International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Corporate Communication Network and Stock Price Movements: Insights from Data Mining

M.S.S.N.R vamsi[1], M.chaitanya varma[2], M. karthik chowdary[3], vidhya prakash [4]

[1,2,3] *Department of Computer Science and Engineering,R.M.K. Engineering College,TamilNadu,India*

[4] *associate Professor,Department of Computer Science and Engineering,R.M.K. Engineering College, Tamil Nadu, India*

Abstract: *The aim of the project is to know the stock price movements through data mining of previous emails exchange of the employees. Now we have to know whether there is any frequency exits in the emails exchange of the employees which going to reflect in the stock price movement. If such relationships do exist, we would also like to know whether or not the company's stock price could be accurately predicted based on the detected relationships. To know the exact relationship between them, an data mining algorithm is used to mine communication records and previous stock prices so that based on the executed result we can able to the predict changes in stock prices. Using the data-mining algorithm and a set of publicly available Enron email corpus and Enron's stock prices recorded during the same period, we discovered the existence of interesting, statistically significant, association relationships in the data. By the result we also discovered that the result can also predict stock price movements with also an exact accuracy of 80%. Given the increasing popularity of social networks, the mining of interesting communication patterns could provide insights into the development of many useful applications in many areas.*

I. Introduction

Latest researches has shown the presence of interesting communication patterns among various users of various social network platforms. These arrangements have been shown to be useful in prognosticating the sales of the commodity and their shares of the stock. Correlated to a social network, which can considered for showing connections among various users in the social,a corporate network associates only members in a corporation. While users of a social network can express of several interests, representatives of a corporate communication network are mainly supposed to talk about company specific business. If human communication patterns can be discovered in the social networks to forecast products.

Different social networks, in a corporate communication network, e-mails have long been used as a tool for inter organizational and intra organizational information swap.

Unlike social networks, in a corporate communication network, e-mails have long been used as a tool for interorganizational and intra organizational information exchange. In the same way, a social network platform is able to capture participants' behavior and their opinions about various issues and events. Thus, we argue that a corporate communication network in the form of an e-mail ecosystem also contains insightful information, such as organizational stability and robustness [4], about a company's development.

We believe our argument is in line with corporate communication theory [5], which suggests that "employee communications can mean the success or failure of any major change program" resulting from a merger, acquisition, new venture, new process improvement approach, or other management issues. In other words, employee communication can serve a critical "business function that drives performance and contributes to a company's financial success" [6].

Based on these broad corporate communication theories, we hypothesize that every company has its own communication approach with identifiable patterns. We believe that these communication patterns can reflect how a company man-ages major corporate activities (such as mergers, acquisitions, new ventures, new process improvement approaches, going concerns, or bankruptcy) that may subsequently affect the company's performance in the stock market.In this paper, we propose that a company's performance, in terms of its stock price movement, can be predicted by internal communication patterns. To obtain early

warning signals, we believe that it is important for patterns in corporate communication networks to be detected earlier for the prediction of significant stock price movement to avoid possible adversities that a company may face in the stock market so that stakeholders' interests can be protected as much as possible. Despite the potential importance of such knowledge about corporate communication, little work has been done in this important direction.

It is for this reason that we are proposing in this paper to make use of a computational approach to determine if patterns detected in a corporate communication network are related to corporate performance. There has been some effort to use computational methods to mine large-scale social media data for public sentiment [3] toward the stock market.

Specifically, we determine whether or not there exist any association relationships between the frequency of e-mail exchange of the key employees in a company and the performance of the company as reflected in its stock prices. To the best of our knowledge, there has not been any attempt to investigate possible linkages between corporate communications data and a firm's share price. If such relationships do exist, we would also like to know whether or not the company's stock price could be accurately predicted based on the detected relationships.

To test our hypothesis, we chose to look into the case of Enron. The reason why Enron is chosen is that an Enron email corpus has been made publicly available. By mining the data set to determine if the communication patterns discovered have any association with Enron's stock prices recorded during the same period, we managed to confirm that interesting, statistically significant, association relationships do exist in Enron's corporate communication network. In addition, we also discovered that the detected relationships could predict stock price movements with relatively high levels of accuracy. Such results confirm our belief that corporate communication has identifiable patterns and such patterns can reveal meaningful information about corporate performance.

