Department of Applied and Computational Mathematics and Statistics

Amirali Aghazadeh

Department of Electrical Engineering
and Computer Science
University of California, Berkeley

Inferring Biological Functions with Explainable Algorithms

Data-driven machine learning models that infer biological functions from sequences are replacing the costly experimental measurements in a number of application areas in biology including protein, small-molecule, and genome engineering. These data-driven models largely owe their success to the recent advancements in over-parameterized models in machine learning such as deep neural networks (DNNs). However, the number of labeled sequences available for training such models has remained small compared to the vastness of the combinatorial sequence space. In addition, these biological functions are typically complex, manifesting as rugged landscapes with high-order nonlinear (epistatic) interactions. The combination of these two factors makes the biological inference problem statistically challenging.

In this talk, I view the problem of inferring biological functions from a statistical signal processing perspective. I first discuss a fundamental interpretation-computation tradeoff in explaining DNNs in terms of their epistatic interactions. I then discuss how to develop a new hybrid algorithm that blends techniques from optimization and coding theory to regularize DNNs for inducing a biologically-relevant prior into their architecture. Our combinatorial method enables DNNs to predict protein functions using up to three times less number of sequences and explains them in terms of their higher-order epistatic interactions.

Amirali Aghazadeh is a postdoctoral researcher in the Electrical Engineering and Computer Science department at the University of California, Berkeley, working with Kannan Ramchandran and Jennifer Listgarten. Prior to that, he was a postdoctoral researcher at Stanford University working with David Tse. He received his PhD degree in Electrical and Computer Engineering from Rice University with Richard Baraniuk in 2017. His research interest is at the intersection of machine learning, signal processing, inverse problems, and computational biology. He is the recipient of the Hershel M. Rich Invention Award for his thesis on rapid methods for DNA sensing as well as the Texas Instruments Fellowship for his graduate studies. He received his Bachelor's degree in Electrical Engineering from Sharif University of Technology.

Monday, April 19, 2021 | 4:30 PM

Zoom Link:
https://notredame.zoom.us/j/91485916854?pwd=WHN0VkdtTFRwZmVHaDhBNUc4bUVEQT09

Meeting ID: 914 8591 6854
Passcode: 718308