Leveraging Ensemble Models in Credit Scoring of Japanese Small and Medium Companies

Michiko Miyamoto (Akita Prefectural University)

Masakazu Ando (Chiba Institute of Technology)

Credit scoring models have used logistic regression and linear discriminant analysis for identification of potential defaults. Ensemble models combine two or more models are enable a more robust prediction, classification, or variable selection. Ensemble models are the product of training several similar models and combining their results in order to improve accuracy, reduce bias, reduce variance, and provide robust models in the presence of new data. We will describe three types of ensemble models: boosting, bagging, and model averaging.

Chopra and Bhilare (2018) conduct empirical analysis on publicly available bank loan dataset to study banking loan default using decision tree as the base learner and comparing it with ensemble tree learning techniques such as bagging, boosting, and random forests. The results of the empirical analysis suggest that the gradient boosting model outperforms the base decision tree learner, indicating that ensemble model works better than individual models. Guo, et al. (2019) presents a novel multi-stage self-adaptive classifier ensemble model based on the statistical techniques and the machine learning techniques to improve the prediction performance and applied to credit scoring to test its prediction performance. Their results show that compared to single classifier and other ensemble classification methods, the proposed model has better performance and better data adaptability.

We will conduct empirical analysis on Japanese Small and Medium size companies using ensemble models. The data, we use, provides information on each firm's balance sheet and income statement; its default information; the firm's characteristics, including industrial classification, and demographic characteristics of each firm's primary owner, including age, real estate ownership, whether he/she has a successor or not,. Those data were consisting of nearly 480,000 Japanese SMEs after omitting missing values, collected by credit guarantee corporations, as well as government-affiliated and private financial institutions involving SME business in 2010.

References

Chopra, A. and Bhilare, P. (2018). Application of Ensemble Models in Credit Scoring Models, Business Perspectives and Research, https://doi.org/10.1177/2278533718765531

Guo, S., He, H. and Huang, X. (2019). A Multi-Stage Self-Adaptive Classifier Ensemble Model With Application in Credit Scoring. IEE Access, Volume7, 78549-78559.