Computers in Human Behavior 64 (2016) 782–792

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Towards computational discourse analysis: A methodology for mining Twitter backchanneling conversations

Carlo Lipizzi [b], Dante Gama Dessavre [b], Luca Iandoli [a, b, *], Jose Emmanuel Ramirez Marquez [b]

[a] Dept. of Industrial Engineering, University of Naples Federico II, Piazzale Tecchio 80, 80125 Napoli, NA, Italy
[b] School of Systems and Enterprises, Stevens Institute of Technology, Castel Point on Hudson 1, 07030 Hoboken, NJ, USA

ARTICLE INFO

Article history:
Received 19 March 2016
Received in revised form
3 July 2016
Accepted 22 July 2016

Keywords:
Online discourse analysis
Conversational analysis
Social media mining
Social representations
Big data
Online communities
Online conversation
Computer-mediated communication
Collective intelligence
Semantic analysis
Sentiment analysis
New product launch

ABSTRACT

In this paper we present a methodology to analyze and visualize streams of Social Media messages and apply it to a case in which Twitter is used as a backchannel, i.e. as a communication medium through which participants follow an event in the real world as it unfolds. Unlike other methods based on social networks or theories of information diffusion, we do not assume proximity or a pre-existing social structure to model content generation and diffusion by distributed users; instead we refer to concepts and theories from discourse psychology and conversational analysis to track online interaction and discover how people collectively make sense of novel events through micro-blogging. In particular, the proposed methodology extracts concept maps from twitter streams and uses a mix of sentiment and topological metrics computed over the extracted concept maps to build visual devices and display the conversational flow represented as a trajectory through time of automatically extracted topics. We evaluated the proposed method through data collected from the analysis of Twitter users' reactions to the March 2015 Apple Keynote during which the company announced the official launch of several new products.

1. Introduction

The large-scale adoption of Social Media (SM) is one of the most relevant technological and social trends in the history of the Internet. According to the Pew Research Center (Duggan, 2015), about three quarters of adult Internet users in the US spend considerable time on Social Network sites such as Facebook, Twitter, Instagram, Pininterest, and LinkedIn, with 65% of online adults using social media sites (Perrin, 2015), with a tenfold increase in ten years (this percentage goes up to 70% and 59%, respectively, for the adults who use daily Facebook and Instagram). SM are thus creating interaction spaces of unprecedented size and breadth in which a large number of users generates and shares contents through hyper-connected social networks while trying to accomplish various types of tasks such as voicing their opinion, providing or asking for help, sharing information, contributing to a cause, reaching out to friends and acquaintances, or applying for memberships to a community.

Much of this interaction happens through online conversations that can be tracked and mined to extract online analytics for different applications. A broad practical question for media analysts is the availability of reliable tools to summarize large conversational flows into effective representations to answer a number of "what" and "who" questions such as: what the users are talking about? What matters most to them? What is trending right now? Who is talking about what? Who is talking to whom? Etc. The answers to such questions are of immediate use in multiple applications in fields as diverse as marketing (Hoffman & Fodor, 2010, pp. 1–11; Smith, Rainie, Shneiderman, & Himelboim, 2014), politics (Bartlett, Froio, Littler, & McDonnell, 2013; Parker, 2014), or national security

* Corresponding author. Dept. of Industrial Engineering, University of Naples Federico II, Piazzale Tecchio 80, 80125 Napoli, NA, Italy.
E-mail addresses: carlo.lipizzi@stevens.edu (C. Lipizzi), ddessavre@stevens.edu (D.G. Dessavre), luca.iandoli@unina.it, luca.iandoli@stevens.edu (L. Iandoli), jmarquez@stevens.edu (J.E. Ramirez Marquez).

http://dx.doi.org/10.1016/j.chb.2016.07.030
0747-5632/

(Brachman, 2014; Klausen, 2015). Moreover, thanks to the possibility to track online interaction in an objective fashion, online conversations offer plenty of interesting empirical materials for the analysis and modelization of learning in virtual social networks, although predominantly in research on use of SM in education (Greenhow & Greenhow, 2011; Yu, Tian, Vogel, & Kwok, 2010).

In this paper our approach to answer "what" and "who" questions is based on tracking and visualizing the process through which large-scale interaction favors the emergence of shared meaning through a process of participatory sense-making (Jaegher & Paolo, 2007). As we show later in the paper (Section 2), the analysis of this process has received limited attention in the literature on online conversations, which instead could benefit from the application of computational methods based on natural language processing, automated semantic analysis, and social network analysis as well as from theoretical and methodological insights from studies in conversational analysis and discourse psychology (Section 3).

