

年報文字基礎溝通價值與公司信用 評等預測：機器學習模型之應用

陳宗岡 廖咸興 官穎 林育均 郝昀*

在以往信用評等預測相關文獻中，大多以統計模型來進行信用評等的分類預測，並財務特徵變數作為主要的解釋變異來源。然而，根據標準普爾的信用評等準則可知，企業信用評等乃先以企業經營風險及財務風險決定一初值後，再經由多面向的非量化因素調整而得，是故財務變數並無法捕抓信用評等資訊的全貌。不同於以往文獻，本研究以機器學習模型（隨機森林、支持向量機、極限梯度提升、羅吉斯回歸）為基礎，並在既有財務會計特徵變數為標竿設定下，額外引入年報文字基礎溝通價值變數（e.g. 可讀性與語意）來捕抓信評公司所考量的非量化調整因素，特別是不完全會計資訊的部份。本研究利用美國市場 1994 年至 2017 年的公司信用評等資料來進行分析，實證結果顯示：在額外投入年報文字基礎溝通價值變數後，模型預測效力（e.g. F1 分數）均有一定程度的提升，其中又以隨機森林及極限梯度提升模型總體表現最好（F1 分數可達 0.76~0.77，增額提升約 6%）。此顯示年報文字基礎溝通價值資訊對信用評等

「政策與管理意涵」

本研究發現年報文字基礎溝通價值變數與傳統財務變數兩者對信用評等的分類預測效力具有互補性。如對信用品質較差的公司而言，年報文字基礎溝通價值變數所能額外補抓的信用風險資訊效果較財務變數為多。對主管當局來說，應更留意信用品質較差公司的年報文字基礎溝通價值。此亦可作為未來年報文字表達規範建議之政策參考依據。

* 通訊作者：郝昀，國立陽明交通大學管理科學系博士生，通訊地址：新竹市東區大學路 1001 號管理一館 212 室，電話：(03)5712121 ext 57124，E-mail: peterhao.c@nycu.edu.tw；陳宗岡為國立陽明交通大學管理科學系教授暨國立臺灣大學計量理論與應用研究中心特約研究員，E-mail: vocterchen@nycu.edu.tw；廖咸興為國立臺灣大學財務金融學系教授，E-mail: hliao@ntu.edu.tw；官穎為國立陽明交通大學管理科學系碩士，E-mail: kuanhao.861029@gmail.com；林育均為國立陽明交通大學管理科學系學生，E-mail: magic.mg09@nycu.edu.tw。

作者感謝期刊主編鍾惠民教授、兩位匿名審查人、以及 2023 年臺灣財務金融學會年會的評論人葉宗穎教授及其他與會者提供的寶貴建議。作者感謝教育部高等教育深耕計畫補助國立臺灣大學計量理論與應用研究中心之研究經費（編號 112L900201）。

陳宗岡 廖咸興 官顥 林育均 郝昀

有不同於傳統財務變數的額外解釋能力。此外，本研究亦發現年報文字基礎溝通價值資訊能進一步降低非投資級公司被誤判為投資級公司的比率，亦即對非投資級公司的評等分類有更高的預測效力。因此，本研究驗證年報文字基礎溝通價值變數可捕抓到信評公司的非量化因素調整之資訊內涵。

關鍵詞：年報文字基礎溝通價值、信用評等、機器學習、不完全資訊。

Text-Based Communicative Value of Annual Reports and Corporate Credit Rating Predictions: Using Machine Learning Models

Tsung-Kang Chen

Department of Management Science, National Yang Ming Chiao Tung University
& Center for Research in Econometric Theory and Applications,
National Taiwan University

Hsien-Hsing Liao

Department of Finance, National Taiwan University

Hao Kuan

Department of Management Science, National Yang Ming Chiao Tung University

Yu-Chun Lin

Department of Management Science, National Yang Ming Chiao Tung University

Yun Hao*

Department of Management Science, National Yang Ming Chiao Tung University

Different from the previous literature, this study employs American firm credit rating data from 1994 to 2017 to explore whether additionally introducing the text-based communicative value (TCV) variables of annual reports (e.g. readability and tones) improves the effectiveness of credit rating predictions based on the machine learning models only with financial characteristic variables as input ones. Empirical results show that after additionally introducing the TCV variables of annual reports,

* Corresponding author. Yun Hao, E-mail: peterhao.c@nycu.edu.tw; Tsung-Kang Chen, E-mail: vocterchen@nycu.edu.tw; Hsien-Hsing Liao, E-mail: hliao@ntu.edu.tw; Hao Kuan, E-mail: kuanhao.861029@gmail.com; Yu-Chun Lin, E-mail: magic.mg09@nycu.edu.tw.

We are greatly indebted to Prof. Huimin Chung (the editor), two anonymous reviewers, and the conference discussant (Professor Chung-Ying Yeh) and participants at 2023 International Conference of Taiwan Finance Association for their insightful suggestions and comments on earlier drafts of the paper. We gratefully acknowledge the financial support from the Center for Research in Econometric Theory and Applications (Grant no. 112L900201) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

the model prediction effectiveness has a certain improvement, and the random forest and XGBoost models perform the best overall (e.g. F1 score reaches 0.76-0.77, which increases by about 6%). This shows that the TCV information of annual reports has an additional explanatory power for credit ratings compared to the traditional financial variables. In addition, this study also finds that the TCV information of annual reports can further reduce the ratio of non-investment grade firms being misjudged as investment grade ones. That is, TCV information of annual reports has the greater predictive power for non-investment grade firms. Therefore, this study verifies that the TCV information of annual reports captures the information contents of the non-quantitative factor adjustments of credit rating agencies.

Key Words: Text-based communicative value of annual reports, Credit rating, Machine learning, Incomplete information.