INTERNSHIP OFFER

LEARNING MODELS WITH GAN (Generative Adversarial Networks) FOR SIMULATING TRAJECTORIES OF OBJECTS DRIFTING WITH OCEAN CURRENTS

1. CONTEXT

Mercator Ocean, the European ocean forecasting centre, describes and forecasts the evolution of the ocean by developing, and maintaining in operational condition, several numerical models for ocean analysis and forecasting.

Thanks to the gridded current fields produced by these models, it is possible to simulate the trajectories of various objects drifting at the surface of the ocean or at depth: icebergs, plastic debris, oil slicks, marine organisms drifting purely passively (plankton) or more actively (marine animals moving with the currents plus their swimming speed).

However, we know that even in the simplest case of drifting buoys specifically designed to follow currents, the simulated and observed trajectories often compare quite poorly (statistically): movements and their variability are often poorly estimated.

2. OBJECTIVE

The objective of this internship is to develop and implement a trajectory simulation tool based on artificial intelligence techniques, and more particularly deep learning techniques. The explanatory variables used will be chosen among the many variables simulated by ocean models, starting with currents. We will exploit not only the values of currents along the trajectories, but also in their (spatial and temporal) vicinity.

Initially, we will develop a tool capable of simulating the trajectories of simple drifting buoys tracked by satellites. The learning data set available for this purpose is huge: since 2003, between 600 and 1600 buoys have been monitored daily in all oceanic basins ([1], https://www.aoml.noaa.gov/global-drifter-program/).

Depending on the progress made, it will then be possible to tackle the simulation of more complex trajectories such as those of marine turtles (also tracked by satellite). Turtles, like floats, drift with the current and swim to maintain themselves in productive and sufficiently warm waters. It should therefore be possible to simulate their trajectories using, as explanatory variables, current speed, water temperature and primary production or the concentration of micro-nekton (a proxy for prey density). The ultimate objective of this type of research is to predict the areas where marine turtles are most likely to concentrate so that appropriate measures can be taken to minimize the risks of accidental capture by industrial fishing fleets.

From a methodological point of view, GAN (Generative Adversarial Network) type approaches will be explored for the conditional simulation of trajectories [2]. After a state of the art phase, an important part of the work will consist in exploring and evaluating different types of representation and modelling of trajectometric data fully integrating their stochastic character (e.g., latent representation, stochastic differential equations, variational formulations,...) [3,4,5]. Another important aspect will concern the implementation of an

experimental framework for the quantitative and qualitative evaluation of the models and algorithms proposed for the different envisioned case studies.

3. SKILLS

- 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)
- Master 2 / Engineer in Applied Mathematics, Data Sciences, Artificial Intelligence and/or Signal Processing
- Good knowledge of statistical learning and neural networks
- Interest in environmental issues.
- Rigour in the organisation of work, data and results management. The datasets to be handled are diverse and large.
- - Good knowledge of computer tools: this internship requires knowledge of the LINUX environment and the Python programming language.
- - Appreciated knowledge of Python environments for deep learning (keras, tensorflow and/or pytorch).

4. TUTORSHIP

Tutorship will be shared between Carlos Granero-Belinchon (IMT Atlantique), Simon van Gennip an Philippe Gaspar (Mercator Océan). Interested students shall send their CV and a motivation letter to carlos.granero-belinchon@imt-atlantique.fr , pgaspar@mercator-ocean.fr et svangennip@mercator-ocean.fr

Depending on the COVID situation, the internship will take in the Mercator-Ocean offices (near Toulouse), at IMT Atlantique (Brest) or teleworking.

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.