

Titre : Inférence statistique en présence de non-réponse et de méthode d'apprentissage automatique : Quelques travaux récents

Auteur : David Haziza (Université d'Ottawa)

Résumé : Ces dernières années, les méthodes d'apprentissage automatique ont suscité un intérêt croissant au sein des instituts nationaux de statistique. Elles permettent de produire des prédictions précises en analysant de vastes ensembles de données et en identifiant des modèles et relations complexes. Ces prédictions peuvent être exploitées à différentes étapes d'une enquête, notamment pour le traitement des données manquantes et l'estimation pour des petits domaines.

Dans cette présentation, nous exposerons les résultats de travaux récents ou en cours sur l'inférence en présence de non-réponse partielle. Nous commencerons par discuter de l'inférence statistique lorsque les forêts aléatoires sont utilisées pour imputer les valeurs manquantes. Ensuite, nous aborderons les méthodes d'estimation doublement robustes qui intègrent des probabilités de réponse prédites et des valeurs imputées, obtenues au moyen de n'importe quelles méthodes d'apprentissage automatique, dans la construction des estimateurs. Nous soulignerons les avantages des méthodes doublement robustes dans le cadre des méthodes d'apprentissage automatique et discuterons de leur implémentation en pratique. Nous montrerons comment la variance des estimateurs doublement robustes. Enfin, nous présenterons les résultats d'études de simulation visant à évaluer la performance des estimateurs ponctuels et de variance.