Synthetic Accessibility Evolutionary Algorithm Paper

Steven Bennett, Kim E. Jelfs

Paper Draft November 2019

1 Abstract

Using a database filtered for synthetically viable reactions, porous organic cages were systematically constructed using stk, and its associated evolutionary algorithm, the chemical space of cages were explored. Using training data obtained from an expert, synthetic accessibility was defined as a classification problem, and trained on the fingerprints of greater than 10,000 molecules. Feature selection was used to identify the greatest contributing factors to synthetic accessibility. Finally, three several synthetic accessibility models were used as part of the fitness function in the evolutionary algorithm to filter unsynthesisable, ensuring the proposed porous organic cages are synthetically accessible.

2 Introduction

One of the main bottlenecks in the high-throughput materials screening workflow is predicting new materials that are synthetically viable.

3 Methods

3.1 Classification Model

3.2 Evolutionary Algorithm

4 Results and Discussion

4.1 Shape Persistent Cages

4.2 Largest Pore Volume

4.3 Identifying Synthesisable Molecules

5 Conclusions

It is possible to envisage a high-throughput experimental synthesis workflow, learning from the results from previous pathways for synthetically viable compounds.