An Event-Driven SIR Model for Topic Diffusion in Web Forums

Jiyoung Woo*
Graduate School of Information Security
Korea University
Seoul, Korea
jywoo@korea.ac.kr

Hsinchun Chen
Eller College of Management
University of Arizona,
Tucson, Arizona, USA
hchen@eller.arizona.edu

***Abstract*—Social media is being increasingly used as a communication channel. Among social media, web forums, where people in online communities disseminate and receive information by interaction, provide a good environment to examine information diffusion. In this research, we aim to understand the mechanisms and properties of the information diffusion in the web forum. For that, we model topic-level information diffusion in web forums using the baseline epidemic model, the SIR(Susceptible, Infective, and Recovered) model, frequently used in previous research to analyze disease outbreaks and knowledge diffusion. In addition, we propose an event-driven SIR model that reflects the event effect on information diffusion in the web forum. The proposed model incorporates the effect of news postings on the web forum. We evaluate two models using a large longitudinal dataset from the web forum of a major company. The event-SIR model outperforms the SIR model in fitting on major spikey topics that have peaks of author participation.**

Keywords-information diffusion; epidemic model; web forum

I. INTRODUCTION

The Internet supplies new forms of interactivity such as online forums, chat rooms, and Weblogs [1]. The interactivity makes people disseminate information by posting blog entries and discussing their opinions in web forums. According to previous research, information diffusion on the web influences social opinion and has the power to affect society economically and politically [2]. Moreover, views on the web are sometimes surprisingly accepted and transmitted to others without much critical examination [3]. This highly contagious property accelerates the information diffusion on the web.

As social media becomes more prevalent, it is necessary to understand the mechanisms and properties of information diffusion through these new publication methods for political, economic, and marketing purposes. Kleinberg [4] stated that a rumor, a political message, or a link to a web page are all examples of information that can spread from person to person, contagiously, in the style of an epidemic. In this study, we will study the information diffusion process using an epidemic model. We will adopt the SIR (Susceptible, Infective, and Recovered) model [5], the popular epidemic model to interpret information diffusion in the social media using the same principles of the epidemic model. In addition, we will consider the effect of online news media to the social media and propose the event-driven SIR model to test this effect.

Among popular social media, web forums have become new sources for information diffusion where people who have common interests express and discuss their opinions and thereby affect others. The interaction occurs when authors post threads and other authors reply to the threads; this interaction causes authors to infect others' opinions. Web forums where participants both disseminate and receive information online and form self-contained communities, are promising for modeling information diffusion using the epidemic model. Despite its attractiveness, the high volume of data and the casual writing styles of participants make using web forum opinions difficult. Significant efforts and advances are needed to collect and analyze highly complex user interactions in the web forum. The examination on information diffusion using the SIR model and the event-driven SIR model will be performed on a large web forum of a major retail company.

II. RELATED WORKS

Information diffusion includes the spread of technological innovations, the word-of-mouth effect in marketing, the spread of news and opinions, etc. It occurs through the interaction between people, i.e. social contagion. The epidemic models that have been designed for disease contagion have also been used for social contagion, with minor modifications to explain social phenomenon. The innovation diffusion models have been designed to explain product/innovation adoption, and they have been used for explaining social contagion. While the epidemic model mainly reflects the interaction between people, the innovation diffusion incorporates the self-evolving innovation and decision making process. The aim of the diffusion model is to understand how the spread of new diseases/ideas/products works, to predict their success or failure in the early stages, and to shape the underlying process to increase or reduce the chances of diffusion. These diffusion models are recently used for modeling information diffusion in social media.

Pioneering contributions to information diffusion modeling using epidemic models have been made using the study on the spread of scientific ideas. Goffman and Newill [6] suggested that the transmission and development of scientific information within a population can be examined in

terms of infectious disease. Bettencourt et al. [7, 8] studied how a physical theory has been adopted by physicists using the SIR model and suggested the competency model to describe the diffusion process of two competing theories. Epidemic models were also tested for fitness to rumor propagation [9]. Network-based epidemic models were applied to the propagation of computer viruses in the late 1990s and early 2000s [10], [11].