II. Related work

There have been some attempts by researchers to empirically investigate the impacts of a wide range of communication variables (such as communication patterns, feedback, culture, and specific behaviors) on performance-related variables (such as communication satisfaction, job satisfaction, individual performance and productivity and organizational performance, and productivity and change) [9]. For instance, the productivity of an individual in a corporation was found to be explainable by certain communication patterns that cover supervisory communication, subordinate communication, personal feedback, organizational integration, media quality, corporate information, communication climate, and coworker communication correlation [10]. More specifically, the downward form of communication was found to exert the greatest influence on productivity while co-workers' communication rendered the least impact. These findings are in line with other empirical experiments [11], [12] that produce results indicating that communication flow is one of the key determinants of corporate performance.

In some specific domains, research [13], [15] seems to indicate that communication patterns can have significant impact on a corporation's research and development (R&D) performance. In marketing services and research domain, upward communication, i.e., from customer-contact personnel to top management, has been consistently shown to provide valuable information to an organization about performance of products and services [16].

In addition to internal communication among employees, there have been some suggestions that there exists a direct association between public communication strategy and share prices. For instance, corporate communication with key stakeholders has been shown in [19] to enhance potential payoffs, while stock appreciation and corporate performance have been found to be uncorrelated if communication efforts (with outsiders) are faulty. Such findings echoed studies on the strategic role of investor relations in which communication of corporate information is crucial for the perception and evaluation of the financial community including analysts, investors, and potential investors [20], [21].

To summarize, according to a broad body of research, there has been some evidence to indicate that corporate communication is linked with organizational performance. We argue that each corporation has its own communication patterns for both its internal and external stakeholders. All the communication activities capture critical events (such as merger, acquisition, new venture, new process improvement approach, going concern, or bankruptcy) that can occur in a firm and so may contribute to the performance of the firm's stock on the stock market. In other words, this suggests that a firm's stock performance may be linked with merger, acquisition, R&D, marketing, and public communication patterns. However, little work has been done to investigate this phenomenon. As a result, discovering communication patterns in a corporate network to predict share prices remains to be a problem that is largely investigated.

III proposed work

In parallel with the above research, researchers examined the correlation between web buzz and stock market. Minimum *et al.* [46] developed a simple model to analyze the communication dynamics in the blog sphere by covering the number of posts, the number of comments, length and normalized response times of comments, strength of comments, size of the early responder and outliers. These communication dynamics were then used to determine their correlations with stock market movement. More recently, Gilbert and Karahalios [14] used over one million posts from the website http://www.livejournal.com/ to create an index of the U.S. national mood, namely, the anxiety index. These studies have provided rich insights on how to better predict the stock market movement. However, the algorithms used in the above studies ignored the correlations between a company's stock price and the communication network within the company, which may be even more influential and meaningful to observe the stock price movement of a company. It is also worth highlighting the challenge of mining temporal association rules from stock price.

A. Notations and Definitions

Definition 1 *(Communication Networks):* Let G = {G_1, $G_2 \ldots G_t, \ldots G_p$} represents the communication networks (p is the total time points). The one of networks is represented as a graph with a 2-element tuple $G_t = (V_t, R_t)$, with the following characteristics.

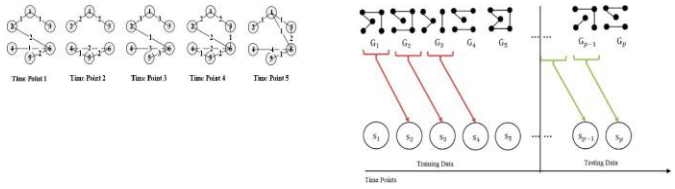

Fig. 1. Example of description data.

1) $V_t = \{v_1, v_2, \ldots v_n\}$ represents a finite set of nodes (employees in the communication network), n is the total number of nodes in the network, and v_c represents the central node in the network. In the Enron case, we choose the CEO as the central node and examined the communication network between the CEO with other employees.

B. Proposed Algorithm

In this section, the details for discretization of communication network and stock price are introduced first. Then, in order to predict the stock price using communication network, a pattern discovery problem of network data is formulated. Finally, prediction rules are constructed to predict stock price using e-mail communication network of Enron without content of email.

