

Predicting Stock Movements using Social Network

Sunil Saumya, Jyoti Prakash Singh, and Prabhat Kumar
Email id: {sunils.cse15, jps, prabhat}@nitp.ac.in

National Institute of Technology Patna,
Bihar, India

Abstract. According to "Wisdom of Crowds" hypothesis, a large crowd can perform better than smaller groups or few individuals. Based on this hypothesis, we investigate the impact of online social media, a group of interacting individual, on financial market in Indian context. The interaction of different users of www.moneycontrol.com, a popular online Indian stock forum, is put to a social graph model and several key parameters are derived from that social graph along with the user suggestion such as (*Buy, Sell or Hold*) related to a stock. The users impact in that forum is then calculated using the social graph of the users. Stock price movement is then predicted using user's suggestions and their impact in that forum. As per our knowledge, this is the first paper which consider the impact of www.moneycontrol.com user's suggestions and social relation to predict the stock prices.

Key words: Wisdom of Crowd; Page Rank; Stock Price Movement; Social Network; Social Graph

1 Introduction

Stock market investment decisions are mainly driven by the market information available to investor. In earlier days, the main source of new information was by news articles containing information related a company, such as the company fundamentals, future plans and so on. The stock price of a company were driven by these publications. The rise of internet and finance related websites and applications changed the scenario as the new information about a company is readily available. Some of the websites are developed as a forum where a number of interested users interact with each others and give their opinion on different stocks. A system which can utilize this data and peoples opinion to predict future changes in prices is highly required to support the decision making of investors and traders.

There are two popular hypothesis regarding stock price prediction (i) Efficient Market Hypothesis (EMH) [1] and (ii) Random Walk [2]. The EMH [1] states it is impossible to "beat the market" because stock market efficiency causes existing share prices to always incorporate and reflect all relevant information. The random walk hypothesis states that stock market prices evolve according

to a random walk and hence cannot be predicted which is consistent with the EMH.

In spite of both the hypothesis stating stock price can not be predicted, some automated systems to use the financial news to help in decision making of investors and traders have been proposed [3]. Another research [4] has shown that there is a strong relationship between stock price fluctuations and publications of relevant news. Schumaker and Chen [5] successfully employed the to prove the effect of news items on the stock prices using stock's history data. The other notable works using historical data to predict the stock prices are by Patel et al. [6,7] using Auto Regressing model and Moving Average model, Zuo et al. [8] using Bayesian network and Auto Regressive Moving Average model. All these works used historical stock prices or company related news to predict future prices using some statistical techniques.

"Wisdom of Crowd" (WoC) [9] hypothesis states that a diverse and independent "crowd" can make more precise predictions than a few people. They can even beat professionals. Crowd is defined as "potentially large and unknown population" [10]. The WoC can also be viewed as collective intelligence [11]. With the rise of social networking sites, this phenomenon has got a boost as diverse people from different places in the world interact with each other on websites, blogs and message boards. It has attracted a lot of researcher to use this crowd's intelligence to predict the stock related information. Ruiz et al. [12] studies the impact of twitter activity on financial market. Latoeiro et al. [13] tried to predict the stock market activity using Google search queries. Although, there have been some studies involving financial industry using crowd's opinions for their investments but there is hardly any practical evidence. The user-generated contents were also used in [14–16] to identify the share returns. But very few work have been done which show a good prediction result on a huge consistent data for long duration of time [17].

All the previous studies were done on developed country markets. Whether those studies are really useful in developing country like India seems an interesting research problem. Mukherjee [18] compared the Indian stock market with international markets and points out a lot of differences. That motivated us to investigate the impact of online social forum on Indian stock market involving Indian stock exchanges such as National Stock Exchange (NSE) and Bombay Stock Exchange (BSE). In this paper, we have considered an Indian financial forum www.moneycontrol.com as our data source and two representative stocks namely State Bank of India (SBIN) and Reliance Communication (RCOM) for our study.

