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

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.

© 2016 Elsevier Ltd. All rights reserved.

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).