Deep neural network models of genomic sequence

John Reid

June 7, 2017

The state-of-the-art in many genomic prediction tasks is advancing quickly using deep neural networks applied to genomic sequences even though their history in this field is relatively short [1, 2]. However, there are several different neural network architectures in use today and it is not clear what each architecture's advantages and drawbacks are. Convolutional layers are prevalent. Low level convolutional layers model fixed width patterns in the genome such as transcription factor binding sites. Combinatorial interactions between transcription factor binding sites can be represented by stacked convolutional layers or alternatively as recurrent layers.

To date some limited evaluations have been performed that compare the architectures in use. The aim of this project is to conduct a systematic evaluation of the architectures in use today on a set of benchmark tasks. Furthermore when this benchmark is implemented, it will enable us to evaluate novel architectures which may progress the state-of-the-art in this fast moving field.

References
