

PROPOSITION DE STAGE

APPRENTISSAGE DE MODELES GAN (Generative Adversarial Networks) POUR LA SIMULATION DE TRAJECTOIRES D'OBJETS DERIVANTS

1. CONTEXTE

Mercator Océan, centre européen de prévision océanique, décrit et prévoit l'évolution de l'océan en développant, et en maintenant en condition opérationnelle, plusieurs modèles numériques d'analyse et de prévision océanique.

Grâce aux champs de courants produits par ces modèles, il est possible de simuler les trajectoires de différents objets qui dérivent dans les océans, en surface ou en profondeur : icebergs, débris plastiques, nappes de pétrole, organismes marins dérivant passivement (plancton) ou plus activement (animaux marins se déplaçant avec les courants plus leur vitesse de nage).

On sait cependant que, même pour les objets les plus simples, à savoir les bouées dérivantes spécifiquement conçues pour suivre les courants, les trajectoires simulées et observées se comparent souvent assez mal (statistiquement) : les déplacements et leur variabilité sont souvent mal estimés.

2. OBJECTIF DU TRAVAIL

L'objectif du stage proposé est de développer et mettre en œuvre un outil de simulation de trajectoires basé sur des techniques de l'intelligence artificielle, et plus particulièrement d'apprentissage profond. Les variables explicatives utilisées seront choisies parmi l'ensemble des variables simulées par les modèles océaniques, à commencer par les courants. On exploitera, non seulement les valeurs de courants le long des trajectoires, mais également dans leur voisinage (spatial et temporel).

Dans un premier temps on s'attachera à développer un outil capable de simuler les trajectoires de bouées dérivantes simples. Le jeu de données d'apprentissage disponible pour ce faire est énorme : depuis 2003, entre 600 et 1600 bouées sont suivies quotidiennement sur l'ensemble des océans du globe ([1], et <https://www.aoml.noaa.gov/global-drifter-program/>). En fonction des progrès réalisés, on pourra dans un second temps, s'attaquer à la simulation de trajectoires plus complexes comme celles de tortues marines (aussi suivies par satellite). Les tortues, comme les flotteurs, dérivent avec le courant et nagent pour se maintenir dans des eaux riches en nourriture et suffisamment chaudes. On devrait donc pouvoir simuler leurs trajectoires en utilisant, comme variables explicatives, la vitesse du courant, la température de l'eau et la production primaire ou la concentration en micro-necton (un proxy de la densité de proies). L'objectif ultime de ce type de recherche est de prévoir les zones où les tortues marines sont les plus susceptibles de se concentrer afin de pouvoir prendre des mesures adéquates pour minimiser les risques de captures accidentelles par les flottes de pêches industrielles.

D'un point de vue méthodologique, on explorera des approches de type GAN (Generative Adversarial Network) pour la simulation conditionnelle de trajectoires [2]. Après une phase d'état de l'art, une part importante du travail envisagé consistera à explorer et évaluer

différents types de représentation et modélisation des données trajectométriques intégrant pleinement leur caractère stochastique (e.g., représentation latente, équations différentielles stochastiques, formulations variationnelles,...) [3,4,5]. Un autre aspect important portera sur la mise en œuvre d'un cadre expérimental d'évaluation quantitatif et qualitatif des modèles et algorithmes proposés pour les différents cas d'étude envisagés.

3. COMPÉTENCES REQUISES

- Formation initiale de type Master 2/Ingénieur en Mathématiques Appliquées, Sciences des données, Intelligence Artificielle et/ou Traitement du Signal
- Bases en apprentissage statistique et réseaux de neurones
- Intérêt pour les problèmes environnementaux.
- Grande rigueur dans l'organisation du travail, la gestion des données et des résultats. Les données à manipuler sont diverses et représentent des volumes importants.
- Bonne maîtrise de l'outil informatique : ce stage nécessite la connaissance de l'environnement LINUX, et du langage de programmation Python.
- Connaissances appréciées des environnements Python pour l'apprentissage profond (keras, tensorflow et/ou pytorch)

4. ENCADREMENT

Le stage sera co-encadré par Carlos Granero-Belinchon (IMT Atlantique), Simon van Gennip et Philippe Gaspar (Mercator Océan). Les étudiants intéressés sont invités à envoyer leur CV et une lettre de motivation à carlos.granero-belinchon@imt-atlantique.fr , pgaspar@mercator-ocean.fr et svangennip@mercator-ocean.fr

En fonction des conditions sanitaires du moment, le stage pourra se dérouler à Ramonville Saint-Agne (Toulouse) sur le site de Mercator-Océan, à l'IMT Atlantique ou en télétravail.

Reference

- [1] Elipot et al. (2016), "A global surface drifter dataset at hourly resolution", J. Geophys. Res. Oceans,121 doi:10.1002/2016JC011716
- [2] Goodfellow et al. (2014). "Generative Adversarial Networks". NIPS, <https://arxiv.org/abs/1406.2661>
- [3] Chen et al (2018). Neural Ordinary Differential Equations. NIPS.
- [4] Ouala et al (2020). Learning Latent Dynamics for Partially-Observed Chaotic Systems. Chaos.
- [5] Fablet et al. (2020). Learning Variational Data Assimilation Models and Solvers. arXiv, 2007.12941.