Name: Sara Mostafavi Email: saramos@cs.washington.edu

Sequence-based deep learning models for predicting molecular phenotypes

Sara Mostafavi

University of Washington

Deep learning (DL) methods have recently become state-of-the-art in a variety of regulatory genomic tasks. By adapting convolutional neural networks (CNNs), these models take as input sub-sequences of genomic DNA and predict as outputs functional properties such as epigenomic modifications and gene expression. Once trained, an extraordinary promise of these sequence-based DL models is to make predictions from personalized genomes to assess the joint impact of genetic variation on molecular phenotypes. Thus, these models promise to serve as an important tool in interpreting the full spectrum of genetic variations in personal genomes. In this talk, I will describe our efforts in training sequence-based DL models, and exploration into when and how we can robustly interpret them to derive new biological insights about gene regulation. I will also discuss our systematic benchmarking effort to evaluate sequence-based DL methods as personal genome interprets, which highlights areas of challenges for current models. In summary, this talk will explore the promise and current limits of sequence-based DL models when applied to human genetics datasets.