Investigating the Effect of Sentiment in High-Frequency Financial Markets

Darragh Mc Kay

Supervisor: Professor Khurshid Ahmad
MAI (Computer Engineering) 2018

Abstract

High-frequency trading (HFT) is a financial trading strategy that is growing in popularity among major trading firms. At the turn of the 21st century, HFT trades had an execution time of several seconds, whereas by 2010 this had decreased to milli- and even microseconds. This recent growth in HFT is paralleled by a massive increase in user-generated social media data, where every day hundreds of thousands of opinions are shared in the form of discussion, reviews and social media statuses. This research aims to leverage the rapid growth in social media, specifically Twitter, to investigate the effect of sentiment in high-frequency financial markets.

This thesis presents a system that automatically retrieves and aggregates firm-specific Twitter data for the purpose of sentiment analysis. Leveraging computational linguistics, it computes a domain-specific affect dictionary from the corpus of tweets for use during sentiment analysis. It then extracts a negative sentiment time series using the Rocksteady affect analysis system at a daily, 5-minute and 1-minute frequency. The resultant time series are then aggregated with firm-specific financial stock returns at their respective frequencies. Finally, the time series are examined using four different vector autoregressive (VAR) models, which individually examine the independent effect of different sentiment variables on returns.

The analysis of the system, using Ryanair as a case study, found that sentiment extracted from social media, specifically Twitter, plays a small but significant role in explaining the returns of an asset but its explanatory power varies at different frequencies. The research shows that the VAR models fit better at a daily frequency than at high-frequency due to the volatile nature of the time series. However, a week-by-week analysis shows that high-frequency sentiment extracted from tweets contributes significantly to modelling the returns in certain situations, in particular during periods of consistent volatility. Thus, negative sentiment found in tweets can be very appropriate for use in high-frequency trading strategies but the circumstances under which it is significant need to be thoroughly examined in future work.