ANITI
ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE

PhD Proposal in Artificial Intelligence and Machine Learning

Artificial Intelligence for Ecosystem Monitoring using Remote Sensing and Digital Agriculture Data

Advisors:
- Mathieu Fauvel, CESBIO, INRA, Toulouse, France
- Jordi Inglada, CESBIO, CNES, Toulouse, France
- Mickaël Savinaud, CS, Toulouse France

Net salary: XXXX euros per month with some teaching (64 hours per year on average)
Duration: 36 months.

Context

In the last years, the advent of Earth observation satellite missions with short revisit time and increased spatial resolution has led to an unprecedented amount of remote sensing images of heterogeneous physical nature (*e.g.*, optical & radar Sentinel time series ...) at various scales (*e.g.*, submetric, decametric ...). Furthermore, satellite image archives, such as *Spot Heritage*, are made available by many space agencies. Such massive data extend existing satellite and *in-situ* acquisition system used to understand, explain and predict the states and trends of our environment.

At the same time, digital agriculture gathers more and more data with the development of include sensors, robotics and machinery. Such data, combined with weather information, provide a richfull and complementary information to remote sensing data. A join use of these two sources of valuable information is a crucial task to enhance the knowledge of our environment and our ability to make decisions. However, the novel complexity of the data makes the conventional analytical methods not adapted, and therefore not suitable for extracting and for processing all the relevant information from the massive flow of data.

In order to address challenges raised by such applicative domains, the interdisciplinary institute in artificial intelligence of Toulouse, named the Artificial and Natural Intelligence Toulouse Institute (ANITI) from which the CNES is partner, has been proposed to develop innovative solutions using theoretical advances in core AI scientific areas. The CESBIO lab, with J. Inglada and M. Fauvel, is part of the ANITI Chair entitled "*Data-driven approximate Bayesian computation for fusion-based inference from heterogeneous (remote sensing) data*" hold by Prof. N. Dobigeon. Two main challenges issued from the ANITI core tracks have direct application for this PhD proposal, co-funded by CS:

1. **Integration of massive multi-source/scale satellite image** (optical & radar image time series, very high spatial resolution) **and in-situ/field data** (digital agriculture, meteorological or crowd-sourced data) **in learning algorithm through large scale distributed optimization**.
2. **Explainable and interpretable model**.

PhD Objectives

The PhD thesis objectives are two-fold. First it aims to integrate into learning algorithms multi-scale&source data from Earth observation systems and from digital agriculture. Second, the definition of spatially constraints and interpretable models will be considered.

Theoretical foundation of the PhD work will be based on last advances in Gaussian Processes (GP) and kernel algorithms. Such methods have regained attention from the machine learning community thanks to last developments in optimization techniques. They allows GP to be used in large scale scenario, with many millions of points [1, 2]. Current researches are bridging the gap between Deep Neural Network and GP, by adding theoretic results in large scale learning [3] and interpretability from the Bayesian modeling.

From the current algorithms, it is not possible to integrate point data into the processing of satellite image time series because of the different spatial sampling (point versus grid sampling). The construction of appropriate latent subspaces will be considered to properly use heterogeneous data by means of appropriate vector-valued kernel function and multi-output GP.

The construction of spatially interpretable model will be considered by a constrained spatial stratification. Large scale analysis of remote sensing often resort to such stratification (e.g., eco-climatic stratification): several models, one for each strata, are learned independently. However, no constraints are imposed and the models could behave differently at the spatial region boundaries. The objective is to include specific spatial constrains in the learning step to ensure a smooth transition between two (or more) spatial regions, i.e., to ensure similar prediction for models at boundaries.

One key step of the proposed methodology is the learning step, i.e., the optimization of the various parameters of the model. Using GP with spatial constraints on massive data set requires to solve large scale non-convex problems, which is not a trivial task. Hence, specific developments on computational statistic and optimization will be conducted to solve such problems efficiently. In particular, distributed optimization strategy will be considered to cope with the possible dissemination over multiple data center of the data [4].

The validation context will be the land-cover production chain of CESBIO, iota^2, and its annual land cover map OSO (http://osr-cesbio.ups-tlse.fr/~oso/). The proposed models will be integrated and validated with respect to the current standard of large scale land cover map production. In terms of products, land-cover maps as well as biodiversity indices will be considered.

