International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 1, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Stocks and Social: An In-Memory Concept to Diversity in Understanding Social Impact on Stocks

Prabal Mahanta
TIP CD & SP, SAP LABS Pvt. Ltd, Bangalore, India
Email: p.mahanta@sap.com

Suhas Sudheendra
TIP CD & SP, SAP LABS Pvt. Ltd, Bangalore, India
Email: suhas.s@sap.com

***Abstract* – This paper deals with the prospect of looking into the world of stocks and social media and conceptualize the landscape in the technical domain and how the concept can generate the idea of correlation of the two domains and present the theory of making interpretation in a newly designed concept to predict the behavior of market and in turn understand the social media in terms of the interaction with the stock market and also few conceptual landscape advancements to introduce the in memory technology to the field of stock trading and social media**

***Keywords* – Stocks, Social, Landscape, In-memory Computing, Trading, Techniques.**

I. Introduction and Survey

The world of finance and the stock market has been a mystery to the world of prediction and it is remarkably gaining a lot of interest in terms of research to study the stock market prediction. The efficient-market and random walk hypothesis (Fama, 1965; Malkiel, 1973)[1][2], shows that there many techniques that were invented in order refute the validity of the denial of predictability of the financial markets.

Historical prices and other indicators if taken into consideration can reveal correlations and patterns of movement of the stock price and in a way predict future market standing of the stock in terms of its value. Advanced computational techniques like support vector machines, genetic algorithms, etc. have contributed to the understanding of this problem of predicting the future of the market stock prices. (Mackinlay, 1997) [3] demonstrated that the stock prices being assumed to be highly correlated and sensitive to sudden events adapts to learning the scenario for finding correlations in such random events and impact of the past events in predicting the future as devised by (Fung, Yu and Lam, 2002; Lavrenko et al. 2000)[4][5].

Reliable information holds the key as the primary objective of a financial market is to channel funds between firms or individuals. This further leads to investment opportunities for individuals. (Fama, 1965)[1] asserted in the effective-market hypothesis that the stock prices always depict the known information and the movement of the prices of the stocks in future are the impact of the external factors such as the random events of news or any unknown information that surrounds the market and the stock in particular.

Another aspect for the same problem comes from the field of machine learning where support vector machine (SVM) is a kernel-based learning algorithm introduced by Boser et al. (1992)[6] and Vapnik (1995)[7] which in a scenario of classification it then obtains the maximum-margin hyperplane as the decision boundary pushed against by those support vectors and thus become capable of extracting the global optimal solutions regardless of the sparsity of the training data and less over-fitted to it which is technique predominantly applicable to the stock market analysis as various insights about the stock market and the future trending of the stocks can be approximated and classified. There are various other techniques in the field of computing and mathematics which are aligned to such problems and recent advances in the fields have also created a wide scope of interpreting the data in a sensible fashion taking into consideration of the interactions in the open world and the impact of such interactions on the point of interest.

(Travers and Milgram, 1969)[8] in their work suggested that digital world is the root cause of the information being able to propagate faster and explore into various domains which is supported by the advances in the social media and the web. This is not only contributing to the larger extent of transforming data but also the outreach of the data to the various extreme extents via the social media utilizing the interactions between users. The fact that the community driven interaction between users also contributes to the discovery of the interrelationship between the data and the chain of reactions that can impact a certain product or stock in our case due to the interactions.

Many have researched in this field and it is proven that the impact of Social media data in various field of interest such as Bollen, Mao and Zeng (2010) [9] collected and classified tweets to forecast daily closing values of the Dow Jones Industrial Average and also many other such research implementations have been put forward by (Koch and Schneider, 2002; Antweiler and Frank, 2004; Wang, Jank and Shmueli, 2008;)[10][11][12]

The aim of this paper is to present a concept of a theory where one can simulate the stock market in virtual domain and an environment of social buzz and study the impact of such interactions in the environment on the stock prices. The research is also to suggest the architecture for the stock market simulation which adheres to social media and past stock market data analysis and computation.

II. The Experiment

The idea of this research is to understand the behavior of a stock exchange in terms of the flow of information, the changes that affect the stock market and the market behavior along with its impact on stock prices in future. The behavior of market not only limited to the corporate/public announcements but also the random event that affects the sentiment of the user and in turn affects the stock market.

