Predicting the stock market using Twitter

Complex Systems Seminars, January 22, 2015
A question

Is it possible to predict the price of stocks by analysing a lot of tweets?

"tweet":

*a short message of maximum 140 characters posted on Twitter*
Example tweet

ML_tweet.png
Twitter mood predicts the stock market
Johan Bollen, Huina Mao, Xiaojun Zeng
Journal of Computational Science, 2011
Outline

- Introduction
  - Twitter
  - Stock market
- Components of the method
  - Granger causality analysis
  - Mood tracking tools
  - Fuzzy neural networks
- Results and discussion
Twitter

- Launched in 2006
- 500 million users
- Hundreds of millions of tweets / day
- Twitterbots
Introduction

Twitter

	twitterstart.jpg
"Wall Street" = New York Stock Exchange (NYSE)
daily trading: $ hundreds of billions
Dow Jones Index Average (DJIA) = 30 companies
Introduction

Stock market

djiashort.jpg
Motivation

Why?

- Stock prices do not perform a random walk?
- Emotions play a significant role in human decision-making
- Surveys expensive and time-consuming
Method

method1.jpg
Components of the method
Granger Causality Analysis

- Test if mood time series has predictive information about Dow Jones
- Linear model
- Mood correlates with Dow Jones, 3 days lag
Tweets are analysed and converted into a mood

Train a neural network with data from mood indicators and past stock index to predict future stock index

Trade on mood, not news
Components of the method
Mood indicators

- Extract data from Twitter into 2 different lexicons
- 1. OpinionFinder
  - Count positive and negative keywords
  
  Example: **happy** = positive keyword
Components of the method
Mood indicators

2. Google-Profile of Mood states (GPOMS)
   - n-gram = phrase of n connected words
   - 6 mood dimensions: Calm, Happy, Vital, Sure, Kind, Alert
   - Example: **Veni vidi vici** = Sure 3-gram

→ In total 7 mood time series

Mood validated with presidential election and Thanksgiving
Components of the method

Fuzzy systems

- Simple, interpretable
- Demands linguistic rules
- Membership function
Components of the method

Fuzzy systems example: student classification
Components of the method

Fuzzy systems example: student classification

table.png
Components of the method

Fuzzy neural networks (FNNs)

- Learning ability of neural network
- Tunes the fuzzy system
  - Input: mood and past stock index
  - Output: future stock index
- Non-linear model
Components of the method

Fuzzy neural networks (FNNs)
Method revisited
Results

- Baseline prediction = past 3 days of stock prices
  - 73 % accuracy up/down (!)

- 1. OpinionFinder (positive/negative) $\rightarrow$ no improvement
2. GPOMS, mood dimension **Calm** best improvement for direction
   - 87 % accuracy up/down

**Calm** + **Happy** $\rightarrow$ smallest Mean Average Percentage Error

Non-linear relation confirmed
Discussion

- Correlation does not imply causation, a philosophical question
- More advanced model incorporate news and other indicators
- At the time of study, Twitter mostly in the US
- Introduction of strategies leads to changes in trader behaviour
Is it possible to predict the price of stocks by analysing a lot of tweets?

**According to the article, yes. But...**

- "Secrets" in methodology
  - GPOMS Lexicon
- Unclear results
  - Overfitting?