

Research project

From: Céline Béji

Establishment: Project-team Miles, data science department, LAMSADE, Paris Dauphine-PSL University, CNRS.

Supervisor: Ph. Jamal Atif, Florian Yger

Subject: Revisiting the problem of causal inference in light of the recent advances in machine learning.

Date: March 2017 – now

Causality is central in many application domains, such as health care, economics and marketing [1]. For example, in personalized medicine, being able to target the responders of a treatment whose recovery might be “caused” by a given medical drug could have significant public health consequences.

In the continuity of Neyman-Rubin’s framework [2], which considers causality as a missing data problem, we reformulate causal inference as a density estimation. Consider the example where we would like to evaluate the effectiveness of a medical treatment on a group of sick person. Let $X \in \mathfrak{M}_{n,p}(\mathbb{R})$ be a features matrix representing n patients and her p characteristics, like her age, weight, medical history and monitoring indicators. The outcome, displaying the patient’s reaction to treatment, is noted as $Y_i(1)$ if the treatment is assigned to the individual i and $Y_i(0)$ otherwise. Knowing the outcome in both treatment, fourth causal populations are identified. The most intuitive is when the treatment is beneficial. The patient recovers if he receives the treatment and remains sick otherwise ($Y_i(1) = 1, Y_i(0) = 0$). However, a patient can recover by himself, with or without treatment ($Y_i(1) = 1, Y_i(0) = 1$). There is no impact. In the same way, he can remain ill with or without treatment ($Y_i(1) = 0, Y_i(0) = 0$). The latter case, because of side effects, occurs when a patient can recover without treatment and remains ill with treatment ($Y_i(1) = 0, Y_i(0) = 1$). The aim of my thesis is to identify the causal populations to identify on each patient whether the outcome was caused by the treatment and to be able to establish a treatment assignment policy for new patients. Unlike current tests that assess the effectiveness of a medical treatment on average, patients who respond positively to treatment could be targeted, even if they represent only a minority of the population. However, the both outcomes with or without treatment, are not simultaneously observable. Their estimation remains the fundamental problem in counterfactual prediction.

Our initial contribution was to model causal inference as a problem of estimating the probability density of a mixture of separate populations; each defined by outcomes with and without treatment. We extend the interpretation of the causal populations in multi-treatment and create causal constraints introduced in the model as partial information. For example, if a person is treated and has successfully responded to treatment $Y_i(1) = 1$, either the effect of the treatment

is positive for her or the treatment has no effect (the patient would have recovered whatever). Since the other two populations are excluded from the convenient choices and their probability can be constrained to zero. We then demonstrated the individual effect of treatment, and its average can be estimated by the distribution of these populations.

A second contribution was to build an adaptation of the EM algorithm [3] to estimate the parameters of populations distribution under the causal constraints. We provide a lower bound for the log-likelihood and prove the convergence of this algorithm to a global maximum. The convergence is the consequence introduced constraints, that produce unique identifications of the estimated distributions to each causal population. We conduct experiments on a mixture of Gaussian distributions which demonstrate its efficiency. In addition, we show on datasets including categorical variables that a mixture of independent Gaussian and Multinomials improves the results. We propose a derived variational algorithm does not maximize the likelihood, but its lower bound when the marginal probability is not directly estimated.

Last contribution provides a non-parametric estimation approach. We use an auto-encoder [4] enhanced by a causal prior, materialized by a mask in the intermediate layer of our network. The features are reconstructed after being reduced to the hidden latent space characterizing the causal of each population which captures the causal distribution. We discuss the number of optimal units on the latent variables manifold and demonstrate the efficiency of our model compared to the baselines on both synthetic and real-life datasets.

Our approach has the advantage of being extendable to multi-treatments. It allows to obtain a confidence interval on the prediction of the outcome for each treatment and estimate the combination of treatments that maximizes the global effect. In future works, we plan to investigate the correlation between treatments on the causal effect, which remains unexplored to date.

References

- [1] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.
- [2] D.B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.
- [3] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.
- [4] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.