

Prédiction de séries temporelles via des modèles hybrides

Simon Malinowski
IRISA/LinkMedia

Thomas Guyet
AgroCampus Ouest

Vincent Lemaire
Orange

`{simon.malinowski}`, `{thomas.guyet}@irisa.fr`

Dans le domaine de l'apprentissage automatique (machine learning), les méthodes basées "ensembles" (combinaisons de plusieurs algorithmes d'apprentissage pour améliorer les performances) sont de plus en plus utilisées depuis plusieurs années. Beaucoup de méthodes de ce type ont été proposées notamment pour des tâches de classification.

Ces dernières années, les recherches en apprentissage automatique ont fait apparaître (au moins) deux grandes familles de méthodes permettant de combiner des classifieurs :

- 1) L'utilisation d'un ensemble de modèles. Les techniques les plus connues sont le bagging et le boosting par exemple.
- 2) L'utilisation d'un algorithme hybride permettant d'associer nativement l'apprentissage simultané de deux (ou plus) algorithmes « locaux » au sein d'un algorithme unique « global ». Contrairement à la première approche, les classifieurs « hybrides » apprennent comment combiner au mieux les sorties des différents algorithmes. On citera par exemple LMT (Logistic Model Tree) [1] ou EML-Tree (Extreme Learning Machine Tree) [2].

Dans le cadre de ce stage, nous souhaitons explorer la deuxième voie énoncée ci-dessus et ainsi examiner l'apport des modèles hybrides pour une tâche de prédiction de séries temporelles. On étudiera notamment comment adapter les algorithmes existants pour cette tâche particulière.

Ce travail sera mené dans le cadre d'une application industrielle de prédiction d'utilisation de ressources de cloud computing et la minimisation de leur infrastructure (maîtrise de la consommation d'énergie).

[1] "Logistic Model Trees", Niels Landwehr and Mark Hall and Eibe Frank, Machine Learning, 2005

[2] "Learning ELM-Tree from big data based on uncertainty reduction", Ran Wang et al. FSS 2014