Environmental Computational Science
and Earth Observation Laboratory

Master thesis subject

Species Distribution Modelling with Meta-learning

Context

Describing and understanding the geographic distribution and environmental suitability of species is a central question in ecology and biogeography. It has become increasingly meaningful in the face of anthropogenic pressure on biodiversity and has occasioned the development of **species distribution models (SDMs)**. Such models relate species occurrence data with environmental variables and are used to understand and predict species' distributions across landscapes (Beery et al., 2021).

At the same time, **machine learning** and especially **deep learning** methods have become increasingly popular in many fields, including ecology. However, their application to SDM has not yet been widely adopted. This is partly due to the fact that the datasets of species observations used in SDMs are small and suffer from high selection biases, making the use of naive deep learning approaches difficult.

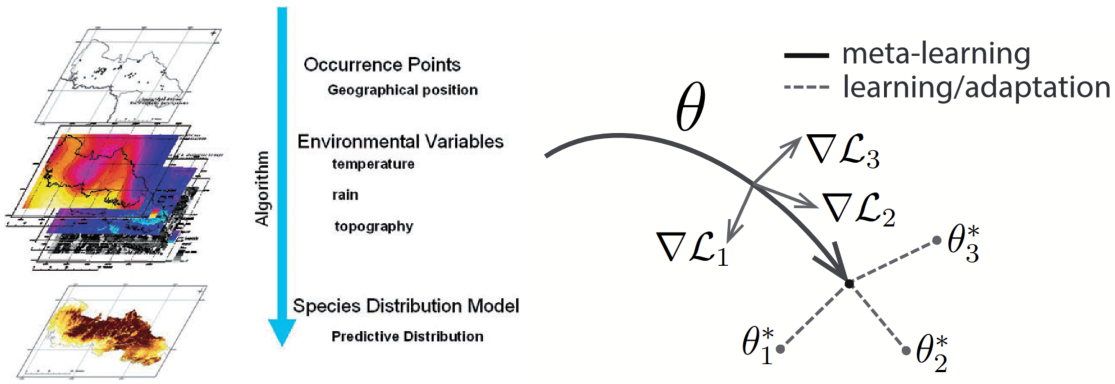

*Figure: (Left) An example of species distribution modelling. Presence only data is used to sample environmental covariates of where a species might occur. A machine learning model then extrapolates out the correlative relationship between presence and environment to generate a habitat suitability map. (Fook et al, 2009). (Right) Meta-learning approaches aim to "learn-to-learn" by training a model on a variety of tasks (here species), such that it can quickly adapt to new tasks (species) with only a small number of training samples (*Finn et al., 2017*).*

ECEO
EPFL ENAC IIE
Devis Tuia, Prof.
Rue de l'Industrie 17
Case Postale 440
CH – 1951 Sion
Téléphone : +41 21 69 382 83 (secr).
E-mail : devis.tuia@epfl.ch

Project

To address these issues, we would like to explore the use of meta-learning methods for SDMs. Meta-learning is a promising approach when the amount of data is small (Hospedales et al., 2022), and could be effective when modelling multiple species simultaneously. We therefore aim to improve current species distribution models by integrating ideas from meta-learning. Using a well-established tabular benchmark dataset of species observations (Elith et al., 2020), this project consists of the implementation and comparison of several meta-learning methods, including MAML (Finn et al., 2017) and simple transfer learning.

Requirements

- Experience in machine learning
- Proficiency in Python and relevant libraries such as Scikit-learn and Pytorch
- Familiarity with ecology and SDMs is a plus
- Knowledge of meta-learning is a plus
- Strong willingness to learn and ability to work independently

Literature

- Sara Beery, Elijah Cole, Joseph Parker, Pietro Perona, and Kevin Winner. 2021. Species Distribution Modeling for Machine Learning Practitioners: A Review. In ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '21). Association for Computing Machinery, New York, NY, USA, 329–348. https://doi.org/10.1145/3460112.3471966
- Elith Jane, Catherine Graham, Roozbeh Valavi, Meinrad Abegg, Caroline Bruce, Simon Ferrier, Andrew Ford, Antoine Guisan, Robert J. Hijmans, Falk Huettmann, Lucia Lohmann, Bette Loiselle, Craig Moritz, Jake Overton, A. Townsend Peterson, Steven Phillips, Karen Richardson, Stephen Williams, Susan K. Wiser, Thomas Wohlgemuth, and Niklaus E. Zimmermann. 2020. “Presence-Only and Presence-Absence Data for Comparing Species Distribution Modeling Methods”. *Biodiversity Informatics* 15 (2):69-80. https://doi.org/10.17161/bi.v15i2.13384
- Finn Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." *International conference on machine learning*. PMLR, 2017.
- Timothy Hospedales, Antreas Antoniou, Paul Micaelli and Amos Storkey, "Meta-Learning in Neural Networks: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5149-5169, 1 Sept. 2022, https://doi.org/10.1109/TPAMI.2021.3079209

Contact

Prof. Devis Tuia, devis.tuia@epfl.ch