Abstract Title: Classification and predictive data modeling in ecology.

Alla Sapronova

[1] *Centre for Big Data Analysis (Uni Research , Bergen, Norway)*

Abstract

Classification, the process of assigning data into labelled groups, is one of the most common operation in data mining for ecological and biological systems. Classification can be used in predictive modeling to learn the relation between desired feature-vector and labelled classes. Models predicting the distribution of live organisms are used widely in ecology and biology and usually based on environmental data. Performance of such models based on the true positives rate when predicting occurrence of the organisms at test locations.
Two major pitfalls in using such predictive data modelling in ecology are: unspecified number of parameters influencing the system of interest and little amount of observations carrying reliable and/or labelled data for supervised training and validation of the model. As a result, the successfully validated models with relative high score of true positives often are not suitable for real world problem application: high rate of false negatives make those models not reliable or even useless.
When dealing with ecological or biological systems where the data set contains arbitrary big number of missed data and/or the amount of data samples is not adequate to the data complexity, it is important to define a strategy that allows to reach the model's desired accuracy without increasing the number of false positives.
In this work the author presents classification-based predictive model, that connects pelagic fish occurrence with environmental data. Model's optimization includes three different strategies: input pruning, semi-auto selection of various classification methods, and data volume increase. It is shown that even with limited number of samples the presented model was able to reach 92% of accuracy in predicting the fishes occurrence.

Keywords—*artificial intelligence, predictive modeling, classification, data mining in biology*

Professional Biography (100-150 words)

Alla Sapronova is an experienced data scientist with a demonstrated history of work in data science and machine learning applications for both academic and industrial sectors. Dr. Sapronova completed her PhD in Physics and Mathematics at the age of 29 from Moscow State University, Russia and postdoctoral studies from UniFob, University of Bergen, Norway. Currently she is the Head of Data Science at Center for Big Data Analysis, Uni Research, a multidisciplinary research institute in Bergen, Norway. Last 5 years she has published more than 15 papers in reputed journals and has been serving as an external censor for University of Bergen, Norway. Her scientific interests lie in the areas of Big Data Mining, Machine Learning, Knowledge Extraction, Time Series Analysis, Classification and Predictive Modeling.



Email Id: alla.sapronova@uni.no
Contact Number:+47 41488643

Any Comments: