Technology Review

CS410: Text Information Systems

University of Illinois at Urbana-Champaign

Author: David Kim. davidk11@illinois.edu

Professor: ChengXiang Zhai

Natural Language Processing in the Stock Market

Introduction

This paper explores research and applications of natural language processing in the stock market. Artificial Intelligence (AI) in general has become a hot topic in finance and one of the most affected areas in finance is asset management (Bartram et al., 2020). Natural Language Processing (NLP), a subset of AI, while has smaller research interest than other AI methods, has still also become increasingly popular, along with other machine learning techniques such as artificial neural networks, decision trees and random forests, evolutionary algorithms, and support vector machines (Bartram et al., 2020).

Applications of Natural Language Process in the Stock Market

Investors have traditionally employed financial models, such as Discounted Cash Flow (DCF) or Dividend Discount models, when forecasting future stock prices. These models are typically based on numerical data such as interest rate and income statements contained in financial reports submitted by companies. While these financial models are widely used on Wall Street, they do not capture all available information, especially qualitative and textual information.

Models using only numbers have been used on Wall Street traditionally, but now with advances in the field of text retrieval and text mining, NLP will be incorporated more in asset management.

Applications of NLP includes text mining and sentiment analysis from social media, websites, new articles, and corporate SEC filings (Bartram et al., 2020). With the phenomena such as the short squeeze of GME in 2021, investors and traders have started to monitor and mine social media to potential trading opportunities (Xu, 2021). One of reasons for the increase in application and research popularity is that there are plenty of text data from websites, social media, and filings (Bartram et al., 2020). Fisher et. at. (2016) has found that NLP is used to text to detect frauds, make inferences, and forecast asset prices.

Research in applications of NLP to financials markets have been existed for over two decades. For example, Fawcett and Provost (1999) have postulated a relationship between stock price and news articles, and Lavrenko et al. (2000) similarly researched about a language model and stock forecasting.

Schumaker and Chen studied the impact of financial news on different textual representation: Bag of Words, Noun Phrases, and Named Entities, and found that their SVM model performed better at predicting stock prices than traditional linear regression.

Conclusion

As AI techniques, especially NPLs, are increasingly becoming employed in asset management, investors will be able complement the current methods that mostly use numerical data with more qualitative analysis embedded in text data. For example, Kloptchenko et al. (2002) used both quantitative and qualitative parts of their financial reports to predict asset prices.
This will help in areas such as portfolio management and risk management.

References

Bartram, Söhnke M. and Branke, Jürgen and Motahari, Mehrshad, *Artificial Intelligence in Asset Management* (September 14, 2020). CFA Institute Research Foundation Literature Reviews, August 2020, ISBN 978-1-952927-02-7, Available at SSRN: https://ssrn.com/abstract=3692805 or http://dx.doi.org/10.2139/ssrn.3692805

Xu, Muxi, NLP for Stock Market Prediction with Reddit Data (2021), available at https://web.stanford.edu/class/cs224n/reports/final_reports/report030.pdf

Fung, Gabriel & Yu, Jeffrey & Lam, Wai. (2002). News Sensitive Stock Trend Prediction. Advances in Knowledge Discovery and Data Mining. 481-493. 10.1007/3-540-47887-6_48.

T. Fawcett and F. Provost. Activity Monitoring: Noticing Interesting Changes inBehavior. In Proceed ing s of th e 5th I ntern ati ona l Con fere nce on K DD, San Diego,California, 1999.

V. Lavrenko, M. Schmill, D. Lawire, P. Ogilvie, D. Jensen and J. Allan. Min-ing of Concurrent Text and Time Series, In Proceeding s o f the 6th I nte rna ti ona lConference on KDD, Boston, MA, 2000

Schumaker, Roobert P. and Chen, Hsiinchun, Textual Analaysis of Stock Market Prediction Using Financial News Articles, The University of Arizona, Available at https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.333.3366&rep=rep1&type=pdf