Deep MR: Using Mendelian Randomization to Interrogate Neural Networks[2]

Stephen Malina

March 3, 2020

[2] Title is a work in progress.

Outline

Deep learning is amazing!

Meta: I use 'deep learning' as a synonym for 'differentiable model with lots of modular components'. If you want to silently substitute that every time I say 'deep learning', feel free!

- Who would've thought one family of models could learn so many different complex functions **in practice**?
- Convolutional neural nets have gotten so easy to fit!
- You're telling me that components designed to recognize features in images just **happen** to be good at identifying features in DNA sequences?

Deep learning is horrible!

1. How am I going to afford all these GPUs?
2. It works in practice, but does it work in theory?
3. Default settings in the streets, hyperparameter tuner in the sheets
4. **Why does my model make the predictions it does?**
5. **Does my model actually learn meaningful features?**

This talk: **??** and **??**.

Deep learning is great but has limitations

Deep learning is very good at learning nonlinear functions to predict interesting things directly from (DNA/RNA/text/etc.) sequence.

But, in application areas like biology we want to:

1. Understand **why** our models make the predictions they do
2. Test whether our models learn causal relationships that we know about from experiments

This requires new methods like the one we'll talk about today!

Sub-perspectives on Expressive, 'Black-Box' Models

Sub-perspective 1

Complex models are great, but need better ways to give biologists confidence that they've learned the mechanisms they care about

Sub-perspective 2

How can we go from answering pre-defined questions to generating new hypotheses?

Sub-perspective 1: Increasing Confidence in our[9] Models

[9] ML folks'

Sub-perspective 2: From Hypothesis Verification to Generation

Existing interpretability methods limit themselves to:

1. Answering human-defined queries about individual processes
   - Does mutating this nucleotide increase the probability of binding?
   - Does binding probability for this sequence respond to changes in any nucleotide within it?
2.

Mendelian Randomization

Deep learning for DNA/RNA sequence-specific prediction

Questions we hope our method will answer

- Do jointly trained models learn[14] causal relationships?
- According to this model, does binding of this transcription factor causally influence chromatin accessibility (and vice versa)?

[14] Where 'learn' is defined in terms of whether they can generate data that reflects them.

Terminology & Notation

Step-by-step Overview

1. Jointly train DL model to predict candidate cause(s) ('exposure') & effect(s) ('outcome')
2. Randomly sample sequences from held-out set and do in-silico saturation mutagenesis on each
   - Generate predictions
3. Treat difference between probability of (e.g.) binding and between each mutation and the reference sequence as effect size, analogous to a GWAS summary statistic
4. Use MR to estimate average causal effects for each sequence
5. Use meta-analysis method to aggregate causal effects across sequences

Binding & Chromatin Accessibility Model

Using DeepSEA [**?**]

Our deep net is deeply mis-calibrated

Near-term Work

- Figure out whether we can calibrate DeepSEA without re-training it
-

If you only remember three things...

1.

Bibliography