

REC curves for visual evaluation of sugarcane yield machine learning models

Thiago da Silva Siqueira*, Felipe Ferreira Bocca, Luiz Henrique Antunes Rodrigues.

Abstract

Model validation often is performed with metrics unsuitable for the task. Also, no metric should be used alone as criterion. One alternative is the use of Regression Error Characteristic Curve (REC). The use of REC curves was able to replace the results of single metric evaluation while providing information about trade-offs and model variability. Considering the limitations of plotting several curves, REC curves should be used for final steps of model validation.

Key words:

Data Mining Models; Model Validation; Visual Analysis

Introduction

Modeling with machine learning techniques requires several steps. Model validation is one of the last steps, and it is required for deciding if a model should be deployed. Both the use of Root Mean Square Error (RMSE) or the Coefficient of Determination (R^2), and the use of only one metric, instead of a set of metrics, are often criticized for model validation^{1,2}.

Regression Error Characteristic Curves (REC) - analogous to Receiver Operating Characteristic Curves (ROC) for classification models - are an alternative for model validation. The characteristics of REC curves allow for the assessment of several characteristics required² for model validation. The objective of this work is to implement REC curves and evaluate their use for model validation.

Results and Discussion

The algorithm for REC curves³ was implemented in the R language. Vertical error bars were implemented following recommendations for ROC curves^{3,4}. The error bias (E), the mean absolute error (MAE), the agreement index (D) and the modeling efficiency (EF) were evaluated to address the need of comprehensive model validation² of models previously fitted⁵.

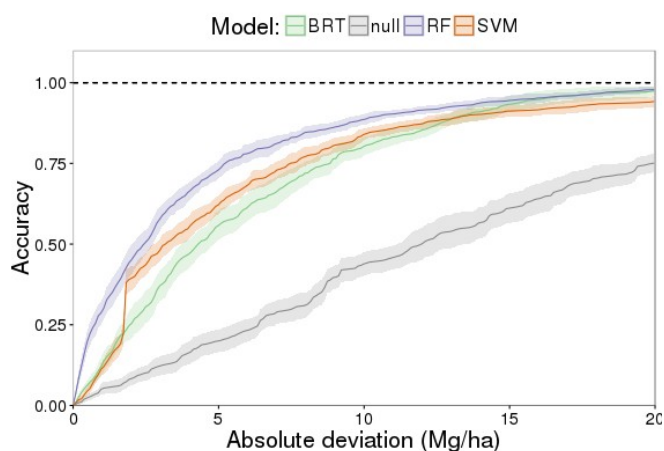
Representing model variability with error bars was troublesome due to the high number of points. The use of a shaded area resulted in better visualizations. No model evaluated had abnormal variability (Figure 1). Also, all models performed better than the null model.

Table 1. Model performance for different metrics.

Model	E	MAE	D	EF
RF	-0.29	4.11	0.96	0.85
SVM	-0.69	5.69	0.93	0.74
BRT	-0.59	5.94	0.93	0.77
null	-0.87	13.68	0.08	0

Comparing several models in one plot was not feasible, requiring that only a few curves were plotted for analysis. The analysis of models using REC curve dominance leads to the same results as the MAE evaluation. The dominant model, Random Forest (RF) presented the lower MAE (Table 1). REC curve analysis was able to highlight trade-offs in the models, which can be seen in the metrics presented in Table 1, such as for MAE and EF for Boosted Regressions Trees (BRT) and Support Vector Machines (SVM).

Figure 1. REC curves for model validation.



Conclusions

REC curve analysis resulted in similar results as a metric based analysis, although being capable of highlighting model trade-offs and allowing for variability assessment. Due to limitations of visual analysis, only a few models can be compared simultaneously, and we suggest the use of REC curves only in the final steps of model validation when few candidate models are being evaluated.

Acknowledgement

This work was partially supported by Bioen/Fapesp and Odebrecht Agro-industrial (Process # 12/50049-3).

¹ C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005.

² J. M. Yang, J. Y. Yang, S. Liu, and G. Hoogenboom, "An evaluation of the statistical methods for testing the performance of crop models with observed data," *Agric. Syst.*, vol. 127, pp. 81–89, 2014.

³ J. Bi and K. P. Bennett, "Regression Error Characteristic Curves," *Proc. Twent. Int. Conf. Mach. Learn.*, pp. 43–50, 2003.

⁴ T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

⁵ F. F. Bocca, "Produtividade de cana-de-açúcar: caracterização dos contextos de decisão e utilização de técnicas de mineração de dados para modelagem," p. 86, 2014.