Domain Adaptation by Using Causal Inference to Predict Invariant Conditional Distributions

Sara Magliacane1, Thijs van Ommeren2, Tom Claassen1, Stephan Bongers3, Philip Versteeg3 and Joris M. Mooij1

1 IBM Research, New York 2 Institute for Informatics, University of Amsterdam, The Netherlands 3 Radboud University Nijmegen, The Netherlands

Abstract
In both domain adaptation and causal inference, an important goal is to make accurate predictions in an unseen target domain, where the distribution is different from the source domain(s). We consider causal domain adaptation problems, where the domains correspond to different interventions of a single system. The approach we propose exploits causal inference and does not rely on prior knowledge of the causal graph, or of intervention types/targets.

Problem setting: Causal domain adaptation
Unsupervised multi-source domain adaptation with an underlying causal graph, potentially with latent confounders (ADMG). We are modelling a generic setting in which the experimenter decides having access to the measurements.

Overview of our approach
Use conditional independences that can be tested on the available data, to infer enough about the unknown causal graph (G) to find separating sets (C). Predictions using such feature sets will transfer across domains, while other predictions may suffer arbitrarily large loss when transferred.

Joint Causal Inference (JCI)
JCI [Mooij, Magliacane and Claassen, 2018] is a meta-algorithm for systematically pooling data from multiple domains, even when intervention types and targets are possibly unknown, reducing causal discovery from different distributions to causal discovery of a single joint causal graph with auxiliary context variables. We distinguish:

• System variables X, representing the system in each distribution
• Context variables C, describing the changes between distributions

JCI [Mooij, Magliacane and Claassen, 2018] is a meta-algorithm for systematically pooling data from multiple domains, even when intervention types and targets are possibly unknown, reducing causal discovery from different distributions to causal discovery of a single joint causal graph with auxiliary context variables. We distinguish:

• System variables X, representing the system in each distribution
• Context variables C, describing the changes between distributions

JCI assumptions:
1. No system variable directly causes any context variable, and
2. No system variable is confounded with a context variable, and
3. Each pair of context variables is purely confounded (i.e. $C_i \perp \perp C_j \mid \mathcal{C} \neq \emptyset, C_i \neq C_j$).

Intuition: We are modeling a generic setting in which the experimenter decides on the performed interventions before the measurements are performed (or without having access to the measurements).

Causal domain adaptation algorithm
A brute-force strategy to select the feature set $A$ with the best asymptotic guarantee on the prediction error:

1. For all $A \subseteq \{X \cup Y \cup C\}, \{Y, C\}$ do
   $L_A \leftarrow$ estimate of generalization error when using $A$ to predict $Y$ in source domain
2. For all $A \subseteq \{X \cup Y \cup C\}, \{Y, C\}$, in increasing order of $L_A$ do
   if we can infer that $Y \perp \perp C$ then
      return $Y$ in the target domain using feature set $A$
   end if
end for
return abstain from making a prediction

Experimental results
We evaluated our method on simulated and real-world data.

• Simulated data: from randomly generated causal graphs
• Real-world data: hematoma data from CRM Causal Inference Challenge (phenotype data for wild-type and single-gene knockout mice)
• In both cases, 2 context and 3 system variables
• Baseline method: Feature selection + regression (random forests), does not try to detect separating sets

Dealing with missing data
Due to the missing data, some inferences cannot be tested. Some of those can be inferred based on our transfer assumptions. For an independence $A \perp B | S$, if $Y \not\in A \cup B \cup S$, independence is testable in data
If $Y \in A \cup B \cup S$ and $C \in S$, follows from the transfer assumptions, $A \perp B | S \Rightarrow A \perp B | \{S \setminus C\} | C = 0$

• $Y \in A \cup B \cup S$ and $C \in S$, is untestable and does not follow from the assumptions, e.g. $Y \perp \perp C | S$

Logic-based causal discovery method
Task: “Is $Y \perp \perp C_i \mid A$?” given all available conditional independences
We modify the method by [Hyttinen, Eberhardt and Järvisalo, 2014] combined with the scores from [Magliacane, Claassen, Mooij, 2016]

• Input: list of weighted conditional (in)dependence statements (in our case: some are missing)
• Output: a measure of confidence that the statement is true (or false)

References
2. Magliacane, T. Claassen and J. M. Mooij, Artificial causal inferences. MIP*2018