Plant recommendation using environment or biotic associations - GeoLifeCLEF 2019 Working Notes

Sara Si-Moussi[1,2,3], Mickael Hedde[1], and Tanguy Daufresne[1]

[1]*Eco&Sols, UMR 1222 INRA-Montpellier SupAgro, Montpellier*
[2]*Loria, UMR 7503 Inria, Vandœuvre-lès-Nancy*
[3]*LECA, UMR 5553 CNRS-Université Grenoble Alpes, Saint-Martin-d'Hères*

1 Introduction

Predicting the most likely species in a given location is of great importance in biodiversity studies. In particular, knowledge of the dominant plants provides a description of the floristic habitat, thereby offering hints on the microorganisms and animals we would expect to see. Besides, plants among other autotrophs are at the center of the energy and biomass creation processes that drive ecological networks, providing many crucial services for human well-being such as food provision.

To predict the biological community composition in a given location is an age-old problem in biogeography, commonly solved through Species Distribution Modeling (SDM) or Habitat Suitability Modeling (HSM). The problem consists on learning a density function of the species over the geographic space from a set of observed geolocalized occurrences. In practice, due to sampling bias, limited examples and local habitat heterogeneity, geographic coordinates are not used directly as predictors. Instead, SDMs predict species abundance as a function of the environmental conditions at the given locations.

The local environment is usually described by abiotic features such as climate and pedology. Recently, more studies include biotic covariates in the form of other living species abundance. The latter use is motivated by the need to account for the dependencies between species that can affect their co-distributions at local scale. Indeed, despite suitable physical habitat a species may not be observed if it has been excluded by a competitor. Conversely, the species might be observed in an area with less favorable abiotic conditions if the microhabitat created by a facilitating species allows her to sustain. On the other hand, two species distribution may be correlated indirectly through a latent unmeasured abiotic variable or directly if they interact in some way that creates a dependency between their respective populations as in the case of plant-pollinators, host-parasites, predator-prey, etc.

The set of locations with suitable abiotic (respectively biotic) conditions for a given species define its Grinnellian (resp. Eltonian) niche. The intersection of the two previous sets produces the species potential or *Fundamental niche*. Finally, the subset of the potential niche that is accessible to the species by means of dispersal delimits its *Realized niche* [6]. The latter is what is observed in practice. Consequently, SDMs attempt the reverse process of inferring the species habitat preferences (Grinnellian niche) and their mutual dependencies or interactions (Eltonian niche) from their realized niches. Thanks to that, they can be used to predict the species abundance in any other location.

For the GeoLifeClef 2019 task, we train two SDMs on recommending the most probable plants. The first model relies purely on abiotic features. It exploits the feature extraction abilities of deep convolutional networks for end-to-end supervised learning from environmental rasters. The second proposal only uses regional level co-occurrences of expert-selected organisms. Inspired from word2vec Continuous Bag of Words, the model learns pairwise associations between target plants and chosen non-plant species. We describe both solutions and discuss results obtained during training phase on validation set as well as their performance on the competition blind test.

2 Task and dataset description

What env features, how many occurrences of plants/non plants, uncertainty

3 Proposals

We evaluate two architectures for solving the task:
- *GrinnellNet*: A convolutional neural network architecture using environmental rasters.
- *EltonNet*: A species embedding network leveraging associations with nearest non-plant occurrences.

3.1 GrinnellNet: a CNN with categorical rasters embedding

3.1.1 Convolutional neural network

The observation of a species in a given location does not necessarily translate a suitability of the abiotic environment. Indeed, some species can sustain in locations with unfavorable conditions (sink) as long as new individuals continuously join the population from a nearby suitable habitat (source), through seed dispersal for instance via wind currents. These so called source-sink dynamics [3] ensure an indefinite sustainability of the sink populations despite unfavorable abiotic conditions. Consequently, failing to account for these stochastic spatial processes may result in overestimation of the species environmental niche's breadth. Beyond stochastic assembly processes, georgraphic coordinates measured by various smartphone devices present an uncertainty that needs to be accounted for.

For the previous reasons, it is necessary to consider a landscapewise rather than a pointwise description of the environment around the observation geolocation by using environmental patches instead of local values. This requires appropriate feature extraction techniques that are able to harness the spatial structure of such inputs. Convolutional neural networks[4] constitute an ideal choice as they allow to extract features from multidimensional image-like inputs within an end-to-end learning process. Finally, convnets have been previously shown to provide substantial improvements in predicting species abundance [1, 2].

3.1.2 Embedding categorical features

Part of the soil physico-chemical properties are described by categorical or pseudo-ordinal features such as texture, land cover, erodibility and crusting class. Therefore, for each categorical feature F_c, before feeding it to the CNN layers, we replace each of its N_c categories by a real vector representation of size K_c. In practice, this is implemented by a feature-specific embedding lookup layer parameterized by a (N_c, F_c) matrix E_c, such that $E_c[i,]$ is the K_c sized embedding of the ith value of F_c. We apply this transformation batchwise in parallel to all patch cells. For an input of dimension $(BatchSize, PatchRadius, PatchRadius, 1)$, the embedding layer of F_c returns a tensor of dimension $(BatchSize, PatchRadius, PatchRadius, K_c)$. Vector representations are learnt along with other network parameters whereas K_c is a tunable hyperparameter, typically chosen in the interval $[\![2, \frac{N_c}{2}]\!]$.

