Neurocomputing
Volume 264, 15 November 2017, Pages 20-28

Stock portfolio selection using learning-to-rank algorithms with news sentiment

Qiang Song, Anqi Liu, Steve Y. Yang

https://doi.org/10.1016/j.neucom.2017.02.097
Get rights and content

Abstract

In this study, we apply learning-to-rank algorithms to design trading strategies using relative performance of a group of stocks based on investors' sentiment toward these stocks. We show that learning-to-rank algorithms are effective in producing reliable rankings of the best and the worst performing stocks based on investors' sentiment. More specifically, we use the sentiment shock and trend indicators introduced in the previous studies, and we design stock selection rules of holding long positions of the top 25% stocks and short positions of the bottom 25% stocks according to rankings produced by learning-to-rank algorithms. We then apply two learning-to-rank algorithms, ListNet and RankNet, in stock selection processes and test long-only and long-short portfolio selection strategies using 10 years of market and news sentiment data. Through backtesting of these strategies from 2006 to 2014, we demonstrate that our portfolio strategies produce risk-adjusted returns superior to the S&P 500 index return, the hedge fund industry average performance - HFRIEMN, and some sentiment-based approaches without learning-to-rank algorithm during the same period.

Introduction

Impacts of investors' sentiment to financial market have been well documented in a number of behavioral finance studies [3], [9], [12], [25], [33]. However, since investors' sentiment is an abstract concept, researchers have put a lot of effort in finding adequate and reliable proxies to represent its underlying mechanism. These proxies include mutual fund flows, closed-end funds prices, news and social media messages aimed to extract investors' attitude towards financial market [20], [33]. In recent years, due to its broad digital accessibility, news media has been increasingly becoming a critical source of information that supports investors' investment decisions. The arrival of news continually updates investors' understanding and knowledge of the market and influences investor sentiment as a result. Most notably, the primary financial data vendors, such as Bloomberg and Thomson Reuters, now publish news analytic solutions to traders and investors in an almost real-time fashion. According to the enhanced news feeds delivery, many recent studies have focused on parsing news articles to calibrate investor sentiment through text mining and machine learning techniques [24], [29], [37].

In this paper, we employ a machine learning ranking method, learning-to-rank algorithms, to construct equity portfolios based on news sentiment. Many machine learning algorithms have been used for financial market prediction and trading strategy development [11], [17], [19], [23], [34], these efforts have been mostly focused on constructing portfolios based on return forecasting. We propose a stock portfolio construction approach from the viewpoint of ranking investors' relative views on stocks' performance. Learning-to-rank is a class of algorithms that apply supervised machine learning approaches to solve ranking problems, and it is a task to automatically construct a ranking model using training data, such that the model can sort new objects according to their degrees of relevance, preference, or importance [22]. We argue that many investment or trading decisions can be naturally cast into a ranking problem. For example, to identify which assets will outperform in the future. Making an investment decision is essentially to rank all the assets in the investable universe and buy the top ones or short the bottom ones. With increasing data availability, it is natural to leverage machine learning algorithms for such ranking problems. Machine learning algorithms are proven to be effective in fitting parameters automatically, avoiding over-fitting, and being capable of combining multiple inputs. Ranking investors' sentiment hence provides a natural way to select stocks based on the "portrayed performance" in news media.

The advantages of using learning-to-rank algorithms in portfolio selection are twofold. First, unlike the traditional machine learning algorithms that predict a value for one input based on past information, learning-to-rank provides ranks for a group of inputs. In other words, learning-to-rank targets on relative orders instead of absolute values. This property is particularly useful for investment applications. Secondly, we can combine different performance indicators, and the parameters tuning can be done automatically. Traditionally, a ranking model is built without training. Items are ranked by either relevance or importance, such as Boolean model [2], BM25 [28] and PageRank [27]. These conventional ranking models require parameter tuning and face over-fitting problems. In addition, combining different models to get better results is not straightforward. Learning-to-rank algorithms are particularly suitable for investment decisions. Unlike traditional machine learning algorithms, which only make one prediction at a time, learning-to-rank predicts a ranked list. This approach presents a relative performance among stocks, and it is more reliable than an absolute performance forecast. For the majority of the money managers, their task is not to achieve absolute return, but to outperform a specific benchmark. Moreover, there are often a plethora of indicators for predicting stock returns, and sometimes they give conflicting signals. The learning-to-rank algorithms can automatically combine a range of indicators and weigh them to produce optimal results.

The primary contribution of this study is to demonstrate the outperformance and robustness of a relative performance based portfolio construction method: learning-to-rank algorithms in designing trading strategies using a relative view of investors' sentiment. It bridges the gap of predicting relative performance for a group of stocks through multiple information sources. We argue that learning-to-rank algorithms are effective in producing reliable ranking models to predict the best and the worst performing stocks based on investor sentiment and market information. In a previous study, we designed sentiment shock and trend indicators to investigate firm-specific stock price movements and generate trading signals for individual stocks. This trading strategy is based on the hypothesis that ranking of news sentiment reflects expected returns during the near future. In this study, we use the sentiment shock and trend indicators introduced in previous studies [31], [38] to develop stock selection rules of holding long positions of the top 25% stocks and short positions of the bottom 25% stocks according to the stock rankings produced by learning-to-rank algorithms. In the experiments of portfolio strategies, we apply ListNet and RankNet in stock selection processes and test long-only and long-short strategies. Applying backtesting of the models for the period from 2006 to 2014, we show that the selected portfolios using our learning-to-rank methods have superior profitability to the S&P 500 index. Moreover, the long-short strategies produce robust Sharpe ratio in both high volatility and low volatility regimes. This study demonstrates two key features of the proposed methods. First, learning-to-rank algorithms produce accurate predictions of expected return rankings, and the proposed stock selection approach works robustly under different financial market conditions. Secondly, the sentiment indicators [31], [38] support ranking predictions consistently in reflecting individual stock's future performances under different market conditions.

The rest of the paper is organized as follows. In the next section, we review existing studies on sentiment analysis and application of machine learning methods to investment decisions. Section 3 introduces sentiment shock and trend indicators, ListNet and RankNet algorithms, and the stock selection approach. Section 4 presents data sources of market information and news sentiment along with their statistical relationships. Section 5 describes an application of portfolio strategy with the proposed stock selection process. Section 6 discusses the key findings of the portfolio management experiments based on learning-to-rank technique. Section 7 concludes the discussion and proposes some future work.

Section snippets

Trading with financial news sentiment

Financial sentiment has been widely explored in both academic and industrial research. To evaluate investor sentiment, researchers have used a variety of information sources to form sentiment proxies. Financial news has been one primary source of investor sentiment analysis. In general, research on financial news impacts to markets targeted on explaining two questions: (1) Does news information lead financial market activities? (2) Can special patterns of news sentiment form indicators that…

News sentiment indicators

The correlation between news sentiment and market return indicates that sentiment can be an indicator to market movements. However, this time series property may not be effective on cross-sectional analysis of a group of assets. Absolute sentiment scores do not accurately represent relative performance of different stocks. Therefore, we examine the structure of sentiment time series and design two sentiment indicators. The sentiment shock scores normalize firm-specific sentiment based on the…

Data

Market data and financial news sentiment data are obtained from Bloomberg terminal and Thomson Reuters News Analytics (TRNA), respectively. We collect data from January 2003 to December 2014.…

Sentiment indicators parameter optimization

To determine the look-back window N in sentiment shock and trend indicators, we apply Spearman rank correlation (see Eq. (9)) to measure quality of predicting subsequent return ranks. Higher rank correlation means stronger prediction power. As assigning one parameter for each stock is computationally intensive, we categorize the stock universe according to GICS sectors and run an optimization for each sector. The optimization rule is to maximize Spearman rank correlation between sentiment shock …

Results

We run experiments of long-only and long-short strategies using ListNet and RankNet algorithms. The backtest period is from 2006 to 2014, which covers a high volatility regime during the 2008 financial crisis and the economic recovery period from 2011. To justify the performance of our strategies, we choose S&P 500 index as a benchmark for comparison. All these strategies, especially the two using ListNet algorithm for stock selection, outperform the benchmark according to their higher…

Conclusion

In this paper, we design a stock portfolio selection approach utilizing learning-to-rank algorithms on two news sentiment indicators to capture relative performance of stocks. ListNet and RankNet algorithms are applied in the experiments of portfolio strategies. Our data include features, such as historical market returns, financial news sentiment, and sentiment shock and trend scores. Our methodology combines this information to derive ranking models that facilitate stock selection. The…

Qiang Song received his Bachelor degree in Electronic science and technology from Shandong University, China. He is currently a Ph.D. candidate majoring in Financial Engineering at Stevens Institute of Technology, US. His research interests include algorithmic trading strategy, applications of machine learning algorithms in finance and financial news analytics. He also holds a CFA level 3 certificate. He has interned at several financial asset management firms. He is currently working for Citi…

References (39)

N. Barberis *et al.*
A model of investor sentiment
J. Financ. Econ. (1998)

M.A. Dempster *et al.*
An automated fx trading system using adaptive reinforcement learning
Expert Syst. Appl. (2006)

A. Khadjeh Nassirtoussi *et al.*
Text mining of news-headlines for FOREX market prediction: a multi-layer dimension reduction algorithm with semantics and sentiment
Exp. Syst. Appl. (2015)

KimK.-j.
Financial time series forecasting using support vector machines
Neurocomputing (2003)

LeeC.M.C. *et al.*
Investor sentiment and the closed-end fund puzzle
J. Financ (1991)

MaoH. *et al.*
An automatic news analysis and opinion sharing system for exchange rate analysis
Proceedings of the 2014 IEEE 11th International Conference on e-Business Engineering (2014)

A. Shashua *et al.*
Ranking with large margin principle: two approaches
Adv. Neural Inf. Process. Syst. (2003)

WuQ. *et al.*
Adapting boosting for information retrieval measures
Inf. Retr. (2010)

W. Antweiler *et al.*
Is all that talk just noise? The information content of Internet stock message boards
J. Financ. (2004)

R. Baeza-Yates *et al.*
ModernInformation Retrieval
(1999)

View more references

Cited by (67)

Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory
2023, Expert Systems with Applications

Show abstract

A comprehensive decision support system for stock investment decisions
2022, Expert Systems with Applications

Show abstract

The role of investor attention in global asset price variation during the invasion of Ukraine
2022, Finance Research Letters

Citation Excerpt :

…Future scope of the research may contain the study of volatility spillover effect among markets, as our results showed that for 36 countries, conflict attention variable is significant, which can be caused by contagion, similar to studies Pericoli and Sbracia (2003), Diebold and Yilmaz (2009, 2012), Lyócsa et al. (2020a) and Okorie and Lin (2021). Our findings may also be applied in the context of portfolio asset allocation, which utilizes methods of Google Trends as Kristoufek (2013) and Maggi and Uberti (2021) or with adding the information regarding sentiment, the methods like Dumas et al. (2009), Song et al. (2017) and Yu et al. (2022). Finally, our findings may prove to be useful in the domain of option pricing as a better volatility estimate leads to a more appropriate option price (Cretarola et al., 2020).…

Show abstract

Learning to rank complex network node based on the self-supervised graph convolution model
2022, Knowledge-Based Systems

Citation Excerpt :

…The higher the score, the greater the relevance to the query. Learning to rank has been successfully applied in many fields, such as collaborative filtering [17], object detection [18], sentiment analysis [19], etc. Approaches of learning to rank can be divided into three categories: pointwise approach, pairwise approach, and listwise approach.…

Show abstract

Stock portfolio selection balancing variance and tail risk via stock vector representation acquired from price data and texts
2022, Knowledge-Based Systems

Show abstract

Predicting abnormal trading behavior from internet rumor propagation: a machine learning approach
2023, Financial Innovation

View all citing articles on Scopus

Recommended articles (6)

Research article

A study of contagion in the financial system from the perspective of network analytics
Neurocomputing, Volume 264, 2017, pp. 42-49

Show abstract

Research article

The repo market in Mexico: Empirics and stylized facts
Neurocomputing, Volume 264, 2017, pp. 2-19

Show abstract

Research article

A local adaptive learning system for online portfolio selection
Knowledge-Based Systems, Volume 186, 2019, Article 104958

Show abstract

Research article

Machine learning in finance—Guest editorial
Neurocomputing, Volume 264, 2017, p. 1

