International Journal of Information Management 35 (2015) 490–503

Contents lists available at ScienceDirect

International Journal of Information Management

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

Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams

Carlo Lipizzi[a,*], Luca Iandoli[b,a], José Emmanuel Ramirez Marquez[a,c]

[a] School of Systems and Enterprises, Stevens Institute of Technology, USA
[b] Department of Industrial Engineering, University of Naples Federico II, Italy
[c] Graduate School, Tecnologico de Monterrey Campus Guadalajara, Mexico

ARTICLE INFO

Article history:
Received 3 October 2014
Received in revised form 2 April 2015
Accepted 3 April 2015
Available online 13 May 2015

Keywords:
Social media
Twitter Case study
Consumer electronics industry
Competitive analysis
Competitive intelligence
Competitor intelligence
Actionable intelligence
Text mining
Content analysis

ABSTRACT

In this paper we use Twitter data to assess customers early reactions to the launch of two new products by Apple and Samsung by analyzing the streams generated in a 72 h window around the two events. We present a methodology based on conversational analysis to extract concept maps from Twitter streams and use semantic and topological metrics to compare the conversations. Our findings show that there are significant differences in the structural patterns of the two conversations and that the analysis of these differences can be highly informative about early customers perceptions and value judgments associated with the competing products.

1. Introduction

Social media and Twitter in particular are increasingly popular. Twitter reached 18% of web users in May 2013 (eMarketer, 2013), with a peak of 30% in the 18–29 years old population bracket. This population generates more than 500 million messages ("tweets") per day (Forbes, 2013). The analysis of this massive amount of data may provide a valuable opportunity for companies to elicit market information by learning directly from the customers' voice.

In domains in which the opinion of a large population matters, such as marketing, the analysis of social media data is becoming a popular subject. Social media create new opportunities for companies to interact with their customers through online campaigns and mining social media data is increasingly common to support digital marketing initiatives as well as a variety of business intelligence applications (Osatuyi, 2013; Royle & Laing, 2014; Rud, 2009). Among current marketing applications of social media mining, monitoring the launch of a new product can help companies to assess the immediate market reactions and obtain clues about how to adjust marketing actions to increase their effectiveness of the launch strategy (Banerjee et al., 2012), to get new insights for new products design and improvements (Marcus et al., 2011), or to monitor the sentiment and reputation of a particular product or brand (Chamlertwat, Bhattarakosol, & Rungkasiri, 2012; Melville, Sindhwani, & Lawrence, 2009).

The analysis of Twitter streams to evaluate consumers' reactions to the launch of a new product requires the content to reflect their collective judgment on the product or brand. Several commercial platforms are available to retrieve and assess to some extent this collective judgment through sentiment analysis (Cognizant, 2014; Crimson Hexagon, 2014). Sentiment, however, just measures how the "crowd" feels about a product, but does not offer insight on the structure and the determinants of the customers' preferences. Our challenge is instead to dig deeper into Twitter streams to capture and assess structured contents that are embedded in them through an analytical and quantitative approach. From the practical point of view, the elicitation of such organized content could support market analysts to achieve a better understanding of

* Corresponding author. Tel.: +1 2017366460.
E-mail address: clipizzi@stevens.edu (C. Lipizzi).

http://dx.doi.org/10.1016/j.ijinfomgt.2015.04.001
0268-4012/

consumers' perceptions and to better manage the reach-richness trade off between qualitative and quantitative market analyses. An additional practical use of our study can be to find a way to ascertain whether the presence of given semantic patterns in Twitter streams can be predictive of early market success for a new product.

In order to contribute to these research objectives, we offer a perspective based on conversational analysis to mine Twitter streams and get insights directly from the consumers' voices about early reaction, satisfaction drivers, and possibly improvements for the product. We consider the special case in which Twitter is used as a backchannel by customers to post tweets in a short time window around the launch of a new product that has been previously announced and advertised through other media or official presentations. In this special case the public's attention is channeled toward a joint object of interest and use of the otherwise disorganized Twitter streams is more likely to get closer to a multiple voices conversation around the event being followed.

Starting from this idea, we model the generation of Twitter streams as a process of progressive accumulation of shared knowledge, although in a fragmented and unsystematic fashion. We present a methodology combining social network and semantic analysis to extract in an automated way this knowledge base as it is generated from Twitter streams created in a short time window surrounding the launch of a new product. Since we represent this knowledge through a dynamic concept map, we suggest a set of semantic and topological metrics to assess the content generated in different conversations or in the same conversation at different points in time. We apply the proposed method to two cases of new product launches that were executed in the Fall of 2013 by Apple and Samsung. Our analysis shows that the reactions to the launches as measured through the above approach are significantly different, with Apple exhibiting richer and more structured content than Samsung.

