
Sonam Bhardwaj* and Er. Sumit Sharma

1Computer Science Department, Chandigarh University, Rohtak - 124001, Haryana, India; sonambhardwaj98@gmail.com
2Computer Science Department, Chandigarh University, Ludhiana - 141001, Punjab, India; sumit_sharma@mailingaddress.org

Abstract

Objective: In this paper, we propose a framework for automatic incorporation of fine-grained news data to generate S&P BSE Sensex trading strategies. Method/Statistical Analysis: Financial news containing information about companies' events, share prices, sales, Government and economic policies etc., are fetched from web. For fine-grained news data, text mining is applied to pre-process the news. Sentiment analysis is done to estimate the exact sentiment of news messages containing contradictory words and input at right place into the framework. News entities specific to stock market behavior prediction are extracted from fine-grained news data. Pearson Correlation characterizes the relation between stock returns and expert defined Impact Factor (IF). Findings: The purposed framework incorporates the news dataset specific to 30 companies listed under S&P BSE Sensex stock index. Stock returns corresponding to news entities generate the buy or sell signal. The results depicting positive value of Pearson's correlation describes that market moving in right direction. It is proved that the specified news entities incorporated at right place into the framework have significant effect on stock movement and generate higher magnitude of returns as compared to other entities used in previous work. The framework gives approximately 94% accuracy which is greater than other news based trading model. Application/Improvements: The accuracy of purposed framework can be improved by combining a number of technical trading indicators with news entities. Also optimal trading strategies can be generated by using optimal programming.

Keywords: Impact Factor, News Preprocessing, News Entities, Sentiment Analysis, Trading Strategies

1. Introduction

Financial data becomes an important consideration to predict the stock market moves to take an investment decision. In companies take the investment decisions on the basis of a company's events, press releases, customer interests and feedbacks, political and economic events, government policies, raw material policies, interest rates, managerial decisions etc. Informative reports published by media agencies are considered by investors to take the investment decision. Along with it; reviews of both categories of users like technical and non-technical users can also be accumulated from Twitter and Facebook. News announced by media channels are taken as recommendation to take trading decision. Technical trading indicators like Simple Moving Average (SMA), Bollinger Bands (BB), Rate of Change (RoC), Momentum, Moving Average Convergence Divergence (MACD) etc., are used to generate the buy or sell signal. All these technical indicators consider the historical stock prices to estimate the stock market movement. The performance of technical indicators depends on their widespread use in trading signal. Nowadays, in the era of internet, financial information is globally available. The information about a

to predict the stock returns. The text mining techniques for method and given as learning input to neural networks to predict the complex market trading strategies. There are approximately 3000 news reports which are published in the newspaper on monthly basis. This leads to generation of high volume of news data which is very difficult to be analyzed manually. Therefore, it is necessary to collect and analyze the financial information as it contains the investment related decisions in an automated manner. Various machine learning tool and techniques are available for extracting the knowledge from databases. Neural networks are significantly used to analyze the non-linear behavior of financial market. The real time series data can be decomposed using Discrete Wavelet Transform method and given as learning input to neural networks to predict the stock returns. The text mining techniques for preprocessing of text allows the researcher for automatic processing of news data instead of manual reading and understanding each news individually.

News data have impact on stock market prices. Thousands of records corresponding to thousands of companies can be extracted from News data source and processed in very less time by approaching text mining techniques. For fine-grained news data, a plenty of unstructured news messages are converted into structured trading information by text processing in order to filter out unnecessary information. Text preprocessing steps are to be applying on unstructured news data i.e., parsing, tokenization, stop words removal, synonym merging, stemming, and impact feature extraction. The paper describes the text mining steps in brief:

- **Parsing:** The unstructured news data fetched from web present in the form of XML, Json or RSS needs to be parsed to extract the news messages with their corresponding date on which news was announced.

- **Tokenization:** It means breaking the unstructured text of each news message into words, symbols or meaningful entities called tokens. These tokens are further separated by whitespace.

