Impact of Social Media on Stock Price

Kexin Hui, Dongbo Wang, Zhaoheng Hu
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
{khui3, dwang49, hu61}@illinois.edu

Abstract

This document serves as the proposal of our final project for ECE 544NA Fall 2017. Our team has three members and our project focusing on discovering the impact of social media on stock price.

1. Introduction

There are numerous stories about someone made a huge fortune overnight through investing on stock market, which draws tremendous attention from people trying to make money. Yet, it is always hard to tell what is going to be a good time to invest on the stock market, simply because there are too many factors that potentially have huge impacts on the behaviour of the market. Nowadays, the influence of social media is undoutedly of particular importance. People like to post everything on social media platforms like Facebook and Twitter, not along to those great men who run a Fortune 500 company or are nominated as the Worlds Billionaires. What they posted will more or less influence the company stock price they are related to. For example, on August 13th 2013, billionaire Carl Icahn, of Icahn Capital Management, announced his Apple position over Twitter with the accompanying comment of it being significantly undervalued. Within seconds Apple stock spiked and within minutes gained $17 billion in market cap. This cannot just be a coincidence. Therefore, at this point, we want to explore more about the impacts of social media on stock market. The focus is on technology stocks, which attracts a lot of attention and exposure on the social media. We start from Teslas stock TSLA(NASDAQ) and tweets of Elon Musk, CEO of Tesla Inc.

2. Proposed Method

In the scope of this project, our main goal is to interpret the relationship between for example twitter posts and the corresponding stock price a certain time after the posts were made. We propose the following methods to handle certain tasks for the project.

2.1. Signal Processing

The stock price [3] of certain company is basically a 1-dimensional signal in time domain. To extract features out of this signal, we might want to apply some simple signal processing method to make the trend more clear or reduce some random patterns in between.

2.2. NLP

The task of extracting features out of the posts on social media falls naturally into the topics of Neuro-linguistic programming (NLP) [1]. We will apply some simple model or methods from NLP to extract meaningful features from the posts.

2.3. Deep Learning

After the features are extracted, the majority of the learning work will be done be deep learning. We planned to build some neural network [2], train and tune it and see if it is capable of learning the subtle relationship between posts and stock price.

3. Dataset Description

To dig into the impacts of social media on stock market, we will use the Elon Musk Tweets dataset, and the Tesla Stock Price dataset from Kaggle. As is known to all, Elon Musk is the CEO of Tesla Inc. and what he tweeted everyday may more or less pose some influence on the stock price of Tesla. As a consequence, it's meaningful for us to figure out the structure and content of each dataset so we can fully make use of them to train a proper model which can tell us the impact of Elon Musk's tweets on Tesla's stock price.

After carefully reading the explanation and example of datasets, we have a comprehensive understanding about these datasets.

The first one is Elon Musk's tweets in past 7 years. It contains text and relevant information of all tweets which are posted by Elon Musk, CEO of Tesla, during 2010 to

2017. Mainly, the following features will be used in our project:
1. Tweet id: representing the unique id of each tweet, it contains tweet-stamp.
2. Timestamp: the date and time of the day that corresponding tweet was posted, each record is in 24hr representation format.
3. Tweet text: the text of tweets, additionally, b is removed.

Secondly, Teslas stock price in past 7 years is another essential dataset in the project. This dataset is telling us the historical price start from 2010, the IPO (initial public offering) year of the incorporation, to early 2017. There are several important features provided:
1. Date: the date of each price record.
2. Open: the opening price of the stock.
3. High: the high price of that day.
4. Low: the low price of that day.
5. Close: the closed price of that day.
6. Volume: The amount of stocks traded during that day.
7. Adj close: The stocks closing price that has been amended to include any distributions/corporate actions that occurs before next day open.

4. Proposed Experiments

Currently, since the available data resources are Elon Musks tweets and historical stock prices from 2010 to 2017. We will choose first five years data (2010 - 2015) to train the model and use data of 2016 for validation. Finally, data generated in 2017 will be applied for testing. For future work, if we have enough time and could find other incorporations CEOs posts and stock prices, we will perform training by using whole dataset of Elon Musk and Tesla to get a model and apply this model to other incorporations and their CEOs to test. In this way, we will probably be able to find the connection between CEOs social media activities and stock price of the company.

5. Resource Feasibility

The following data are available:
1. On twitter posts: `https://www.kaggle.com/kingburrito666/elon-musk-tweets`
2. On stock price: `https://www.kaggle.com/rpaguirre/tesla-stock-price`

The data for certain stock price is easily reachable, either from multiple dataset available through Kaggle, or could be obtained from Google stock api. The text data for twitter should be reachable as well. There are tools available on the internet, but we need to investigate further more and be careful no to break any regulations.

Weeks	Steps
$1 \sim 2$	Data reading and cleaning
$2 \sim 3$	Feature Extraction via NLP and DSP
$2 \sim 3$	Apply Deep Learning and Training
$1 \sim 2$	Validation and Testing
$0 \sim 1$	Finalization and Documentation

Table 1. Tentative timeline and necessary steps

6. Tentative timeline and the necessary steps

We propose the tentative timeline (approximately 11 weeks) and necessary steps for this project, as shown in table 1. There are still details that we need to nail down for the project, so the contents of this table is subject to changes.

References

[1] W. J. Corvey. Twitter in mass emergency: what nlp techniques can contribute. *Association for Computational Linguistics*, 2010.

[2] X. Ding. Deep learning for event-driven stock prediction. *Ijcai*, 2015.

[3] Kim, Kyoung-jae, and I. Han. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert systems with Applications*, 19(2):125–132, 2000.