Expert Systems with Applications 42 (2015) 3381–3395

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A new model to quantify the impact of a topic in a location over time with Social Media

J. Bernabé-Moreno [a], A. Tejeda-Lorente [a], C. Porcel [b,*], E. Herrera-Viedma [a]

[a] *University of Granada, Department of Computer Science and Artificial Intelligence, Granada, Spain*
[b] *University of Jaén, Department of Computer Science, Jaén, Spain*

A R T I C L E   I N F O

Article history:
Available online 20 December 2014

Keywords:
Social Media impact
Topic Engagement
Topic Exposure
Social Media sensor
Geo-located Social Media
Social network analysis
RFM
Information extraction

A B S T R A C T

Social Media can be used as a thermometer to measure how society perceives different news and topics. With the advent of mobile devices, users can interact with Social Media platforms anytime/anywhere, increasing the proportion of geo-located Social Media interactions and opening new doors to localized insights. This article suggests a new method built upon the industry standard Recency, Frequency and Monetary model to quantify the impact of a topic on a defined geographical location during a given period of time. We model each component with a set of metrics analyzing how users in the location actively engage with the topic and how they are exposed to the interactions in their Social Media network related to the topic. Our method implements a full fledged information extraction system consuming geo-localized Social Media interactions and generating on a regular basis the impact quantification metrics. To validate our approach, we analyze its performance in two real-world cases using geo-located tweets.

1. Introduction

The usage of Social Media (SM) is an ever-growing phenomenon (White, 2013). Media consumers are increasingly shifting from classic (printed) media to digital platforms. As a result the communication stops being one-way with clearly defined *author/reader* roles. With the advent of the web 2.0, the definition of *author* started to blur. The blogosphere empowered readers to make their own contributions to the content published by a given author, which radically increased the information richness, adding further perspectives and points of view. Simultaneously, media started to be democratized, as anybody could start a blog and the visibility of the blog in the search engines was determined a priori by the number of people that considered the blog to be relevant outside the realm of the paid search (Page, Brin, Motwani, & Winograd, 1998).

The SM platforms based on the concept of *micro-blogging* took it to the next level, as everybody could be an author and a reader anytime. The *push-first, comment-later* paradigm so popular in the blogosphere started to look old-fashioned. Rather, anybody was empowered to initiate a communication, enrich an existing thread, jump from a thread to another one, ignore, criticize, share richer content like pictures, videos, etc. The ease of publishing, sharing and consuming content boosted the adoption of these Social Media platforms as the place to talk anytime about everything with everybody. The best example is Twitter, which has become a communication platform for almost all the digital world (Kwak, Lee, Park, & Moon, 2010). By March 2012 the platform counted 140 million active users creating an average of 340 million tweets a day (Bennett, 2012). The night of November 7th, during 8:11 and 9:11 pm when the world wanted to share the results of the US elections, an average of 9965 Tweets per Second (TPS)[1] resulted in the creation of more than 35 million tweets within one hour.

With the advent of wireless internet technologies based on WiFi hot-spots and mobile communication networks, the Social Media content creation became more pervasive. The access to the digital media was no longer exclusive to desktops; the rise of the smartphones and mobile data packages enabled the always-on era and opened the door to a new set of insights based on the location where the user interacted with the social network. As the proportion of geo-located SM interactions increased, the *geo-fencing* or delimitation of the location boundaries where the SM dialog took

* Corresponding author.
E-mail addresses: juan.bernabe-moreno@webentity.de (J. Bernabé-Moreno), atejeda@decsai.ugr.es (A. Tejeda-Lorente), cporcel@ujaen.es (C. Porcel), viedma@decsai.ugr.es (E. Herrera-Viedma).

[1] https://blog.twitter.com/2012/bolstering-our-infrastructure

http://dx.doi.org/10.1016/j.eswa.2014.11.067
0957-4174/

place became more accurate. These new capabilities led to more meaningful and representative analytic results, to the point that the SM activity could be taken as a good indicator of what is happening anytime anywhere.

The influence and impact in the SM channel has been matter of research almost from the advent of the modern SM networks and platforms. Yet, the research community mainly focused on understanding and modeling the impact of a particular user or a particular group of users on their own and foreign social networks. Our intent here is to prove that the impact of a given topic can be measured, quantified and monitored over time. Obviously, this topic centric geo-located impact measuring would open a new window of possibilities in different domains, such as understanding the performance of marketing campaigns on a given area, or understanding the affinity of local communities to certain marketing offers. Likewise, the generated insights can be used in the area of recommender systems and applied in different scopes, especially in e-commerce and digital media (Porcel, Tejeda-Lorente, Martnez, & Herrera-Viedma, 2012; Tejeda-Lorente, Porcel, Peis, Sanz, & Herrera-Viedma, 2014). Unlike the metrics typically used to assess the SM influence of a particular user, which mainly rely on well-defined entities and parameters present standard-wise in social network platforms (like *User*, *Friend*, *Follower*, etc.), there's no entity to represent a *topic*. Thus, modeling techniques need to be applied, which introduces a new level of complexity entering in the realm of semantic web (Berners-Lee, Hendler, & Lassila, 2001) and Natural Language Programming (NLP) (Manaris, 1998). Although our aim was not solving any NLP problem, we implemented a system to extract the required information from the Social Media networks and to apply the quantification methodology for a topic which relies on a whole set of NLP components.

