Forecasting stock markets with Twitter

Argimiro Arratia
argimiro@lsi.upc.edu

Joint work with **Marta Arias** and **Ramón Xuriguera**

LARCA. Laboratory for Relational Algorithmics, Complexity and Learning
UNIVERSITAT POLITÈCNICA DE CATALUNYA

To appear in: *ACM Transactions on Intelligent Systems and Technology*, 2013, as part of larger study by the title "Forecasting with Twitter data".

Motivations and Goals

Does a public Sentiment Indicator extracted from daily Twitter messages can indeed improve the forecasting of social, economic, or commercial indicators, based on time series models?

Answer:

Experiment with all possible machine learning models reinforced with a Sentiment Index built from Twitter data, and compare their performance when trained with and without the Twitter index.

Ref	Event	Models	Corpus	Conclusion
[Wolfram 2010]	NASDAQ stocks	SVM	Edinburgh Corpus, English, Relevant to stocks	Works with high freq. data. No sentiment analysis, but direct count of frequency of words
[Zhang et al. 2010]	DJIA, S&P500, NASDAQ, VIX	n/a	English, with mood keywords	Finds correlations of tweet's emotions (hope, fear, worry) and the direction of the DJIA stock index.
[Bollen et al. 2011]	DJIA	SOFNN	~10M tweets, Stock market prices	An index of the calmness of the public is predictive of the DJIA and predictions can be significantly improved using a SOFNN.
[Mishne and Glance 2005]	Movie sales	n/a	Blog posts with links to IMDB, IMDB sales data	Considering the sentiment of blog posts improves the correlation between references to movies and their financial success.
[Asur and Huberman 2010]	Movie sales	Linear regression	~2.9M tweets for 24 movies	The model built with the tweet rate time series outperforms the baseline that uses the Hollywood Stock Exchange (HSX).
[O'Connor et al. 2010]	U.S. polls	n/a	10^9 tweets (omitting non-English), Public opinion polls	The evolution of Twitter sentiment correlates to periodical public polls on the presidential election and on the presidential job approval.
[Tumasjan et al. 2010]	German 2009 election	Logistic Regression	100K tweets	Additional information is not provided to predictive models. Only a comparison of the share of voice and the election results.
[Gruhl et al. 2005]	Book sales	Custom *Spikes* predictor	Blog posts, Amazon sales rank	Correlation detected between the number of blogs refering to a book and its sale spikes.
[Wakamiya	TV Ratings	n/a	Japanese tweets	No predictions or correlations are

Some technical difficulties

There are no public data sets available and Twitter have imposed several restrictions on retrieving on-line posted tweets.
As of April 2010 Twitter forbids 3rd parties to redistribute tweets.
Hence we have to create our own data set, with limitations.
Begun on 22 March 2011. Use a Streaming API:

- One HTTP connection is kept alive to retrieve tweets as they are posted
- Filter the stream by keyword or user

The Twitter Gold Mine

You can BUY data from official Twitter reseller. Very, very expensive!

The Twitter-Hedge fund

July 2011: Derwent Capital, a UK based hedge fund, in partnership with Bollen et al. began trading a $40 Million Hedge Fund using the Twitter Predictor.
http://www.derwentcapitalmarkets.com/
http://mashable.com/2011/05/17/twitter-based-hedge-fund/

How we did it

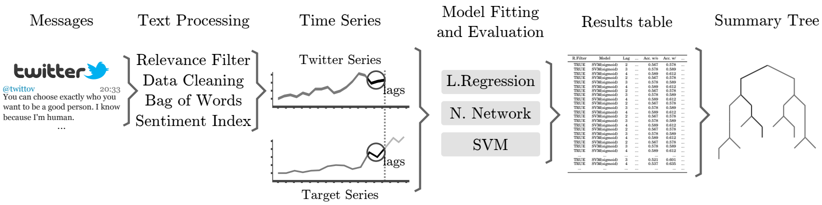
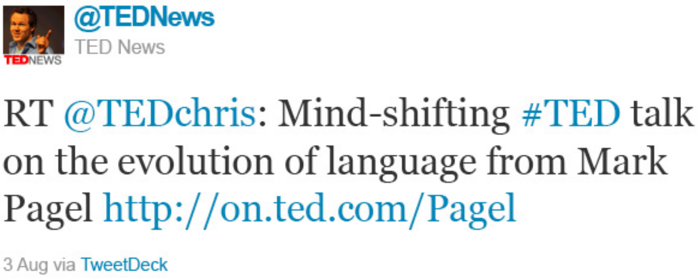


Figure: Overview of data collection, preprocessing, forecasting and final analysis processes.

