

Nom et Prénom : TAJMOUATI SAMYA

Date de soutenance : 25/06/2022

Directeur de Thèse : BOUAZZA EL WAHBI

Sujet de Thèse :

Prévision des séries temporelles par l'algorithme des k-plus proches voisins

Abstract:

Time series forecasting domain has been studied for many years. Traditionally, time series forecasting has been performed using statistical models, such as autoregressive integrated moving average and exponential smoothing models. Over time, machine learning models have emerged and competed the statistical models. Compared to statistical models, the main purpose of machine learning models is to increase predictive performance rather than to study, analyze, and interpret the relationship between explanatory and explained variables. In this thesis, we focus on k-nearest neighbors (k-NN). Motivated by the lack of applying k-NN in time series setting, this thesis contributes to developing new automatic methods for forecasting time series with k-NN. First, we develop two new approaches based on cross-validation, allowing the optimization of k-NN hyperparameters. Such optimization is necessary to improve the predictive performance of k-NN. Next, we extend the application of conformal prediction in time series setting, where we use k-NN to build reliable and efficient prediction intervals. Finally, we study the convergence of the two new approaches and propose a stopping condition to reduce their time complexity without loss of precision.

Keywords : Time series forecasting, Machine learning, Hyperparameters, Predictive performance, k-nearest neighbors, Cross validation, Conformal prediction, Prediction intervals, Convergence, Stopping condition, Time complexity.

Résumé:

La prévision des séries temporelles est un domaine qui est étudié depuis de nombreuses années. Traditionnellement, les modèles statistiques comme les modèles autorégressifs à moyenne intégrée ou lissage exponentiel sont utilisés dans la prévision des séries temporelles. Au fil du temps, des modèles d'apprentissage automatique sont apparus et devenus les concurrents potentiels des modèles statistiques. Par rapport aux modèles statistiques, le but principal des modèles d'apprentissage automatique est l'augmentation de la performance prévisionnelle plutôt que l'étude, l'analyse, et l'interprétation de la relation entre les variables explicatives et la variable expliquée. Dans cette thèse, on s'intéresse à un modèle d'apprentissage automatique peu appliqué dans le domaine des séries temporelles : les k-plus proches voisins (k-NN). Motivée par le manque d'application de l'algorithme k-NN dans le contexte des séries temporelles, la contribution de cette thèse s'inscrit dans le cadre du développement de nouvelles méthodes automatiques de prévision des séries temporelles par k-NN. Dans un premier temps, nous développons deux nouvelles approches basées sur la validation croisée, permettant d'optimiser les hyperparamètres de k-NN. Une telle optimisation est nécessaire pour renforcer sa performance prévisionnelle. Dans un second temps, nous proposons une nouvelle approche permettant de quantifier ses erreurs de prévision. L'approche proposée étend l'application de la prédiction conforme aux séries chronologiques et utilise l'algorithme k-NN pour construire des intervalles de prédiction fiables et efficaces. Finalement, nous étudions la convergence des nouvelles approches de réglage d'hyperparamètres et proposons une condition d'arrêt permettant de réduire leur complexité temporelle sans perte de précision.

Mots clés : Prévision des séries temporelles, Apprentissage automatique, Hyperparamètres, Performance prévisionnelle, k-plus proches voisins, Vali-