

Cycle de Conférences du Laboratoire MIPS
2014-2015

21 mai 2015 à 14h00

Petit-Amphithéâtre – ENSISA-Lumière

CHORDALYSIS - APPRENTISSAGE AUTOMATIQUE DE
MODÈLES GRAPHIQUES

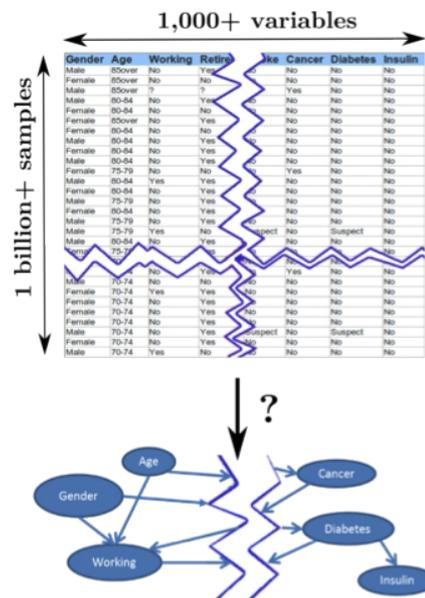
Dr François Petitjean

Centre for Data Science - Monash University – Melbourne - Australia

Les modèles graphiques sont à la base de nombreux outils d'analyse de données. Ils combinent les théories des graphes et des probabilités afin de représenter des distributions de probabilités de façon compacte, efficace et interprétable. Ils sont par exemple utilisés en sciences sociales ou en médecine afin de comprendre les corrélations mises en jeu entre différentes variables, et ce afin de répondre à des questions du type :

- « Le risque de crise cardiaque est-il indépendant du niveau de cholestérol étant donné la prise de médicaments pour la régulation du cholestérol ? » ou encore
- « La probabilité de voter à l'élection présidentielle est-elle indépendante de l'âge étant donné le genre et le milieu socio-culturel ? »

La dernière décennie a vu la quantité de données collectées augmenter de façon exponentielle, décuplant ainsi le potentiel de découverte des méthodes d'analyse de données. Cependant, les méthodes d'apprentissage automatique de modèles graphiques n'ont pas suivi cette évolution, réduisant ainsi directement l'intérêt de ces modèles pourtant cruciaux.



Dans ce séminaire, je présenterai les résultats de plusieurs années de recherche consacrés à ce passage à l'échelle. Notre solution, appelée Chordalysis, combine de nombreuses avancées en théorie des graphes (graphes chordaux), en statistiques et en fouille de données [1,2,3]. Le logiciel libre associé [4], permet de découvrir la structure de graphes associés à des jeux de données comportant des milliers de variables et des millions d'exemples, le tout sur un ordinateur standard.

Biographie : François Petitjean est chercheur en analyse de données au sein du « Centre for Data Science » de Monash University à Melbourne. Il a rejoint l'Australie en 2012, après avoir obtenu sa thèse financée par l'agence spatiale française (CNES), et pour laquelle il a reçu le prix de thèse de l'Université de Strasbourg ainsi que le prix de thèse EGC. Plus d'information à www.francois-petitjean.com

[1] F. Petitjean, G. Webb and A. Nicholson, "Scaling log-linear analysis to high-dimensional data," IEEE International Conference on Data Mining, 2013.

[2] F. Petitjean, L. Allison, G. Webb, "A statistically efficient and scalable method for log-linear analysis of high-dimensional data," IEEE International Conference on Data Mining, 2014.

[3] F. Petitjean and G. Webb, "Scaling log-linear analysis to datasets with thousands of variables," SIAM International Conference on Data Mining, 2015.

[4] <https://github.com/fpetitjean/Chordalysis/releases/tag/v2.0>