

1173-92-46

David Murrugarra* (murrugarra@uky.edu), Department of Mathematics, University of Kentucky, Lexington, KY 40506. *RNA secondary structure prediction via state inference with machine learning techniques*. Preliminary report.

RNA state inference is the task of determining which nucleotides of an RNA sequence are paired or unpaired in the secondary structure of an RNA, which can be studied via different machine learning techniques. The state inference is a binary classification task on each nucleotide which is different from determining the full secondary structure consisting of sets of nested base pairs. Successful state inference of RNA sequences can be used to generate auxiliary information for data-directed RNA secondary structure prediction. In this talk I will highlight the challenges from using methods such as hidden Markov models (HMM) and variational autoencoders (VAE) for state inference. I will also discuss how the imbalance of RNA families in training data sets can create prediction issues such as overfitting. Finally, I will discuss different approaches for improving RNA secondary structure prediction via the Nearest Neighbor Thermodynamic Model (NNTM) using machine learning techniques for state inference. (Received September 11, 2021)