Contribution ID: **107** Type: **not specified**

Deciphering Random Forest models through conditional variable importance

Tuesday, 14 September 2021 11:40 (20 minutes)

In many data analytics applications based on machine learning algorithms, the main focus is usually on predictive modeling. In certain cases, as in many applications in manufacturing, understanding the data-driven model plays a crucial role in complementing the engineering knowledge about the production process. There is therefore a growing interest in describing the contributions of the input variables to the model in the form of "variable importance", which is readily available in certain machine learning methods such as random forest (RF). In this study, we focus on the Boruta algorithm, which is an effective tool in determining the importance of variables in RF models. In many industrial applications with multiple input variables, it becomes likely to observe high correlation among these variables. It is shown that the correlation among the input variables distorts and overestimates the importance of variables. The Boruta algorithm is also affected by this resulting in a larger set of input variables deemed important. To overcome this, in this study we present an extension of the Boruta algorithm for the correlated data by exploiting the conditional importance, which takes into consideration the correlation structure of the variables for computing the importance scores. This leads to a significant improvement of the variable importance scores in the case of a high correlation among variables and to a more precise ranking of the variables that contribute to the model significantly. We believe this approach can be used in many industrial applications by providing more transparency and understanding of the process.

Keywords

Random Forest, Conditional Importance, Feature selection, Boruta Algorithm

Special/invited session

Primary authors: ROTARI, Marta; KULAHCI, murat

Presenters: ROTARI, Marta; KULAHCI, murat

Session Classification: Modelling 1