The paper is structured as follows: in the next section we review current methodological approaches to social media mining in business applications and explain how our method is positioned in the existing literature. We then describe a theoretical perspective based on the Theory of Common Ground (Brennan & Clark, 1991) to model the generation of Twitter streams in backchanneling. We then present the methodology and its application to two launches through the analysis of more than 60,000 tweets downloaded in the US in 72 h around the events and compare the two conversations through a set of metrics and data visualization tools. In the discussion section we provide a detailed description of the findings and offer ideas about the practical and scientific relevance of the approach.

2. Current approaches to social media mining in business applications

2.1. Overview

The use of Twitter to comment on the launch event of a new product is a particular case in which the micro-blog platform is used as a backchannel. In backchanneling a medium is employed to sustain a real-time, online conversation alongside a primary activity or event; e.g., to comment or retweet about a specific event that is object of public attention. The Twitter data flow generated by specific events through backchanneling provides a more focused opportunity for social media mining. Examples include monitoring TV audience reactions (Harrington, Highfield, & Bruns, 2013; Nielsen, 2014; Wakamiya, Lee, & Sumiya, 2011), understanding ad predicting traders' reactions in the stock market (Evangelopoulos, Magro, & Sidorova, 2012), or measure the diffusion of ideas in online discussions (Tsur & Rappoport, 2012). In all these cases the objective is to extract meaningful and representative feedback from online conversations and exploit this feedback to help decision makers to identify effective ways to influence attitudes and behavior.

The methodological approaches that are most commonly used in social web mining are based on Information Diffusion in Social Networks (Bird, Gourley, Gertz, Devanbu, & Swaminathan, 2006; Wasserman, 2011). Another important category of methods is based on semantic analysis and includes applications such as sentiment and reputation analysis (Brown, 2012; Saif, He, & Alani, 2012) or on text mining of the content published on social media site (He, Zha, & Li, 2013).

Diffusion theory provides conceptual and methodological tools to assess the rate of information percolation throughout a social network, with applications ranging from the adoption of a new idea or practice (Rogers, 1976) to computer viruses (Vespignani, 2005) or viral marketing in online networks (Leskovec, Adamic, & Huberman, 2007). The unit of analysis is the topology of connections between users in a social network and how it favors or hinders the diffusion of a message. Social networks are formally described as graphs $G = (U, V)$, where U is a non-empty set of nodes (the subjects in the network) and V is a set of edges between pairs of nodes representing a relationship. For example, in Facebook U could be a set of users, V the set "friendship" links in U, in Twitter U could be "senders", V the "re-tweet" links. Social graphs can be analyzed through social network analysis (SNA) models based on structural properties of the network including centrality (Perry-Smith, 2006), links distribution (Cowan & Jonard, 2001), relational embeddedness (Reagans & McEvily, 2003; Uzzi, 1997), ties strength (Granovetter, 1973), and role of brokerage (Burt, 2004).

SNA models are typically content-neutral; i.e., they assume that information meaning is not problematic and that it is preserved when information is passed across social links. Moreover, these approaches assume that the communication process is a one-way flow of information and pay little attention to the recursive and cyclic nature of the interactions taking place between users. This view of information as a neutral token to be passed between an active sender and a passive receiver is probably due to the remarkable and long lasting intellectual influence of Claude Shannon's Information theory (Sloane & Wyner, n.d.).

Approaches based on semantic analyses focus on the automated extraction of meaning from the messages propagated in a network, using a variety of techniques based on text mining and sentiment analysis.

Text mining techniques involve the automatic indexing, classification, and extraction of information from online data streams. Techniques like term frequency weighting are used to extract important terms out of documents modeling to rank the results according to their relevance (Manning, Raghavan, & Shutze, 2009). For instance, tf–idf (term frequency–inverse document frequency) is a well-known method to calculate the relevance of a term within a repository of documents. This method computes the importance of a term in a document based on its occurrences in the document as well in all the documents of the whole collection and has been applied primarily to large documents (Salton & Buckley, 1988). Tf–idf can be used to select features in a collection or in a corpus analyzing the frequency of the word. While td–idf and its variations are used in several applications such as search engines as a central tool in scoring and ranking a document's relevance given a user query (from the tfidf.com website and (Croft, Metzler, & Strohman, 2010)), the frequency may not be representative of the semantic relevance of the term (Qian, Zhou, Zhang, & Zhao, 2013).

Creating topological representations from words in text, based on word co-occurrence, has been applied to extract information on the structure of the text and eventually its content (Introne &