Authority-Based Conversation Tracking in Twitter

Twitter is undoubtedly one of the most widely used data sources to analyze human communication. The literature is full of examples where Twitter is accessed, and data are downloaded as the previous step to a more in-depth analysis in a wide variety of knowledge areas. Unfortunately, the extraction of relevant information from the opinions that users freely express in Twitter is complicated, both because of the volume generated—more than 6000 tweets per second—and the difficulties related to filtering out only what is pertinent to our research. Inspired by the fact that a large part of users use Twitter to communicate or receive political information, we created a method that allows for the monitoring of a set of users (which we will call authorities) and the tracking of the information published by them about an event. Our approach consists of dynamically and automatically monitoring the hottest topics among all the conversations where the authorities are involved, and retrieving the tweets in connection with those topics, filtering other conversations out. Although our case study involves the method being applied to the political discussions held during the Spanish general, local, and European elections of April/May 2019, the method is equally applicable to many other contexts, such as sporting events, marketing campaigns, or health crises.

1. Introduction

The widespread use of social networking services (SNS) and the fact that users increasingly trust the news shared by their contacts or friends over those in traditional media has significantly changed our information and communication habits. In particular, when contacts considered "opinion leaders" [1] publish a story within their sphere of influence, their contacts/followers tend to give it more credibility than they would give the mainstream media.

User activity in SNS somehow reflects actual news and everyday-life trends. Analyzing trends is highly relevant for understanding public opinion [2], as it allows researchers to examine public conversations and debates in detail, evaluate the behavior of a product after a marketing campaign, or analyze reactions to a particular event, among others. Since trending topics to some extent describe the opinion of a community and provide the means to analyze it, knowing where public attention is at a certain point in time becomes a topic of interest for researchers and professionals.

Not all SNS serve the same purpose in any case. Twitter, which according to Gadek [3] is the reference in microblogging platforms, allows users to post short messages called tweets, and interact with the tweets posted by others replying to them, quoting or retweeting them. User messages often include hashtags as a way of explicitly marking the relevant topics, easing such monitoring and analysis. Such a simple mechanism makes Twitter especially convenient for information retrieval and automatic processing purposes, as vast amounts of data are generated every day; data which provides valuable information for many different domains, such as political communication, consumer behavior, or disaster management, just to name a few. These data are freely and publicly available through Twitter's application programming interface (API), allowing for real-time monitoring of users' preferences, opinion, and behavior.

2. Twitter

Twitter is a window open to the spontaneous communication model that takes place in the real world, a model rooted in the fact that user behavior and occurrences are unpredictable and dynamic. Therefore, as messages in Twitter reflect real-time news and everyday-life trends, and given that many events cannot be accurately foreseen—e.g., natural disasters or accidents—it is really hard for analysts to anticipate the wording that will be used by users in their hashtags. Of course, if manual monitoring was carried out, new hashtags could be added as they appear as qualifiers of hot topics of interest, and yet this approach would have important shortcomings:

- It would introduce a significant delay between the moment when a new topic emerges and the starting point of its tracking. Such a delay would translate into losing relevant information or interaction;

- Human supervision, which is not always possible (e.g., late at night), would be necessary. Unfortunately, depending on humans can cause the introduction of failures derived from fatigue, incorrect interpretation of the information, inability to detect and track relevant changes, and others.

Moreover, it is also worth noting that, through hashtags, users pretend to label their content in a word or expression that summarizes what they are talking about; this is, however, a linguistic construct, and therefore, it carries its limitations. For an averagely informed person, the meaning of a hashtag can be relatively easy to understand or infer, but for an automatic process, this can become a complex task, as even the most famous hashtags such as #MeToo or #StayAtHome are not self-explanatory, and thus, need to be considered in their own context. Hashtag wording is part of the game in Twitter and often corresponds to specific (and deliberate) communication strategies.

The mechanism we propose consists of two separate parts, each governed by an algorithm. The first algorithm identifies the hottest topics of conversation of a group of authorities. It does this by continuously monitoring any tweet from the authorities, producing as an output an always-updated list of hashtags that includes the most interesting topics of conversation. Simultaneously, a second algorithm uses the current hashtag list as input, extracting and storing all the tweets tagged with any of the hashtags in the list. The whole mechanism must be initialized by setting up the set of authorities as well as some other parameters such as the time window frame. Once running, the software—which aims to minimize the loss of information—works in unattended mode without human intervention.

It is well known that political information is one of the most shared types of information on social media. In fact, literature on the use of Twitter for political activities abound, such as those studies on the effect of social media, especially Twitter, as a facilitator in political campaigns [4] and protests [5] worldwide. The fact is that two-thirds of social media users show some kind of political engagement by, for instance, following candidates, posting thoughts about political issues, or pressing friends to vote [6]. This behavior is especially evident in Twitter opinion leaders, who consistently show a higher involvement in political processes [7]. Besides, politically engaged young people integrate social media use into their existing organizations and political communications [8]. This prominence of the use of Twitter for politics, the interest that the study of political information awakens, and the rising concerns about the effect of false stories (or "fake news") on social media [9] are what inspired us to apply our model to the analysis of political discussions during the Spanish general elections as a use case.

References

1. Jason Turcotte; Chance York; Jacob Irving; Rosanne M. Scholl; Raymond J. Pingree; News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking. *Journal of Computer-Mediated Communication* **2015**, *20*, 520-535, 10.1111/jcc4.12127.

2. Fang, Y.; Chen, X.; Song, Z.; Wang, T.; Cao, Y. Modelling propagation of public opinions on microblogging big data using sentiment analysis and compartmental models. In Natural Language Processing: Concepts, Methodologies, Tools, and Applications; IGI Global: Hershey, PA, USA, 2020; pp. 939–956.

3. Guillaume Gadek; Alexandre Pauchet; Nicolas Malandain; Laurent Vercouter; Khaled Khelif; Stéphan Brunessaux; Bruno Grilhères; Topological and topical characterisation of Twitter user communities. *Data Technologies and Applications* **2018**, *52*, 482-501, 10.1108/dta-01-2018-0006.

4. Shannon C. McGregor; Rachel R. Mourão; Logan Molyneux; Twitter as a tool for and object of political and electoral activity: Considering electoral context and variance among actors. *Journal of Information Technology & Politics* **2017**, *31*, 1-14, 10.1080/19331681.2017.1308289.

5. Jost, J.T.; Barberá, P.; Bonneau, R.; Langer, M.; Metzger, M.; Nagler, J.; Sterling, J.; Tucker, J.A.; How social media facilitates political protest: Information, motivation, and social networks. *Polit. Psychol.* **2018**, *39*, 85–118, .

6. Rainie, L.; Smith, A.; Schlozman, K.L.; Brady, H.; Verba, S.; Social media and political engagement.. *Pew Internet Am. Life Proj.* **2012**, *19*, 2–13, .

7. Park, C.S.; Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. *Comput. Hum. Behav.* **2013**, *29*, 1641–1648, .

8. Ariadne Vromen; Michael A Xenos; Brian Loader; Young people, social media and connective action: from organisational maintenance to everyday political talk. *Journal of Youth Studies* **2014**, *18*, 80-100, 10.1080/13676261.2014.933198.

9. Allcott, H.; Gentzkow, M.; Social media and fake news in the 2016 election. *J. Econ. Perspect.* **2017**, *31*, 211–236, .