Causality and Autoencoders in the Light of Drug Repurposing for COVID-19

Caroline Uhler (MIT)

IST Lisbon
Drug development against COVID-19

- Given the urgency, most viable approach is drug repurposing
- Many drugs are inhibitors (bind to a protein so that it cannot perform its downstream role)
- Want to identify a drug that pushes the system back to normal state
- Available data: Drug signatures (screens with ~1000 FDA approved drugs) and their targets, disease signatures, Protein-protein interaction networks (20,000 nodes, 200,000 edges)
- How to determine drug candidates for repurposing against a particular disease?

→ Learn causal graph!
Overview - causal transport problems

Predicting the effect of an intervention:

1. Genomic interventions such as knockout experiments (few mostly known targets)

   ⇒ Theoretical and algorithmic framework for learning causal networks from observational and interventional data

2. Transport drug intervention (many unknown targets) to new cell type

   ⇒ Use inductive bias of autoencoders for synthetic interventions
Framework: Structural equation models

- Introduced by Sewell Wright in the 1920s
- Major contributions by Judea Pearl, Jamie Robins, Don Rubin, Peter Spirtes since 1970s
- Represent causal relationships by a directed acyclic graph (DAG)
- Each node is associated with a random variable; stochasticity is introduced by independent noise variables $\epsilon_i$

\[
\begin{align*}
X_1 &\leftarrow f_1(X_3, \epsilon_1) \\
X_2 &\leftarrow f_2(X_1, \epsilon_2) \\
X_3 &\leftarrow f_3(\epsilon_3) \\
X_4 &\leftarrow f_4(X_2, X_3, \epsilon_4)
\end{align*}
\]
Framework: Structural equation models

- Introduced by Sewell Wright in the 1920s
- Major contributions by Judea Pearl, Jamie Robins, Don Rubin, Peter Spirtes since 1970s
- Represent causal relationships by a directed acyclic graph (DAG)
- Each node is associated with a random variable; stochasticity is introduced by independent noise variables $\epsilon_i$

$$
\begin{align*}
X_1 &\leftarrow f_1(X_3, \epsilon_1) \\
X_2 &\leftarrow f_2(X_1, \epsilon_2) \\
X_3 &\leftarrow f_3(\epsilon_3) \\
X_4 &\leftarrow f_4(X_2, X_3, \epsilon_4)
\end{align*}
$$

- Structural equation model also defines *interventional distribution*:
  - **Intervention** on $X_2$: $\text{do}(X_2 = c)$
  - $p(X_3 \mid \text{do}(X_4 = c)) = p(x_3) \neq p(x_3 \mid x_4)$, but $p(X_4 \mid \text{do}(X_3 = c)) = p(x_4 \mid x_3) \neq p(x_4)$
From causal graphs to independence relations

- Missing edge \((i, j)\) encodes \textit{conditional independence (CI) relation}:

\[
X_i \perp X_j \mid X_{\text{ancestors}(i,j) \setminus \{i,j\}}
\]

- \textbf{Markov equivalence}: different causal graphs can encode same CI relations and are generally indistinguishable from observational data

\(\xRightarrow{\text{Skeleton and immoralities (i\rightarrow j\leftarrow k) are identifiable}}\)

Verma & Pearl, 1992

\(\xRightarrow{\text{Interventional Markov equivalence classes have been characterized}}\)

Hauser & Buehlmann, JMLR 2012
Yang, Katcoff & Uhler, ICML 2018
Permutation-based search

GES: Greedy search over Markov equivalence classes: [Chickering, 2012]

- Large search space
- No consistency guarantees in the presence of interventional data
Permutation-based search

GES: Greedy search over Markov equivalence classes: [Chickering, 2012]

- Large search space
- No consistency guarantees in the presence of interventional data

**Idea:** DAG defined by ordering of vertices (permutation) and skeleton

- For \( p = 10 \) search space is of size \( 10! = 3,628,800 \) versus \( 10^{18} \)
- For each permutation \( \pi \) construct a DAG \( G_{\pi} = (V, E_{\pi}) \) by

\[
(i, j) \in E_{\pi} \iff X_i \not\perp\!\!\!\perp X_j \mid X_{\text{ancestors}_{\pi}(i,j) \setminus \{i,j\}}
\]

**Theorem (Uhler \& Raskutti, Stat 2018)**

*Under weak conditions any sparsest DAG \( G_{\pi} \) is Markov equivalent to the true DAG (as sample size \( n \to \infty \)).*
Greedy sparsest permutation (GSP) algorithm

edges in polytope of permutations (i.e., permutohedron) connect neighboring transpositions, e.g. 
\((3, 1, 4, 2) \rightarrow (3, 4, 1, 2)\)

Theorem (Solus, Wang & U., 2018)

Greedy sparsest permutation (GSP) algorithm is consistent (as sample size \(n \rightarrow \infty\)), i.e., every local minimum is a global minimum.
 Greedy sparsest permutation (GSP) algorithm

- Our Python package [https://github.com/uhlerlab/causaldag](https://github.com/uhlerlab/causaldag) has code for all methods, pre-processed perturb-seq data, etc.