- **Stop Word Removal:** These are grammatical words used to construct the sentences in any language. The stop words are to be removed from news messages to focus on companies’ related information. For example, consider a news message related to CIL listed under S&P BSE SENSEX “Due to divestment decision CIL miss output target for fiscal year 2015 by over 10 million ton”. In this news message, the terms “due” “to” “for” “by” “over” are commonly used in the English language. By discarding these terms we left with the important keywords “divestment” “decision” “CIL” “miss” “output” “target” “fiscal” “year” “2015” “10” “million” “ton”. A list of stop words need to created which includes articles (example a, an, the), prepositions (example before, after, over, of, at, on), noun (example this, that, day name). Adjectives should not be add into stop words list because sentiment analysis includes adjective to predict the actual emotion related to news message.

- **Synonym Merging:** Different words dictate the same meaning. For example, “up” “raise” “increase” “high” “surge” conveys the same meaning. Therefore, merge the synonyms and aggregate by their meaning. This method is termed as pseudo word generation.

- **Stemming:** It is the process of reducing the derived word to their root word. Porter algorithm can be used to implement the step.

After processing the text, Semantic analysis is to be performed to estimate the polarity of news in terms of positive, negative or neutral. All news messages do not contain the financial information. Sentiment analysis is needed to analyze the actual emotion contained in the news messages to generate the accurate trading signal.

As news data possesses high dimensional textual features and each feature has some possible interaction with other features therefore, it is difficult to analyze exact sentiment corresponding to each news messages which contain contradictory words. An individual news can contain more than one contradictory word like “Loss is less than previous year’s loss” or “Opposing investment decision”. The first news statement is actually conveying about the company profit in current year. Also, the second statement is not about business expansion. Therefore, sentiment analysis technique is necessary to be implemented to analyze the actual meaning conveyed by news in order to predict the more accurate stock prices.
2. Challenges in Automation of News Information Extraction

In the era of globalization of stock market, the velocity with which information move and the slightest national and international Government policies may have a significant impact on future stock returns. Government change policies and impose abrupt changes like excise duty, sales tax may affects investor's choice for the investment in the country. They can shift their capital abroad. Also, increase in Repo Rate results in lower the capital availability. The actions taken by Government build the sentiment of investors whether invest or divest in the market. All the related news should be incorporated to generate the optimum trading strategies.

In an automated framework for incorporating news data into stock trading strategies, a number of challenges are explained in this paper.

2.1 Government Policies

Government may impose some new or slightly varied policies on financial companies. A number of policies on public spending, taxation, interest rates, exchange rates, the size of the debt and the deficit, income redistribution and capital movement affect the future stock returns. In past years, Government has much more freedom to impose any financial policy which they considered appropriate and regulate the financial market. But now the situation has been changed entirely. In very short duration, financial market has grown rapidly and the influence of Government policies has now been reduced. First of all, taxation power is restricted because companies can invest in different country where they can be more profitable. Second, regulation over financial companies is also restricted in case when companies have billions of electronics transactions. Finally, the power to print money and borrow to finance a budget, no longer exists because market forces can neutralize any monetary or tax policy that they consider unwise. The policies of Government affect the stock market in following ways:

2.1.1 Monetary Policy: The Printing Press

Monetary policy is the process of controlling the supply of money by monetary authority of a country (mostly a central bank like Reserve Bank of India). These policies have large impact on future stock prices and therefore catch the investor's attention. The aim of monetary policy is price stability and financial growth. The market value depends on this policy but this is not permanent effect on the degree of change in interest rate. The stock market prices will not be affected if RBI increase the interest by same rate as market is expecting. Instead, stock market prices will be affected in case when RBI raises interest rate by unexpected percentage. Therefore, impact of this policy depends upon the difference between expected and actual interest rates announced by RBI. In case if RBI raises the interest rates more than expected, it will lead to lower the value of future cash flow in today's rupees. Another impact of higher interest rate is reducing the supply of bank loans which ultimately leads to less investment and lower profits. Also, tightening monetary policy reduces the inflation rate which results in less money in investor's hands to buy the stock and thus make the future stock returns downwards.