2 Data Collection

The forum data was collected from mmb.moneycontrol.com for two stocks namely SBIN and RCOM. We used data scraping technique to collect the data. Data scraped from mmb.moneycontrol.com were then filtered and pre-processed to collect the following fields, which were (i) User_Name, (ii) User_Level, (iii)

Num_Messages, (iv) Followers, (v) Following, (vi) Stocks_Tracked, (vii) Time, (viii) Message_Text, (ix) Rating, and (x) Replies. The details of the fields are given in Table 1. A typical message on mmb.moneycontrol.com is given here in the Figure 1 and the details of the user is given in Figure 2.

Table 1. The details of fields of a user message

Sl. no.	Field	Description
1	User_Name	The screen name of the user participating in the forum
2	User_Level	The levels of the user
3	Num_Messages	Total messages written by the user
4	Followers	The number of other users who follows this user
5	Following	The number of other users this user is following
6	Stocks_Tracked	The number of stocks tracked by a user
7	Time	The date and time on which the message was written
8	Message_Text	The message about a stock written by the user
9	Rating	The rating given by other users to a message
10	Replies	The number of replies given to a message

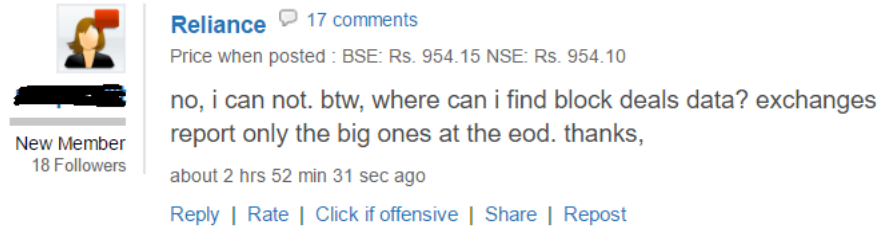


Fig. 1. A general format of message on mmb.moneycontrol.com

Some of the messages were discarded as they do not convey any sentiment about the stocks. Those are just informative or enquiry messages as shown in Figure 1. After discarding those messages, we worked with 490 reviews for SBIN and another 389 reviews for RCOM. The custom-developed program in Python was then used to extract the suggestions given by users in terms of "Buy", "Sell", or "Hold".

3 Proposed Method

We extracted the suggestions (*Buy, Sell, or Hold*) of the users through their messages using Natural Language Toolkit (NLTK) [19] package in Python. The



Fig. 2. A general format of Member information on mmb.moneycontrol.com

message text was tokenized and Part of Speech (POS) tagging was done. The verbs from those tagged words were taken and if they match *Buy, Sell, or Hold*, then they were associated with the tag. In some messages, users have not directly written whether to *Buy, Sell or Hold* but they have given the next price (target) of the stock. We have extracted the current price of the stock using *nsetools* which is a Python library for extracting real time data from National Stock Exchange (India). If the current price of the stock is less than the target given by the user then the suggestion is taken as *Buy* else *Sell*. After extracting the suggestions of the user from their messages, we created a network of users and their suggestions. A sample network showing the interaction of different users and their suggestions such as *Buy, Sell, or Hold* is shown in Figure 3.

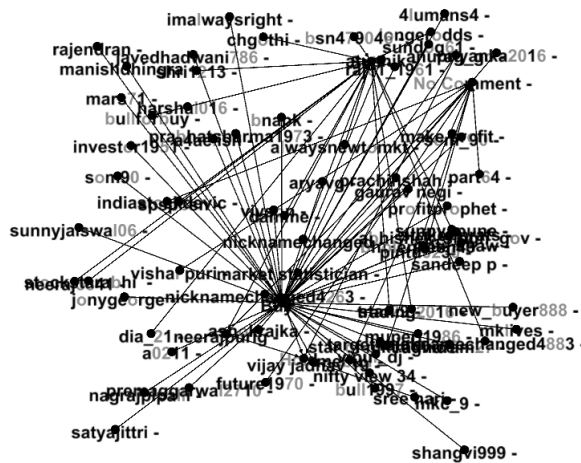


Fig. 3. The user interaction graph

The network was made using tool known as "Gephi" [20] which supports calculation of features such as degrees, page rank etc. Degree refers to how many links a user has with another users. PageRank [21] algorithm estimates the importance of a website by counting the number and quality of links to another websites. This concept is used here to find how important a user is by counting the number and quality of links with another users of the system. The importance of the users (weight) is a real number ranging between 0 and 1 with 0 being less important and 1 being most important. The user's suggestions (*Buy, Sell, Hold*) are encoded with $(+1, -1, 0)$ respectively. The stock movement is then decided by finding the weighted average of encoded suggestions with respect to user's weight. The results are shown in Table 2.

