

Comparison of Computer Simulation Methods for Predicting Chemical Reactions

Graham Gibson

IBM, Austin TX, 78723, USA.

(Dated: December 29, 2015)

In this paper we compare two primary methods of predicting basic organic chemistry reaction predictions. We analyze two types of models, an NLP based Neural Network and an agent based model. We compare and contrast the complexity, accuracy, and generalizability of both models as applied to predicting organic chemistry reactions. Two basic reaction mechanisms are explored, Elimination and Addition reactions. These two mechanisms are simple but fundamental to the set of organic reactions, as many more complicated reactions use these mechanisms as intermediates. We first verify the model on an alkene halogen addition reaction and then investigate the models generalizability to the elimination reaction.

PACS numbers: PACS numbers go here. These are classification codes for your research. See <http://publish.aps.org/PACS/> for more info.

I. INTRODUCTION

The problem of reliably predicting organic chemical reactions is vital in computation chemistry and biology. A model that is able to describe 3-dimension conformations of reactions would go a long way towards solving the more complicated protein folding problem. Finding such a solution would have extremely beneficial consequences in the area of targeted drug development and computational medicine. Such a solution would also drastically reduce drug development time and help eliminate clinical trial risk.

Traditional computational chemistry approaches have involved analyzing the potential energy surface of the atoms in the n-molecule system. The energy surface contains an array of parameters from 3-d cartesian coordinates, to electronegativity, to electron-electron repulsion. Finding the most likely product is equivalent to finding minima in the energy surface using traditional techniques like gradient descent. This method is often termed molecular dynamics. A second approach that is frequently taken is to go deeper into the reaction and look at the driving quantum mechanical principles that are at play. These methods use a combination of approximation techniques to guess the Hamiltonian of an atom and analytically solve for the orbital energy. This allows one to identify reactive orbitals very precisely. The two primary methods are called Hartree-Fock and Density Functional Theory. Although these approaches are invaluable in their description of the energy of chemical reactions, they are both quite complicated and computationally involved. The quantum mechanical approaches are extremely resource intensive as they involve the approximation of complicated integrals and matrix multiplications. It also appears that both methods are essentially overkill when trying to predict the result of organic reactions. The reasoning behind this is that chemistry students begin studying reaction mechanisms and gain the ability to predict basic reactions long before they learn about solving Hamiltonian equations. Organic chemistry textbooks focus on what we call the "grammar" of organic chemistry. This involves studying common reaction mechanisms and patterns of behavior of common functional groups such as OH, COOH, alkenes etc. Organic chemistry students learn chemistry as a combination of a language of interacting words and they verify reaction products as syntactically and semantically valid under such language. Using this view we can define two separate higher level approaches to predicting chemical reactions.

1. First Approach: Language Based Modeling

Continuing with the idea of modeling organic reactions as language over the alphabet of atoms, predicting a chemical reaction has a natural analogy between verifying if a sentence is valid in a given language. This problem has been studied in depth in the field of computer science and has been most successfully modeled by using recursive neural networks (RNNs) such as the Long Short Term Memory network architecture. This architecture maintains state of the neural network by copying the hidden state H_{t-1} to the hidden state H_t using a series of memory gates. LSTM networks have gained a lot of popularity in the computer science community for their ability to remember context from many time steps previous to the current because of this copying of hidden layer state. A common LSTM network approach to the problem of machine translation is depicted below as in [?] -sequence-to-sequence-learning-with-neural-networks)

II. RESULTS

Including figures, tables, and equations is easy. Latex also permits easy reference to document elements (figures, tables, sections) with the

`\ref`

command2. Citations are made with the

`\cite`

command[1].

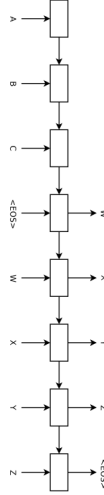


FIG. 1: You will need to include the package graphicx to be able to make figures like this.

FIG. 2: You will need to include the package graphicx to be able to make figures like this.

A simple table.

TABLE I: $X(3872)$ Discovery Modes.

mass	width	production/decay mode	events	significance	experiment
$3872.0 \pm 0.6 \pm 0.5$	< 2.3 90% C.L.	$B^\pm \rightarrow K^\pm X \rightarrow K^\pm \pi^+ \pi^- J/\psi$	25.6 ± 6.8	10σ	Belle
$3871.3 \pm 0.7 \pm 0.4$	resolution	$p\bar{p} \rightarrow X \rightarrow \pi^+ \pi^- J/\psi$	730 ± 90	11.6σ	CDFII
$M(J/\psi) + 774.9 \pm 3.1 \pm 3.0$	resolution	$p\bar{p} \rightarrow X \rightarrow \pi^+ \pi^- J/\psi$	522 ± 100	5.2σ	DØ
3873.4 ± 1.4	—	$B^- \rightarrow K^- X \rightarrow K^- \pi^+ \pi^- J/\psi$	25.4 ± 8.7	3.5σ	BaBar

And a sample equation (Eq. 1).

$$\Gamma(X \rightarrow \alpha\beta D) = \int \frac{d^3 Q}{(2\pi)^3} \Gamma(C \rightarrow \alpha\beta) \frac{|\tilde{T}(Q)|^2}{(M(X) - E_{CD}(Q))^2 + \Gamma_C^2/4} \quad (1)$$

III. CONCLUSIONS

Man, latex is great!

Acknowledgments

The author is grateful to Donald Knuth for inventing tex, and making publication quality typesetting a reality for scientists around the world.

-
- [1] *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*, Leslie Lamport [1994] (ISBN: 0-201-52983-1) pages: xvi+272.
- [2] I.M. Smart *et al.*, J. Plumb Phys. **50**, 393 (1983).