Technological Forecasting & Social Change 133 (2018) 220–228

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Identifying the features of reputable users in eWOM communities by using Particle Swarm Optimization

M.R. Martínez-Torres[a,*], F.J. Arenas-Marquez[a], M. Olmedilla[b], S.L. Toral[a]

[a] Facultad de Turismo y Finanzas, Universidad de Sevilla, Avda. San Francisco Javier, s/n, 41018 Sevilla, (Spain)
[b] Léonard de Vinci Pôle Universitaire, Research Center, 92 916 Paris La Défense, France

ARTICLE INFO

Keywords:
Particle Swarm Optimization
Electronic Word-of-Mouth communities
Reputable users
Classification rules

ABSTRACT

Electronic Word-of-Mouth communities have become popular over the last several years as websites where people can share their online reviews about any type of product or service. As a mechanism to improve trust, posted reviews can also be scored by the rest of the community in terms of helpfulness, so users can reach a high level of reputation through their interactions with other users and can thus increase their credibility. The aim of this paper is to investigate the main patterns of activity that characterize reputable users by using a set of classification rules. However, the class of reputable users is only a fraction of the total number of users. Due to the imbalance between the classes, i.e., reputable and non-reputable users, and the high dimensionality of the problem, an evolutionary computation algorithm such as Particle Swarm Optimization (PSO) is applied to obtain the main activity patterns of reputable users. Obtained results can help us better understand the mechanism of trust in eWOM communities and avoid the undesirable manipulation of reputations by false accounts.

1. Introduction

Electronic word of mouth (eWOM) communities are a manifestation of the increasing popularity of user-generated content (Fu et al., 2015). With the advent of the Internet, users have plenty of opportunities to share information with other people all over the world. Shared information can be videos, photos, or stories but also recommendations about products and services (Martinez-Torres et al., 2015). Sharing recommendations through the Web is a modernized version of traditional WOM with its own characteristics (Hajli and Sims, 2015; Vilpponen et al., 2006). eWOM websites facilitate the exchange of information among users about a wide variety of products and services. Similar to traditional WOM, consumers seek and give opinions about different products, services, brands, companies and experiences. However, unlike traditional WOM, eWOM communications are frequently anonymous and asynchronous (Jeong and Jang, 2011), directed to multiple people without time and location constraints. This information is also persistent and publicly available, so it can be collected from consumer opinion websites and further analysed (Dellarocas et al., 2007).

The success of such communities can be explained because people tend to prefer information and experiences shared by other consumers rather than information provided by manufacturers (Lee et al., 2013; Parmentier and Mangematin, 2014). As a result, they are starting to influence the consumers' decision-making process in a decisive way (Jones et al., 2009). Following this trend, companies are increasingly monitoring social media activity to pick up all the information related to their products, services and brand image. As with other online communities, there are opinion leaders that represent a minority within the community (Jiyao and Reynolds, 2010). They not only influence the opinions of other consumers but also disseminate positive and negative opinions. Therefore, they represent a target group for companies that are interested in monitoring user preferences or developing viral marketing campaigns.

However, the anonymity of users in eWOM websites also introduces some credibility issues. Hidden behind anonymity, users can post fake reviews to downgrade products or promote certain products. Recent studies estimate that approximately 25–30% of all online reviews are deceptive reviews (Roberts, 2013; Li et al., 2014). Although there is a consensus in the fact that online reviews will continue growing in the future, the presence of fake reviews is a threat that can undermine consumer confidence on shared opinions. For this reason, eWOM websites such as Amazon, TripAdvisor or Ciao typically include some mechanisms to obtain and publicly show the reputation of users, such as the ratings that posted reviews receive from the rest of the community. Therefore, people can judge the credibility of the communicator based on the associated cues through these online reputation systems (Cheung and Lee, 2012). In fact, these associated cues are often

* Corresponding author.
E-mail addresses: rmtorres@us.es (M.R. Martínez-Torres), fjarenas@us.es (F.J. Arenas-Marquez), maria.olmedilla_fernandez@devinci.fr (M. Olmedilla), storal@us.es (S.L. Toral).

https://doi.org/10.1016/j.techfore.2018.04.017
Received 18 November 2016; Received in revised form 7 March 2018; Accepted 9 April 2018

cited as an important determinant of information sharing behaviour (Constant et al., 1996). The main disadvantage of using ratings given by other users is that they can be easily manipulated through the creation of fake accounts to artificially boost the reputation of a fraudster. More elaborate detection methods are mainly based on supervised approaches such as linear/logistic regression models, naive Bayesian models, SVM or nearest-neighbour algorithms (Jiang et al., 2016). However, using classifiers to predict reputed users has the problem that reputed and non-reputed classes are not balanced, as reputed users are typically a fraction of the total community, and classifiers do not perform well when working with unbalanced classes. Additionally, most of them rely on features that users receive from the rest of the community, so they can be boosted using fake accounts.

This paper advances the identification of reputable users in two ways: first, we propose using the patterns of activity and interactions of users within eWOM, so our methodology does not rely on metrics generated by other users; second, our proposed methodology based on evolutionary computation can classify users, even with very unbalanced classes, as is the case with reputed users.