2.1.2 Currency Inflation

The Government policy of currency inflation has a greater relevance with the stock trends. Both positive and negative impacts can be seen due to inflation on future stock returns. In case of high inflation, a company starts selling off its products. Therefore, an increase in supply of stock results in decrease in stock prices which results in economic downturn. On the other hand, a positive impact is also possible between inflation and stock prices as unexpected inflation raises the company’s equity value if they are debtor. Government inflates the money to provide short-term financial boost and companies raise the prices of products. Hence, stock returns will be greater. Therefore, inflated money is a good Government policy for the investors to shooting up the stock prices as a short-term policy. On the other hand, its long-term impact would be negative as it would punish the savers due to lower stock prices of products.

2.1.3 Fiscal Policy: Interest Rates

Interest is the additional amount which lenders charge when they give loans to the borrowers. In case of rising interest rates, companies not take the loans and step back on spending. This lead to business contraction and eventually earnings fall down. Thus future stock prices will also fall. On the flipside, decrease in interest rates leads to generation of more buying power and results in business expansion. The market move towards positive future stock returns.
2.1.4 Subsidies and Tariffs
Subsidies are taxation system imposed by Government on general public. The money earned via subsidies is given to Government chosen company to make it stable and more profitable. Tariffs are also kind of taxes which Government imposes on foreign products to make them more expensive. Both the policies have direct effect on the stock market. The financial support to a company by the Government makes it globally competitive company and results in higher stock returns.

2.1.5 Regulations and Corporate Taxes
High taxes can proved as restrictions to invest the capital by the investors into the country. Investors may shift their capital in some other country where they found less corporate tax.

2.2 News Filtering
In the era of digitization, data is growing with peta-bytes via social networking sites, e-papers, blogs etc. A large number of media channels publish thousands of financial news every day. Beside media channels, billions of people upload posts on social networking sites like Facebook and Twitter. According to surveys conducted by various financial companies, the big data market has grown from US$3.5 billion in 2010 to US$ 17.5 billion in 2015. To process such a high volume of unstructured data is a big difficult task. One major challenge for stock market analyst will be the regulating the frequency of data generation on daily basis and delivery in financial services companies like investment, banking and IT sectors. The second main challenge is news filtering. News filtering involves following tasks: 1. Extraction of company’s relevant news, 2. Removing duplicate news messages, 3. The news announced on the day when stock exchange is closed should not be incorporated; 4. Rare news messages should not be included.

The third main challenge is real-time analysis of news to see immediate impacts on future stock prices. Finally, sentiment analysis is also a challenge to analyze the actual meaning of news and incorporate at the right place in the framework.

2.3 International Factors
The international stock indices like Dow Jones of USA, DAX of Germany, Nikkei of Japan, Hang Seng of Hong Kong etc. also affect the stock market returns of another country. If recession came in foreign country, it will certainly affect the markets of other countries due to divestment, lower export and import of products. Another factor is value of currency, if currency value of any country gets depreciated, stock market would fall. Also, policy changes by Government of any country like abrupt changes in excise duty, sales tax leads to variations in investment or divestment. Interest rates in international market also decide the purchasing capability of financial companies. All the related news data should be incorporated to generate the more accurate buy or sell signal.

2.4 Raw Material Policies
A significant amount of raw materials and energy is required for manufacturing the products. The variations in the cost of energy affect the company like oil and natural gas stock prices. The financial companies are inter-dependent and can affect the stock prices of one another. Also, fluctuations in the prices of raw materials primarily affect the market prices of company like Sun Pharmaceuticals affecting their share prices from period to period. In one case, a steel company fixed prices for users without capping vendor contracts, and as a result, was not able to bear increase in raw material prices. The situation leads to dropping in production and market capitalization fall by approximately 35%. The news related to raw material policies should be included in the automated framework to generate the optimum trading strategies. In current year, Prime Minister of India signed the agreement on procurement of energy resource Uranium with Canada. The investment made by Indian Government leads to higher stock returns in future.