The purpose of this paper is to define a new method to quantify the impact of a topic during a period of time on a given place based on how the users located in this place are exposed to the topic over their social network and how they actively interact or engage with the topic themselves. In other words, we want to turn the Social Media platform Twitter into a topic impact thermometer. Our method relies on the well-established industry-standard Recency, Frequency and Monetary schema (RFM) (Bult & Wansbeek, 1995). RFM models have been employed in the industry for almost 30 years to identify and segment the customer base in countless companies across industries based on following questions: *How recently? How often? How much value?*. In our case we rely on the same RFM components to make the value modeling for topic impact dependent on the time and on the number of interacting users. Each component consists of a set of metrics based on the number of users interacting with the topic in a location, the Engagement of these users with the topic – computed by the share of the content they produce related to the topic –, and their Exposure to the content their network creates related to the topic, with the option of creating an aggregate index as well.

This paper is structured as follows: firstly the background information where we briefly review the related work is presented. Then, we introduce our method together with metrics to quantify the impact of a topic on the Social Media channel. After that, we present a system that implements our metrics and then we show some practical examples of topic impact quantification. Finally, we share our conclusions and point out future work on this topic.

2. Background and related work

In this section we provide all the background information and related work to base our research, starting with the review of impact modeling and topic diffusion, introducing the RFM model and finally discussing the approaches to topic modeling and information extraction in Social Media.

2.1. Topic diffusion and Social Media impact

The diffusion of news or topics in the social networks has been subject of intense research especially in the last years (Cavusoglu, Hu, Li, & Ma, 2010; Centola, 2010; Stieglitz & Dang-Xuan, 2013). Although the methodology we propose in this article is not intended to explain the dynamics of the topic propagation in the social networks, rather to provide a measure for the impact, there are common elements used in both researching lines to understand the contribution of a given user based on how active she/he is, the handling of the variation over the time of the topic-related activity and the semantic definition of the topic. Guille and Hacid (2012) defined three dimensions playing a role in the propagation of a topic: social, semantics and temporal to model the probability of dispersion. The social dimension is defined taking into account the users' activity index, the ratio of directed tweets to the user, the mentioning rate and whether the user being mentioned is directly related to the mentioner. On the other hand, the semantics is based on the presence of a keyword in the message being propagated. The temporal dimension is provided as a computation of the user activity in 6 partitions of the day, but probably leaving the door open to finer time granularity.

Rajyalakshmi, Bagchi, Das, and Tripathy (2012) demonstrated the role of the strong links in the virality of the topics by modeling the diffusion with a stochastic approach, identifying as driving parameters the users activity time and the fading out effect – represented as a weight decay for a topic as time passed by. In their work, two cases are clearly separated: users creating instances of a global topic or users copying it from their network – local social network effect vs. the overall trending effect. Romero, Meeder, and Kleinberg (2011b) established a mechanism relying on Exposure curves to quantify the impact Exposure to other users in making them adopt a new behavior (e.g.: turning them from passive to active contributors or to start using a hash-tag, etc.). In addition, there have been several approaches to model the influence of a particular user in his/her own and in the global Social Media network. Ye and Wu (Ye & Wu (2010)) defined 3 different metrics to quantify the social influence: followers influence – the higher the number of followers, the higher the influence –, reply influence – the more replies one user receives, the more influential the user is –, and re-tweet influence – the more re-tweets, the more influent. Kwak (Kwak et al., 2010) suggested also 3 metrics but substituted the reply influence by one inspired by the Google Search PageRank algorithm (Page et al., 1998) to allow the propagation of influence. Depending on the metric applied the ranking of the top users varied. Romero, Galuba, Asur, and Huberman (2011a) demonstrated that influent users are those whose contributions are not just consumed but also forwarded and therefore overcome the so called passivity and more interestingly, that the popularity of an user and its influence do not quite often correlate. Cha, Haddadi, Benevenuto, and Gummadi (2010) differentiated 3 kinds of influence for a Social Media user: due to the size of the user's audience or social network indegree influence –, due to the generated content with pass-along value retweet influence, which is also aligned with the passivity activity work presented by Romero et al. (2011b) and due to the Engagement in others' conversation mention influence – and all of them are present as component for either Exposure or Engagement when applicable in our approach. The use of geo-localized SM interactions to provide information about local communities is a field of incipient research. In Scellato, Noulas, Lambiotte, and Mascolo (2011) the authors provide an extensive description of the social spatial properties of location based social networks. In Backstrom, Sun, and Marlow (2010), the authors rely