Package 'MaxentVariableSelection'

July 21, 2025

Type Package

Title Selecting the Best Set of Relevant Environmental Variables along with the Optimal Regularization Multiplier for Maxent Niche Modeling

Version 1.0-3

Date 2018-01-23

Author Alexander Jueterbock

Maintainer ``Alexander Jueterbock" `<Alexander-Jueterbock@web.de>`

Description Complex niche models show low performance in identifying the most important range-limiting environmental variables and in transferring habitat suitability to novel environmental conditions (Warren and Seifert, 2011 <DOI:10.1890/10-1171.1>; Warren et al., 2014 <DOI:10.1111/ddi.12160>). This package helps to identify the most important set of uncorrelated variables and to fine-tune Maxent's regularization multiplier. In combination, this allows to constrain complexity and increase performance of Maxent niche models (assessed by information criteria, such as AICc (Akaike, 1974 <DOI:10.1109/TAC.1974.1100705>), and by the area under the receiver operating characteristic (AUC) (Fielding and Bell, 1997 <DOI:10.1017/S0376892997000088>). Users of this package should be familiar with Maxent niche modelling.

License GPL (>= 2)

Imports ggplot2, raster

Depends R (>= 3.1.2)

Suggests knitr, rmarkdown

VignetteBuilder knitr

SystemRequirements maxent.jar file

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2018-01-23 21:56:31 UTC

Contents

MaxentVariableSelection-package

Selecting the Best Set of Relevant Environmental Variables along with the Optimal Regularization Multiplier for Maxent Niche Modeling

---

Description

Complex niche models show low performance in identifying the most important range-limiting environmental variables and in transferring habitat suitability to novel environmental conditions (Warren and Seifert, 2011 <DOI:10.1890/10-1171.1>; Warren et al., 2014 <DOI:10.1111/ddi.12160>). This package helps to identify the most important set of uncorrelated variables and to fine-tune Maxent's regularization multiplier. In combination, this allows to constrain complexity and increase performance of Maxent niche models (assessed by information criteria, such as AICc (Akaike, 1974 <DOI:10.1109/TAC.1974.1100705>), and by the area under the receiver operating characteristic (AUC) (Fielding and Bell, 1997 <DOI:10.1017/S0376892997000088>). Users of this package should be familiar with Maxent niche modelling.

Details

Package:	MaxentVariableSelection
Type:	Package
Version:	1.0-3
Date:	2018-01-23
Depends:	R (>= 3.1.2)
Imports:	ggplot2, raster
Suggests:	knitr, rmarkdown
VignetteBuilder:	knitr
License:	GPL (>= 2)
Literature:	Akaike H (1974)
	A new look at the statistical model identification
	IEEE Transactions on Automatic Control **19**:6 716–723.
	Fielding AH and Bell JF (1997)
	A review of methods for the assessment of prediction
	errors in conservation presence/absence models
	Environmental Conservation **24**:1 38–49.
	Jimenez-Valverde A (2012)

Insights into the area under the receiver operating characteristic curve
(AUC) as a discrimination measure in species distribution modelling
Global Ecology and Biogeography **21**:4 498–507.

Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F and De Clerck, O (2012)
Bio-ORACLE: a global environmental dataset for marine species distribution modelling
Global Ecology and Biogeography **21**:2 272–281.

Warren DL, Glor RE, and Turelli M (2010)
ENMTools: a toolbox for comparative studies of environmental niche
models
Ecography **33**:3 607–611.

Warren DL and Seifert SN (2011)
Ecological niche modeling in Maxent: the importance of model
complexity and the performance of model selection criteria
Ecological Applications **21**:2 335–342.

Citation

To cite the package 'MaxentVariableSelection' in publications use:

Jueterbock A, Smolina I, Coyer JA and Hoarau, G (2016)
The fate of the Arctic seaweed *Fucus distichus* under climate change:
an ecological niche modelling approach
Ecology and Evolution **6**(6), 1712-1724

Author(s)

Alexander Jueterbock

Maintainer: Alexander Jueterbock, <Alexander-Jueterbock@web.de>

Backgrounddata *CSV file with background/pseudoabsence data*

Description

Longitude and latitude values, as well as values of four environmental variables (from the Bio-
ORACLE dataset; Tyberghein et al., 2012) for each of 10,000 background points. The background
points were selected randomly along shorelines of all continents in the northern hemisphere.

