Using Machine Learning To Dynamically Filter Nanopore DNA Sequencing Data

Philip, Aaron (School: Los Alamos High School)
Svyatsky, Leonid (School: Los Alamos High School)

In this project we use machine learning (ML) to dynamically filter long reads of data to remove noise and apply our method to nanopore DNA sequencing. Although nanopore sequencing presents many advantages over alternate sequencing methods due to its efficiency and low cost, the method is relatively less accurate than other sequencing methods. Effective filtering increases the viability of this method. We computationally modeled circuit filters in Python and wrote our own optimizer with a custom loss function that was capable of choosing filtration parameters such that a noisy signal could be filtered to optimally match a ground truth signal. To verify that the optimizer was working correctly, we constructed the circuit filters and created voltage waveforms using a Raspberry Pi which was then filtered through the constructed circuits. By comparing our collected voltage data to computational filters with the same parameters, we verified that our optimizer converged correctly. We built our own database of optimal circuit filters by using our filtration and optimizer code then trained neural networks to predict filter parameters when passed time-series input by filtering different segments differently based on noise present. Currently, our networks achieve a variety of performances. We will apply to data from a known sequence and determine whether a nanopore basecaller can more accurately predict nucleotide sequences after our dynamic filtration. Additionally, we will rework our ML architectures to learn from frequency data rather than time-series data.

Awards Won:
King Abdulaziz &amp
his Companions Foundation for Giftedness and Creativity: Full Scholarship from King Fahd University of Petroleum and Minerals(KFUPM) (and a $400 cash prize)
King Abdulaziz &amp
his Companions Foundation for Giftedness and Creativity: NOT TO BE READ -- $400 cash prize for each Full Scholarship from King Fahd University award recipient