Alla Sapronova et al., J Comput Eng Inf Technol, 6:5
DOI: 10.4172/2324-9307-C1-014

Prune the inputs, increase data volume, or select a different classification method – A strategy to improve accuracy of classification

Alla Sapronova[1] and **Mark Sytse Ybema**[2]
[1]Uni Research Computing, Norway
[2]Sustainovate AS, Norway

Classification, the process of assigning data into labeled groups, is one of the most common operations in data mining. Classification can be used in predictive modeling to learn the relation between desired feature-vector and labeled classes. When 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 reaching highest possible classification accuracy. In this work, authors present results on classification-based predictive model's accuracy for three different strategies: input pruning, semi-auto selection of various classification methods, and data volume increase. Authors suggest that a satisfactory level of model's accuracy can be reached when preliminary input pruning is used. The presented model is connecting fishing data with environmental variables. Even with limited number of samples, the model is able to resolve the type of the fish with up to 92% of accuracy. The results of using various classification methods are shown and suggestions are made towards defining the optimal strategy to build an accurate predictive model, opposed to common trial-and-error method. Different strategies for input pruning that assure information's preservation are described.

Biography

Alla Sapronova has completed her PhD at Moscow State University, Russia and Post-doctoral studies at UniFob, University of Bergen, Norway. She is the Head of Data Science at Center for Big Data Analysis, Uni Research, Norway. She has published more than 15 papers in reputed journals.

alla.sapronova@uni.no

Notes: