Trading by Charts

A Multivariate and Multimodal CNN System to Predict Retail Investor Trading

Chaojie (Jay) Liu
PhD Researcher
University of Bristol Business School

Jay.liu@bristol.ac.uk
+44 7491 598135
+86 15611761856

Abstract

Understanding retail investor behavior is interesting and important, but difficult; the prediction of retail investor trading is even harder. This study sheds some light on predicting retail investors' holding on stocks by building an accessible time-series prediction system using the convolutional neural network (CNN) technique for pre-COVID and COVID-19 periods. A multivariate and multimodal CNN is built for the first time with both numerical and graphical data. Focusing on the U.S. stock markets between 2018 and 2020, this study utilises component stocks of the S&P 500 index as the sample with relative data on stock characteristics, retail investor holdings, and retail investor ownership. The pioneering CNN system performs great (with R^2 from 0.8 to 0.95) in predicting aggregate retail investor trading behaviour and outperforms the random forest models built, which only apply numerical data. The results support previous studies on the performance of deep learning techniques like CNN and investor trading behaviour and sentiment. Besides, retail investor holding contains little predictive information for stock price movement. This study contributes to the economic and financial literature by filling the gap in the predictions of retail investor behavior using cutting-edge machine learning techniques based on novel applications of data. In addition, this prediction system can improve social welfare by helping retail investors make less biased decisions, informing financial institutions to better engage with retail investors, and assisting financial authorities to better monitor and manage risks caused by retail investors in the market.

Keywords: Retail Investment; Investor Behaviour; Machine Learning; Deep Learning; Time-Series Prediction

JEL Codes: G11; G17; G41; G50

Notes: This work benefits a lot from the discussions with Dr. Manuela Pedio, Prof. Vincent Han, Prof. Chris Brooks, and all conference participants at the Alan Turing Institute, PhD Writing Circle at the University of Bristol, 4th Digital Economy Forum of the China Society of World Economics, 9th PKU-NUS Annual International Conference on Quantitative Finance and Economics, 2025 International Symposium on Forecasting, and 2025 International Doctoral Forum at CUFE. All mistakes are my own.