S0503. Use and limitations of various metrics to assess the quality of extreme sparse datasets in geotechnics

Matthias Hahn (TU Graz & Montanuniversität Leoben),
Alla Sapronova (Graz University of Technology),
Marlene Villeneuve
Room: R8 2023-08-08 11:20

1 INTRODUCTION

Like the comparison of the round time in Formula 1 car racing, metrics show which machine learning algorithm best fits the data prediction requirements. The question is, are all metrics suitable for every algorithm and data set in geotechnics? This research gives a first recommendation of the stability of different metrics for regression and classification predictors for a wide variety of geotechnical data sets.

2 METHODOLOGY

Training and performance measurements of different algorithms on various data sets are used to show the stability of the metrics to eliminate the problem of subjectivity of parametrization. For this reason, a large number of different data sets are needed. Geotechnical data for machine learning are raw and often sparse, so to eliminate subjectivity the solution was to split every data set randomly into training and prediction subsets and repeat this procedure 300 times per data set. In this way the normalized standard deviation can be calculated, and the stability of the metric can be estimated. To make the metrics comparable, one probability distribution is used for every algorithm over all applied metrics. The stability of a metric is rated by Kolmogorov-Smirnov p-value > 0.05 and normalized standard deviation < 0.5.

3 CONCLUSIONS

The used methodology shows, in a limited way, how stable different metrics are, addicted to the algorithm. A recommendation for a small set of algorithms and metrics is possible. To validate these results, more data sets are needed. The best case would be, that the 300 times repeated training and performance measurement cycle can be substituted by a sufficiently large number of different data sets. Python provides more machine learning algorithms and metrics, which have not been examined in this research. If enough computational power is available, the algorithm-metric recommendation in this research can be done for all implemented algorithms and metrics in Python/Scikit-learn package.