Format

A data frame that specifies environmental conditions and geographic locations of 10,000 background sites.

species The species name is here set to 'bg', which stands for background

longitude longitudinal coordinate

latitude latitudinal coordinate

calcite Mean calcite concentration (mol/m3)

parmean Mean photosynthetically active radiation (Einstein/m2/day)

salinity Mean salinity (PSS)

sstmax Maximum sea surface temperature (degree celsius)

References

Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F and De Clerck, O (2012)
Bio-ORACLE: a global environmental dataset for marine species distribution modelling
Global Ecology and Biogeography **21**:2 272–281.

Examples

```
backgroundlocations <- system.file("extdata",
                                   "Backgrounddata.csv",
                                   package="MaxentVariableSelection")
backgroundlocations <- read.csv(backgroundlocations,header=TRUE)
head(backgroundlocations)
```

Occurrencedata *CSV file with occurrence data*

Description

Longitude and latitude values, as well as values of four environmental variables (from the Bio-ORACLE dataset; Tyberghein et al., 2012) for each of 98 occurrence sites (locations where a species was recorded).

Format

A data frame that specifies geographic locations and environmental conditions of 98 occurrence sites.

species The species name is here set to 'bg', which stands for background

longitude longitudinal coordinate

latitude latitudinal coordinate

calcite Mean calcite concentration (mol/m3)

parmean Mean photosynthetically active radiation (Einstein/m2/day)

salinity Mean salinity (PSS)

sstmax Maximum sea surface temperature (degree celsius)

References

Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F and De Clerck, O (2012)
Bio-ORACLE: a global environmental dataset for marine species distribution modelling
Global Ecology and Biogeography **21**:2 272–281.

Examples

```
occurrencelocations <- system.file("extdata",
                                   "Occurrencedata.csv",
                                   package="MaxentVariableSelection")
occurrencelocations <- read.csv(occurrencelocations,header=TRUE)
head(occurrencelocations)
```

VariableSelection	*Selecting the best set of relevant environmental variables along with the optimal regularization multiplier for Maxent Niche Modeling*

Description

This is the core function of the package in which a set of environmental variables is reduced in a stepwise fashion in order to avoid overfitting the model to the occurrence records. This can be done for a range of regularization multipliers. The best performing model, based on AICc values (Akaike, 1974) or AUC.Test values (Fielding and Bell, 1997), identifies then the most-important uncorrelated environmental variables along with the optimal regularization multiplier.

Usage

```
VariableSelection(maxent, outdir, gridfolder, occurrencelocations,
backgroundlocations, additionalargs, contributionthreshold,
correlationthreshold, betamultiplier)
```

Arguments

maxent
: String specifying the filepath to the maxent.jar file (download from here: https://www.cs.princeton.edu/~s The package was tested with maxent.jar version 3.3.3k.

outdir
: String specifying the path to the output directory to which all the result files will be written.Please don't put important files in this folder as all files but the output files of the VariableSelection function will be deleted from this folder.

gridfolder
: String specifying the path to the directory that holds all the ASCII grids (in ESRI's .asc format) of environmental variables. All variables must have the same extent and resolution.

occurrencelocations
: String specifying the filepath to the csv file with occurrence records. Please find the exact specifications of the SWD file format in the details section below.

backgroundlocations

> String specifying the filepath to the csv file with background/pseudoabsence data. Please find the exact specifications of the SWD file format in the details section below.

additionalargs   String specifying additional maxent arguments. Please see in the details section below.

betamultiplier   Vector of beta (regularization multipliers) (positive numerical values). The smaller this value, the more closely will the projected distribution fit to the training data set. Overfitted models are poorly transferable to novel environments and, thus, not appropriate to project distribution changes under environmental change. The model performance will be compared between models created with the beta values given in this `betamultiplier` vector. Thus, providing a range of beta values from 1 (the default in Maxent) to 15 or so, will help you to spot the optimal beta multiplier for your specific model.

correlationthreshold

> Numerical value (between 0 and 1) that sets the threshold of Pearson's correlation coefficient above which environmental variables are regarded to be correlated (based on values at all background locations). Of the correlated variables, only the variable with the highest contribution score will be kept, all other correlated variables will be excluded from the Maxent model. Correlated variables should be removed because they may reflect the same environmental conditions, and can lead to overly complex or overpredicted models. Also, models comiled with correlated variables might give wrong predictions in scenarios where the correlations between the variables differ.

contributionthreshold

> Numerical value (between 0 and 100) that sets the threshold of model contributions below which environmental variables are excluded from the Maxent model. Model contributions reflect the importance of environmental variables in limiting the distribution of the target species.

Details

For further details on the model selection process and the variable settings, please have a look at the vignette that comes with this package.

Value

The following result files are saved in the directory specified with the `outdir` argument.