Through the development of a methodology that combines ideas and tools from conversation analysis, online analytics, and data visualization (Section 4), we provide a methodology and a tool to better handle the trade-off between in depth, small-scale, qualitative analysis and processing of large amount of data generated in conversations. We then show that SM-enabled conversations can support sense-making about new events by favoring the accumulation of common ground generated by multiple participants when following a same event in real time (Clark & Brennan, 1991) (Section 5).

Among the several available SM, we focus on Twitter for a number of reasons. Twitter is popular and widespread among Internet users (Duggan, 2015) and the access to its data is relatively easy and low cost with most users making their content public. Tweets lack the complexity and richness of more structured messages, such as those that can be found in online forums, thanks to the well-known 140 characters constraint. Finally, Twitter is increasingly used as a backchannel to follow up with events unfolding in the real world (Dork, Gruen, Williamson, & Carpendale, 2010). Twitter backchannel feeds are also frequently used to support virtual participation to a real event (McNely, 2009) by providing participants with a parallel communication channel to exchange comments about what is going on.

Backchanneling represents an interesting case for the empirical analysis of SM because the short duration of the events being followed helps to create a strong focus in the discussion, unlike the more unstructured and chaotic feature of uneventful micro-blogging. More importantly, backchanneling is akin to virtual participation to a real world event, which participants are trying to make sense of and to which they are attracted for a variety of individual motivations. We suggest that this sense-making process is easier to observe in those cases in which the participants are generally ignorant about the ways the event is going to unfold, although they may have some pre-existing knowledge and expectations about its development and the subject in general. This is the case for events such as a TV political debate (Shamma, Kennedy, & Churchill, 2009), a new episode of a TV show (Harrington, Highfield, & Bruns, 2012), and the use of SM to coordinate in emergency situation (Hughes & Palen, 2009; Mills, Chen, Lee, & Raghav Rao, 2009).

In the next section we show that the dominant approaches to SM mining are biased towards the analysis of user-generated content rather than of the process through which the same content is generated. Our challenge is instead to dig deeper into Twitter streams to observe the emergence of shared meaning in online, large-scale conversations by using a quantitative methodology inspired to ideas and theories from studies in discourse analysis. The proposed methodology is applied to extract social and concept maps from micro-blogs streams and to compute a set of conversational analytics for the design of visual devices able to display the conversational flows.

Since the maps are generated as the conversation unfolds, our approach helps to analyze how the structure of these graphs changes in time and reflect not only *what* the participants are talking about but also *how* they are talking about the event. Structural analysis is then used to visualize the conversational flow and identify emergent conversational patterns and the process that generates them. The combination of social, semantic, and more traditional traffic metrics helps to observe conversational interaction and its outcome in a dynamic fashion, possibly offering insights on the role of social cognition on meaning generation and transformation (Jaegher & Paolo, 2007) and empirical grounding to better justify the use collaborative/social technologies to support informal and continuous learning (Siemens, 2005).

We report empirical findings obtained from one case study related to the launch of new products, the Apple Keynote presentation that took place in March 2015, a 100 min event during which the company launched the new MacBook and the much-awaited AppleWatch. We selected this event as an example of a social happening that is highly focused, that is object of wide attention and ample coverage from other media, and that is scheduled regularly by Apple twice per year. The application to a marketing case is also used to offer an example of practical applications, but the proposed method can be applied to other events that exhibit similar characteristics in terms of focus, popularity, and expectations.

2. Current approaches to twitter streams analysis

2.1. Overview

The availability of reliable and efficient methods for the collection and analysis of SM data is an increasing concern for companies and policy makers (Hoffman & Fodor, 2010, pp. 1–11; Smith et al., 2014). In particular, without the ability to listen to and measure SM flows, media analysts miss a valuable opportunity to exploit the abundant and informative data that is incessantly generated by millions of online users and transform this information into actionable analytics to support more responsive and data-driven decision-making (Culnan, McHugh, & Zubillaga, 2010).

These needs bring new online analytic challenges due to the growing variety of online media to be monitored and the consequent production of massive quantity of data to be analyzed. For instance, different types of SM enable different types of conversational tasks and allow users to assume multiple online identities, especially among younger users (Bolton, Aksoy, van Riel, & Kandampully, 2013; Williams, Crittenden, Keo, & McCarty, 2012). Facebook, for instance, can be used for sharing personal contacts, news and experiences with friends, family and other acquaintances. Twitter is born with the idea of giving people a simpler tool to answer the question: what are you doing now? Lately, it has evolved into a public space in which people broadcast their opinions or whatever they think is worthwhile of public attention. SM conversations are also very different from online forums, which are often directed towards a common topic and support longer lasting conversations that are attended by a smaller number of participants.

Consequently, alternative ways to visualize these new types of social streams have been proposed recently while methods developed to analyze more "traditional" types of online conversations such as in online forum (Melville, Sindhwani, & Lawrence, 2009) or email exchanges (Viégas, Golder, & Donath, 2006) are not