More recently, as online information diffusion has become a major topic for diffusion studies due to the growth of email, web, and social media. New attempts to apply diffusion models to email, blogs, and forums have emerged. Among other social media, blog is intensively studied. Blog communications as the proxy of information diffusion in online networks have attracted great attention from researchers. The intensive literature on social networks led to the studies on information diffusion in blogs with the assumption that the blog implies a real-world social network. People reflect their social relationships in blogs and build online social networks through blogs.

Much previous research on information diffusion in blogspace employed the independent cascade model (ICM) [12] and showed that the diffusion process takes place through contact between neighbors in the social network. Gruhl, Guha, Liben-Nowell and Tomkins [13] defined the characteristics of diffusing topics in the blog and proposed a method for estimating the transmission probability for ICMs. Saito, Nakano and Kimura [14] used expectation maximization to estimate the transmission probability for ICMs. Recent blogspace research studies the cascading effect. Close examination of the cascading behaviors of customers' influence can lead to viral marketing applications. Leskovec, Admic and Huberman [15] analyzed the pattern of cascades using a large blog dataset and suggested the cascade generation model under the SIS(Susceptible, Infected, Susceptible) framework with fixed transmission probability.

The research on web forums differs from that on blogs in that it focuses on the diffusion vector such as topics, news, and documents. Kubo, Naruse, Sato and Matubara [16] showed the analogy between the disease propagation model, the SIR model, and posting data in web forums. Zhou, Guan, Zhang and Zhang [17] predicted the tendency of topic discussion in online social networks using a dynamic probability model that embeds individual interest factor, behavior factor, and time lapse factor. Bampo, Ewing, Mather, Stewart and Wallace [18] applied the SIR model on various ideal networks to measure the efficiency of email marketing campaigns. In previous work, we adopt the SIR model to model violent topic diffusion in the Jihadi forum [19].

Research on information diffusion through social media has been limited to blogs, and the studies of information diffusion in blogs have focused on how the network structure affects information flow. There has been little research on information diffusion in web forums. In addition, previous research lacks consideration of the influence of external events on the diffusion process.

III. EPIDEMIC MODEL FOR WEB FORUMS

A. SIR Model for web forums

The dynamics underlying the diffusion of ideas/opinions hold many qualitative similarities to disease infections [7]. Disease infection spreads through contact; it starts when a few individuals are infected. Each infected person then has contact with others. Then individuals who had contact with infected people become infected themselves at a certain rate. These infected individuals come into contact with additional individuals. This infection process continues until no more susceptibles or infectives exist.

In the web forum context, an initiating author begins a discussion on a topic by posting a thread; the author becomes an infective. Other users who have a certain level of interest in a topic, said to be susceptibles, will read and post comments on the thread. Some commenters and readers will post other threads, thereby infecting others with their posts. These authors belong to the infective class. After a certain period of time, some authors will stop participating in discussions and lose the power to influence others, said to be recovered. This interactive process leads to the flow of discussion on a topic from one author to another.

The possible interaction between future authors and current authors is expressed as the multiplication of the number of susceptibles, S(t) and the number of infectives, I(t), S(t)I(t). Among all possible interaction, some portion is effective in infecting susceptibles. The ratio of effective interactions over all possible interactions is the infection rate, α. The interaction transfers $\alpha S(t)I(t)$ amount of authors from the susceptible class to the infective class per time period. Infectives are recovered at the rate of β. $\beta I(t)$ amount of authors are recovered per time, thus moving from the infective class to the recovered class. The total population, N(t) is the sum of the susceptible class, the infective class, and the recovered class. The total population consists of the people directly involved in the diffusion process.

Since we aim to model the long-term diffusion process not an instant diffusion, we consider the population growth. The plausible model that reflects the growth of forum users and the growth of users' interests must allow the total population to vary in time. We adopt logistic growth model where population grows based on population size and there is an upper limit to the population size. The upper limit is called the carrying capacity, K. The population grows proportionally to its size, N(t), and the remaining capacity, (K-N(t)). Because incoming users are initially susceptible, we add the logistic growth term into the differential equation of the susceptible class. The logistic growth of susceptible authors is represented as $\mu(\frac{K-N(t)}{K})N(t)$. The interaction between authors, the growth of forum users, and the decay of influence of past authors are formulated using differential equations as shown in equation (1). S(t), I(t), and R(t) are the number of susceptible, infective, and recovered respectively, and s(t), i(t), and r(t) are their derivatives.