Categorical embeddings capture richer relationships than raw categories. They are also considered as a dimensionality reduction technique, more practical than one-hot-encoding when dealing with high-cardinality yet sparse features (a typical example in our case is land cover). Note that each embedding dimension could have multiple meanings that do not necessarily line up with ordinal dimensions. Categories with similar representations translate a similar effect on the target variable.

3.1.3 GrinnellNet architecture

Continuous features are input to the neural network as batches of multi-channel images of dimension equal to $(BatchSize, PatchRadius, PatchRadius, N_n)$, N_{num} being the number of continuous channels/features. To this tensor, we concatenate the output of categorical features embeddings resulting in a tensor of dimension: $(BatchSize, PatchRadius, PatchRadius, Nchannels)$ where $Nchannels = N_{num} + \sum_{c \in C} N_c$. C refers to the set of categorical features. Retained embedding sizes, patch radius and resulting number of channels are shown in GrinnellNet architecture: Figure 1.

GrinnellNet comprises a 2-block VGG-like architecture[7] to extract features from the environmental tensor. Each block contains two 3x3 convolution layers set to extract 256 features, followed by MaxPooling then a Leaky Rectified Linear Unit activation. The latter function choice prevents the model from falling into a dying ReLu problem (experienced during first tests with ReLu activation)[5]. The last block is followed by 3 fully connected layers with respectively (8192,4096,3353) neurons. Finally, a softmax activation is applied on the last layer's output.
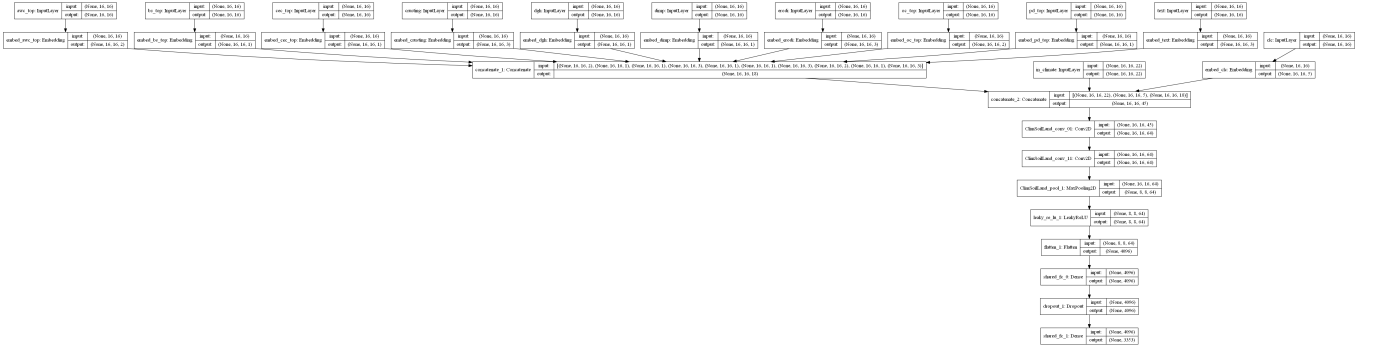
Figure 1: GrinnellNet architecture

3.1.4 Optimization and evaluation metrics

Focal softmax loss, qu'est-ce que ça apporte ? Une vue du déséquilibre des classes dans le jeu de données Correction par pondération des classes + focal loss Stratified sampling

3.1.5 Training and validation

Preprocessing, cross-val split, graphes tensorboard

3.2 EltonNet: a species embedding network

Protocol followed with tanguy and Mike for each taxonomic group Distance calculation Visualiser les associations apprises

References

[1] BOTELLA, C., JOLY, A., BONNET, P., MONESTIEZ, P., AND MUNOZ, F. A deep learning approach to species distribution modelling. In *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*. Springer, 2018, pp. 169–199.

[2] DENEU, B., SERVAJEAN, M., BOTELLA, C., AND JOLY, A. Location-based species recommendation using co-occurrences and environment-geolifeclef 2018 challenge. In *CLEF 2018* (2018).

[3] HOLT, R. D. Population dynamics in two-patch environments: some anomalous consequences of an optimal habitat distribution. *Theoretical population biology 28*, 2 (1985), 181–208.

[4] LECUN, Y., BENGIO, Y., ET AL. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks 3361*, 10 (1995), 1995.

[5] MAAS, A. L., HANNUN, A. Y., AND NG, A. Y. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (2013), vol. 30, p. 3.

[6] PETERSON, A. T., SOBERÓN, J., PEARSON, R. G., ANDERSON, R. P., MARTINEZ-MEYER, E., NAKAMURA, M., AND ARAÚJO, M. B. *Ecological niches and geographic distributions (MPB-49)*, vol. 56. Princeton University Press, 2011.

[7] SIMONYAN, K., AND ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).