Deep Mendelian Randomization: Using Mendelian Randomization to Detect Learned Causal Relationships in Deep Learning Models

Anonymous Authors[1]

Abstract

1. Introduction

1.1. Background & Related Work

1.2. Mendelian Randomization

Connect MR to IVs. Briefly describe IVs and assumptions. Link to more comprehensive references. Introduce idea of using multiple instruments to weaken assumptions. State assumptions Egger allows to relax and what it adds.

1.3. Monte Carlo (MC) Dropout

Need this to estimate uncertainty to feed to Mendelian randomization. Allows us to get predictive means and standard errors for our predictions.

2. Methods

2.1. Algorithm Overview

Introduce notation that you'll use – *exposure*, *outcome*, *sequence*.

Describe input / output. Input: test set of one-hot encoded sequences for each exposure. Output: causal effect estimates for individual sequences and each exposure-outcome pair.

Recall that our goal is to determine whether data our model generates reflects known causal relationships. To accomplish this (see for visual depiction), we do the following for each transcription factor:

- Randomly sample sequences for which experiments showing transcription factor binding ("reference sequences").
- Perform *saturation in-silico mutagenesis* for each reference sequence to generate sequence length $\times$ number of nucleotides $- 1$ mutated sequences per original sequence.
- For each reference and set of mutated sequences, use MC-dropout to generate predictive means and standard errors of binding probabilities for the (reference—mutated) sequences.
- Generate sequence length $\times$ number of nucleotides $- 1$ *effect sizes* by taking the differences between each mutated sequence's predictive mean and the corresponding reference sequence's predictive mean and the standard error of these differences.
- Apply Mendelian randomization to each reference sequence's effect sizes and their standard errors to estimate a per-transcription factor, per-sequence region causal effect.
- Estimate overall per-transcription factor causal effects.

This leaves us with estimates of local (transcription factor and sequence level) and global (transcription factor level) causal effects.

2.2. Saturation In-Silico Mutagenesis

3. Experiments & Results

3.1. Dataset & Model

Model. We use a pre-trained version of DeepSEA provided by the Kipoi library to generate binding and chromatin accessibility predictions.

Dataset. To generate predictions from DeepSEA, we use sequence regions from DeepSEA's held-out test set that had binding in ChIP-seq experiments for the relevant transcription factors. This data is provided as part of the ENCODE project.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review at the ICML 2020 Workshop on Computational Biology (WCB). Do not distribute.

3.2. DeepSEA transcription factor & chromatin accessibility experimetn

Ran Deep MR on DeepSEA's HepG2 TFs paired with chromatin accessibility. Only use sequences where binding occurred to make linear effect more plausible.

Figure X depicts the distribution of causal effects of each of the N TFs on chromatin accessibility.

As you can see, Deep MR predicts positive causal effects for the N TFs on which we ran it, occasionally finding quite large causal effects.

Deep dive into causal effects for FOXA1.

4. Discussion

4.1. MR finds uniformly positive causal effects

4.2. MR finds significant causal effects with minimal pleiotropy

4.3. Limitations

Mendelian randomization assumptions.

- MR assumptions brief sentence

Software and Data

References

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.