![Graph showing expected neighborhood size vs. proportion of simulations for various methods.

![Graph showing computation time (secs) vs. number of variables.

[Solus, Wang & Uhler, 2018]

Learning from interventions and with latent variables

- **GIES**: perfect intervention adaptation of greedy search on space of Markov equivalence classes
  - In general not consistent  
    - [Hauser & Bühlmann, 2012]

- **IGSP**: interventional adaptation of GSP: provably consistent algorithm that can deal with interventional data
  - for hard interventions  
    - [Wang-Solus-Yang-Uhler, NIPS 2017]
  - for soft interventions  
    - [Yang-Katcoff-Uhler, ICML 2018]
  - for unknown intervention targets  
    - [Squires-Wang-Uhler, UAI 2020]

- **GSPo**: greedy search over posets to deal with latent confounders
  - sparsest poset is consistent  
    - [Bernstein-Saeed-Squires-Uhler, AISTATS 2020]
  - no consistency proof of greedy search yet
Predicting the effect of an intervention:

1. Genomic interventions such as knockout experiments (few mostly known targets)

   ⇒ Theoretical and algorithmic framework for learning causal networks from observational and interventional data

2. Transport drug intervention (many unknown targets) to new cell type

   ⇒ Use inductive bias of autoencoders for synthetic interventions
Style transfer and transporting causal effects

Latent space arithmetics for style transfer using autoencoders / GANs:

Lotfollahi, Wolf & Theis, Nature Methods 2020

Is this a general phenomenon? How does this fit in with work by Bareinboim, Pearl and co-authors on necessary and sufficient conditions for causal transportability?
Predicting the effect of a drug on a different cell type

- CMap: 1.2mio samples (1000-dim expression vectors), 1000s of perturbations (knockouts, overexpression, small molecules including ~800 FDA-approved drugs)
Overparameterized autoencoders align drug signatures

Correlation between drug signatures of A549 and MCF7 cells

Reconstruction accuracy

Under-parameterized autoencoder

Over-parameterized autoencoder

Ok reconstruction

Good reconstruction

Bad reconstruction

Inductive bias of over-parameterized autoencoders

- Given training examples $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^k$, $n < k$, autoencoders are typically trained using gradient descent initialized $\approx 0$ to solve
  \[
  \arg \min_{\psi \in \Psi} \sum_{i=1}^{n} \|\psi(x^{(i)}) - x^{(i)}\|_2^2
  \]

- Over-parameterized linear setting: solutions range from the identity map to the projection onto $\text{span}(x^{(1)}, \ldots, x^{(n)})$
Given training examples $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^k$, $n < k$, autoencoders are typically trained using gradient descent initialized $\approx 0$ to solve

$$\arg \min_{\psi \in \Psi} \sum_{i=1}^{n} \|\psi(x^{(i)}) - x^{(i)}\|^2_2$$

Over-parameterized linear setting: solutions range from the identity map to the projection onto $\text{span}(x^{(1)}, \ldots, x^{(n)})$

In the extreme case of $n=1$ the training example is memorized
Inductive bias of over-parameterized autoencoders

- Over-parameterized autoencoders have many ways to interpolate training data

- View autoencoder as discrete dynamical system: \( x_{t+1} = f(x_t) \)

- We proved that standard over-parameterized autoencoders (without additional regularizers) are self-regularizing: they learn maps, where training examples are attractive fixed points!

[Radhakrishnan, Belkin & Uhler, PNAS 2020]
Correlating disease and drug signatures

Correlate disease and drug signatures to obtain list of drug candidates

RNA-seq data from: Blanco-Melo et al., Cell, 2020
Validating drug targets using a causal analysis

- RIPK1 has most downstream differentially expressed genes based on inferred (from single-cell RNA-seq data) causal graph in A549 cells and also in AT2 cells.
- While role of other targets is similar, RIPK1 becomes peripheral in causal graph without taking ageing into account.
Role of RIPK1 linking SARS-CoV-2 replication & ageing?

Activation of NF-κB, immune response, & survival pathways

Apoptosis, necroptosis; fibrosis & blood clotting

Conclusions

- **Transporting between interventions:** Developed a theoretical and algorithmic framework for integrating observational and interventional data for causal inference.

- **Transporting intervention effects between populations:** Over-parameterized autoencoders show implicit bias that may be of great interest for causal transportability.

- **A principled causal framework is critical for drug discovery**

  Belyaeva, Cammarata, Radhakrishnan, Squires, Yang, Shivashankar & Uhler, Nature Communications (in press)
Acknowledgments

PhD students:
• Raj Agrawal
• Anastasiya Belyaeva
• Louis Cammarata
• Saachi Jain
• Adityanarayanan Radhakrishnan
• Chandler Squires
• Karren Yang
• Jiaqi Zhang

MSc/undergraduate students
• Josh Amaniampong
• Sathwik Karnik
• Eshaan Nichani
• Neha Prasad
• George Stefanakis
• Annie Yun

Postdocs:
• Daniel Bernstein
• Jan-Christian Huetter
• Neriman Tokcan

Collaborators:
• Mikhail Belkin
• G.V. Shivashankar

Funding: