
Logiciel Weka et Notebook - L'apprentissage automatique pour la classification textuelle

Patrice Bellot^{*1}

¹Laboratoire d'informatique et systèmes – Aix-Marseille Université - AMU, Université de Toulon (France), Centre National de la Recherche Scientifique - CNRS – France

Résumé

Présentation de l'atelier

L'objectif de l'atelier est de présenter, sous forme de démonstrations et de notebooks partagés, deux environnements logiciels permettant la classification automatisée de données textuelles selon des approches d'apprentissage automatique incluant les réseaux neuronaux. Cela permettra aux auditeurs de comprendre la nature et les objectifs des traitements mis en oeuvre et d'estimer l'effort nécessaire pour expérimenter les approches les plus actuelles sur ses propres données.

Programme détaillé

- Manipulation des outils d'apprentissage automatique de classification supervisée (catégorisation) ou non supervisée (partitionnement en classes proches) à partir de textes.
- Classification supervisée de document (des textes annotés et des représentations vectorielles des mots et des documents) - Expérimentation avec l'environnement Weka.
- La mise en oeuvre d'une approche neuronale pour la classification automatique - Expérimentation à partir d'un Notebook en Python.
- De nombreuses tâches de la fouille de textes vues comme des problèmes de classification.

Vidéo de présentation (ANF TDM 2020)

<https://www.youtube.com/watch?v=bThk-Akl6RA>

^{*}Intervenant