Machine Learning in the Service of Policy Targeting: The Case of Public Credit Guarantees

by M. Andini, M. Boldrini, E. Ciani,
G. de Blasio, A. D'Ignazio, A. Paladini

Discussion: Andrea Tamoni

London School of Economics

26-27 Nov, 2018
Modelling with Big Data and Machine Learning, BOE

Outline of the discussion

- Main findings and contributions of the paper.
- Some comments on empirical results.
- Some extensions.
- Conclusions.

The paper in a nutshell

- Demonstrate effectiveness of machine learning based policy assignment for public guarantee programs
- Resource allocation problem: "How to distinguish those firms which actually need financing from those which would find financing anyway?"
- This paper:
  - Uses ML to gauge credit-worthiness AND credit constraints
  - Argues that typical naive assignment rules are biased against credit constrained firms
  - Shows that ML-based assignment is more effective in targeting firms that out-perform ex-post relative to current assignment rules
  - Advocates use of tree-based methods: decision tree / random forest

Decision Tree vs. Random Forest

- Transparency is a common criticism of ML methods: "*ML is a black box*"
- Simple decision trees are as transparent as possible: clear data-driven and rule-based predictions
- Prediction accuracy of simple decision tree vs random forest is not drastically different

	Misclassification Rate	
	Credit-Constraints	Credit-Worthiness
Decision Tree	31.85%	20.02%
Random Forest	32.09%	17.66%
Logistic LASSO	33.83%	18.55%

- **Suggestion:** Focus on the simple decision tree

Decision Tree vs. Random Forest

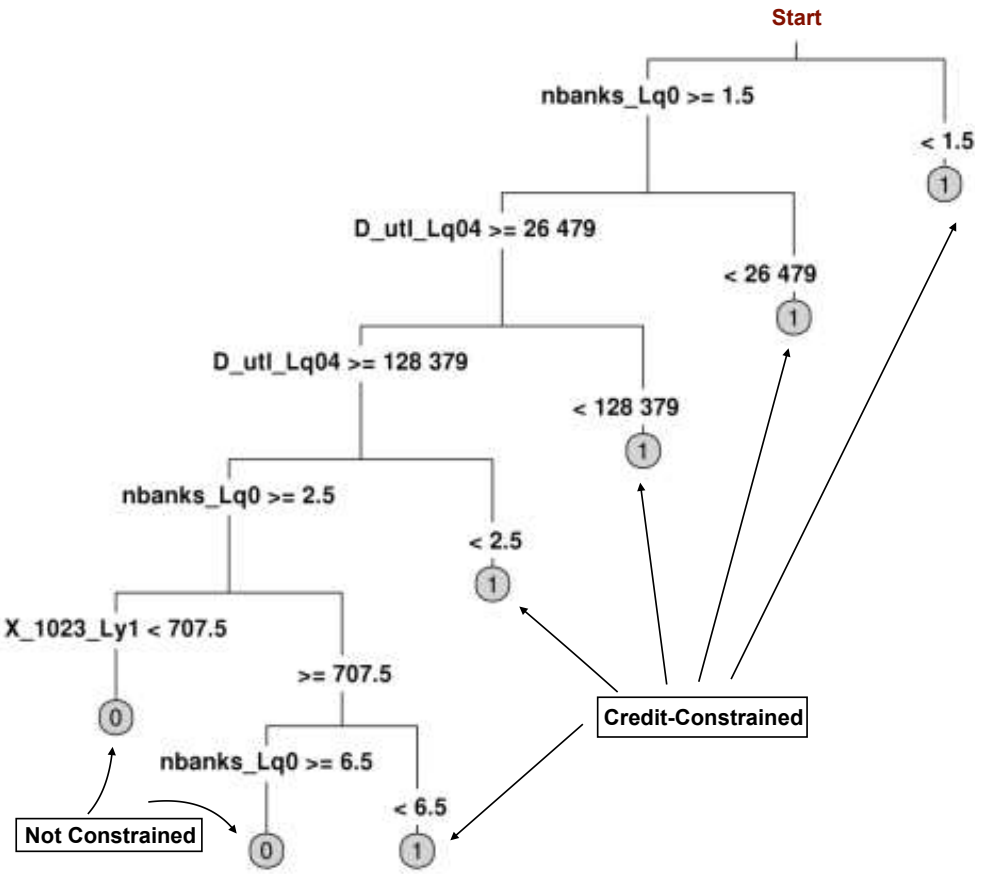

Figure: **Decision Tree: Credit-Worthiness Exercise.** `nbanks_Lq0`: number of banks lending in previous quarter, `D_utl_Lq4`: Change in total amount of granted loans in past year, `X_1023_Ly1`: long-term debts.

Making sense of random forests / boosted trees

- In general, predictive power of tree ensembles (random forests / boosted trees) is higher than that of simple decision trees
- Challenge: **Interpretation!**
- Current ML research is starting to address this issue:

Making Tree Ensembles Interpretable

Satoshi Hara SATOHARA@NII.AC.JP
National Institute of Informatics, Japan
JST, ERATO, Kawarabayashi Large Graph Project

Kohei Hayashi HAYASHI.KOHEI@GMAIL.COM
National Institute of Advanced Industrial Science and Technology, Japan

Abstract

Tree ensembles, such as random forest and boosted trees, are renowned for their high prediction performance, whereas their interpretability is critically limited. In this paper, we propose a post processing method that improves the model interpretability of tree ensembles. After learning a complex tree

rules for prediction. Non-expert people cannot understand such tremendous number of rules. A decision tree, on the other hand, is well known as one of the most interpretable models. Despite weak prediction ability, the number of regions generated by a single tree is drastically small, which makes the model transparent and understandable.

Obviously, there is a tradeoff between prediction performance and interpretability. Eto et al. (2014) proposed a

Further comments

- It would be interesting to see how a gradient boosted tree or a neural network performs in comparison to the techniques used. Particularly, since the sample is rather large in size, hence lending itself to treatment with a big data technique like neural networks.
- Joint prediction of credit-constraints and credit-worthiness would be nice. First attempts in the paper are quite promising (roughly the same accuracy). Might help in resolving the puzzle that firms of equal risk exhibit different credit rationing in current model.

Conclusions

- Interesting and carefully executed paper!
- ML will change policy making / assignment
- Transparency challenges will become less substantial as ML evolves