Towards Automated Reaction Kinetics with Message Passing Neural Networks

Lagnajit Pattanaik

September 8, 2022

Predictive chemistry holds great promise in accelerating scientific discovery and innovation. Only by understanding the key reactions and pathways present in a given chemical system can scientists hope to manipulate them for the benefit of society. An approach towards predictive chemistry involves decomposing systems into kinetic mechanisms consisting of elementary reactions and quantitatively describing each of those reactions. Incredibly, the immense progress in computational methods and compute power now allows the calculation of thermodynamic and kinetic parameters of elementary reactions at an accuracy necessary for predictive chemistry. Real systems can consist of tens of thousands of elementary chemical reactions, so it is infeasible to calculate these parameters by hand. Furthermore, general-purpose parameter estimators are not yet accurate across diverse areas of chemical spaces, which necessitates the use of expensive and labor-intensive computational methods to obtain them.

This thesis focuses on computing kinetic parameters by both automating and accelerating the computational pipelines used to generate them, relying on modern machine learning frameworks—specifically, message passing neural networks—to facilitate these calculations.

Noting that in the framework of automated kinetic parameter calculation, transition state search is a primary bottleneck, this thesis first devises a method to generate transition state geometries with deep learning. The new method achieves improvements in both accuracy and speed compared to existing alternatives, demonstrating that high-throughput quantum chemistry pipelines can provide ripe data to develop machine learning algorithms to, in turn, accelerate these pipelines. This thesis next investigates a fundamental limitation of message passing neural networks to capture tetrahedral chirality and proposes several fixes to address this limitation.

While generating a single transition state structure is an important goal, accurate calculation of kinetic parameters often requires investigating multiple conformations. Hence, this thesis builds a generative framework to predict multiple low-energy conformations directly from the molecular graph. The method is demonstrated for stable species conformer generation and outperforms existing baselines in a number of benchmark studies. Integrating all the developed models together. this thesis next develops an end-to-end pipeline to generate transition state conformers directly from the atom-mapped reaction SMILES. The pipeline embeds both the transition state guess method and the stable species conformer generation method from previous chapters. It further develops a new method for transition state generation using equivariant graph networks and uses semi-empirical quantum mechanics to quickly interrogate multiple transition state conformers.

Although most of presented work investigates reactions in the gas phase, reactions in condensed phase require additional solvation corrections. To facilitate calculation of these corrections, this thesis constructs a large dataset of solution free energies across a range of solvents. It then develops a model to predict relevant conformations of the solute for any given solute-solvent pair.

The tools developed in this thesis will become an integral part of modern computational chemistry pipelines. Undoubtedly, the future of automated predictive chemistry will heavily rely on these and similar deep learning models for fast and accurate parameter estimation.