Research article

Genetic programming optimization for a sentiment feedback strength based trading strategy
Neurocomputing, Volume 264, 2017, pp. 29-41

Show abstract

Research article

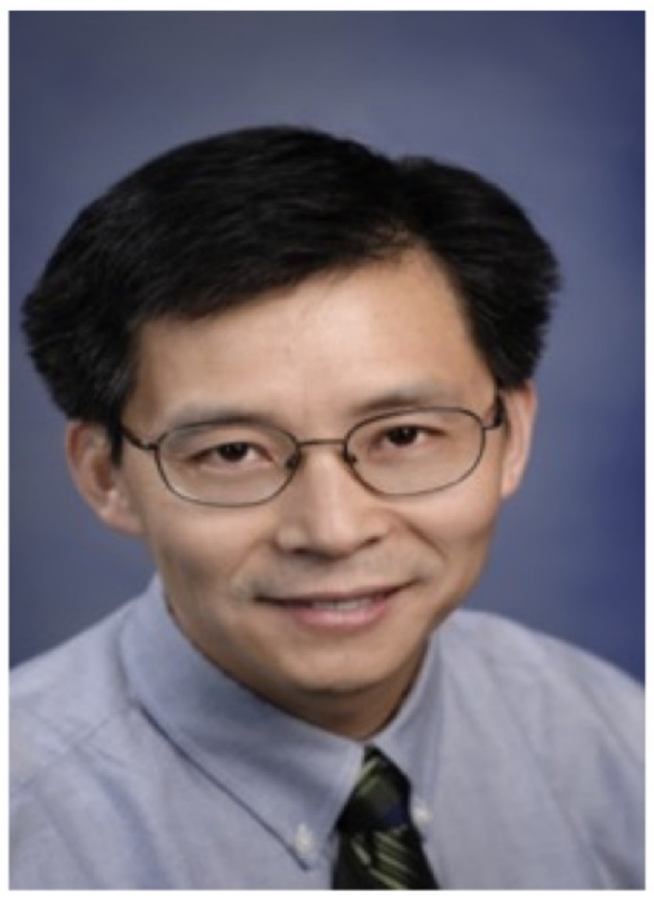
An investor sentiment reward-based trading system using Gaussian inverse reinforcement learning algorithm

Expert Systems with Applications, Volume 114, 2018, pp. 388-401

Show abstract

Qiang Song received his Bachelor degree in Electronic science and technology from Shandong University, China. He is currently a Ph.D. candidate majoring in Financial Engineering at Stevens Institute of Technology, US. His research interests include algorithmic trading strategy, applications of machine learning algorithms in finance and financial news analytics. He also holds a CFA level 3 certificate. He has interned at several financial asset management firms. He is currently working for Citi Group in the area of financial analytics in NY.

Anqi Liu received a Bachelor of Science degree in applied mathematics from Northwest University, PR China in 2011, and a M.S. degree in financial engineering from Stevens Institute of Technology in 2013. She is currently a Ph.D. candidate in the School of Systems and Enterprises, Stevens Institute of Technology. Her research interests include behavioral financial, sentiment analysis and agent-based artificial market simulation. She has been collaborating with a number of financial researchers in the area of quantitative finance and computational finance, and has published a series of papers in international journals and conferences. In addition, she received Stevens Innovation and Entrepreneurship Scholarship for 4 years.



Steve Y. Yang is an assistant Professor of Financial Engineering at the School of Systems and Enterprises in Stevens Institute of Technology. He holds a Ph.D. in Systems and Information Engineering from University of Virginia with concentration on Financial Engineering. His current research interests include behavioral finance and its application in trading strategies, modeling algorithmic trading behavior, financial trading and information fraud detection, and agent based financial market simulation. Dr. Steve Yang was elected as a Distinguished Visiting Scholar by the U.S. Securities and Exchange Commission in 2016. As an expert in studying High Frequency Trading strategies, he serves as a Research Fellow at the Division of Enforcement at the Commodity Futures Trading Commission in Washington DC. Dr. Steve Yang has a number of impactful academic papers published in Quantitative Finance, Journal of Banking & Finance, Environment Decision Systems, etc. He is currently an active committee member of the Computational Financial and Economics Technical Committee in IEEE Computational Intelligence Society.

View full text