Accelerating geometry optimisations with machine learning

Robert A. Shaw‡, Adam N. Hill*, J. Grant Hill
Department of Chemistry, University of Sheffield, S3 7HF, UK
‡rashaw1@sheffield.ac.uk, *anhill1@sheffield.ac.uk

Introduction

Often one of the first hurdles to performing geometry optimisations is the question of what is a reasonable starting geometry? This is followed quickly by the arduous process of building the initial guess. What if we could take a drawn chemical structure and turn it into a reasonable starting guess for use in further computational optimisation?

Computational geometry optimisation
Generated starting guess
Structure optimisation

Image recognition

The wealth of structural data now available means that machine learning is ideally suited to speeding up the process of generating a first guess. We demonstrate how it can be used to generate a reasonable 3D structure from a simple chemical graph. This method is lightweight and rapid enough to be turned into a smartphone app, with both practical and outreach uses.

Method

1. Take a set of known molecular structures (ANI1 dataset[, up to 7 heavy atoms]

2. Convert to molecular graph/adjacency matrix

3. Train recurrent neural net (RNN) on RMSD error in 3D geometry

4. Fine tune hydrogen placement using reinforcement learning with a simple force field

Training

The RNN was trained to minimise the root-mean-square deviation (RMSD) between the generated and target geometries. Reducing the features to the first two principal components shows good agreement.

Reinforcement

Most of the latent error was due to misplaced hydrogens. This was considerably improved by rewarding any adaptions that improved the balance of forces between atoms, as measured using a simple force field.

Validation

Structures for a separate validation set were compared to those made conventionally using Avogadro[]. The graph-generated structures were much closer to the DFT-optimised targets, needing a quarter of the steps on average.

Extension

While not possible to extrapolate, the method works for similar but much larger systems.

Future Work

The above is only a proof-of-concept demonstration. There are still a number of issues to tackle:

- include atoms beyond C, H, N, O;
- introduce stereochemistry;
- develop the computer vision side to read hand drawn chemical structures.

The work also raises interesting questions about minimal representations of molecules.

References