$$
\begin{aligned}
&s(t) = \mu(\frac{K - N(t)}{K})N(t) - \alpha S(t)I(t) \\
&i(t) = \alpha S(t)I(t) - \beta I(t) \\
&r(t) = \beta I(t) \\
&s(t) = \frac{dS}{dt} \quad \text{at time t} \\
&i(t) = \frac{dI}{dt} \quad \text{at time t} \\
&r(t) = \frac{dR}{dt} \quad \text{at time t} \\
&\text{S(t) : the number of future authors at time t} \\
&\text{I(t) : the number of current authors at time t} \\
&\text{R(t) : the number of past authors at time t} \\
&\text{N(t) = S(t) + I(t) + R(t) : the number of total authors at time t} \\
&\text{K : carrying capacity} \\
&\mu \text{ : population growth rate} \\
&\alpha \text{ : infection rate} \\
&\beta \text{ : recovery rate}
\end{aligned}
\tag{1}
$$

B. Event-driven SIR Model

Some discussion topics come from news media and the spread of a topic in the web forum is affected by news media. Previous research has proven that news media influences public opinion [20, 21]. When an event related to a topic occurs, news media will report about the event immediately. To detect an event from news posting, we employ the event definition that is usually used in event detection in text mining. An event is represented by a centroid aggregated from the contents of related documents or is defined as a collection of documents within a recent time window [22]. In the web forum context, an event is defined as the excessive news postings. The importance of the event is proportional to the number of news postings about it. This suggests that we can detect the event impact on topic diffusion from the number of news posting on the topic. When the number of postings on a topic exceeds the average number of postings on the topic, we assume that there is an event at that time. This event ramps up people's interest in the topic due to their exposure to news media. The event could accelerate the infection rate, meaning that the infection rate of posts increases as the result of the growth of authors' interest. The event attracts potential new authors on a topic by frequent postings, so it increases the number of susceptibles.

In the proposed event-driven model, we incorporate the acceleration of the infection rate and expansion of the susceptible class and the infective class. Event influence terms are added into differential equations of i(t) and s(t), the growth rate and the infection rate. Incorporating the effect of events on topic diffusion, we therefore have the model:

$$
\begin{aligned}
&s(t) = (\mu + \delta_1 e(t))(1 - N(t)/K)N(t) - (\alpha + \delta_2 \text{e(t)})S(t)I(t) \\
&i(t) = (\alpha + \delta_2 \text{e(t)})S(t)I(t) - \beta I(t) + \delta_3 e(t)I(t) \\
&r(t) = \beta I(t)
\end{aligned}
\tag{2}
$$

These differential equations combine the SIR model and the event effect. In Equation (2), e(t) that we introduce for the event-driven SIR model is the dummy variable that indicates the event occurrence. If an event occurs at time t, then e(t) is 1, otherwise it is 0. δ_1 represents the coefficient of the event effect to susceptible growth. The growth of the susceptible class depends on the coefficient of event effect and the number of susceptibles at each time. This is expressed as $\delta_1 e(t)S(t)$. δ_2 is the coefficient of the event effect to the acceleration of the infection rate. The event accelerates the infection rate as much as δ_2 as shown ($\alpha + \delta_2 e(t)$) in the equation of i(t). δ_3 denotes the coefficient of the event effect to infective growth. The infective class increases as much as the event effect, δ_3, depending on the number of infectives. This effect term is expressed as $\delta_3 e(t)I(t)$ in the equation of i(t).

IV. Experiments

A. Research Testbed

Experiments are conducted on the Yahoo! Finance - Walmart message board (YAHOO) and Walmart-related news in the Wall Street Journal (WSJ). As one of the largest international companies in the world, Walmart draws attention from investors, analysts, and activists, especially on various social media sites [23]. The proposed event-driven SIR model was tested on the Walmart message board and the Wall Street Journal. A news outlet like the Wall Street Journal provides a more balanced view of the company and has a lesser potential to bias news reports [24].

On the Yahoo! Finance -Walmart message board, we extracted key topics according to their frequencies of appearance. The topics with intermittent bursts are excluded in key topics since they have low volume. We excluded chatter topics that have ongoing patterns from the analysis because these topics without epidemic patterns are not said to be contagious and cause the contagion between users.