ModelPerformance.txt

> A table listing the performance indicators of all created Maxent models
>
> `Model` Unique model number
>
> `betamultiplier` Maxent's regularization multiplier
>
> `variables` Number of environmental variables
>
> `samples` Number of occurrence sites
>
> `parameters` Number of parameters estimated from the model
>
> `loglikelihood` log likelihood value
>
> `AIC` Akaike Information Criterion

`AICc` sample size corrected AIC

`BIC` Bayesian information criterion

`AUC.Test` Area under the receiver operating characteristic fro test data

`AUC.Train` Area under the receiver operating characteristic fro training data

`AUC.Diff` Difference between AUC.Test and AUC.Train

The information criteria (AIC, AICc, and BIC) are set to 'x' if the number of parameters is lower than the number of variables in the model.

`ModelSelectionAICc_MarkedMaxAUCTest.png`

A figure showing the AICc values of all models, which are ordered along the x-axis based on the applied beta-multiplier. The number of environmental variables included in each model is coded by dot color and size. The model with highest AUC.Test value is marked in red.

`ModelSelectionAICc_MarkedMinAICc.png`

A figure showing the AICc values of all models, which are ordered along the x-axis based on the applied beta-multiplier. The number of environmental variables included in each model is coded by dot color and size. The model with highest minimum AICc value is marked in red.

`ModelSelectionAUCTest_MarkedMaxAUCTest.png`

A figure showing the AUC.Test values of all models, which are ordered along the x-axis based on the applied beta-multiplier. The number of environmental variables included in each model is coded by dot color and size. The model with highest AUC.Test value is marked in red.

`ModelSelectionAUCTest_MarkedMinAICc.png`

A figure showing the AUC.Test values of all models, which are ordered along the x-axis based on the applied beta-multiplier. The number of environmental variables included in each model is coded by dot color and size. The model with highest minimum AICc value is marked in red.

`ModelWithMaxAUCTest.txt`

Subset of the table `ModelPerformance.txt`, which shows only the model with the highest AUC.Test value.

`ModelWithMinAICc.txt`

Subset of the table `ModelPerformance.txt`, which shows only the model with the lowest AICc value.

`VariableSelectionProcess.txt`

Table listing model contributions for and correlations between each of the environmental variables for all created Maxent models. The numbers of the models refer to the unique model numbers in the table `ModelPerformance.txt` (see above). The following entries describe the content row-wise (not column-wise)

`Test` Either 'Contributions' or 'Correlation. Informs if the numbers for each of the environmental variables refers to model contribution coefficients or to correlation coefficients.

`Model` The unique model number (the same unique model number as in `ModelPerformance.txt`.)

`betamultiplier` The (regularization multipliers) used to compile the respective model.

X 'X' stands here for the name of an environmental variable. The `Test` row
above informs whether the values in this row refer to the model contribu-
tion of this environmental variable or to its coefficient of correlation with
another environmental variable. The variable to which it is compared is rec-
ognizable by a correlation coefficient of 1. If this environmental variable
was excluded from the model, the value in this row is 'NA', which stands
for 'Not Available'.'

`VariableSelectionMaxAUCTest.txt`

Subset of `VariableSelectionProcess.txt` that shows only those models which
lead directly to the model with the highest AUC.Test value.

`VariableSelectionMinAICc.txt`

Subset of `VariableSelectionProcess.txt` that shows only those models which
lead directly to the model with the lowest AICc value.

Warning

Depending on the number of environmental variables and the range of different betamultipliers you
want to test, variable selection can take several hours so that you might want to run the analysis
over night.

Author(s)

Alexander Jueterbock, `<Alexander-Jueterbock@web.de>`

References

Akaike H (1974)
A new look at the statistical model identification
IEEE Transactions on Automatic Control **19**:6 716–723.

Fielding AH and Bell JF (1997)
A review of methods for the assessment of prediction
errors in conservation presence/absence models
Environmental Conservation **24**:1 38–49.

Examples

```
## Not run:
# Please find a workflow tutorial in the vignette of this package. It
# will guide you through the settings and usage of the
# 'VariableSelection' function, the core function of this package.

## End(Not run)

VariableSelection(
maxent="C:/.../maxent.jar",
outdir="OutputDirectory",
gridfolder="BioORACLEVariables",
occurrencelocations=system.file("extdata", "Occurrencedata.csv", package="MaxentVariableSelection"),
backgroundlocations=system.file("extdata", "Backgrounddata.csv", package="MaxentVariableSelection"),
```

```
additionalargs="nolinear noquadratic noproduct nothreshold noautofeature",
contributionthreshold=5,
correlationthreshold=0.9,
betamultiplier=seq(2,6,0.5)
)
```

Index