Flexible Learning Reading Group @TU Berlin
1st Session: 24th of July 2019

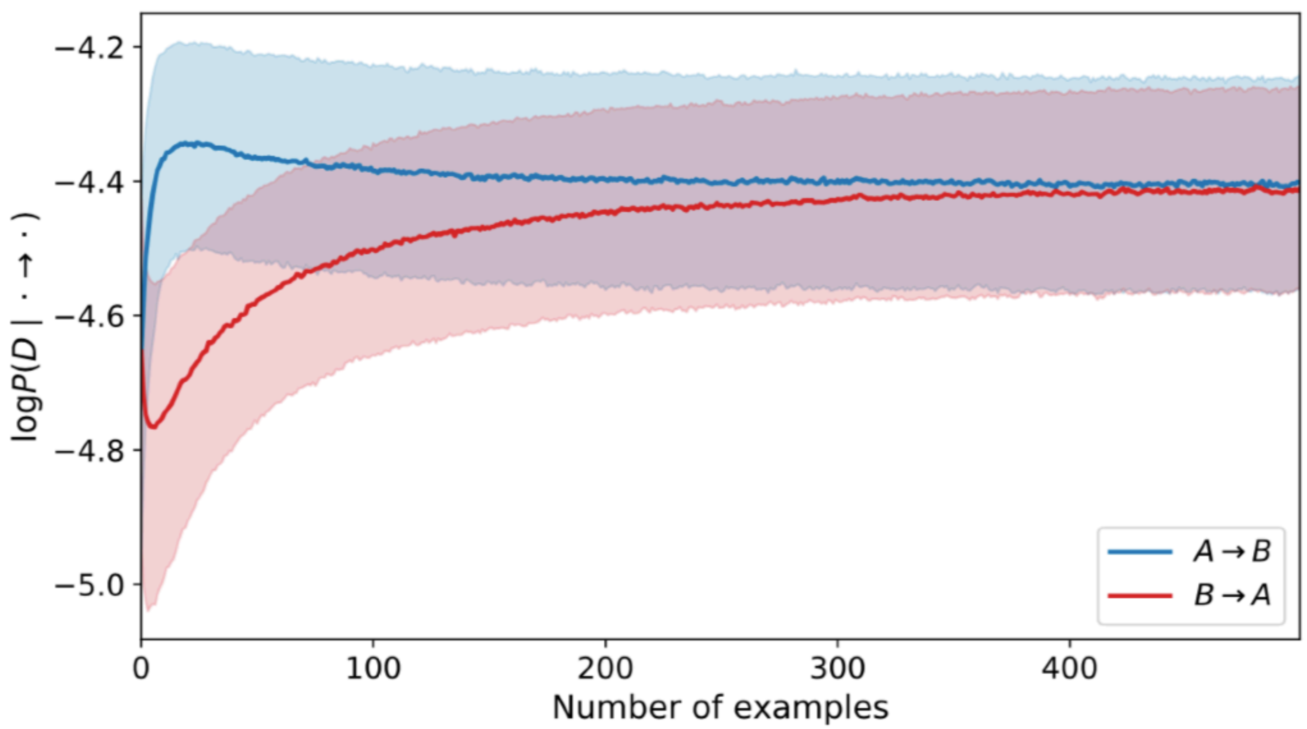
A Meta-Transfer Objective for Learning to Disentangle Causal Mechanisms

Yoshua Bengio[1,2,5], **Tristan Deleu**[1], **Nasim Rahaman**[4], **Nan Rosemary Ke**[3], **Sébastien Lachapelle**[1], **Olexa Bilaniuk**[1], **Anirudh Goyal** [1] **and Christopher Pal**[3,5]
Mila, Montréal, Québec, Canada
[1] **Université de Montréal**
[2] **CIFAR Senior Fellow**
[3] **École Polytechnique Montréal**
[4] **Ruprecht-Karls-Universität Heidelberg**
[5] **Canada CIFAR AI Chair**

Abstract

We propose to meta-learn causal structures based on how fast a learner adapts to new distributions arising from sparse distributional changes, e.g. due to interventions, actions of agents and other sources of non-stationarities. We show that under this assumption, the correct causal structural choices lead to faster adaptation to modified distributions because the changes are concentrated in one or just a few mechanisms when the learned knowledge is modularized appropriately. This leads to sparse expected gradients and a lower effective number of degrees of freedom needing to be relearned while adapting to the change. It motivates using the speed of adaptation to a modified distribution as a meta-learning objective. We demonstrate how this can be used to determine the cause-effect relationship between two observed variables. The distributional changes do not need to correspond to standard interventions (clamping a variable), and the learner has no direct knowledge of these interventions. We show that causal structures can be parameterized via continuous variables and learned end-to-end. We then explore how these ideas could be used to also learn an encoder that would map low-level observed variables to unobserved causal variables leading to faster adaptation out-of-distribution, learning a representation space where one can satisfy the assumptions of independent mechanisms and of small and sparse changes in these mechanisms due to actions and non-stationarities.

