Heterogeneous Treatment Effects : When Machine Learning meets multiple treatments regime

Naoufal Acharki*[1], Josselin Garnier[1,2], and Antoine Bertoncello

[1]Centre de Mathématiques Appliquées - Ecole Polytechnique – Ecole Polytechnique, Centre National de la Recherche Scientifique : UMR7641 – France

[2]École Polytechnique – CMAP, École Polytechnique – France

Résumé

With the rapid development of Machine Learning and its efficiency in predicting outcomes, the question of counterfactual prediction arises. Engineers may want to know how the outcome would be affected when a feature is changed to aspecific value, not only on average but also within a smaller scale to personalize treatments at these levels. A set of statistical tools has been developed by epidemiologists and statisticians based on the Potential Outcomes theory (Rubin, 1974). They aim to make causal inference and estimate the effects of a treatment on the outcome, whether on average among the whole population or inside different sub-groups using meta-learning algorithms (Kunzel et al.,2019). These methods have successful applications in many fields (medicine, economics etc.) but seldom used in the industry. Furthermore, most existing methods and studies are limited to the setting of a binary treatment. Instead, we consider a multiple discrete treatment regime where the treatment belongs to a finite set and we aim to estimate heterogeneous treatment effects. We develop the framework of meta-learning algorithms to estimate the so called Conditional Average Treatment Effects (CATE). We evaluate the performances of each meta-learner, combined with a machine learning model, on randomized and observational studies, then on a real-world large-scale experiment.

*Intervenant

sciencesconf.org:jps-2021:368979