The most important implication of this study is the improvement of trust and credibility in public online environments such as review sites. Although opinion websites try to improve mechanisms to increase the credibility of shared opinions, the patterns of fake accounts also change even faster than detection algorithms (Jiang et al., 2016). Therefore, it is necessary to proceed with more complex metrics and even non-linear decision rules, which are more difficult to scam. The proposed algorithm can be used by review websites to predict and filter malicious reviews prior to their publication. It would also be possible to implement it as an online software tool, so it can adapt to the changing patterns followed by deceptive reviewers.

The rest of the paper is structured as follows: the next section details some related work. Section 3 describes the research hypotheses. Section 4 formulates the problem as an optimization problem and introduces the proposed methodology as well as the variables used in this study. Section 5 shows the obtained results from the application of evolutionary computation to the identification of reputable users. Section 6 discusses the results obtained and their implications. Finally, Section 7 concludes the paper.

2. Related work

The success of eWOM is based on the credibility of shared reviews. Online reviews influence the attitude towards the product and the purchase intention whenever they come from a credible source (Shan, 2016). However, the meaning of a credible source in eWOM is different from that in traditional WOM, where the source is someone belonging to the inner circle of the information seeker. Online reviews are posted by anonymous users only identified by an alias. This is why eWOM websites not only display the content and the author of the review, but they also make available other self-generated profiles and system-generated profiles about reviewers. In eWOM, the most salient cue about the source credibility is the reviewer's reputation, which is rated by other eWOM users (Cheung et al., 2009).

Online reputation systems help users to identify the reputation of others by collecting, distributing, and aggregating feedback about participants' past behaviour (Resnick et al., 2000). The system-generated profile includes a brief information about the author posting the review as well as the community-rated reputation of reviewers, indicating the perceived usefulness of previously posted reviews and other products purchased or rated (Wu, 2013). Using this information, people can judge the credibility of reviews by considering the reputation or previous activity of the author. In this line, many eWOM websites such as Amazon, TripAdvisor or Ciao, allow users to vote on the “helpfulness” of posted reviews (Arenas-Marquez et al., 2014; Gonzalez-Rodriguez et al., 2016). In some websites such as Ciao, a trust network is also made explicit, created by those other members of the community who specifically trust a user as a reviewer (Ku et al., 2012). The quality of this information is assured through the peer evaluations given by other community members (King et al., 2014). This is a key issue for eWOM communities, as their success is based on the credibility of shared opinions. As this is publicly available information, it can be used to distinguish reputable reviewers from the rest of the community. However, online reviews may be manipulated by interested parties through fake reviews (Hu et al., 2012), taking advantage of the anonymity of the Internet. Imposters can post fake reviews and try to downgrade or discredit competitors. Many of the posted fake reviews can be determined using some basic reputation mechanisms. However, even the online reputation systems can be manipulated, so community managers are searching for more complex mechanisms to filter malicious behaviours.

Some previous studies have specifically studied the profiles of reputable reviewers in eWOM. Cheung et al. (2012b) focus on review credibility and conclude that people rely not only on argument quality to evaluate online consumer reviews but also on other review cues, such as source credibility, review consistency, and review sidedness. They explicitly tested that source credibility has a positive effect on review credibility. Previous works have also found that community members rated reviews containing identity-descriptive information as more helpful than anonymous reviews (Forman et al., 2008). Such online identity disclosure behaviour may facilitate forming the relationships, common bonds, and social attraction that online community members value (Ren et al., 2007). The reputation mechanism was analysed by Bolton et al. (2004). They suggest that relying solely on a cumulative measure of trustworthiness may not be appropriate because it hides information critical to the members' decision to trust.

The role of the users' trust network was studied by Ku et al. (2012). As a novel contribution, these authors considered the trust network as a 2-hop network. They define the trust intensity given by the size of the trust network as well as the average intensity of the 2-hop neighbours, which is the trust intensity of the members of the trust network. Both relate positively to the reputation of the member. This study also considers several other factors related to review behaviour, such as the number of reviews and the degree of review focus on a given category of products. When a member has a high level of expertise in some category, he or she often writes more reviews within this category (Huang et al., 2010), and the degree of review focus was supported in the study by Ku et al. (2012). However, the positive effect of the number of reviews in a given category was not supported.

In addition to posting reviews, reviewers can also score products using a numerical scale, and this is another feature related to reputable reviewers' behaviour (Baek et al., 2015). Finally, the reviewers' experience, as measured by the number of previously posted reviews, has also been considered in other studies (Hussain et al., 2017; Racherla et al., 2012)

The analysis of online recommendation systems with reputation mechanisms faces two important challenges. The first challenge is collecting the data. eWOM communities receive posts from thousands of users whose profiles are continuously updated. Therefore, it is necessary to check the system-generated profiles one by one, collecting the data from the HTML code and storing the information. This task requires the development of a specific crawler to access the specific required data within the webpages. Fortunately, many software platforms, such as Python and R, have modules specifically designed to access web data, and they can be used by a software programmer to build the required crawlers (Martínez-Torres, 2013; Olmedilla et al., 2016). The second challenge is related to the problem of unbalanced classes when dealing with reputable reviews. Frequently in online communities, only a small percentage of users are actually reputable reviewers. This is a consequence of the participation inequality exhibited in many online communities, which means that only a fraction of potential users are active within the community (Martínez-Torres, 2014). In addition, from this fraction, only a small percentage achieve a high reputation. Therefore, if we try to formulate a classification