TITLE OF THE THESIS

PREDICTION OF STOCK MARKET TRENDS BY SENTIMENT FARMING USING OPINION MINING

A

RESEARCH PROPOSAL

SUBMITTED TO THE

SHRI RAMDEOBABA COLLEGE OF ENGINEERING AND MANAGEMENT,

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

IN

(COMPUTER SCIENCE)

By

ARTI VISHRAM BUCHE

Registration No <TBD>

UNDER THE GUIDANCE OF

DR. M. B. CHANDAK

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SHRI RAMDEOBABA COLLEGE OF ENGINEERING AND MANAGEMENT,
NAGPUR, MAHARASHTRA - 440015
Year 2015
About this proposal
This doctorate research proposal document describes the working title of the research proposal and general overview of the area. The research plan mentioned in this document may be modified based on the approval on this documented research proposal.
CONTENTS

1. RESEARCH TITLE........................................................................................................ 1
2. ABSTRACT.................................................................................................................... 1
3. LITERATURE SURVEY.............................................................................................. 2
4. PROBLEM DEFINITION............................................................................................. 4
5. PROPOSED METHODOLOGY.................................................................................... 4
6. APPLICATIONS......................................................................................................... 6
7. REFERENCES............................................................................................................. 8
1. Research Proposal Title

Prediction of Stock Market Trends by Sentiment Farming using Opinion Mining

2. Abstract

Stock market has become one of the major components of economy in developed countries. News and stock prices are closely related and news usually has a great influence on stock market investment. Stock market prediction is an area of extreme importance to an entire industry. Stock price is determined by the behavior of human investors, and the investors determine stock prices by using publicly available information to predict how the market will act or react. Financial news articles can thus play a large role in influencing the movement of a stock as humans react to the information.

Making decision in stock market is not really easy because a lot of factors are involved with every choice we make. Therefore, a lot of analysis is required to make an optimal move on stock market which may involve price trend, market's nature, company's stability, different news and rumors about stocks etc.

Recently, the web has rapidly emerged as a great source of financial information ranging from news articles to personal opinions. Text mining technology is becoming increasingly important as a means of rapidly analyzing massive texts in order to derive significant information intelligently. Text mining and analysis of such financial information can aid stock market predictions.

The objective of this study is to extract fundamental information from relevant news sources and use them to analyze or sometimes forecast the stock market from the common investor's viewpoint. We surveyed the existing business text mining researches and proposed a framework that uses our text parser and analyzer algorithm with an open source natural language processing tool to analyze (machine learning and text mining), retrieve (natural language processing), forecast (compare with historic data) investment decisions from any text data source on stock market.

We show that our method can be used to understand unstructured data and we also reveal that news sentiment can be used to make predictions about stock price fluctuations, whether up or down as well as stock market movements.
3. Literature Review and related work

3.1 Opinion Mining
Opinion mining, as a sub-discipline with text mining and computational linguistics, is referred to as the computational techniques use to extract, classify, understand, and assess the opinions expressed in various online news sources, social media comments, and other user-generated content. ‘Sentiment’ analysis is often used in opinion mining to identify sentiments, subjectivity and other emotional states in the online text.

In order to extract more accurately opinions and sentiments from a text, it is very important to build up a lexicon of opinion mining. If a lexicon such as a sentimental word dictionary is developed properly, the results of opinion mining will be good. Furthermore, sentimental word dictionaries are more effective when domain specific characteristics are taken into consideration.

A basic task in opinion mining is classifying the polarity of a given text at the document, sentence, or feature/aspect level - whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.[1]

3.2 Stock market prediction using news
Many works over the years have continued to prove that the news is closely related to stock prices. In particular, with the recent explosive increase in the amount of unstructured text data from the internet, mobile channels, and SNS (Social Network service), there have been attempts to predict stock movements using such text data. Using the stock prediction system NewsCASTS (i.e. the News Categorization and Trading System, which consists of three engines, namely, news pre-treatment, categorization and trading), analyzed media news on specific companies, and experimented with a comparison between news and stock price flows. [2]

News contents are one of the most important factors that have influence on market. Considering the news impact in analyzing the stock market behavior, leads to more precise predictions and as a result more profitable trades. So far various prototypes have been developed which consider the impact of news in stock market prediction. The challenge of textual financial prediction is to process the large amounts of textual information which exist for securities. The material not only includes
required reports such as periodic SEC (Securities and Exchange Commission) filings, but also a wealth of financial news articles reporting unexpected events and routine news alike. Financial news articles can be automatically capitalized on by using Natural Language Processing (NLP) and text-processing technique. [3]

### 3.3 Prediction Methods

#### 3.3.1 Polarity of text

To exactly predict the stock price is very complex task till the date. A prediction based system was proposed by on news articles using one of the Text Mining concepts like sentiment analysis can be proposed for Indian Stock market. Implementation steps to be followed to make a prediction system are:

1. Gathering of news articles.
2. Perform sentiment analysis on news articles
3. Get Polarity of the text
4. Make a prediction based on current stock price and calculated polarity of the text. [4]

#### 3.3.2 Linear Regression

Regression predicts a numerical value. Regression performs operations on a dataset where the target values have been defined already. And the result can be extended by adding new information. The relations which regression establishes between predictor and target values can make a pattern. This pattern can be used on other datasets which their target values are not known. Therefore the data needed for regression are two parts, first section for defining model and the other for testing model. First, the data is divided data into two parts of training and testing. Then the training section is used for starting analysis and defining the model. Scatter plot of 80% out of data with taking this into consideration that the (Average) parameter is the mean of the prices of Open, Low, High and close. [5]
4. **Problem Definition**

One area of limited success in Stock Market prediction comes from textual data. Information from quarterly reports or breaking news stories can dramatically affect the share price of a security.

