

Distinguished Lecture Series

How much can machines learn finance from Chinese text data?



10 December 2021 (Friday)

11:00 a.m. - 12:00 noon GMT+8 (Hong Kong Time)



Online via Zoom


(Meeting ID: 973 9570 8109)



ABSTRACT

Most studies on equity markets using text data focus on English-based specified sentiment dictionaries or topic modeling. However, can we predict the impact of news directly from the text data? How much can we learn from such a direct approach? We present here a new framework for learning text data based on the factor model and sparsity regularization, called FarmPredict, to let machines learn financial returns automatically. Unlike other dictionary-based or topic models that have stringent pre-screening processes, our framework allows the model to extract information more fully from the whole article. We demonstrate our study on the Chinese stock market, as Chinese text has no natural spaces between words and phrases and the Chinese market has a very large proportion of retail investors. These two specific features of our study differ significantly from the previous literature that focuses on English-text and the U.S. market. We validate our method using the literature on the Chinese stock market with several existing approaches. We show that positive sentiments scored by our FarmPredict approach generate on average 83 bps stock daily excess returns, while negative news has an adverse impact of 26 bps on the days of news announcements, where both effects can last for a few days. This asymmetric effect aligns well with the short-sale constraints in the Chinese equity market. As a result, we show that the machine-learned sentiments do provide sizeable predictive power with an annualized return of 116% with a simple investment strategy and the portfolios based on our model significantly outperform other models. This lends further support that our FarmPredict can learn the sentiments embedded in financial news. Our study also demonstrates the far-reaching potential of using machines to learn text data.

Sponsored by:

 Centre for Mathematical Imaging and Vision



Joint Research Institute for Applied Mathematics

 SRCC

Statistics Research & Consultancy Centre

Professor Jianqing Fan

Princeton University

Jianqing Fan, is a statistician, financial econometrician, and data scientist. He is Frederick L. Moore '18 Professor of Finance, Professor of Statistics, and Professor of Operations Research and Financial Engineering at the Princeton University where he chaired the department from 2012 to 2015. He is the winner of The 2000 COPSS Presidents' Award, Morningside Gold Medal for Applied Mathematics (2007), Guggenheim Fellow (2009), Pao-Lu Hsu Prize (2013) and Guy Medal in Silver (2014), Noether Senior Scholar Award (2018) and the election to the fellow of American Association for the Advancement of Science, Institute of Mathematical Statistics, and American Statistical Association. He also got elected to Academician from Academia Sinica in 2012. Fan is interested in statistical theory and methods in data science, statistical machine learning, finance, economics, computational biology, biostatistics with particular skills on high-dimensional statistics, nonparametric modeling, longitudinal and functional data analysis, nonlinear, survival analysis, time series, wavelets, among others. He has been consistently listed among the top 10 highly-cited mathematical scientist since the existence of such a ranking.