Decision Support Systems xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Analyzing firm-specific social media and market: A stakeholder-based event analysis framework

Shan Jiang [a,*], Hsinchun Chen [a], Jay F. Nunamaker [a], David Zimbra [b]

[a] Department of Management Information Systems, University of Arizona, McClelland Hall 430, 1130 E. Helen St., P.O. Box 210108, Tucson 85721 AZ, United States
[b] Department of Operations Management and Information Systems, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053, United States

ARTICLE INFO

Article history:
Received 18 June 2013
Received in revised form 27 February 2014
Accepted 6 August 2014
Available online xxxx

Keywords:
Social media
Stakeholder
Community identification
Market prediction

ABSTRACT

Discussion content in firm-specific social media helps managers understand stakeholders' concerns and make informed decisions. Despite such benefits, the over-abundance of information online makes it difficult to identify and focus on the most important stakeholder groups. In this study, we propose a novel stakeholder-based event analysis framework that uses online stylometric analysis to segment the forum participants by stakeholder groups, and partitions their messages into different time periods of major firm events to examine how important stakeholders evolve over time. With this approach, we identified stakeholder groups from a sample of six companies in the petrochemical and banking industries, using more than 500,000 online message postings. To evaluate the proposed system, we conducted market prediction within the identified groups, and compared the prediction performance with traditional approaches that did not account for stakeholder groups or events. Results showed that some stakeholder groups identified by our system had stronger relationships with firms' market performance, compared to the entire set of web forum participants. Incorporating event-induced temporal dynamics further improved the prediction performance.

1. Introduction

Firm-specific web forums (e.g., Yahoo! Finance Message Board) serve as a specialized social media platform for exchanging investment opinions. A large number of participants in these web forums are company stakeholders with interests in a particular firm's market performance. They express personal opinions and share investment information in such platforms [25]. Using firm-specific web forums as a proxy, managers may better understand the stakeholders' concerns and make more informed decisions. When major firm events occur, firm-specific web forums can also become an essential information source for evaluating the impact of events on company stakeholders, allowing more effective managerial responses to be made. However, it is often difficult to identify and focus on information from the most important stakeholders due to the vast amount of information posted online.

This study addresses this problem by proposing a stakeholder-based event analysis framework to support managerial decisions regarding firms' stakeholders and major events. In the proposed framework, forum participants were clustered into stakeholder groups based on their activity characteristics, topics of interest, and stylometric clues. In different stages of events, important stakeholder groups were identified based on their predictive relationships with the company's market performance. We tested the resulting system on six companies' discussion boards and tracked how the importance of different stakeholder groups evolved in response to relevant firm events.

The remainder of the paper is organized as follows. We first provide a review of relevant literature, based upon which research gaps are identified. Next, our stakeholder-based event analysis framework on web forums is presented, which is designed to address these gaps. We then describe two experiments used to evaluate our system, and discuss the results. Finally, conclusions and implications of the study are provided.

2. Related works

2.1. Stakeholder theory and identification of stakeholder groups

Since the rise of the concept of "stakeholder," there has been a debate on its definitions. In its narrowest definition, stakeholders are defined as groups of people on whom an organization's survival is dependent [14]. Usually, the narrow definition of stakeholders adopts a role-based stakeholder identification that focuses more on groups that hold functional relationships with the firm, such as customers, employees, and shareholders. In contrast, the broadest definition of a stakeholder is "any group or individual who can affect or is affected by the achievement of the organization's objectives" [13]. Stakeholders are identified by "their interest in the corporation, regardless of whether the organization has any corresponding functional interest in them" [28]. From this perspective, participants in firm-specific web forums

* Corresponding author. Tel.: +1 480 2055717; fax: +1 520 6212433.
E-mail address: jiangs@email.arizona.edu (S. Jiang).

http://dx.doi.org/10.1016/j.dss.2014.08.001
0167-9236/

are all stakeholders. Through analyzing their discussions, managers of a firm can understand various stakeholder concerns and take informed actions to manage a firm's reputation in case of events [44].

Traditionally, stakeholder groups are categorized based on their functional roles with the firm. For example, researchers have used multiple classifiers to classify stakeholders into functional groups such as customers, employees, and shareholders in web pages [7]. However, within a specific functional group, interests and concerns may vary greatly [37]. It has been argued that role-based stakeholder groups may be overly simplistic and does not adequately address the crucial issues managers face in reality [15]. Therefore, researchers have also advocated unsupervised clustering approaches that allow people with similar interests to be grouped together [37]. To directly address their interests and concerns, content-based features such as the terms used in their forum posts can be utilized to capture the major topics of discussions. Also, the association of a stakeholder with a group may be characterized by adoption of similar online communication practices, such as frequency of creating new threads and participating in an existing discussion thread. Stylometric analysis literature also suggests that authors with similar backgrounds and interests tend to have similar writing styles [3]. For these reasons, content-based, forum activity-based and writing style features have been used for identifying and segmenting online groups in social media [1,43].