Four topics have major peaks in their time-series patterns as shown in Figure 1. The topics that embed major peaks include the healthcare-related topic with the keywords of "healthcare" and "health insurance", the minimum wage-related topic with "minimum wage" and "minimum pay", the stock-related topic with "stock index" and "stock price", and the product price-related topic with the "low price." We tested the models on data for the major outbreak periods using the major topics.

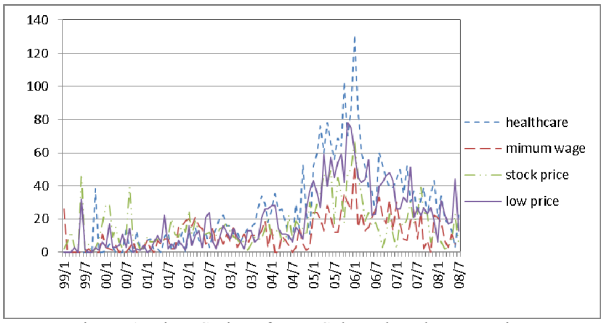

Figure 1. Time Series of Four Selected Walmart Topics.

The number of distinct authors including commenters was counted monthly for the four topics. The threads that include the pre-defined keywords were accumulated and the authors involved in the threads were counted. We used the aggregated data, 3-month moving average value, to reduce the effect of fluctuation of time series. For the event-driven SIR model, the number of news postings on the four topics was derived and the event indicator variable was calculated on a monthly basis from the number of news postings.

B. Evaluation Setting

The parameters describing the initial conditions (S(0), I(0), R(0)) and the dynamic parameters (α, β, μ) for the SIR model and (α, β, μ, δ1, δ2, δ3) for the event-driven SIR model are fed to the estimation process with the observed variable I(t). The models are parameterized with the non-linear square estimation that parameterizes the model with observations to minimize the sum of residuals. Simulated annealing is employed as the optimization algorithm. We compare the goodness of fit in terms of mean squared error (MSE) and the R-square value of two models.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(I_i - \hat{I}(t_i, \hat{\theta}))^2$$

$$R - square = 1 - \frac{\sum_{i=1}^{n}(I_i - \hat{I}(t_i, \theta))^2}{\sum_{i=1}^{n}(I_i - \bar{I})^2} \quad (3)$$

where I_i is the number of infectives at time i,

$\hat{I}_i$ is the estimated number of infectives at time i

$\bar{I}$: the average of I_i

n : the number of samples

t : time

$\hat{\theta}$: the estimated parameter set

C. Fitting Results

Table I compares the fitting performances of the SIR model and the event-driven SIR model.

TABLE I. PARAMETER ESTIMATI ON RESULTS OF THE SIR MODEL AND THE EVENT-DRIVEN SIR MODEL.

Topic	SIR model		Event-driven SIR model	
	MSE	***R^2***	***MSE***	***R^2***
#1	289	0.52	90	0.85
#2	28	0.64	26	0.66
#3	80	0.57	25	0.86
#4	106	0.66	35	0.89

Topic 1 - healthcare, insurance
Topic 2 - minimum wage, pay
Topic 3 - stock price
Topic 4 - low price

In Figure 2, the left figure shows that the SIR model reproduced the general shape of the time series with the fair adjusted R-square of 0.52 and MSE of 289 for the time series of the discussion on healthcare/insurance. The right figure shows the estimated value and real value of I(t) with the estimated parameters of the event-driven SIR model. The event-driven SIR model outputs a better overall fit compared to the SIR model with MSE of 90 and R-square of 0.85. In this graph, we found that there was a slight change in the slope of the estimated infectives when events happen while the SIR model generates a smooth estimation value that doesn't capture the fluctuation of real data.

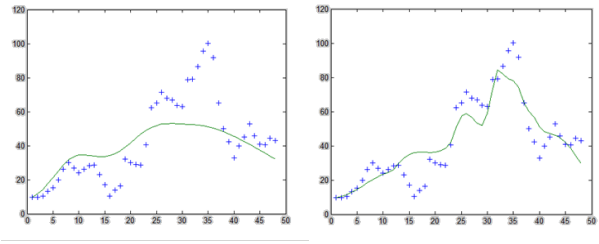

Figure 2. The Curves of the SIR moodel and the Event-driven SIR Model on the Topic of Healthcare/Insurance.

Table II summarizes the results of the parameter estimation of the event-driven SIR model.

TABLE II. PARAMETER ESTIMATION VALUES OF THE EVENT-DRIVEN SIR MODEL.

