

Twitter Data Mining for Situational Awareness

Marco Vernier

University of Udine, Italy

Manuela Farinosi

University of Udine, Italy

Gian Luca Foresti

University of Udine, Italy

INTRODUCTION

The most recent catastrophic events, from the 2010 Haiti earthquake to the devastating 2013 Colorado floods, have shown a strong adoption of social media platforms by ordinary people. The data and meta-data produced by users during and after extra-ordinary situations could have enormous potentialities if integrated with traditional systems for emergency management and used for hyperlocal situational awareness (Foresti et al., 2015a). The great majority of current literature is focused on Twitter for several reasons, in particular, the architectures and practices of the platform itself. In the following section, an overview of social media use in extra-ordinary contexts will be presented, with a specific focus on the Twitter social media platform. Following this, in a section dedicated to the Twitter systems classification for situational awareness, a classification of the existing systems based on the analysis of Twitter data will be provided on the basis of the following three categories: 1) semantic systems; 2) meta-data systems; and 3) smart self-learning systems. Finally, in the future research trends section, an innovative and smart solution will be proposed for future development.

BACKGROUND

Social Media Use in Extra-Ordinary Contexts

Social media platforms are built from the beginning to be used socially, and oriented around collaboration and sharing. These potentialities are emphasized in extra-ordinary contexts, when ordinary people adopt these tools to provide or search for first-hand and real-time information regarding a certain event (i.e. an earthquake, flood, etc.) (Lindsay, 2011; Taylor et al., 2012). The most recent catastrophic events, from the 2010 Haiti earthquake to the devastating 2013 Colorado floods, in fact, have shown that these platforms have been strongly used both during and after disasters (Figure 1), allowing a real-time dissemination of information to the wider public, an effective situational awareness, and an up-to-date picture of what is happening on the ground (i.e. Farinosi & Micalizzi, 2013; White et al., 2014).

According to Kotsiopoulos (2014), in extra-ordinary situations, social media enable citizens to play at least three roles: 1) first responders/volunteers; 2) citizen journalists/reporters; and 3) social activists. Oftentimes citizens on the scene experience the event first-hand and are able to

provide updates more quickly than disaster response organizations and traditional news media (Sweetser & Metzgar, 2007; Procopio & Procopio, 2007; Farinosi & Micalizzi, 2013).

Given the increasing availability of data and meta-data produced and/or distributed on these online platforms, it is pivotal to understand how they should be used and integrated with traditional systems for situational awareness, supporting in this way the work of Civil Protection, Red Cross, Fire Department, and other agencies.

A Focus on Twitter

The great majority of existing research is focused on the social media platform Twitter. Twitter is a popular microblogging network that allow users to share up to 140 character text messages called tweets. Twitter was founded in 2006 by Jack Dorsey, Evan Williams, Biz Stone, and Noah Glass. Today it is one of the most used social platforms with about 250 million active users and about 500 million tweets shared per day (source: <https://dev.twitter.com/>). The platform is mainly based on the exchange of messages between networks of contacts. Each user can create his/her own custom networks, following other users' feeds. In turn, each user can be followed by other users called followers (Marwick and Boyd, 2011). Besides the possibility to share a short text content, Twitter also offers the opportunity to share visual content, such as videos and images, and URLs to web sites (Java et al., 2007; Huberman et al., 2008; Zhao & Rosson, 2009).

The reason why the majority of research is focused on Twitter is due to many factors. First of all, given the instantaneous nature of communication on Twitter, the platform is particularly suitable for real-time communication. Furthermore, the architecture and some specific features of Twitter seem to facilitate the widespread dissemination of information. Among the most important features offered by the Twitter platform is the possibility to share content using hashtags. This is an annotation format represented by the “#” symbol, used to

indicate with a single word (or a combination of words) the core meaning of a tweet. Conversations centered on specific hashtags promote focused discussions, even among people who are not in direct contact with each other.

In addition, the choice to analyze Twitter is also motivated by the prevailing public nature of the great majority of the accounts (only a small percentage of accounts are private), a feature that distinguishes this platform to other social networking sites (Bruns & Burgess, 2013). On the one hand, this peculiarity promotes public conversations, even among users that were not previously in contact with one another (Twitter in fact offers the possibility to interact with other users, to share content with them, and to reply or mention someone in your own tweet simply using the symbol “@” followed by the username of the person whom you want to tag). On the other hand, it makes it easier to conduct analysis that aims to reconstruct the spread of communication flows within the platform. Finally, yet importantly, this characteristic makes the use of tweets for research purposes less critical from an ethical point of view.

For example, the analysis of the 2011 floods in Queensland provides a detailed mapping of the general dynamics of Twitter use during an emergency and offers useful general indications (Bruns, 2012). Their findings highlight that the space-time variables represent a crucial element to obtain relevant data and to improve situational awareness during disasters. For instance, the physical distance of the Twitter users from the site of the catastrophe can reflect different types of needs to be met and a different perception of danger. In addition, given that a social platform like Twitter is structurally connected to forms of activation just in time, the time variable plays a fundamental role. Previous research demonstrates that immediately after the event there is a greater presence of forms of instinctive response, while tweets containing links to official news sources tend to arrive later (Acar & Muraki, 2011; Farinosi & Micalizzi, 2013).

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/twitter-data-mining-for-situational-awareness/183919

Related Content

Improving Knowledge Availability of Forensic Intelligence through Forensic Pattern Warehouse (FPW)

Vivek Tiwari and R. S. Thakur (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 1326-1335).

www.irma-international.org/chapter/improving-knowledge-availability-of-forensic-intelligence-through-forensic-pattern-warehouse-fpw/112531

An Effective Analysis Method of Discussions in Bulletin Board Sites

Shigeaki Sakurai (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 2032-2041).

www.irma-international.org/chapter/an-effective-analysis-method-of-discussions-in-bulletin-board-sites/112610

Trusted and Trustworthy Information Technology

Piotr Cofta and Hazel Lachée (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 4436-4444).

www.irma-international.org/chapter/trusted-and-trustworthy-information-technology/112885

Hindi Text Document Classification System Using SVM and Fuzzy: A Survey

Shalini Puri and Satya Prakash Singh (2018). *International Journal of Rough Sets and Data Analysis* (pp. 1-31).

www.irma-international.org/article/hindi-text-document-classification-system-using-svm-and-fuzzy/214966

Factors Influencing the Adoption of ISO/IEC 29110 in Thai Government Projects: A Case Study

Veeraporn Siddoo and Noppachai Wongsai (2017). *International Journal of Information Technologies and Systems Approach* (pp. 22-44).

www.irma-international.org/article/factors-influencing-the-adoption-of-isoiec-29110-in-thai-government-projects/169766