2.5 Managerial Factors
Management change events can also have negative or positive effects on stock returns. The change events like appointing new Chief Executive Officer (CEO), chairman of the board, president can result in either positive or negative direction of companies’ stock prices graph. There is no direct relation between management change and stock returns. If performance of changed management is worse than negative effects can be seen. Otherwise, if a company is in net debts and perform bad, then management change can prove as positive news. In this case, new visions and strategies deployed by change management could revise the future stock returns.
3. Proposed Methodology

As the size of the data increases, news filtering becomes difficult. News filtering means incorporation of news related to user interest. Stock market movement depends upon multiple variables like financial, economic and political events. Therefore, to calculate the future stock returns of stock index, we need news related to company events, press releases, and headlines published in various newspapers. For this, one solution is to use Google “Personalize” to filter the news of relevant companies only. The Google search engine generates personalized recommendations related to news topic searched by user by using collaborative filtering approach. The filtered news can be fed inside the news automation framework using RSS Feeder. But the solution provides limited number of news related to company. To access a large amount of news database, EINNEWS offers historical and live news data about world and region, a large variety of topics like banking, finance, investment, business, economy, politics, religion, retail, social media, technology, companies, conferences and trade fairs, energy, international organizations, Government law, Natural Disasters, Real Estate and Property Management etc. The website provides a search engine to filter the news of user interests. To search the news related to Coal India Limited (CIL) user can put the following query into website search engine with custom date “Coal India and (sale* | profit* | share* | interest* | Government policy* | joint* | international* | agree* | collaborate* | target* | up* | down* | product* | tax* | repo*)”. The output will contain all the news of CIL related to its share prices, profit values, government policies, international policies etc.

To automate the knowledge being conveyed by news data we use four-step approach consisting of as shown in Figure 1.

- Fetching news messages present in free-text format and perform text mining.
- Sentiment analysis to estimate the emotions contained in news messages.
- Extracting News Impact Entities from news data.
- Calculate returns corresponding to each entity to generate trading strategy.

3.1 Data Collection

Two data sets are needed to implement the methodology. First data set consists of historical stock prices and another one consists of news messages. We are collecting data from 1st April 2015 to 30th June 2015. The news and stock prices datasets contains the data of all 30 companies listed under S&P BSE SENSEX index. The historical stock prices of stock index are available at Yahoo! Finance website (https://in.finance.yahoo.com). The excel file downloaded from website contains open, close, high, low and adjusted closing prices. The news data can be fetched via RSS Feeder by visiting the website EINNEWS (available at https://www.einnews.com). Researchers can feed the news related to finance, stock, business, and government policies, political and economic events from the website. For removal of unwanted information, text mining steps like parsing, tokenization, stop word removal, synonym merging and stemming are applied. Porter algorithm is used to perform the stemming.

3.2 Sentiment Analysis

To analyze the sentiment of each news message related to companies, we use MeaningCloud Tool (https://www.meaningcloud.com), a multi-lingual text analytics application that provides feature-level sentiment analysis, topic extraction, text categorization and clustering. For news like “profit is less than previous year’s profit”, a sentiment model is created containing keywords related to companies stock market moves with corresponding sentiment in terms of highly positive (P+), highly negative(N+), positive (P) and negative (N). The tool can be downloaded for Excel and parse the news dataset with addition of sentiment model for estimating the actual polarity of news to incorporate the message at right place into the trading framework.

3.3 News Entities

The entities which can influence the companies’ stock prices are to be identified first. To measure the effective-
ness of health information entities like salary of person, concern on health etc. can be taken as. In this paper, we are taking a predefined set of news impact entities as shown in Figure 2. For example sale, share prices, target price, management events, production, company agreements, tax, interest rate etc. These fixed entities are to be extracted from news messages and considered as up or down, increase or decrease rise or fall according to its corresponding news sentiment. In case, if prices of raw materials and interest rate are decreased then company with same capitalization can buy more raw materials for manufacturing the products. This leads to increase in company’s production value and thus in share prices. The news entities can be used to see the impact of news messages on equity. If the news impact entities having positive polarity results in positive stock returns then our strategy is correct to predict stock market on the basis on news data otherwise not.

Figure 2. Frequency of news entities.

The relevant news impact factors are extracted from database on the basis of following criteria:

- Same News announced on multiple times in one day should not be considered as it will lead to duplicate news impact factors.
- Rare impact factors are to be removed on the basis of term frequency value. The factors which have less than 0.5% value of term frequency with respect to all impact factors can be omitted.

