

Analyse exploratoire de données textuelles à l'aide de modèles thématiques probabilistes

Julien VELCIN, Laboratoire ERIC, Université Lyon 2

Mots-clés : Données textuelles, topic modeling

L'analyse exploratoire de vastes corpus textuels nécessite le recours à des techniques d'apprentissage non ou peu supervisés : techniques de projection en faible dimension, (co-)clustering de documents, modélisation thématique. Dans cet exposé, après un bref panorama des différentes techniques à la disposition du scientifique des données, je détaillerai davantage les modèles thématiques probabilistes en cherchant à illustrer l'étendu des utilisations possibles à l'aide de cas concrets. Ces modèles ont par exemple été utilisés pour améliorer la recherche d'information, pour analyser l'opinion ou pour détecter la nouveauté dans les flux de données. Je terminerai en donnant quelques travaux récents qui cherchent à combiner ces modèles avec des techniques de plongement de mots.

Références

Julien VELCIN, ERIC Lab,
5 av. Pierre Mendès-France
69676 Bron Cedex,
France