Bankruptcy prediction in Colombian case

Using Decision trees, Random forest and Adaboost

Iván Andrés Trujilllo Abella

BIT

ivantrujillo129@gmail.com

Review

MDA model Altman 1968, Olhson 1980 (probit and logit, Survival models 1992, Hazard model (Shumway, 2001), and today the Neural networks, has been applied considerably in the field of bankruptcy. The models used to measure the risk, has been evolute to tackle empirical limitations, however the models more used are MDA, logit model by simplicity in spite of better perfomance of other models. (Smaranda, 2014)

Theorical framework

There is not a defined concept about bankruptcy. Notwithstanding, for empirical practice, has been consider the declaration for themselves for the legal entities, and the financial accounts made a important reference, to find patters or associations to predict the event.

Justification

The financial distress is a term related with the situation of company insolvency, usually the term refer to the impossibility of payments, or its legal declaration of default, this situations have a negative impact in the economy, as reducing employment and raising prices. In summary, it could lead to reduce output and the improvements in living conditions of population.

Goals

Determine the accuracy of the decision tree join the bagging and boosting techniques to predict bankruptcy.

- Identifying the Colombian companies that underwent bankruptcy in a longitudinal database with the retrieved information from 2012-2014.
- Estimate the True Positive Rate and Accuracy of the decision tree, random forest and adaBoost model to predict bankruptcy.

Data

Retrieved from Super Intendencia de Sociedades the variables are *capital work/total assets, retained earnings/assets total, net utility/total assets and equity total liabilities* however, either limitations in model or the variables we could get low prediction as consequence but only estimate a pure Altman model due to be comparable results with another works.

Decision trees

k nodes that are constructed with the gain information of a measure of impurity, in this case entropy (spite of there is not the only measure, the other measures not produce significance difference).

Random Forest

Constructed over the concept of boostraping, of **X** select choose n samples with the same length and with replacement and ensemble with the following criteria:
$h(X)$ it is a classification rule, then Majority Vote Classifier is defined as

$$C(x) = mode\{h(X_1)..h(X_n)\} \tag{1}$$

we can assign weights:

$$C(x) = arg \max_i \sum_{j=1}^{N} w_j I(h_j(X) = i) \tag{2}$$

Note that if $w_j = \frac{1}{N}$ then is equal to the mode defined previously.

Adaboost

Constructed over the concept of weak learners, boosting propose learning of error in each training iteration, thus assume that we initialize with $\mathbf{w} = \frac{1}{N}$ the number of data points thus we train the model updating the weights and after normalize in each iteration thus each weight that correspond with a data point classify correctly then $\Delta w_j < 0$ and the miss classify data points $\Delta w_j > 0$.
in general terms we hope a better performance of the Adaboost technique, because unlike random forest that only tackle variance problems, this techniques reduce bias too.

Assessment of models

$$TPR = \frac{TP}{FP + TP} \tag{3}$$

The accuracy (ACC) is a measure of the overall prediction ability.

$$ACC = \frac{TP + TN}{TP + TN + FN + TP} \tag{4}$$

Results

In training data 786 firms were classify as $+1$ and 24040 as -1. We perform a bivariate analysis of each financial ratio and found significant difference with the groups in medians due the variables not are distributed normally the test used was wilcoxon rank sum test. To test model in 2014 roughly the same size of classes were used but the failures were roughly the halve of training.
According to the following table, there not significative difference among the models, however Adaboost outperform the another two techinques.

Table: Perfomance of prediction

Model	TPR	ACC
Decision Trees	0.963	0.962
Random Forest	0.998	0.997
AdaBoost	0.999	0.999

References

1. Ulku, H. RD, Innovation, and Economic Growth; An Empirical Analysis. IMF Work. Pap. (2004).

2. Edward I, A. Financial ratios, disciminant analysisi and the prediction of corporate Bankruptcy. J. Finance 23, 589–609 (1968).

3. Beaver, W. H. Financial Ratios As Predictors of Failure. J. Account. Res. 4, 71 (1966).

4. Son, H., Hyun, C., Phan, D. Hwang, H. J. Data analytic approach for bankruptcy prediction. Expert Syst. Appl. 138, 112816 (2019).

5. Alaka, H. A. et al. Systematic review of bankruptcy prediction models: Towards a framework for tool selection. Expert Systems with Applications vol. 94 164–184 (2018).

6. Hosaka, T. Bankruptcy prediction using imaged financial ratios and convolutional neural networks. Expert Syst. Appl. 117, 287–299 (2019).

7. Qu, Y., Quan, P., Lei, M. Shi, Y. Review of bankruptcy prediction using machine learning and deep learning techniques. in Procedia Computer Science vol. 162 895–899 (Elsevier B.V., 2019).

8. Shi, Y. Li, X. A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms. Heliyon vol. 5 e02997 (2019).

9. González García, L. M., Viga Juárez, C. A. Fierro Martinez, S. D. Prospección del riesgo operativo de las Mipymes en Colombia. Suma Negocios 8, 79–87 (2017).

10. Romero Espinosa, F. Determining financial variables in the business failure to small and medium enterprises in Colombia: analysis on Logit model. Pensam. Gestión 235–277 (2013).

11. Pérez García, J. I., Lopera Castaño, M. Vásquez Bedoya, F. A. Estimation of bankruptcy risk probability in Colombian companies from a model for rare events. Cuad. Adm. 30, 7–38 (2017).

12. Pérez G., J. I., González C., K. L. Lopera C., M. Modelos de predicción de la fragilidad empresarial: aplicación al caso colombiano para el año 2011. Perf. Coyunt. Económica 205–228 (2013).

13. Martinez, O. Determinantes de fragilidad en las empresas colombianas — Banco de la República (banco central de Colombia).

14. Berrío Guzmán, D. Cabeza de Vergara, L. Verificación y adaptación del modelo de ALTMAN a la Superintendencia de Sociedades de Colombia. Pensam. y gestión Rev. la Div. Ciencias Adm. la Univ. del Norte 26–51 (2003).

15. Gómez-González, J. E. Hinojosa, I. P. O. Un modelo de alerta temprana para el sistema financiero Colombiano. Ensayos Sobre Polit. Econ. 62, 123–147 (2010).

16. Gómez, N. E. Z. Determinantes de la Probabilidad de Incumplimiento de las Empresas Colombianas. Borradores Econ. (2007).

17. Rivillas, C. S., Gutiérrez, W. R. Betancur, J. C. G. Credit risk estimation for companies in the manufacturing industry in Colombia. Estud. Gerenciales 28, 169–190 (2012).

18. Rosillo, J. Modelo de predicción de quiebras de las empresas colombianas. Innovar Rev. Ciencias Adm. y Soc. 12, 109–124 (2002).

19. Wang, M. et al. Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction. Eng. Appl. Artif. Intell. 63, 54–68 (2017).