2.2. Social media-based market prediction and analytical techniques

Prior research has extensively studied how stock performance can be predicted based on social media [2,4,8,16,30,40,42]. Common metrics of stock performance indicators include stock return, volatility, and trading volume. These indicators are regressed on various web forum variables in prior time intervals, often in daily basis. The most typical web forum variables include message volume, average message length, sentiment index, and disagreement index. Message volume indicates how many postings are created, indicating the activity level of forum participants. A high message volume in a day is consistently found to result in reduced stock returns the following day [2,8]. Stock volatility, in contrast, has been found to positively correlate to the prior day's message volume [2,8]. Message length denotes the number of words used in a message. A high average message length in a day indicates that forum participants are active. The sentiment index is an aggregated indicator of opinions expressed in discussions, reflecting whether the overall attitude of forum participants is positive or negative. Researchers have found that the relationship between the sentiment index and next day return is insignificant for individual stocks [2,5,8,42]. However, when aggregating industry-wide opinions, sentiments helped predict the return of related aggregated indexes such as the Morgan Stanley High-tech Index [8]. Yu et al. [40] further revealed that the impacts of sentiment on stock return could vary by social media types. The disagreement index, usually defined as the variance of message-level sentiment [2,8], measures the extent to which participants' opinions differ. Through disagreement, a trade occurs between a seller and buyer [17]. Therefore, disagreement index may be indicative of trading volume [5].

Several text mining techniques have been employed to analyze textual data in web forums. Sentiment analysis has been extensively used to assess the valence and intensity of opinions in social media texts [2,8,20]. The techniques can generally be classified as supervised or unsupervised approaches [24]. In supervised approaches, classifiers are trained based on a set of tagged messages, and then applied to messages of unknown sentiments. For example, Naïve-Bayes and SVM classifiers were used to classify the forum messages as buy, sell, or hold positions [2,8,40]. An unsupervised approach does not require a manual labeling process. It leverages sentiment lexicons, such as the Senti Word Net (SWN) lexicon [11], to map terms with scores that indicate the term's sentiment valence and intensity [5,33].

Attempts have also been made to filter out noise and extract the most relevant information from web forums. At the message level, topic analysis or document clustering is often used to group messages of similar topics together [35]. In this approach, the clustering of messages is based on the content-based features, which are mainly term and phrase occurrences. At the user level, forum participants sharing similar characteristics can be grouped by online stylometric analysis. Stylometric analysis extracts stylistic features from texts to construct authorship profiles for users. In addition to content-based features, stylistic features also include lexical, syntactic and structural feature categories [1,18,43]. Lexical features relate to the character and word usage such as frequency of letters and vocabulary richness [41]. Syntactic features include usage of function words and punctuation. Structural features include the text organization and layout, such as the number of paragraphs. After constructing authorship profiles, clustering is used to group similar authorship profiles together.

2.3. Event and temporal dynamics of social media

Online discussions in social media show great variations over time, especially when major events occur. This phenomenon is often referred to as the temporal dynamics of social media [39]. The temporal dynamics of social media can be observed in terms of participants, discussion contents, and relationships between social media activity and the real world. The aim of studying the temporal dynamics of social media participants is to understand how different groups of people react to critical social events. Generally, big events attract more people to participate in discussion and change the composition of participants' constituencies. Robinson (2005) conducted a case study on three online forums for two months starting from September 11 [29]. He found that the disaster of 9/11 led to the formation of groups holding different viewpoints in forums. Also, a large number of casual participants were observed immediately after the beginning of September 11. The temporal dynamics of discussion contents has also been studied [9,39]. These studies revealed how people's concerns were affected by events and changed over time. It has been shown that popularity of contents varies over time. Typically, a topic receives attention immediately upon its appearance and reaches its peak within a day. The peak can last for some time and then popularity decreases. When social media activity is used to predict real world phenomena, the predictive relationships may evolve over time. For example, in a movie's pre-production stage, the number of forum postings about a movie is not significantly correlated with the opening strength of the movie, while this relationship becomes significant during production to release [21]. The changing relationship between social media and the real world can be partially attributed to people's changing attention in response to events [38].

2.4. Research gaps

Several gaps can be identified from the literature. First, prior research has suggested that a firm should address each stakeholder group's concerns individually. Although attempts have been made to classify stakeholder groups into functional groups in business web pages [7], few studies have segmented the participants of firm-specific forums by stakeholder group. As prior research has suggested, significant noise could be introduced by using the entire web forum for analysis [8,40]. Second, major firm events can have a great influence on stakeholder groups. However, few previous studies have addressed the temporal aspects of social media and examined how stakeholder groups and their relationships with the firm's market performance changed in response to events.

3. System framework

Fig. 1 illustrates the proposed stakeholder-based event analysis framework developed to address the research gaps. The major components of the framework will be discussed next.

Please cite this article as: S. Jiang, et al., Analyzing firm-specific social media and market: A stakeholder-based event analysis framework, Decision Support Systems (2014), http://dx.doi.org/10.1016/j.dss.2014.08.001