



Social media analytics – Challenges in topic discovery, data collection, and data preparation



Stefan Stieglitz^{a,*}, Milad Mirbabaie^a, Björn Ross^a, Christoph Neuberger^b

^a University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany

^b Ludwig-Maximilians-Universität München, Oettingenstr. 67, 80538 München, Germany

ARTICLE INFO

Keywords:

Social media analytics
Social media
Information systems
Big data

ABSTRACT

Since an ever-increasing part of the population makes use of social media in their day-to-day lives, social media data is being analysed in many different disciplines. The social media analytics process involves four distinct steps, data discovery, collection, preparation, and analysis. While there is a great deal of literature on the challenges and difficulties involving specific data analysis methods, there hardly exists research on the stages of data discovery, collection, and preparation. To address this gap, we conducted an extended and structured literature analysis through which we identified challenges addressed and solutions proposed. The literature search revealed that the volume of data was most often cited as a challenge by researchers. In contrast, other categories have received less attention. Based on the results of the literature search, we discuss the most important challenges for researchers and present potential solutions. The findings are used to extend an existing framework on social media analytics. The article provides benefits for researchers and practitioners who wish to collect and analyse social media data.

1. Introduction

Social media has evolved over the last decade to become an important driver for acquiring and spreading information in different domains, such as business (Beier & Wagner, 2016), entertainment (Shen, Hock Chuan, & Cheng, 2016), science (Chen & Zhang, 2016), crisis management (Hiltz, Diaz, & Mark, 2011; Stieglitz, Bunker, Mirbabaie, & Ehnis, 2017a) and politics (Stieglitz & Dang-Xuan, 2013). One reason for the popularity of social media is the opportunity to receive or create and share public messages at low costs and ubiquitously. The enormous growth of social media usage has led to an increasing accumulation of data, which has been termed Social Media Big Data. Social media platforms offer many possibilities of data formats, including textual data, pictures, videos, sounds, and geolocations. Generally, this data can be divided into unstructured data and structured data (Baars & Kemper, 2008). In social networks, the textual content is an example of unstructured data, while the friend/follower relationship is an example of structured data.

The growth of social media usage opens up new opportunities for analysing several aspects of, and patterns in communication. For example, social media data can be analysed to gain insights into issues, trends, influential actors and other kinds of information. Golder and Macy (2011) analysed Twitter data to study how people's mood

changes with time of day, weekday and season. In the field of Information Systems (IS), social media data is used to study questions such as the influence of network position on information diffusion (Susarla, Oh, & Tan, 2012).

Many existing research papers are isolated case studies (Kim, Choi, & Natali, 2016; Li & Huang, 2014; Oh, Hu, & Yang, 2016) that collect a large data set during a specific time frame on a specific subject and analyse it quantitatively. Despite the variety of disciplines such projects can be found in, they have much in common. The steps necessary to gain useful information or even knowledge out of social media are often similar. Therefore, the field of “Social Media Analytics” aims to combine, extend, and adapt methods for the analysis of social media data (Stieglitz, Dang-Xuan, Bruns, & Neuberger, 2014). It has gained considerable attention and subsequently acceptance in academic research, but there is still a lack of comprehensive discussions of social media analytics, and of general models and approaches. Aral, Dellarocas, and Godes (2013) presented a framework to organise social media research, and van Osch and Coursaris (2013) proposed a framework and research agenda explicitly limited to organisational social media. Both frameworks are geared towards classifying areas of research and, by extension, research questions, not methods to address these questions. While such frameworks are useful to decide what to research, and to locate individual projects within a larger context, they do not offer guidance

* Corresponding author.

E-mail address: stefan.stieglitz@uni-due.de (S. Stieglitz).

on how to carry out the research, and which challenges might arise. Of course, there is also research that discusses challenges researchers face when employing specific methods for analysing social media data, such as social network analysis (Kane, Alavi, Labianca, & Borgatti, 2014) or opinion mining (Maynard, Bontcheva, & Rout, 2012), and there are literature reviews focused on specific goals such as the identification of users who are influential offline (Cossu, Labatut, & Dugué, 2016) or on specific topics such as social bots (Stieglitz, Brachten, Ross, & Jung, 2017b). Yet social media analytics consists of several steps, of which data analysis is only one. Before the data can be analysed, they have to be discovered, collected, and prepared. An overview of the challenges of social media analytics is needed to be able to manage the complexity of conducting social media analytics.

We therefore carried out a systematic literature review, arguing that the complexity of these equally important steps has not yet been adequately covered in research, and there are no widely accepted standards on how to proceed within each of the steps. We explicitly focused on papers that deal with the challenges researchers face when discovering topics, and when collecting and preparing social media data for analysis, regardless of the method they later use during the analysis.

Our paper focuses on the following research question:

- RQ: What challenges do researchers face when discovering topics, collecting and preparing social media data for further analyses?

The answers to this question will help researchers who have little experience with the analysis of social media data, and still be useful for those who are experienced. Newcomers to the field will find the overview of common challenges and proposed solutions useful, so that difficulties can be considered before they arise, when setting up the research design, instead of encountering problems in an advanced phase of the research. Experienced researchers will get a bird's eye view of the existing research, which helps identify areas that may need further investigation and challenges that have not been addressed adequately yet.