There are two main problems investigated in stock market prediction, one is how to perform opinion mining using feature extraction and other is obtaining structured event information from large-scale news streams. Event prediction using bag of words approach can be used to predict the two problems of stock market prediction.

5. **Methodology**

**Quantitative Research Methodology**

The computational prediction of security markets follows two distinct paths, the first of which parallels information markets where an artificial market is constructed and predictions are made by varying system inputs. These inputs could be as simple as varying the time in which new information is received and acted upon market exchange. The second computational security prediction type is that of a quantitative nature. In real market predictions, quantitative systems, or quants, follow various stock parameters and are essentially automated versions of existing market strategies (e.g., look for high growth, unvalued securities, etc.) except with the ability to follow all stocks in real-time. This advantage has led quants to steadily outperform market averages by 2-3% for the past several years.

The figure shows the overall outline of the system. The first step consists of news gathering. In this step news articles from online economic news boards are collected and stored in database. The second step consists of performing Opinion Mining on the financial textual information gathered from the web. There are a variety of methods available to analyze financial news articles. One of the most common methods is to apply a vector representation where article terms are indexed and then weighted. Selecting article terms can be as simple as tokenizing and using each word in the document. This technique assigns importance to determiners and prepositions which have little contribution to the overall meaning of the article. One method of circumventing these problems is to
use a Bag of Words approach. In this approach, a list of semantically empty stop-words are removed from the article (e.g.; the, a, and for). The remaining terms are then used as the textual representation. The Bag of Words approach has been used as the de facto standard of financial article research primarily because of its simple nature and its ability to produce a suitable representation of the text.

Building upon the Bag of Words approach, one such method using this approach is Noun Phrasing. Noun Phrasing is accomplished through the use of a syntax where parts of speech (i.e., nouns) are identified through the aid of a lexicon and aggregated using syntactic rules on the surrounding parts of speech, forming noun phrases. Named entities technique builds upon Noun Phrases by using lexical semantic/syntactic tagging where nouns and noun phrases can be classified under predetermined categories.

In the third step, as news events affect human decisions and the volatility of stock prices is influenced by human trading, it is reasonable to say that events can influence the stock market. Accurate extraction of events from financial news may play an important role in stock market prediction. However, after performing opinion mining on documents mainly using simple unstructured features, such as bags-of-words, noun phrases, and named entities, it is difficult to capture key events embedded in financial news. For example, representing the event “Apple has sued Samsung Electronics for copying ‘the look and feel of its iPad tablet and iPhone smartphone.” Using term-level features {“Apple”, “sued”, “Samsung”,}
“Electronics”, “copying”, ...} alone, it can be difficult to accurately predict the stock price movements of Apple Inc. and Samsung Inc., respectively, as the unstructured terms cannot indicate the actor and object of the event. Thus, predicting the stock market movements whether to buy, hold or sell.

Therefore, the proposed system uses simple feature extraction as well as structured information to represent events, and develop a prediction model to analyze relationship between events and stock market.

6. Applications

It is rightly said that ‘Sentiment without action is the ruin of the soul’. The applications of opinion mining are described as follows:

6.1 Applications to review-related websites

The same capabilities that a review-oriented search engine would have could also serve very well as the basis for the creation and automated upkeep of review and opinion-aggregation websites. That is, as an alternative to sites that solicit feedback and reviews, one could imagine sites that proactively gather such information. Topics need not be restricted to product reviews, but could include opinions about candidates running for office, political issues, and so forth. [6]

6.2 Applications in business and government intelligence

The field of opinion mining and sentiment analysis is well-suited to various types of intelligence applications. Indeed, business intelligence seems to be one of the main factors behind corporate interest in the field. [6]

6.3 Applications across different domains

One exciting turn of events has been the confluence of interest in opinions and sentiment within computer science with interest in opinions and sentiment in other fields. As is well known, opinions matter a great deal in politics. Some work has focused on understanding what voters are thinking, whereas other projects have as a long term goal the clarification of politicians’ positions, such as what public figures support or oppose, to enhance the quality of information that voters have access to. [6]
7. **References**


7. Jiangjiao Duan ; Sch. of Econ., Fudan Univ., Shanghai, China ; Jianping Zeng, ‘Mining opinion and sentiment for stock return prediction based on Web-forum messages’, Fuzzy Systems and Knowledge Discovery (FSKD), 2013 10th International Conference INSPEC Accession Number:14319735


11. Kari Lee and Ryan Timmons, ‘Predicting the Stock Market with News Articles’.


END OF DOCUMENT