Application

Requirements The candidate must have a solid background at least in one of the following items

- Statistical signal and image processing,
- Machine learning,
- Optimization.

A good knowledge of English and scientific programming (Python, C/C++) is required.

Contact Candidates should send an e-mail to mathieu.fauvel@inra.fr, jordi.inglada@cesbio.cnes.fr and mickael.savinaud@c-s.fr containing:

- Full CV,
- Motivation letter,
- Contact information for 2 references, and/or recommandation letter.

The beginning of the thesis is scheduled on September 2020. The Application is open until the position is fulfilled. The recruit will be registred to the doctoral school ED173 “Geosciences, Astrophysics, Space and Environmental Sciences” or ED475 “Mathematics, Informatics and Telecommunications”.

Practical details

The monthly gross salary is XXXX euros per month with some teaching (64 hours per year on average), for 36 months. The position comes with health insurance and other social benefits. The recruit will be located in the CESBIO lab, in Toulouse and will interact with people involved in the project (CS, ANITI, CNES & INRA). French is not mandatory.

CESBIO Research at CESBIO aims to develop knowledge on continental biosphere dynamics and func- tioning at various temporal and spatial scales and as such participates in the specification of space missions and the processing of remotely sensed data. CESBIO is or has been PI for 2 ESA satellite missions (SMOS, the Soil Moisture and Ocean Salinity satellite, and BIOMASS, a P-band SAR system to be launched in 2020) and for the French-Israeli Venus satellite (2-day revisit, 10 m resolution, optical sensor for vegetation monitoring, launched in 2017). CESBIO has developed the IOTA2 processing chain for the operational production of land-cover maps at the national French scale. It has therefore a strong experience in upscaling learning and classification processes. CESBIO has been committed over the last two years in collecting feed- back, tailoring IOTA2 outputs for various end-users, and disseminating it for several research institutes in France.

ANITI The ambition of the ANITI project is to develop a new generation of artificial intelligence called hybrid AI, combining data-driven machine learning techniques with symbolic and formal methods for expressing properties and constraints and carrying out logical reasoning. This approach will provide better guarantees in terms of reliability, robustness and the ability to explain and interpret the results of the algorithms used, while ensuring social acceptability and economic viability. Such guarantees are required by many applications targeted by the project, such as autonomous vehicles of the future. Starting operations this autumn, ANITI will bring together more than 200 researchers from universities, engineering schools, scientific and technological research organizations, and about thirty companies in the Toulouse region.

More details here: https://en.univ-toulouse.fr/aniti

CS Communication & Systèmes CS Systèmes d'Information is a French ETI of more than 2000 people, a major European player in the integration of systems, including space and simulation systems. Within CS-SI, the *BU Space*, in charge of space-related activities, is composed of more than 300 people working for 35 years for major players in the space sector in France and in Europe: CNES, ESA, Airbus Defense & Space and Thales among others. Within the *BU Space*, the PDA department brings its competences and expertise in image quality control, processing, production and operational exploitation of geospatial data (satellite images, geographic databases, etc.). The department's activities cover almost the entire space value chain, from sensor to applications. This broad positioning makes it possible to have within a single department a multidisciplinary team with different but totally complementary profiles and experiences. As part of this positioning CS SI intervenes on various topics of R& D whose AI is one of the main axes. In addition, CS SI has defined a highly structured open source policy that innervates these different activities of both R& D and development of operational applications. For example, the Remote Sensing and Toolbox team integrated into the PDA department develops open source software such as the state-of-the-art of image processing: Orfeo ToolBox, SNAP, S2P, ...

References

[1] Ke Alexander Wang, Geoff Pleiss, Jacob R. Gardner, Kilian Q. Weinberger, and Andrew Gordon Wilson. Exact gaussian processes on a million data points. *arXiv pre-print arXiv:1903.08114*, 2019.

[2] Marc Deisenroth and Jun Wei Ng. Distributed gaussian processes. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1481–1490, Lille, France, 07–09 Jul 2015. PMLR.

[3] Julien Mairal. *Large-Scale Machine Learning and Applications*. Habilitation à diriger des recherches, UGA - Université Grenoble Alpes, October 2017.

[4] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011.