

Restricted SDM: Predicting habitat suitability from species occurrence only

Boyan Angelov

May 2018

DOI: <https://doi.org/10.5281/zenodo.1248990> (*Journal of Brief Ideas*)

Species Distribution Models are commonly used to predict species habitat suitability. The data needed to build such a model usually consist of a range of environmental variables together with a target variable - the species occurrence itself (often in presence/absence format, or just presence).

An interesting idea that can be applied to this problem is to rely on the species occurrence only. The assumption here is similar to a market basket analysis: similar species should co-occur, much likely to a shopping cart containing frequently purchased together items, such as ketchup and french fries. In this case our features become different species, and our observations are different sites where they are either present or absent. We can use such a model to predict the probability of a new species occurring when some of the others have already been identified. Possible difficulties in using this approach are class imbalance, large dimensionality and sparsity in the data. Dimensionality reduction techniques such as PCA can be useful in this case and improve the prediction performance.

Attached are some preliminary results, predicting the occurrence of the Great horned owl (*Bubo virginianus*) from the ebird reference dataset.

