Improving NLP for Under-resourced Languages

C.M. Downey, PhD

Department of Linguistics and Goergen Institute for Data Science

In this talk, I will survey recent projects of mine focused on developing NLP tools for endangered and otherwise under-resourced languages (those without the large text datasets needed to robustly train machine learning models). Under-resourced languages, including some with millions of speakers, face a methodological gap in computation, given the dominant approach for languages like English is to train large neural networks on vast text datasets — a resource most languages lack. This disparity undermines the vital role machine learning can play in ensuring a diversity of world languages can thrive in the digital era. To overcome this limitation, I specialize in techniques like unsupervised learning, multilingual modeling, and transfer learning, which can be employed successfully in the face of data scarcity. This talk will highlight case studies in which I leverage these techniques for application to acutely under-resourced languages. It is my hope that this work may be applicable to Indigenous and endangered language revitalization, both by developing tools to assist field linguists in the study of these languages, and by promoting equitable language technology access in the digital era.

Application of Molecular Dynamics to Elucidate Key Binding Interactions for the Development of Therapeutics Against Opioid Overdose

Emily Robinson, Department of Biochemistry & Biophysics

Opioid overdose has been a long-standing public health issue in the United States. This crisis has only been exacerbated by the Covid-19 pandemic, with overdose deaths involving opioids increasing from an estimated 70,630 in 2019 to 107,941 in 2022. There are treatments for opioid overdose, however currently no therapeutic tools exist to prevent it. Moreover, treatments are limited to emergency scenarios due to induction of withdrawal and loss of analgesia. Fatal opioid overdoses are primarily attributed to opioid-induced respiratory depression (OIRD). As part of an ongoing collaboration, a class of cysteine esters have been identified that reverse OIRD without blocking analgesic effects or inducing withdrawal. The current hypothesis is that these esters function by binding β-arrestin, a protein that signals downstream of the opioid receptors. The goal of my proposed work is to characterize this binding interaction to rationalize the trends observed in the preliminary data. I will apply molecular dynamics simulation techniques to elucidate the molecular interactions of these cysteine esters. Using these techniques, I have identified preliminary binding sites for some members of this class of cysteine esters to the inactive structures of β-arrestin 1 and β-arrestin 2. I am using alchemical free energy calculations to determine the affinity of the candidate binding sites; the results will be tested experimentally by collaborators using surface plasmon resonance and hydrogen-deuterium mass spectroscopy. These techniques will be repeated looking at the active structures of these proteins to help further characterize these binding interactions and form hypotheses toward mechanism of action.

Friday, October 18, 2024 | 11:30 AM – 1:00 PM | Wegmans 1400

UNIVERSITY of ROCHESTER