Predicting altcoin returns using social media

Lars Steinert¹, Christian Herff

Motivation

• Increasing public and scientific interest in cryptocurrencies
• Previous focus primarily on Bitcoin
• Lack of similar approaches
• Smaller cryptocurrencies are even easier to influence
• Detection of “pump and dump”-schemes

Materials

• Data collection via CryptoCoinCharts and Twitter API using official currency codes e.g. "$LTC" for Litecoin
• Dataset containing prices and the social media activity of 181 altcoins in the form of 426,520 tweets
• Data collection over 71 days in 3-hour intervals:
  • 45 days or 333 samples per cryptocurrency for training
  • 26 days or 190 samples per cryptocurrency for testing

Pre-processing

• Calculation of short-term returns for the upcoming 24 hours due to less autocorrelation
• Filtering out those altcoins that were referred to on Twitter on at least 10 % of all days led to 131 altcoins that were considered in the subsequent steps
• Sentiment analysis using VADER

Methods

• OLS Linear regression analysis to predict short-term returns based on mean VADER scores and number of tweets
• Focus on those five altcoins displaying the highest mean R² value over all time lags
• Model evaluation on test data
• Application of Bonferroni correction on results

Results

• Statistically significant results for 16 out of 40 predictions
• Wide range of results in terms of size of selected altcoins
• No unique picture regarding R² values or time lags with significant results
• Studies related to stock market prediction provide much lower R² values

¹ lars.steinert@uni-bremen.de