International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 1, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

For the experiment there were assumptions from the development perspective:

a. Stocks generated for 20 virtual companies
b. The prices simulated for past 20 years and the next 15 years
c. Buzz Simulated comprised of 15000 opinions per stock
d. Sentiment adjustment for per stock per opinion per company
e. Abrupt events of significant 2X,5X orders done to provide the imbalance in the system
f. Random events every 2yr simulated to crash market by reducing random prices for random companies

III. Architecture and Discussion

The research took considerable amount of energy to come up with the architecture to model the actual market and stock exchange scenario.

The flow architecture shown in Figure i capture the data flow that takes places during the simulation. The Data Service Engine as shown in Figure ii is an in-memory engine which comprises of the generation of market data which was trained against the actual market behavior data and the simulation for market data generation can be said to be 43% similar to the actual market scenario. Here the Exchange simulation is done via study of real stock exchange and modeling it against random event based simulation. The data is generated for every company emotion by scoring algorithm for each one of them and the words generated for sentiment is an approximation to the score with the seed value for pseudo random percent division of the sentiments.

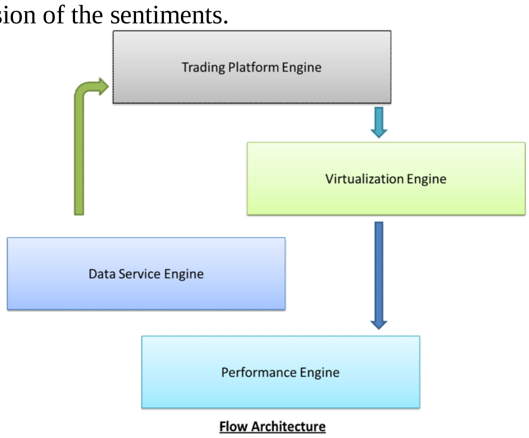

Fig.i. Flow Architecture of the idea

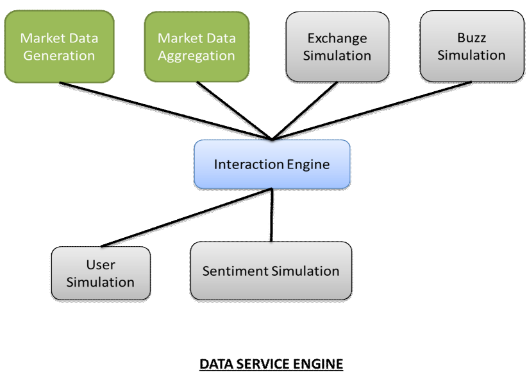

Fig.ii. Data Service Engine

Now the Trading Platform Engine also captures the balance mechanism the stock market proposes for every imbalance that occurs in realtime scenario. This is an actuation of the real-time event and the results were many false positives and the training of the datasets for price per stock, social sentiment per stock per company, adhoc order simulations were tuned for mashing up the actual result for the simulation and risk calculation done to mitigate the market fall due to the discrepancy in the stock prices.

