

STAGE M2 – DATA SCIENCE / IA

Fast Maximum Evaluation Via Neural Network For Efficient Time Series Analysis

Noms et coordonnées des encadrants :

- Vincent RUNGE, MCF. vincent.runge@univ-evry.fr
- Charles TRUONG, PostDoc. charles.truong@ens-paris-saclay.fr

Laboratoires ou équipes :

- LaMME, Université d'Évry
- Centre Borelli, ENS Paris-Saclay

Le stagiaire travaillera entre ces deux laboratoires ou dans l'un des deux selon préférence.

Début du stage : Entre fin janvier et début mars 2025

Durée : 5 mois

Rémunération : 750 euros par mois

Candidature : envoyer **CV** et **notes de M1-M2** à vincent.runge@univ-evry.fr et charles.truong@ens-paris-saclay.fr

Si présélectionné(e), un entretien par zoom aura lieu courant janvier.

Condition :

- être actuellement dans un M2 Data Science / IA / maths appliquées (ou parcours ingénieur)
- La connaissance des réseaux de neurones et de leur implémentation en python est obligatoire.

Description (2 pages) :

Notre projet de stage s'inscrit dans la continuation d'un projet de recherche en cours entre le LaMME d'Évry et le Centre Borelli de l'ENS Paris-Saclay. Le stage intégrera également l'expertise de notre collègue Toby Hocking (Université de Sherbrooke, Canada), en visite académique entre nos laboratoires de janvier à juin 2025 et avec qui nous collaborons déjà activement.

—

Publications et preprint:

A) Duality-Based Pruning Methods For Exact Multiple Change-Point Detection. **Vincent Runge, Charles Truong, Simon Querné, 2024**

B) An Efficient Algorithm for Exact Segmentation of Large Compositional and Categorical Time Series. stats. **Charles Truong, Vincent Runge, 2024**. <http://dx.doi.org/10.1002/sta4.70012>

C) Change Point Detection in Hadamard Spaces by Alternating Minimization. **Anica Kostic, Vincent Runge, Charles Truong, aistat 2025**

—

Nous étudions un algorithme de **détection de ruptures multiples** par vraisemblance pénalisée. L'algorithme **DUST** (cf références A et B) permet une résolution extrêmement efficace du problème, battant les récents algorithmes de l'état de l'art (PELT, Killick 2012 et FPOP, Maidstone 2017). Le code et certains résultats sont disponibles ici :

<https://github.com/vrunge/dust>

Cette stratégie est **utilisable pour de nombreux modèles de données** : multivariés sur modèle exponentielle, modèle linéaire...

Cependant, un calcul interne ralentit nos algorithmes : la recherche du maximum d'une fonction duale pour chaque intégration d'une nouvelle donnée dans l'algorithme. Cette recherche est même effectuée à chaque itération pour chaque indice non élagué qui pourrait être retenu comme début du dernier segment.

La fonction duale est concave et paramétrique $D(x) = D(x; S_1, \dots, S_d)$ avec S_1 à S_d les paramètres : des sommes partielles sur les données.

Le but de ce stage est de **remplacer la recherche du maximum faite par l'algorithme de Quasi-Newton** (voir référence A) **par un apprentissage préalable de la position de ce maximum** en utilisant une base d'apprentissage que nous allons construire :

$$(S_1, \dots, S_d) \Leftrightarrow \operatorname{argmax} D(x)$$

La forme non paramétrique complexe de la position du maximum nous incite à apprendre cette relation à l'aide d'un réseau de neurones.

Cette méthode sera déployée pour les algorithmes déjà accessibles dans DUST (<https://github.com/vrunge/dust>) et **étendue à d'autres modèles plus complexes** (coût robuste, intégration de contraintes de continuité...)

Le package sera bientôt intégré au CRAN pour accroître sa visibilité dans la communauté des utilisateurs. L'étudiant sera encouragé à participer à la promotion des nouvelles méthodes efficaces auprès des spécialistes et utilisateurs (rencontres R, JDS, en juin 2025). Si le temps le permet, un article de recherche sera rédigé.