3.4 Stock Returns Estimation

The effect of news entities on stock index can be calculated using relative returns based on adjusted closing price . For a particular product, a stock return can be calculated as

\[ r_n = \frac{P_{i+n} - P_i}{P_i} \times 100, \quad n > 0 \]  \hspace{1cm} (1)

Where \( i \) represent the day on which news was announced and \( n \) is number of days over which stock index returns are calculated. The returns are calculated on day 1, day 2 and day 5. It is expected that magnitude of return can increase as news become public with time.

For same impact entities extracted on different date regarding same product the return is averaged for the number of different date using following equation:

\[ R_D = \frac{\sum_{n=1}^{D} r_n}{D} \]  \hspace{1cm} (2)

Where \( D \) represents the total number of days on which same impact factor occurred. Also, if actual returns are greater than expected returns then the returns generated are called excess return or abnormal return.

The diagrammatic representation of trading framework is shown in Figure 1. The results are considered to be correct if the positive impact factors generate positive magnitude of stock returns. Also, it is equally important that impact factors of negative polarity generate returns of negative magnitude. In terms of trading signal, a positive return indicates the buy and a negative return indicates the sell recommendation.

4. Results Discussion

The results are based on the hypothesis that news has certain effect on future stock prices. Also, a predefined value of IF is assigned by experts to each news entity. The IF is an additional data to find out the correlation between stock index returns and IF. We calculate absolute returns S&P BSE SENSEX for day one, two and five. For each news entity, we compute two-tail t-test significance called p.

The results predictable from Table 1 conveying that for 94 percent of news entity, the stock index returns are according to expert predefined IF i.e., positive returns in case of positive defined IF and vice-versa. The entity company agreements have negative return on day 1. The estimated returns are considered correct since these news entities can have negative returns in case of poor relationship between shareholders. These news entities can be proved both negative and positive for company’s growth depends on honesty among shareholders.
However, associated IF considered as positive because experts think that these news entities would generate high revenue in future. Also, we cannot see effect on stock index returns on day 1 as companies take processing time to start working on agreements and investor’s wait for the knowing about terms and conditions about the agreements. The returns corresponding to target price lower and stock down are not as expected because in case if different types of news announced on same day that results in different news entities would decrease the magnitude of stock index returns as other entities dominate the returns. The returns calculated by trading framework corresponding to share up, sale up, tax decrease, interest rate decrease, repo rate decrease, management appointment events, production raise, share down, business expand, stock up and sale down are considered as correct as these returns are in same direction with respect to IF value assigned to them.

In Table 1, the relation between stock index returns and their corresponding IF value is concluded by Pearson Correlation Coefficient and Coefficient of determination. The Pearson’s correlation coefficient predicting the expected correlation between the stock index returns and IF.