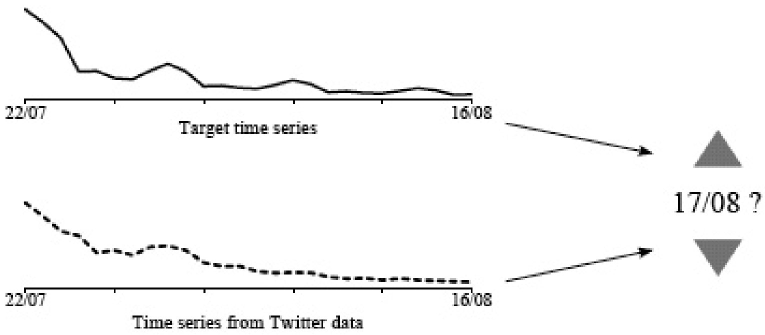
Twitter



- Online service that allows to build social networks based on microblogging
- Messages or tweets up to 140 characters
- Special or reserved symbols: @ for replying; # (hashtags) for topics assignment or keywords; RT (*retweet*) for sharing with followers; URLs.

Twitter social indicators implemented

- Our **Twitter** based indicators: **Volume** and **Sentiment**
- Hypothesis:
  (Volume) More messages $\mapsto$ more variance (Volatility)?
- Hypothesis:
  (Sentiment index) negative/positive $\mapsto$ decrease/increase benefits (returns)?



Sentiment Analysis

Goal: Determine whether a message contains positive or negative impressions on a given subject

$\swarrow$ $\searrow$

| Subjectivity Recognition | | Polarity Detection |

- Begin with creating a labelled corpora for supervised training a sentiment classifier
- We apply a recent pictorial tagging idea [Bifet et al, 2010]: Create a dataset of tweets that are automatically labelled positive if contains a smiley of the form: `:-)` ,`;-D` or negative if contains `:-(`.
  (Formally, consider all regular expressions from `[:=8][ -]?[)D]` for positive, and from `[:=8][ -]?(` for negative.)

Sentiment Analysis

- Expand the training datasets by Feature Lists of words: classify by frequent words with at least 5 occurrences, and topic (e.g. a stock's ticker).
- Clean the data previously:
  - Remove duplicates (mostly in retweets);
  - Remove stopwords (e.g. pronouns, prepositions, many verbs);
  - Stemming (reduce words to their roots by removing suffixes);
  - Negation handling (use tagging: not good $\mapsto$ NOT_good);
- Relevance filtering: text containing some of the keywords is no guarantee of its relevance for the subject we want to classify. E.g.

  "I really love eating an apple"

  "Apple stock soared above $404 today"

Sentiment Classifiers (SC)

- Multinomial Naïve Bayes: Given D the set of all docs. and C set of class labels, a Naive Bayes classifier will assign a doc. d the class with the highest conditional probability given that document. (d is represented as n-dim. vector $(w_1, \ldots, w_n)$ of Bernoulli-distributed variables indicating whether word w_i occurs in d). The naive assumption is that the presence of a particular feature of a class is unrelated to the presence of any other feature.
- We only focus on binary classification (positive/negative)
- 3 classes of data sets to train different SCs: English, Multi-language, Stock. Variations of SC obtained by training on these data sets with different preprocessing, feature words to represent docs, etc.
- Best scoring SC extracted: *C-En*, *C-Ml*, *C-Stk*.
- Accuracy: 76.49% for *C-En*, 79.5% for *C-Ml*, 76% for *C-Stk*

Sentiment Index (or Twitter series)

Sentiment: A time series consisting of the daily percentage of positive tweets (over the total number of tweets posted) for each top-scoring SC.