1) Discretizing Communication Matrix and Stock Price: The value of edge (d_{ij}) between two different nodes for representing communication frequency between two nodes is numerous. In order to reduce and simplify the original data, numerous values are

always replaced by a small number of interval labels, which leads to a concise easy-to-use, knowledge-level representation of mining results [37]. The existing and mature methods on unsupervised discretization are primarily equal frequency discretization and equal width discretization [31]. The equal width method is typically used in every statistic program to produce regular histograms [47]. However, equal width discretization can hardly handle the situation if outliers exist in the data set. Equal frequency can overcome the limitations of the equal width discretization by dividing the domain in intervals with the same distribution of data points [48]. Hence, considering different people may have different habits to exchange emails with each other, we use equal frequency [31], the most represented algorithm, to discretize other nonzero value for the value of weight in the communication network.

o_{lk} (e_{lk})		***Stock Price*** (in the certain position later)					**Total**
		ds_1	…	ds_i	…	ds_I	
$E^c_{j,t-1}$	dr_1	o_{11} (e_{11})	…	o_{i1} (e_{i1})	…	o_{I1} (e_{I1})	o_{+1}
	. . .	. . .		. . .		. . .	. . .
	dr_j	o_{1j} (e_{1j})	…	o_{ij} (e_{ij})	…	o_{Ij} (e_{Ij})	o_{+j}
	. . .	. . .		. . .		. . .	. . .
	dr_J	o_{1J} (e_{1J})	…	o_{iJ} (e_{iJ})	…	o_{IJ} (e_{IJ})	o_{+J}
Total		o_{1+}	…	o_{I+}	…	o_{I+}	M′

B. Experimental Result With Comparisons

1) Different Time Interval: In the application of stock prediction, one may ask if we are to predict tomorrow's stock movement based on three days ago or a week ago? In order to answer this question, we considered how many days we should look back when predicting price movement of the current day.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithm, the Enron Corpus is used as a real-life data set for experiment. We predict stock price movements based on Enron's e-mail communication network documented in 2000–2001. According to the timeline of the email corpus, we chose the amount of increases and decreases of stock price of Enron from January 1, 2000 to December 31, 2001. In order to achieve the best prediction result, we first identify the patterns of the relationships between e-mail communication and stock price and determine the length of time interval through the prediction result. Then, the whole e-mail communication corpus that involved 150 Enron employees is used to predict the stock price

TABLE V
PREDICTION ACCURACY WHEN CONSIDERING DIFFERENT TIME DELAYS

Time Delay	Training Data	Testing Data	The Proposed Algorithm (Acc.)
1 day	650	80	43.75%
	700	30	53.33%
2 day	649	80	**61.25%**
	699	30	63.33%
3 day	648	80	60%
	698	30	**66.67%**
4 day	647	80	56.67%
	697	30	60%
5 day	646	80	55%
	696	30	63.3%

2) Determine the Time Interval: If discovering daily pat-terns does not give a high enough accuracy for one to risk investing money based on the proposed algorithm pre-diction, one may wonder if the algorithm may be able to predict a semiweekly three-day or a weekly five-day average. Hence, in order to determine the most suitable time window (in terms of days) to estimate the impact of stock price change in the future, we predict the stock price movement using different time intervals. The prediction rule is specified in a time interval, saying "one day," "three days" for the short-term perspective, "one week" for the long-term perspective. That is to say, when "one day" is considered as the time interval, the proposed

algorithm focuses on discovering the patterns between communication frequency in the previous day and the stock price in the next day. Similarly, when "one week" is considered as the time interval, the proposed algorithm focuses on discovering the patterns between the average value of communication frequency in the previous week and the average price of the stock in the next week. Table VI shows the prediction result using the proposed algorithm when different time intervals are considered. We also compare these results with the traditional classification algorithm, i.e., Decision Tree, in Table VI, with the details below.