### Table 1. Average S&P BSE SENSEX returns $R_d$ after d days

<table>
<thead>
<tr>
<th>News Entities</th>
<th>Freq</th>
<th>IF</th>
<th>$R_1$</th>
<th>p</th>
<th>$R_2$</th>
<th>p</th>
<th>$R_3$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share Up</td>
<td>94</td>
<td>10</td>
<td>0.43</td>
<td>0.67</td>
<td>1.01</td>
<td>0.33</td>
<td>5.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Sale Up</td>
<td>24</td>
<td>10</td>
<td>0.09</td>
<td>0.92</td>
<td>1.08</td>
<td>0.30</td>
<td>3.37</td>
<td>0.00</td>
</tr>
<tr>
<td>Target Price Raise</td>
<td>28</td>
<td>10</td>
<td>1.52</td>
<td>0.15</td>
<td>2.84</td>
<td>0.01</td>
<td>9.39</td>
<td>0.00</td>
</tr>
<tr>
<td>Management Appoints</td>
<td>5</td>
<td>15</td>
<td>4.52</td>
<td>0.00</td>
<td>6.52</td>
<td>0.00</td>
<td>13.44</td>
<td>0.00</td>
</tr>
<tr>
<td>Production Raise</td>
<td>3</td>
<td>6</td>
<td>4.03</td>
<td>0.00</td>
<td>0.90</td>
<td>0.38</td>
<td>0.70</td>
<td>0.49</td>
</tr>
<tr>
<td>Business Expand</td>
<td>5</td>
<td>15</td>
<td>4.67</td>
<td>0.00</td>
<td>9.34</td>
<td>0.00</td>
<td>14.98</td>
<td>0.00</td>
</tr>
<tr>
<td>Company Agreements</td>
<td>6</td>
<td>6</td>
<td>-2.68</td>
<td>0.21</td>
<td>3.29</td>
<td>0.00</td>
<td>0.26</td>
<td>0.79</td>
</tr>
<tr>
<td>Tax Decrease</td>
<td>17</td>
<td>10</td>
<td>3.88</td>
<td>0.00</td>
<td>1.84</td>
<td>0.09</td>
<td>7.76</td>
<td>0.00</td>
</tr>
<tr>
<td>Interest Rate Decrease</td>
<td>4</td>
<td>6</td>
<td>1.05</td>
<td>0.31</td>
<td>0.98</td>
<td>0.34</td>
<td>5.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Repo Rate Decrease</td>
<td>7</td>
<td>6</td>
<td>1.21</td>
<td>0.25</td>
<td>2.34</td>
<td>0.04</td>
<td>4.88</td>
<td>0.00</td>
</tr>
<tr>
<td>Stock Up</td>
<td>108</td>
<td>15</td>
<td>6.70</td>
<td>0.00</td>
<td>12.92</td>
<td>0.00</td>
<td>12.72</td>
<td>0.00</td>
</tr>
<tr>
<td>Share Down</td>
<td>7</td>
<td>-10</td>
<td>-2.78</td>
<td>0.01</td>
<td>-3.06</td>
<td>0.00</td>
<td>-1.96</td>
<td>0.07</td>
</tr>
<tr>
<td>Sale Down</td>
<td>13</td>
<td>-10</td>
<td>-1.55</td>
<td>0.14</td>
<td>-1.64</td>
<td>0.12</td>
<td>-1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>Target Price Lower</td>
<td>14</td>
<td>-10</td>
<td>-0.11</td>
<td>0.91</td>
<td>-2.31</td>
<td>0.41</td>
<td>1.83</td>
<td>0.09</td>
</tr>
<tr>
<td>Stock Down</td>
<td>24</td>
<td>-15</td>
<td>-1.32</td>
<td>0.21</td>
<td>0.33</td>
<td>0.74</td>
<td>0.95</td>
<td>0.36</td>
</tr>
</tbody>
</table>

News Entities: Companies Stock Effecting Factors.
Freq: Number of times News Entities Occur.
IF: Impact factor decided by experts.
$R_1$: Average Stock Index Returns.
P: 2-tail t-test significance at p < 0.10.

### Table 2. Pearson’s correlation and coefficient of determination between S&P BSE SENSEX stock index returns and IF

<table>
<thead>
<tr>
<th>$R_d$</th>
<th>r</th>
<th>P</th>
<th>$r^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_1$</td>
<td>0.722</td>
<td>0.48</td>
<td>0.522</td>
<td>0.61</td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.738</td>
<td>0.47</td>
<td>0.544</td>
<td>0.59</td>
</tr>
<tr>
<td>$R_3$</td>
<td>0.779</td>
<td>0.45</td>
<td>0.607</td>
<td>0.55</td>
</tr>
</tbody>
</table>

$R_d$: Average stock index returns after d days.
r: Pearson’s correlation coefficient.
$r^2$: Coefficient of determination and.
P: 2-tail t-test significance at p < 0.10.

### 5. Conclusion

The news data used to generate the trading strategy by extracting news entities which provide the real-time recommendations to investors whether to buy, hold or sell a product. Sentiments of news are analyzed to place the news entities at correct place into the framework. The paper discusses all internal and external news factors which can impact the volatility of stock prices of a...
company. Lower tax and interest rate generates the positive stock returns of market. Additionally, this approach provides greater magnitude of returns. In terms of accuracy, 94 percent of news entities generate the S&P BSE Sensex stock returns in the expert-defined direction. At next, News impact entities can be combined with technical trading indicators to generate the higher magnitude of stock returns. Also optimal trading strategies can be generated by using optimal programming.

6. References