Volume: time series of daily number of tweets concerning a subject.

Financial Time Series

Goal

To predict stock's **return** or **volatility**

Focus on following companies and indices:
AAPL, MSFT, GOOG, YHOO
S&P100 (OEX), S&P100's implicit volatility (VXO), S&P500 (GSPC), S&P500 implicit volatility (VIX).

Returns

- Computed from Adjusted Close
- Log–normally distributed
- Log returns

Volatility

- Computed from log returns
- Exponential Weighted Moving Average

$$V(t_n) = (1-\lambda)\sum_{i=1}^{m} \lambda^{i-1} R_{n-i}^2$$

Forecasting time series

Models for prediction

- Linear Regression $y = \sum_{i=0}^{n} w_i x_i$
- Neural Networks (feedforward) $y_j = \varphi\left(\sum_{i=0}^{n} w_{ij} x_{ij}\right)$
- Support Vector Machines (with kernel polynomial, radial or sigmoid)

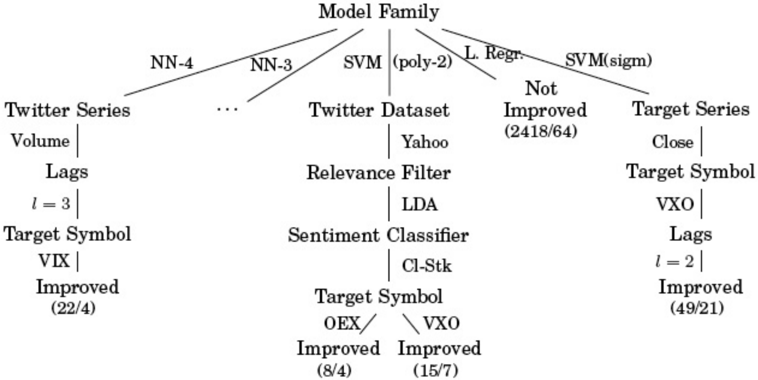
Model Assessment

- Nonlinearity: we use White neural network test for neglected nonlinearity to assess the possible nonlinear relationship between two time series
- Causality: apply Granger test of causality, both parametric and non-parametric, to the two time series and for different lags.
- Prequential evaluation: Our data is from short period, we cannot spare an independent set for training. Use past data to update/retrain the classifier and predict the direction of the time series for the following day.

Combining all different parameters give us approx. 39000 experiments.

Summary trees

Decision trees built with `REPTree` (Weka), by greedy selection of attributes that give a higher performance gain. The tree give an indication of the attributes that most influent the increase on accuracy.



Experimental Results (General)

- Big failure of linear models (only improve 2.5% of the time: 64 improved/ 2418 not improved)
- Nonlinear models with either Twitter series improve as predictors of the trend of volatility indices (VXO, VIX) and historic volatilities of stocks.
- Predicting the trend of benefits is more dependent on the parameters, the input data and Twitter classifier. Best case is SVM with poly-2 kernel for predicting OEX with sentiment index obtained from C-Stk classifier.

Experimental Results (Specific)

Success rates by model family for predicting the VIX index using Tweet Volume and $lags = 3$.

Model family	Successful	Unsuccessful	Success rate
Neural Networks	100	35	74.07%
Support Vector Machines	103	121	45.98%

Success rates of SVM by kernel type for predicting the VXO index when $lags = 2$, plus either Twitter series.

Kernel Type	Successful	Unsuccessful	Success rate
Polynomial Kernels	104	192	35.13%
Radial	5	70	0.07%
Sigmoid	68	7	90.67%

WARNING! Don't play with your home account

The Twitter-hedge fund crash

October 2011: Derwent Capital closes shop with reported returns of 1.86% after a month of trading with Bollen et al Twitter predictor, and is for sale today:
http://www.derwentcapitalmarkets.com/auction/

The only machine model used in Bollen et al is a "Fuzzy Logic Neural Network" and they only care about price direction disregarding volatility. We were not able to test such machine (the program for it was of course not available), but in general Neural Network score low in our tests for predicting direction of returns for stocks (and the results of such experiments are very parameter-dependent).

This is the END