Topic	K	S(0)	α	β	μ	δ_1	δ_2	δ_3
#1	1,490	82	0.003	0.590	0.076	0.054	0.001	0.003
#2	700	50	0.018	0.662	0.046	0.073	0.016	0.000
#3	992	79	0.002	0.661	0.072	0.019	0.001	0.005
#4	1,208	71	0.004	0.562	0.046	0.075	0.001	0.004

For the healthcare/insurance topic, the estimation sample consists of 48 monthly observations from April 2003 to March 2007. We estimated the parameters of the event-driven SIR model. The optimal values of parameters were estimated at 82 for S(0), 0.003 for α, 0.590 for β, 0.076 for μ, and 1490 for K. This means that the number of forum users who might have interest in the topic of healthcare/insurance and possibly become authors is approximately 82, the number of infectives who get infected by the contacts between an infective and 10,000 susceptibles is 82 per month, and the number of infectives who will recover per 100 infectives during a month is 59. The carrying capacity for this topic is approximately estimated at 1490 and the population grows at the rate of 0.076 proportionally to the current population adjusted by the ratio of the remaining carrying capacity over the carrying capacity. The susceptible growth rate (δ_1) indicates that an event that causes the number of postings in the news forum to exceed the average number of posting on a specific topic expands the existing susceptibles at the rate of 0.054 per susceptible. An event accelerates the transmission rate as much as 0.001, which is the optimal infection accelerating rate (δ_2). Also, the optimal infective growth rate (δ_3) implies that an event expands the existing infectives at the rate of 0.003.

According to the estimation results of the parameters common to the two models, It would be fair to conclude that

the data is generally in good agreement with the two models. In addition, we got consistent results, with the event-driven SIR model outperforming the SIR model in terms of fitting results on the major spikey topics.

V. DISCUSSION

In this work, we identified the underlying mechanism in occurrence of the spiky discussion on a specific topic in the web forum. According to the estimation results, we found that the event-driven SIR model is a plausible model for the topic diffusion in the web forum. It implies that the occurrence of spikey topics in the web forum is mainly caused by the interaction between people and the fluctuation is caused by the event.

From this model, we can also estimate how many authors have latent interest on each topic at the initial stage of the diffusion process and how many authors will be involved along the diffusion process eventually. We can predict the outbreak of topics through the examination on initial stage of diffusion process. In the case when the diffusion process follows an epidemic model, this threshold behavior enables us to estimate the likelihood of a peak of the diffusion process. In addition, we can estimate the expected duration and the intensity of diffusion process at an initial stage. Even though these estimation measurements become reliable when it closes to the peak, we can estimate them in an adaptive way. Moreover, we can compare those values for topics that lead the peak.

Since the mathematical model, especially the deterministic model, simplifies diffusion process, it does not provide a complete analysis on that. The usefulness of a mathematical model is to obtain system-level measurements and test hypotheses using them. In this study, we incorporate the news effects on the topic diffusion process to provide a better explanation on the diffusion patterns with minor fluctuations. The most important usefulness of building a mathematical model is that we can incorporate forum characteristics in the diffusion model. We can perform further investigation on other factors that may affect the diffusion dynamics by incorporating those factors to the model. For instance, the sentiment of posts may determine the infectivity of a topic. Then, we can design the model with an infective rate varying depending on the cumulated sentiment score of posts.

VI. CONCLUSIONS

In this research, we examined the feasibility of applying the epidemic SIR model to topic diffusion on web forums and proposed a new model that incorporates the news media effect on social media. The experiment results from applying the two models to a large web forum revealed the following findings. The SIR model performs well in modeling topic diffusion in the web forum. The parameter estimation process of the models with simulated annealing converged to optimal values within finite iterations. The event-driven SIR model outperforms the SIR model.

For future research, we plan to employ other parameter estimation techniques. A genetic algorithm could improve the efficiency of the parameter estimation process, allowing us to skip the initial exploration for initializing parameters. Kalman filter [25] is one of considerations to trace the diffusion curve over longer periods. We also plan to perform experiments with other social media (e.g., twitter, Youtube, virtual worlds) systematically in the next years.

ACKNOWLEDGMENT

This work was supported in part by grant no. HDTRA1-09-1-0058 of the U.S. Defense Threat Reduction Agency, and by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2009-352-D00329).