TABLE VI
PREDICTION ACCURACY USING DIFFERENT EVENT WINDOWS AND DIFFERENT NUMBERS OF TRAINING/TESTING DATA

Event Window	Total Time Point	Training Data	Testing Data	Decision Tree (Acc.)	The Proposed Algorithm (Acc.)
One Day	730	625	105	40%	41.9%
		650	80	47.5%	43.75%
		675	65	38.46%	36.3%
		700	30	46.67%	53.33%
Average				43.16%	43.82%
Three Days	243	170	73	45.21%	63.01%
		180	63	46.03%	61.09%
		190	53	47.17%	66.3%
		200	43	46.51%	62.8%
Average				46.23%	63.3%
One Week	105	80	25	43.33%	**80%**
		85	20	40%	**85%**
		90	15	40%	**86.67%**
		95	10	50%	**80%**
Average				43.33%	**82.92%**

TABLE VII
MOST SIGNIFICANT FACTORS THAT AFFECTS MOVEMENT

$E^c_{j,t-1}$	E_t	
	Up	Down
	$E^c_{i,t-1} = Weak$	$E^c_{i,t-1} = Strong$

We show how the stock price movement [which can take the values of up (U), down (D)] may be affected by the commu-nication frequency between CEO and other staffs [which can take the values of weak (W) and strong (S)]. From Table VII, we can conclude that when the frequency of communication between CEO and other staffs is weak, the stock price would increase in the next week, while if the communication between CEO and the staff is strong, the stock would decrease in the next week. We can explain it as when the stock price of the company is steady, CEO does not need to contact many other staffs frequently, but when CEO contact others frequently, it may mean that there would be some problem of company development, which leads stock price goes down.

3) Results of Predicting the Stock Price Movement: After determining the time interval, we can conclude that the pro-posed algorithm is effective for predicting stock price in the next week using communication frequency in the previous week. Then we consider different levels for weekly pricing limits to conduct the experiments in more detail for weekly stock price prediction.

We classified the changes of stock price into different levels.

1) Two levels covering UP (i.e., the amount of increases of stock price is higher than 0%) and DOWN (i.e., the amount of decreases of stock price is lower than 0%)

2) Three levels considering: a) UP as the stock price increases more than 1%; b) DOWN as the stock price decreases more than −1%; and c) EQUAL as the change level in the interval [−1%, 1%].

3) Five levels considering: a) "large increase" and "large decrease" in which the stock price changes are more than 3% or −3%, respectively; b) "increase" when the changes between 1% and 3%; c) "decrease" when the changes between −3% to −1%; and d) "equal" when the stock price changes are between −1% and 1%.

TABLE VIII
PREDICTION RESULT FOR DIFFERENT LEVELS OF STOCK PRICE W[illegible] DIFFERENT TRAINING DATA SETS (TIME INTERVAL = ONE W

Event Window	Training Data	Testing Data	Decision Tree (Accuracy)			Proposed Algorithm (Accuracy)		
			Two Levels	Three Levels	Five Levels	Two Levels	Three Levels	Five Levels
One week	75	30	43.33%	36.67%	10%	76.67%	73.33%	66.67%
	80	25	40%	34.62%	12%	80%	76.67%	68%
	85	20	40%	30%	15%	85%	80%	70%
	90	15	50%	25%	13.3%	86.67%	80%	66.67%
	95	10	43.33%	33.3%	10%	80%	80%	70%
Average Accuracy			43.33%	31.92%	12.06%	81.67%	78%	68.27%

V. CONCLUDING REMARKS

The findings and theoretical implications from this paper are two fold. On one hand, we captured the communications among nodes in Enron's major corporate communication network and identified employees' communication patterns. This paper demonstrates that a corporate e-mail ecosystem contains meaningful information about employees' communication patterns. Even if we only focus on the communication frequency, a company (Enron in our case) has identifiable pat-terns of e-mail exchange. Such identifiable patterns can reveal important information about major corporate activities and organizational stability that may subsequently influence the focal company's performance in the stock market. Therefore, cooperate communication patterns can serve as a good proxy to predict a company's stock performance. Our experimental results demonstrated the existence of dependence between e-mail communication network and stock price for Enron. This paper extended the existing communication theories to capture the patterns of corporate communication and the focal company's stock performance.

On the other hand, social networks have become a hot topic in the field of data mining recently. In this paper, we not only provided an innovative idea on using data-mining algorithms but also constructed the relationship between social network and finance. Hence, this paper demonstrated great potential to predict the amounts of increases and decreases of stock price based on the weighted rules.

REFERENCES

[1] G. Miller, "Social scientists wade into the tweet stream," *Science*, vol. 333, no. 6051, pp. 1814–1815, 2011.

