

Hidden Markov Models on DNA Sequencing Data

Sherly Yaghoubi

Hidden Markov Models have previously been used to identify single-molecule interactions on Molecular Electronics chips with a great degree of success. Currently, the new problem is how to identify DNA sequences read by these chips. Hidden Markov Models have the possibility of identifying the 5 states present in DNA sequencing data. This is tested in Python using the package `hmmlearn` to create an HMM algorithm with 5 possible states. DNA data was generated at varying levels of noise and standard deviations before being sent through the HMM algorithm. This presentation reviews the results of a HMM algorithm on DNA sequencing data, and analyzes possible future routes.