REFERENCES

[1] A. Robbin and W. Buente, "Internet information and communication behavior during a political moment: The Iraq war, March 2003," Journal of the American Society for Information Science and Technology, vol. 59, pp. 2210-2231, 2008.

[2] J. Habermas, "Political Communication in Media Society: DoesDemocracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research1," *Communication Theory,* vol. 16, pp. 411-426, 2006.

[3] X. Wu and Z. Liu, "How community structure influences epidemic spread in social networks," *Physica A: Statistical Mechanics and its Applications,* vol. 387, pp. 623-630, 2008.

[4] J. Kleinberg, "Cascading behavior in networks: Algorithmic and economic issues," *Algorithmic game theory,* pp. 613-632, 2007.

[5] W. O. Kermack and A. G. Mckendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society,* pp. 700–721, 1927.

[6] W. Goffman, "Generalization of epidemic theory. An application to the transmission of ideas," *Nature*, vol. 204, pp. 225–228, 1964.

[7] L. Bettencourt, A. Cintrón-Arias, D. I. Kaiser, and C. Castillo-Chávez, "The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models," *Physica A: Statistical Mechanics and its Applications*, vol. 364, pp. 513-536, 2006.

[8] L. M. A. Bettencourt, D. I. Kaiser, J. Kaur, C. Castillo-Chávez, and D. E. Wojick, "Population modeling of the emergence and development of scientific fields," *Scientometrics*, vol. 75, pp. 495-518, 2008.

[9] K. Kawachi, "Deterministic models for rumor transmission," *Nonlinear Analysis: Real World Applications*, vol. 9, pp. 1989-2028, 2008.

[10] S. Yang, H. Jin, X. Liao, and S. Liu, "Modeling modern socialnetwork-based epidemics: A case study of rose," *Autonomic and Trusted Computing*, pp. 302-315, 2008.

[11] J. R. C. Piqueira, B. F. Navarro, and L. H. A. Monteiro, "Epidemiological models applied to viruses in computer networks," *Journal of Computer Science*, vol. 1, pp. 31-34, 2005.

[12] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-ofmouth," *Marketing Letters,* vol. 12, pp. 211-223, 2001.

[13] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," presented at the 13th International Conference on World Wide Web, New York, 2004.

[14] K. Saito, R. Nakano, and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," presented at the 12th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems Zagreb, Croatia, 2008.

[15] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web (TWEB),* vol.1, p. 5, 2007.

[16] M. Kubo, K. Naruse, H. Sato, and T. Matubara, "The possibility of an epidemic meme analogy for web community population analysis,"

Intelligent Data Engineering and Automated Learning-IDEAL 2007, pp. 1073-1080, 2007.

[17] Y. Zhou, X. Guan, Z. Zhang, and B. Zhang, "Predicting the tendency of topic discussion on the online social networks using a dynamic probability model," presented at the Pittsburgh, PA, USA, 2008.

[18] M. Bampo, M. T. Ewing, D. R. Mather, D. Stewart, and M. Wallace, "The effects of the social structure of digital networks on viral marketing performance," *Information Systems Research,* vol. 19, pp. 273-290, 2011.

[19] J. Woo, J. Son, and H. Chen, "An SIR model for violent topic diffusion in social media," 2011, pp. 15-19.

[20] D. P. Fan and R. D. Cook, "A differential equation model for predicting public opinions and behaviors from persuasive information: Application to the Index of Consumer Sentiment," *Journal of Mathematical Sociology,* vol. 27, pp. 29-51, 2003.

[21] B. S. Greenberg, "Diffusion of news of the Kennedy assassination," *Public Opinion Quarterly,* vol. 28, p. 225, 1964.

[22] C. C. Chen, M. C. Chen, and M. S. Chen, "An adaptive threshold framework for event detection using HMM-based life profiles," *ACM Transactions on Information Systems (TOIS),* vol. 27, p.9:1-9:35, 2009.

[23] H. Chen, "Business and Market Intelligence 2.0," *IEEE Intelligent Systems, January/February,* pp. 2-5, 2010.

[24] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," *ACM Transactions on Information Systems (TOIS),* vol.27, p. 12, 2009.

[25] J. Xie, X. M. Song, M. Sirbu, and Q. Wang, "Kalman filter estimation of new product diffusion models," *Journal of Marketing Research,* pp. 378-393, 1997.