Table 2. Weighted average Vs. real stock movement

Sl. No.	Stock Name	Weighted Average	Real Stock Movement
1	SBIN	0.01283	Upward
2	SBIN	-0.3276	Downward
3	RCOM	-0.3713	Downward
4	RCOM	-0.00242	Downward

As can be seen from the Table 2, +ve weighted average indicates a upward movement in stock price where as -ve average indicates downwards movement.

4 Conclusion

We have tried to model the financial market on the concept of WoC using data from www.moneycontrol.com. The users are assigned weights based on their social interaction in the forum. The users weight along with their suggestions is then used to predict the stock price movement. The results looks promising on the small datasets we have considered in this work. The finding indicates that the online financial forum can be used as a data source for stock market movement prediction. One of the major limitation of the current work is small data size. The work can be further extended to predict the exact % of swing in stock price as well as the future price of the stock including more parameters such as historical prices, recent news related to company etc.

References

1. Malkiel, B.G.: Efficient market hypothesis. In: The World of Economics. Springer (1991) 211–218
2. Godfrey, M.D., Granger, C.W., Morgenstern, O.: The random-walk hypothesis of stock market behaviora. *Kyklos* **17**(1) (1964) 1–30
3. Mitra, G., Mitra, L.: The handbook of news analytics in finance. Volume 596. John Wiley & Sons (2011)

4. Fung, G.P.C., Yu, J.X., Lu, H.: The predicting power of textual information on financial markets. *IEEE Intelligent Informatics Bulletin* **5**(1) (2005) 1–10
5. Schumaker, R.P., Chen, H.: A quantitative stock prediction system based on financial news. *Information Processing & Management* **45**(5) (2009) 571–583
6. Patel, J., Shah, S., Thakkar, P., Kotecha, K.: Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications* **42**(1) (2015) 259–268
7. Patel, J., Shah, S., Thakkar, P., Kotecha, K.: Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications* **42**(4) (2015) 2162–2172
8. Zuo, Y., Kita, E.: Stock price forecast using bayesian network. *Expert Systems with Applications* **39**(8) (2012) 6729–6737
9. Surowiecki, J., Silverman, M.P., et al.: The wisdom of crowds. *American Journal of Physics* **75**(2) (2007) 190–192
10. Poetz, M.K., Schreier, M.: The value of crowdsourcing: can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management* **29**(2) (2012) 245–256
11. Kittur, A., Kraut, R.E.: Harnessing the wisdom of crowds in wikipedia: quality through coordination. In: *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM (2008) 37–46
12. Ruiz, E.J., Hristidis, V., Castillo, C., Gionis, A., Jaimes, A.: Correlating financial time series with micro-blogging activity. In: *Proceedings of the fifth ACM international conference on Web search and data mining*, ACM (2012) 513–522
13. Latoeiro, P., Ramos, S.B., Veiga, H.: Predictability of stock market activity using google search queries. Working Papers ws130605, Universidad Carlos III de Madrid (March 2013)
14. Antweiler, W., Frank, M.Z.: Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance* **59**(3) (2004) 1259–1294
15. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. *Journal of Computational Science* **2**(1) (2011) 1–8
16. Avery, C.N., Chevalier, J.A., Zeckhauser, R.J.: The “caps” prediction system and stock market returns. *Review of Finance* (2015) rfv043
17. Nguyen, T.H., Shirai, K., Velcin, J.: Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* **42**(24) (2015) 9603–9611
18. Mukherjee, D.: Comparative analysis of indian stock market with international markets. *Great Lakes Herald* **1**(1) (2007) 39–71
19. Bird, S.: Nltk: the natural language toolkit. In: *Proceedings of the COLING/ACL on Interactive presentation sessions*, Association for Computational Linguistics (2006) 69–72
20. Bastian, M., Heymann, S., Jacomy, M.: Gephi: An open source software for exploring and manipulating networks. In: *International AAAI Conference on Weblogs and Social Media*. (2009)
21. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: bringing order to the web. (1999)