Discussion of "Machine Learning for Zombie Hunting: Firms' failures, financial constraints and misallocation"

Gias Uddin[1]

Bank of England Conference on "Modelling with Big Data and Machine Learning"

October 31, 2019

[1] Bank of Canada

Overview

Paper Summary - Data and Method

- Proposes to use machine learning techniques to measure firm delinquency probabilities and use the probabilities to identify zombie firms.
- **Data**
  - Orbis dataset: Financial accounts of manufacturing firms active in Italy for at least one year between 2008-2017. The sample consists of 304,869 Italian firms.
  - Assumes that a company failed in the first year if it is reported as being 'Bankrupt'/'Dissolved'/'In Liquidation'. The share of such exiting firms in the sample is about 5.7% of the entire sample.
- **Method.**
  - First, compute the probability of failure of each time. The value is between 0 and 1. A value of 1 for a firm means that the firm will surely fail. A value close to 1 denotes the firm is of high-risk.
  - Second, *zombie* firms are detected as those that showed a probability of at least 0.8 for three consecutive years.

Paper Summary - Results

- **Machine Learning Techniques**
  - A suite of machine learning techniques are investigated to compute the probability of firm failure: BART (Bayesian Additive Regression Tree), Conditional Inference Tree, Random Forest, Super Learner, and Logit.
  - The performance of each technique is compared against an out-of-sample dataset. The performance metric used is AUC (Area Under the Curve).
  - BART was found to be the best performer, because of its ability to handle missing data in the firm accounting reports.
- **Empirical Results**
  - The detected zombie firms are divided into five hazard levels based on the probability of failure, ranging from high-risk firms (active firms with probability above the 9-th decile for three consecutive years) to low-risk firms (active firms with probability between 8-th and 9-th decile for three consecutive years). The firms that are found in the five levels are called 'Persistently Distressed Firms (PDF)'.
  - The share of PDFs reached the maximum in 2014 (9.94%) and minimum in 2017 (5.71%).
  - There was a negative lagged correlation between the observed distribution of PDFs against the GDP growth in Italy. Therefore, the zombie firms were not the likely cause of financial crisis.

Strengths

The machine learning classifier BART shows good performance.

Table 4: Machine Learning techniques and out-of-sample prediction accuracy

Model\Accuracy	AUC	PR	Obs Train	Ors CV
Logit	0.8601	0.2570	118527	13170
Ctree	0.8450	0.2431	118527	13170
Rforest	0.8779	0.3331	118527	13170
Super Learner	0.8895	0.3364	118527	13170
BART	0.9498	0.6893	118527	13170

Sensitivity of Prediction Analysis

The authors put a lot of efforts to determine the robustness of the selected covariates to detect the target variable. While more firm-level data can be always desirable, the efforts from the authors show that the 40 covariates, i.e., financial indicators used for analysis are good indicators.

- The authors check the robustness of the classification by proposing to analyze the sensitivity in prediction analysis.
- The proposed technique first creates artificial covariates that show higher correlation with the target variable. The technique then trains two classifiers, one on original set of covariates and another on the original covariates + new covariate. The two techniques are trained on repeated samples and their outputs are analyzed to check for statistically significant differences. The difference was significant only 2.16% times.

Rooms for Improvement

Evaluation of Prediction - Metrics

- The dataset seems to highly imbalanced, i.e., around 93% firms are not delinquent. This means that a random model of assigning everyone 0 (i.e., not delinquent) would lead to at least 93% accuracy.
- The paper reports AUC curve to report the classifier performance. The AUC curve can be susceptible to such data imbalanced and thus is normally discouraged while reporting performance on imbalanced data[1].
- A more careful metric would be F1-score, which is calculated as:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{1}$$

  where *Precision* is the positive predictive ratio and recall is the sensitivity.
- Perhaps the authors need to explain more the Precision-Recall Goodness of Fit values.

[1] See for example `http://www.davidsbatista.net/blog/2018/08/19/NLP_Metrics/`

Evaluation of Prediction - Sample

- It was not clear how the out-of-sample data was created to report the predictors.
- In standard machine learning, cross-validation is normally used, where the full set of sample can be used both for training and testing.
- There is also the fixed separation of training, testing and validation sets.
- It's preferable to report the performance using cross-validation.

Evaluation of Prediction - Overfitting

- The best performing model in BART, which is Bayesian Additive Regression Trees, i.e., a set of weak learner (each tree) combined to form a strong prediction. By default such trees can grow quite large on the training data in their attempt to match satisfy each decision path, leading to their overfitting.
- The authors use their sensitivity of prediction checks to analyze the sufficiency of the selected features. However, it is not clear how the authors handle for overfitting on the selected features.
- Additional information on BART could help, such as tree levels, level cut-off thresholds, etc. As an example, check the gini impurity measures and the post and pre-pruning of trees in Decision trees.

Zombie Firm Detection

The paper needs both qualitative and quantitative analyses of the zombie firm detection strategy based on the sample.

- It is not immediately clear why 8-th and 9-th decile in firm delinquent probability could be used for zombie firms and not also, say, 7-th decile. What happens when you use 7-th to 9-th deciles?
- The three year window is also something that needs to be analyzed more carefully. Why not a different and more restrictive window, such as five years?
- While the probabilities are calculated based on all covariates, all of those covariates may not be applicable to a zombie firm. How about choosing a subset of covariates that most likely characterize zombie firms?
- Are there any existing benchmarks of zombie firms that can be used to validate the results?

Current Zombie Indicator Validation

- The paper uses LASSO to select the most informative features that could be useful for zombie firm identification.
- In Machine learning, a trained model should give such information by querying the model for feature importance. Typically, such importance is computed using information gain, i.e., sort the features based on how well they can divide the decision space in the sample.
- Can such information gain principles be applied on the best performing BART model to get insights on the most informative features?

Empirical Analysis of Zombie Firms - Share of PDFs

- The paper reports to have observed negative correlation between share of PDFs and GDP growth in Italy. How does the comparison work when the market cap of the PDFs vs sound farms are taken into account?
- Are the PDFs all small farms or are there big farms?
- Are the PDFs all startups, which are expected to struggle for survival during the first few years of their establishment?