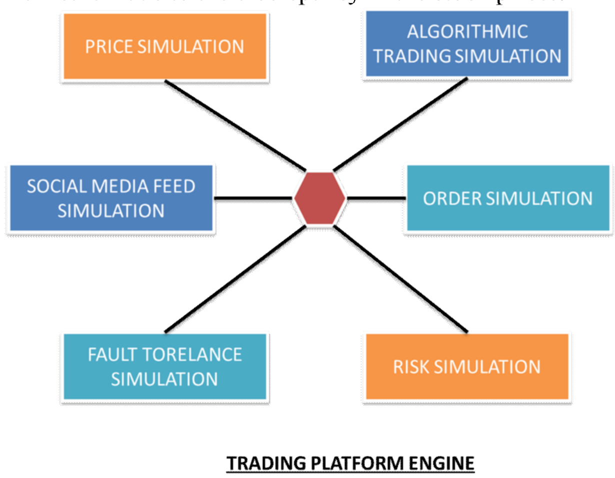

Fig.iii. Trading Platform Engine

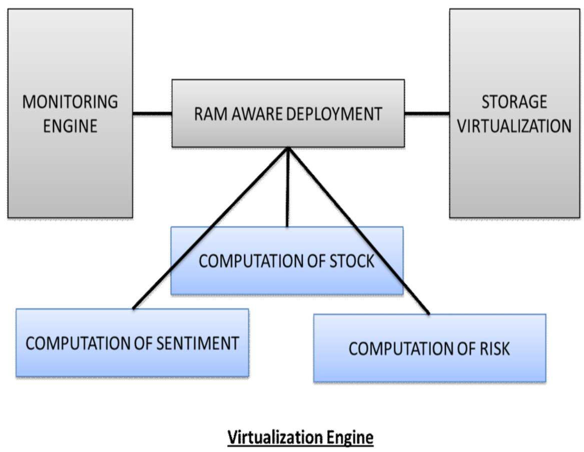

Fig.ii. Virtualization Engine

The Virtualization Engine as shown in Figure iv above is the part of the architecture where the Stock prices, Sentiment Scores and Risk Scores are crunched per stock per company per hour for the final resultant of the stock prices. This is assisted by a Monitoring Engine which keeps track for RAM address per computation scores and then the results are stored in a virtualized environment.

The simulation process is completely in-memory and the sentiment engine is vital for the operation which is based on the calculation formula as follows:

$SS1 = PS/MA100$

$SS2 = NS/MA100$

$SS_{Final} = SS1 - SS2$

SS1= Sentiment Score1, SS2= Sentiment Score2, SS_{Final}= Sentiment Score Final, PS= Positive Sentiment, NS= Negative Sentiment, MA100= 100 day Moving average

IV. Observation

The table 1 shows the generated messages from the system which are simulated via sentiment engine and the analytics are done to understand the intention of the messages related to the company and as shown there are positive and negative message classification along with the unknown classification which analyzed messages for the meaning which rendered to be positive or extremely negative as shown by the negative values.

There are several analytical errors which may have resulted in this computation which we are currently working on to figure out the actual standard against which the error can be reduced. For the part of trading computations, we compute abnormal returns as the difference between raw returns minus returns on a value-weighted portfolio of firms with similar size, book-to-market ratio and past returns inspired from (Daniel et al. 1997) [11][12]. Also abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market/past return-characteristics from t+3 to t+60, where t is the day of article appearance or the ensuing trading day if the article is published on a non-trading day [12].

Table 1: Message Distribution for various sentiments grouped into 3 categories with false positives included

Company	Messages Generated	Positive Sentiment	Negative Sentiment	Unknown
A	121441540	50210302	32623343	17586959
B	82226210	4366543	4114298	252245
C	26424021	7589063	933734	6655329
D	71739782	16559468	5856069	10703399
E	51716662	25152378	7998320	17154058
F	107892865	38315839	28359014	9956825
G	37648885	9235009	10848379	-1613370
H	75734634	4894091	8365201	-3471110
I	53054138	1617230	11082884	-9465654
J	33524374	3035490	5139124	-2103634
K	65511616	28790607	3715604	25075003
L	69013412	25971461	5454162	20517299
M	62866511	19627795	4948861	14678934
N	102751118	40548965	10273600	30275365
O	36005428	7425391	5647328	1778063
P	4107464	1414852	955324	459528
Q	94510925	2001721	485011	1516710
R	42946863	4836432	4104192	732240
S	66709571	3173793	7756932	-4583139
T	7136379	2262700	1411324	851376

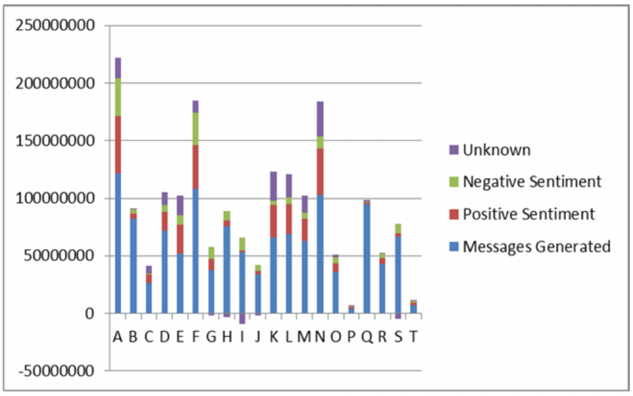

Fig.iii. Sentiment Generation for the 20 companies

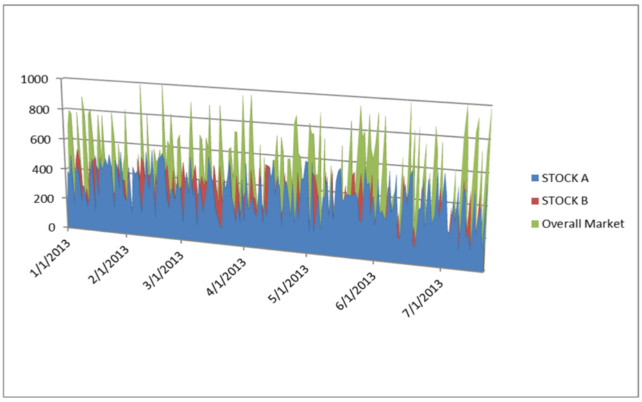

Fig.iv. Comparison of cumulative returns on investments for Year 2013 (simulated for 2 stocks and overall market)

