DSCS 6030 - Research Project Proposal

Your Name
Date

Stock Price Prediction of household name organizations post product launches using sentiment analysis

Household name organizations and other popular organizations receive a lot of attention from the press and social media during product launches such as the launch of an apple device, a Tesla car or a Microsoft product. This results in the online availability of tweets, blogs and press coverage. This material is useful for extracting the general sentiment of the public, further these sentiments may translate into change in stock prices of the organizations. It is the objective of this research project to analyze the public sentiment post product launches by organizations that are household names such as Apple, Samsung, Tesla, Ford, Microsoft etc. to predict the extent to which product launches might impact their stock price in the short-run.

Background

The valuation of a stock is highly complex and is determined by a great number of parameters. However, market sentiment is one of the many non-statistical parameters that influences the price of stocks in the short run and this project would use sentiment analysis of tweets, posts and news articles as the predictor in change to stock prices. The project would establish a positive correlation between social media sentiments and market sentiments post product launches by organizations that are household names or popular and further forecast the range within which the stock price might fluctuate post the product launch.

Data Sources

This project requires two data resources:

1. Text based online posts and news articles for sentimental analysis.
2. Finance data from some reputable websites that would be used to train and test the Machine learning algorithms.

APIs and R packages that have currently been identified for use are:

1. 'mlbench' package for Naïve Bayes algorithm
2. Twitter public REST API for public tweets from twitter. https://dev.twitter.com/rest/public
3. 'Quantmod Package for R' for financial data from Yahoo Finance. http://www.quantmod.com/
4. 'Rbbg Package for R' for financial data from Bloomberg. http://findata.org/rbloomberg/
5. APIs for news articles from, http://www.programmableweb.com/category/News%20Services/apis?category=20250

Machine Learning Algorithms

The following algorithms have been specifically identified for sentimental analysis.

1. Naive bayes - BernoulliNB, GaussianNB, MultinomialNB
2. Support Vector Classifiers - LinearSVC, PolynomialSVC, RbfSVC, NuSVC
3. Maximum Entropy Model - GIS, IIS, MEGAM, TADM

The choice of algorithm(s) requires further evaluation.

Project Evaluation

A set of past product launches will be identified for which online posts, news articles and stock prices before and shortly after the event would be compiled. A portion of the data set would be used to train the model and the remainder of the data-set would be used to test the model. The difference between past events that were used to test the model against the events used to train the model would provide an evaluation of how well the research project succeeded.

References:

1. http://www.rinfinance.com/agenda/2012/talk/Nagar+Hahsler.pdf
2. http://www.forbes.com/sites/davidleinweber/2013/04/24/so-much-for-fund-mining-twitter-sentiment-for-picking-stocks-but-ok-at-the-sec/#68d3857927ff
3. https://www.quora.com/What-are-the-best-supervised-learning-algorithms-for-sentiment-analysis-in-text
4. http://www.thertrader.com/2013/11/08/financial-data-accessible-from-r-part-iii/
5. https://eight2late.wordpress.com/2015/11/06/a-gentle-introduction-to-naive-bayes-classification-using-r/
6. http://stocktwits.com/research/Predictability-of-stock-market-behavior-using-stocktwits-sentiment-and-posting-volume_NunoOliveira.pdf
7. https://www.quora.com/Can-Twitter-sentiment-analysis-guide-stock-market-investment