Re-scaling of model evaluation measures to allow direct comparison of their values

By *A. Márcia Barbosa*

`model` `evaluation` `performance`

Share 3   Tweet 0

Author

A. Márcia Barbosa

Species distribution models are increasingly used in ecology, biogeography and climate change research, and are usually complemented with one or more metrics evaluating their performance. Not all metrics vary within the same scale of measurement: for example, Cohen's kappa and the true skill statistic (TSS) may range between -1 and 1, while most other widely used metrics range only between 0 and 1. Values of different measures are thus not directly comparable, and e.g. a kappa or TSS value of 0.6 does not denote (although it may at first sight suggest) lower discriminative accuracy than an area under the curve (AUC) of 0.8. Yet, these measures are often presented side by side without a clear acknowledgement of this scale difference. I propose clearly acknowledging such difference, or else using a simple formula to standardize these measures so that their values can be compared more directly. The following equation converts an evaluation score that ranges from -1 to 1 into its corresponding value in the 0-to-1 scale: $(score+1)/2$. Conversion can also be done the other way around with $2*(score-0.5)$. This standardization is implemented in the *modEvA* package for R (currently available on R-Forge), both as an independent function and as an option within other functions that compute and compare model evaluation measures.

[NOTE: this is not an abstract, but the whole paper. The *Journal of Brief Ideas* takes only articles with up to 200 words.]