[2] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter?— An empirical investigation of panel data," *Decision Support Syst.*, vol. 45, no. 4, pp. 1007–1016, 2008.

[3] B. Li, K. C. C. Chan, and C. Ou, "Public sentiment analysis in Twitter data for prediction of a company's stock price movements," in *Proc. IEEE Int. Conf. e-Bus. Eng. (ICEBE)*, Nov. 2014, pp. 232–239.

[4] B. Collingsworth, R. Menezes, and P. Martins, "Assessing organizational stability via network analysis," in *Proc. IEEE Symp. Comput. Intell. Financial Eng.*, Mar. 2009, pp. 43–50.

[5] D. J. Barrett, "Change communication: Using strategic employee com-munication to facilitate major change," *Corporate Commun., Int. J.*, vol. 7, no. 4, pp. 219–231, 2002.

[6] K. Yates. (2003). *Effective Employee Communication Linked to Greater Shareholder Returns, Watson Wyatt Study Finds*. Accessed: Nov. 2003. [Online]. Available: http://www.watsonwyatt.com/render.asp?catid= 1&id=12092

[7] M.-C. Wu, S.-Y. Lin, and C.-H. Lin, "An effective application of decision tree to stock trading," *Expert Syst. Appl.*, vol. 131, no. 2, pp. 270–274, 2006.

[8] E. Vamsidhar, K. V. S. R. P. Varma, P. S. Rao, and R. Satapati, "Prediction of rainfall using backpropagation neural network model," *Int. J. Comput. Sci. Eng.*, vol. 2, no. 4, pp. 1119–1121, 2010.

[9] C. W. Down, G. C. Phillip, and A. L. Pfeiffer, "Communication and organizational outcomes," in *Handbook of Organizational Communica-tion*, G. Goldhaber and G. Barnett, Eds. Norwood, NJ, USA: Ablex, 1988.

[10] P. G. Clampitt and C. W. Downs, "Employee perceptions of the relationship between communication and productivity: A field study,"

J. Bus. Commun., vol. 30, no. 1, pp. 5–28, 1993.

[11] G. S. Hansen and B. Wernerfelt,

"Determinants of firm performance: The relative importance of economic and organizational factors," *Strategic Manage. J.*, vol. 10, no. 5, pp. 399–411, 1989.

[12] W. W. Burke and G. H. Litwin, "A causal model of organizational performance and change," *J. Manage.*, vol. 18, no. 3, pp. 523–545, 1992.

[13] R. Katz and M. Tushman, "Communication patterns, project perfor-mance, and task characteristics: An empirical evaluation and integration in an R&D setting," *Org. Behav. Hum. Perform.*, vol. 23, no. 2, pp. 139–162, 1979.

[14] E. Gilbert and K. Karahalios, "Widespread worry and the stock market," in *Proc. 4th Int. AAAI Conf. Weblogs Social Media (ICWSM)*, 2010, pp. 59–65.

[15] R. Katz, "The effects of group longevity on project communication and performance," *Admin. Sci. Quarterly*, vol. 27, no. 1, pp. 81–104, 1982.

[16] V. A. Zeithaml, L. L. Berry, and A. Parasuraman, "Communication and control processes in the delivery of service quality," *J. Marketing*, vol. 52, no. 2, pp. 35–48, 1988.

[17] C. Hargreaves and Y. Hao, "Prediction of stock performance using analytical techniques," *J. Emerg. Technol. Web Intell.*, vol. 5, no. 2, pp. 136–142, 2013.

[18] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA,

USA: Morgan Kaufmann, 1993.

[19] R. B. Higgins and B. D. Bannister, "How corporate communication of strategy affects share price," *Long Range Planning*, vol. 25, no. 3, pp. 27–35, 1992.

[20] R. R. Dolphin, "The strategic role of investor relations," *Corporate Commun., Int. J.*, vol. 9, no. 1, pp. 25–42, 2004.

[21] C. Marston and M. Straker, "Investor relations: A European survey," *Corporate Commun., Int. J.*, vol. 6, no. 2, pp. 82–93, 2001.

[22] A. Corrada-Emmanuel, A. McCallum, P. Smyth, M. Steyvers, and

C. Chaitanya, "Enron Email dataset," in *Proc. Workshop Link Anal., Counterterrorism Secur.*, 2005, pp. 33–44.