The remainder of our paper proceeds as follows: first we provide a status quo of the literature on social media analytics and highlight the theoretical background for our article afterwards. Second, we describe our research design and highlight our findings afterwards. Third, we discuss our results, point out their impact, and discuss a model for social media analytics. Finally, we conclude our article and derive aspects for further research.

2. Theoretical background

The interdisciplinary research field of social media analytics (SMA) deals with methods of analysing social media data. Researchers have divided the analytics process into several steps. We use the steps of discovery, collection, preparation, and analysis, which we adapted from Stieglitz et al. (2014). The particular challenges of social media data, however, have not been addressed comprehensively in the SMA literature. To be able to classify these challenges, we draw on theory from the big data literature instead. In particular, we use the four V's: volume, velocity, variety, and veracity.

2.1. Social media analytics

Since the rise of social media usage in the last decade, people have been seeking to gain information from the crowd as an additional source to traditional media. We use the term social media to refer to "Internet-based applications that build on the ideological and technological foundations of Web 2.0", where Web 2.0 means that "content and applications are no longer created and published by individuals, but instead are continuously modified by all users in a participatory and collaborative fashion" (Kaplan & Haenlein, 2010). Because of the broad definition of social media, its application purposes are manifold.

Despite the large variety of platforms, some characteristics are common to many of them. Because of the amount of the content produced daily and the number of active users on the platforms, organisations are motivated to understand which issues and trends evolve to identify risks and chances in the communication and derive useful implications. Besides the amount of content, it is also relevant for organisations to understand who creates the content and which actors are the most influential drivers in the communication. Both businesses and non-profit organisations seek to collect the data produced by the crowd in order to gain insights into mass communication. The data is often collected with tools which communicate with the respective API of the social media platform, if one exists, and crawl the data.

The term "Social Media Analytics" has gained a great deal of attention. It is defined as "an emerging interdisciplinary research field that aims on combining, extending, and adapting methods for analysis of social media data" (Zeng, Chen, Lusch, & Li, 2010). Whilst the perspective on the system is one important aspect, another aspect is the perspective on the users who create the content. Research that adopts this perspective explores different roles in the communication and the effects a respective role can have on the communication and the diffusion of information (Stieglitz et al., 2017c). Influencers or opinion leaders, for example, can be identified through a social network analysis, and by examining their follower network, one can reveal the reach of such an individual (Mirbabaie, Ehnis, Stieglitz, & Bunker, 2014; Mirbabaie & Zapatka, 2017). Furthermore, the behaviour of the roles is examined in order to understand the causes of a key role in the network and the effects it has on the overall network (Bhattacharya, Phan, & Airoldi, 2015; Kefi, Mlaiki, & Kalika, 2015; Mirbabaie et al., 2014; Zhang, Zhao, Lu, & Yang, 2016). Companies such as media agencies have recognised the importance of influencers and use them e.g. for product placement. Furthermore, the analysis of social media content evolved in the last few years to one of the main research purposes in Information Systems. One research goal might be to identify and analyse the information diffusion (Liu, 2015; Zhang & Zhang, 2016).

Among others, three domains in which social media is important and generates visible benefits are 1) in businesses, in 2) crisis communication, mainly in disaster management, and in 3) journalism and political communication.

In one of the main areas of social media analytics, businesses make use of social media data, for several purposes (Kleindienst, Pflieger, & Schoch, 2015). Social media data can be useful for detecting new trends in the communication or issues which could involve uncontrollable bad publicity (Bi, Zheng, & Liu, 2014). Social media is also used as a channel to communicate with customers (Griffiths & McLean, 2015; Pletikosa Cvijikj et al., 2013). For supporting decision-making processes, companies make use of social media reports, created ex post and based on predefined key performance indicators, or they make use of a dashboard for getting on-going analyses based on real-time social media data (Tsou et al., 2015). Social Media is also used for product placement (Liu, Chou, & Liao, 2015) in the social web.

Crisis communication research is an example of a field where social media data has had an impact. Social media is often used as a channel for emergency management agencies to inform people in an affected area on the current status of the respective crisis or how to behave (Liu, 2015). Social media data in the context of crisis communication can also be analysed to gain additional, previously unknown information, if volunteers e.g. take pictures or videos and spread the information into the crowd. Collected social media data can be also analysed for detecting a specific location or area where the crisis occurs. By analysing GPS data if it is included in the data or by applying the method of Named Entity Recognition the location could be also derived from the text (Alsudais & Corso, 2015; Bendler, Ratku, & Neumann, 2014; Mirbabaie, Tschampel, & Stieglitz, 2016). The spread of a disease can be monitored by mining emotional tweets (Ji, Chun, Wei, & Geller, 2015). Especially for Emergency Management Agencies, it is important to understand the communication behaviour and the current status

Download English Version:

<https://daneshyari.com/en/article/7429038>

Download Persian Version:

<https://daneshyari.com/article/7429038>

[Daneshyari.com](https://daneshyari.com)