The figure vi shows the simulated market behavior for returns on investments for the year 2013 which was simulated for stocks for company A and company B against the overall behavior of the stocks from all the companies and this is highly random and this we are working on presently to reduce the randomness of the events which are the direct outcome of the Risk Engine.

V. Future Work

The present work discussed in this paper is the preliminary study of the field of stock market and the related analysis and the outcome is basic understanding of the dynamics of the field and the results obtained are random in behavior which we are working on presently to modify and generate streamline results and since the market real-time accuracy may not be achieved but we are trying to model the real-time scenario to well understand the behavior of the field and also how we can process the in-memory algorithms in the architecture to converge to the results which will be more valuable.

Acknowledgment

The authors would like to thanks all their team members and the management for the immense support and motivation. The authors would also like to acknowledge the special support from Mr. Ganapathy Subramanian, Vice President, Sap Labs, Bangalore, India.

References

[1] Fama, Eugene F. (September/October 1965). "Random Walks In Stock Market Prices". Financial Analysts Journal 21 (5): 55–59. doi:10.2469/faj.v21.n5.55. Retrieved 2008-03-21.

[2] Malkiel, Burton G. (1973). A Random Walk Down Wall Street (6th ed.). W.W. Norton & Company, Inc. ISBN 0-393-06245-7.

[3] MacKinlay, A. C. "Event Studies in Economics and Finance," Journal of Economic Literature Vol. XXXV, Issue 1 (March 1997).

[4] G. Fung, J. Yu, and W. Lam, "News sensitive stock trend prediction,"Lecture Notes in Computer Science, vol. Volume 233, 2002, p. 481–493.

[5] Lavrenko, V.; Schmill, M.; Lawrie, D.; Ogilvie, P.; Jensen, D.; Allan, J.: Mining of Concurrent Text and Time Series. In: Proceedings 6th ACM SIGKDD Int.Conference on Knowledge Discovery and Data Mining. Boston 2000, pp. 37-44

International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 1, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

[6] Travers, Jeffrey, and Stanley Milgram. "An experimental study of the small world problem." Sociometry (1969): 425-443.
[7] Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." Journal of Computational Science 2.1 (2011): 1-8.
[8] Koch, Stefan, and Georg Schneider. "Effort, co-operation and co-ordination in an open source software project: GNOME." Information Systems Journal 12.1 (2002): 27-42.
[9] Antweiler, Werner, and Murray Z. Frank. "Is all that talk just noise? The information content of internet stock message boards." The Journal of Finance 59.3 (2004): 1259-1294.
[10] Wang, Shanshan, Wolfgang Jank, and Galit Shmueli. "Explaining and forecasting online auction prices and their dynamics using functional data analysis." Journal of Business & Economic Statistics 26.2 (2008).
[11] Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, Journal of Finance 52, 1035–1058.
[12] Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in stock returns, Journal of Finance 52, 1–33.

AUTHOR'S PROFILE

Prabal Mahanta

was born in the city of Guwahati, Assam, India in 1985.

Mr. Mahanta holds a Master's degree in Information Technology (M.Tech) with majors in Embedded Systems from International Institute of Information Technology (IIIT), Bangalore, Karnataka, India. He is currently working as a Developer at SAP Labs, Bangalore, India. He has worked extensively on critical customer projects as well as research projects. He has a strong inclination for research and teaching. He has published many papers in national and international conferences and his research interest lies in in-memory computing, cloud-infrastructure, chaos theory and real time simulations.

Suhas Sudheendra

was born in the city of Mysore, Karnataka, India in the year 1983.

Mr. Suhas holds a Bachelor's degree in Computer Science and Engineering from Visvesvaraya Technological University Belgaum, Karnataka, India. He is currently working as a Senior Developer at SAP Labs, Bangalore, India. He has worked extensively on research projects from the beginning of his career at Infosys which includes performance engineering and application optimization. Current areas of interest include in memory computing and real time analytics.