A novel statistical and machine learning hybrid approach to predicting S&P500 using sentiment analysis

Fionn Murtagh, Rapheal Olaniyan and Daniel Stamate

Department of Computing,
Goldsmiths, University of London

Abstract

The frequent ups and downs are characteristic to the stock market. The conventional standard models that assume that investors act rationally have not been able to capture the irregularities in the stock market patterns for years. As a result, behavioural finance is embraced to attempt to correct these model shortcomings by adding some factors to capture sentimental contagion which may be at play in determining the stock market. The authors address the predictive influence of online expressed sentiment on the stock market returns and volatility by using a non-parametric nonlinear approach that corrects specific limitations encountered in previous approaches. A novel approach to developing sentiment analysis and stock market predictive models based on GARCH, EGARCH and recurrent neural network frameworks is presented, and is compared to previous statistical and/or machine learning approaches addressing this problem, proving its advantages and superiority over the latter. The sentiment information extracted via text mining from online blogs includes variants of indexes expressing relevant sentiment, in particular anxiety, whose predictive value on the dynamic of S&P 500 is rigorously analysed using linear and nonlinear Granger causality and Monte Carlo simulations. Future extensions envisage incorporating the necessary apparatus and efficient mechanism to handle also stream data.