Interpretability of statistical learning models in an industrial context

Clement Benard
Safran Tech, Sorbonne Universite

Supervisor(s): S. Da Veiga (Safran Tech), E. Scornet (CMAP, Ecole Polytechnique), G. Biau (LPSM, Sorbonne Universite)

Ph.D. expected duration: Dec. 2018 - Nov. 2021

Adress: Safran Tech, 1 rue Genevieve Aube, 78114 Magny-les-Hameaux

Email: clement.benard@safrangroup.com

Abstract:

In the manufacturing industry, the core of production processes involves complex physical and chemical phenomena. Their control and efficiency is of critical importance. In practice, data is collected along the manufacturing line, characterizing both the production conditions and its quality. State-of-art supervised learning algorithms can successfully catch patterns of such complex physical phenomena, characterized by non-linear effects and low-order interactions between parameters. However, any decision impacting a production process have long term and heavy consequences, and then, cannot simply rely on stochastic modeling. A deep physical understanding is required and black-box models are not appropriate. Models have to be interpretable, i.e. provide an understanding of the internal mechanism that build a relation between inputs and ouputs, to provide insights to guide the physical analysis. There is no agreement in statistics and machine learning communities about a rigorous definition of interpretability [6]. It is yet possible to define minimum requirements for interpretability: simplicity, stability [8] and predictivity.

Decision tree [2] can model highly non-linear patterns while having a simple structure and is then widely used when interpretability is required. Decision tree is also highly unstable to small data pertubation, which is a very strong limitation to its practical use. Random forest [1] stabilizes decision trees by aggregating many of them, it strongly improves accuracy but the model is a black box. Another class of supervised learning method can model non-linear patterns while having a simple structure: rule models. A rule is a conjunction of constraints on inputs variables that form a hyper-rectangle in the input space, where the estimated output is constant. A collection of rule is combined to form a model. Many algorithms were developped, among them: SLIPPER [3], Rulefit [4], Node Harvest [7] and BRL [5]... They share the same draw brack as trees: instability.

In this work, we design a new classification algorithm which inherits the accuracy of random forests, the simplicity of decision trees while having a stable structure for problems with low-order interaction effects. The principle of random forest is used, but instead of aggregating predictions, we focus on the probability that a given hyper-rectangle (a node) is contained in a randomized tree. The nodes with the highest probabilities are robust to data pertubation and represent strong patterns. They are selected to form a stable rule ensemble model. Our proposed algorithm works as follows:

1. Bin data using empirical quantiles.
2. Generate a large number of rules with the random forest procedure.
3. Select rules based on their frequency of appearance in the random forest.
4. Average the selected rules to form a rule ensemble model.

Many simulations on public datasets of the UCI repository (Asuncion and Newman 2007) show good performance of the procedure in terms of both predictive accuracy and stability.
We use 1 - AUC to measure accuracy. To evaluate stability, a 10-fold cross-validation is run, and, for each pair of folds the relative size of the intersection between the two lists of rules is computed.

Dataset	Random Forest	Rulefit	Node Harvest	CART	Our Method
Diabetes	0.17	0.18	0.19	0.22	0.18
Heart Statlog	0.10	0.11	0.12	0.17	0.14
Heart C2	0.10	0.11	0.12	0.19	0.10
Heart H2	0.11	0.10	0.11	0.18	0.12
Credit German	0.21	0.23	0.26	0.30	0.24
Credit Approval	0.07	0.06	0.07	0.11	0.07
Ionosphere	0.03	0.06	0.07	0.12	0.12
Breast Wisconsin	0.01	0.02	0.02	0.05	0.01

Table 1: Accuracy

Dataset	Node Harvest	CART	Our Method
Diabetes	60%	28%	76%
Heart Statlog	50%	34%	65%
Heart C2	55%	28%	56%
Heart H2	45%	37%	72%
Credit German	47%	39%	73%
Credit Approval	30%	24%	59%
Ionosphere	33%	24%	77%
Breast Wisconsin	60%	57%	82%

Table 2: Stability

References

[1] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. Classification and regression trees. *Wadsworth International Group*, 1984.

[3] William W Cohen and Yoram Singer. A simple, fast, and effective rule learner. *AAAI/IAAI*, 99:335–342, 1999.

[4] Jerome H Friedman, Bogdan E Popescu, et al. Predictive learning via rule ensembles. *The Annals of Applied Statistics*, 2(3):916–954, 2008.

[5] Benjamin Letham, Cynthia Rudin, Tyler H McCormick, David Madigan, et al. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics*, 9(3):1350–1371, 2015.

[6] Zachary C Lipton. The mythos of model interpretability. *arXiv preprint arXiv:1606.03490*, 2016.

[7] Nicolai Meinshausen. Node harvest. *The Annals of Applied Statistics*, pages 2049–2072, 2010.

[8] Bin Yu et al. Stability. *Bernoulli*, 19(4):1484–1500, 2013.

Short biography – Clement Benard is a research engineer at Safran Tech and a first year PhD in statistics, in collaboration with LPSM, Sorbonne Universite.