Learning Representations Using Causal Invariance

Léon Bottou*

* Facebook AI Research
http://leon.bottou.org

Résumé. Les algorithmes d'apprentissage capturent souvent de fausses corrélations présentes dans la distribution des données d'entraînement au lieu de traiter la tâche qui nous intéresse. De telles corrélations parasites se produisent parce que le processus de collecte de données est soumis à des biais de confusion incontrôlés. Supposons cependant que nous ayons accès à plusieurs ensembles de données illustrant le m concept mais dont les distributions présentent des biais différents. Pouvons-nous apprendre quelque chose de commun à toutes ces distributions, tout en ignorant les fausses façons dont elles diffèrent? Ceci peut être réalisé en projetant les données dans un espace de représentation qui satisfait un critère d'invariance causale. Cette idée diffère de façon importante des travaux antérieurs sur la robustesse statistique ou les objectifs contradictoires. Semblable à des travaux récents sur la sélection des caractéristiques invariantes, il s'agit de découvrir le mécanisme réel sous-jacent aux données au lieu de modéliser ses statistiques superficielles.

Summary

Learning algorithms often capture spurious correlations present in the training data distribution instead of addressing the task of interest. Such spurious correlations occur because the data collection process is subject to uncontrolled confounding biases. Suppose however that we have access to multiple datasets exemplifying the same concept but whose distributions exhibit different biases. Can we learn something that is common across all these distributions, while ignoring the spurious ways in which they differ? This can be achieved by projecting the data into a representation space that satisfy a causal invariance criterion. This idea differs in important ways from previous work on statistical robustness or adversarial objectives. Similar to recent work on invariant feature selection, this is about discovering the actual mechanism underlying the data instead of modeling its superficial statistics.