed graph has not exceeded its maximum possible valency of that element type.

Steric strain filter. Valid molecules can still be unrealistic. In particular, it is possible to generate valid molecules that are very sterically strained such that it is unlikely that they will be stable at ordinary conditions. We designed a simple steric strain filter that performs MMFF94 forcefield [1] minimization on a molecule, and then penalizes it as being too sterically strained if the average angle bend energy exceeds a cutoff of 0.82 kcal/mol.

Reactive functional group filter. We also penalize molecules that possess known problematic and reactive functional groups. For simplicity, we use the same set of rules that was used in the construction of the ZINC dataset, as implemented in RDKit.

Reward design implementation. For property optimization task, we use linear functions to roughly map the minimum and maximum property score of ZINC dataset into the desired reward range. For property targeting task, we use linear functions to map the absolute difference between the target and the property score into the desired reward range. We threshold the reward such that the reward will not exceed the desired reward range, as is described in Section 4.1. For specific parameters, please refer to the open-sourced code: `https://github.com/bowenliu16/rl_graph_generation`

References

[1] T. A. Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of computational chemistry*, 17(5-6):490–519, 1996.

*The two first authors made equal contributions.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.