Speed of Adaptation = Quality of Estimated Causal Structure



- Setting: Causal graph - Stylized

$$P_{A\to B}(A,B) = P_{A\to B}(A)P_{A\to B}(B \mid A)$$
$$P_{B\to A}(A,B) = P_{B\to A}(B)P_{B\to A}(A \mid B)$$

- Causal identification is more than simply good transfer - meta!

$$\mathbb{E}[\nabla_{\theta_{A\to B}} R] = 0$$

a $\theta_{A\to B}$: Correctly learned at training

b $\theta_{A\to B}$: Correct set of causal parents

c No change in conditional distr.

- Key idea: Formulate this as a meta-objective function
- Assumption: Intervention is sparse if knowledge represented correct

An E-to-E Optimizable Meta-Objective for Causality

- Model Likelihoods - Different Causal Hypothesis:

$$\mathcal{L}_{A\to B} = \prod_{t=1}^{T} P_{A\to B}(a_t, b_t\,;\,\theta_t) \qquad \mathcal{L}_{B\to A} = \prod_{t=1}^{T} P_{B\to A}(a_t, b_t\,;\,\theta_t)$$

- Meta Regret Objective:

$$\mathcal{R} = -\log\left[\text{sigmoid}(\gamma)\mathcal{L}_{A\to B} + (1 - \text{sigmoid}(\gamma))\mathcal{L}_{B\to A}\right]$$

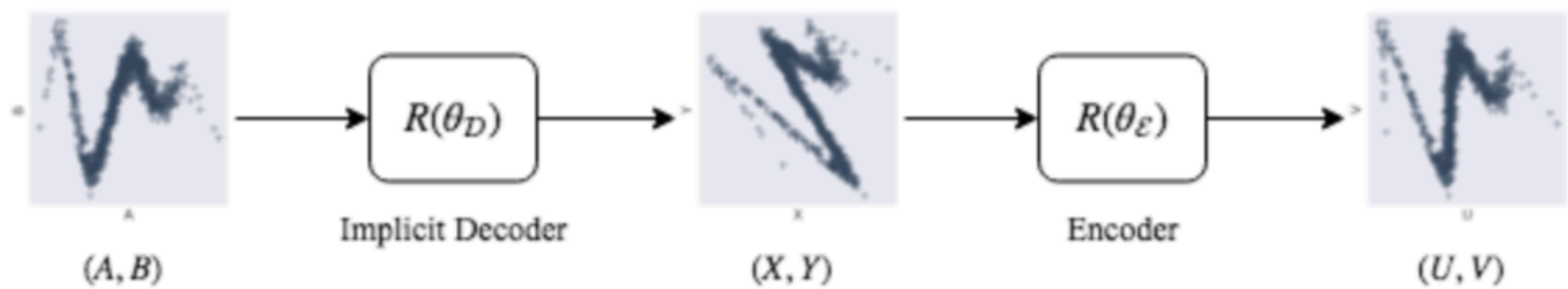
- Key Quantity:

$$\begin{aligned}\frac{\partial R}{\partial \gamma} &= \sigma(\gamma) - P(A \to B|D_2) = \sigma(\gamma) - \sigma(\gamma + \Delta) \\ &= \sigma(\gamma) - \sigma(\gamma + \log \mathcal{L}_{A\to B}(D_1, D_2) - \log \mathcal{L}_{B\to A}(D_1, D_2))\end{aligned}$$

- Inner + outer loop optimization alternation:
  1. Optimize model parameters for different hypothesis
  2. Optimize meta objective

Representation Disentangling

- Assumption: Causal graph is sparse - independent components
  + affected by sparse distributional shifts
- Problem: Not realistic for real-life high-dimensional data
- Bengio solution: Learn enabling representations E-to-E



- Simplified „rotation as intervention“ example