Contents lists available at ScienceDirect

Information Sciences

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

Pollution, bad-mouthing, and local marketing: The underground of location-based social networks

Helen Costa [a], Luiz H.C. Merschmann [a], Fabrício Barth [b], Fabrício Benevenuto [c,*]

[a] Federal University of Ouro Preto, Ouro Preto, Brazil
[b] VAGAS Tecnologia Ltda, São Paulo, Brazil
[c] Federal University of Minas Gerais, Belo Horizonte, Brazil

ARTICLE INFO

Article history:
Received 10 September 2013
Received in revised form 20 February 2014
Accepted 26 March 2014
Available online xxxx

Keywords:
Location-based social network
Social network
Tip spam
Tip analysis

ABSTRACT

Location Based Social Networks (LBSNs) are new Web 2.0 systems that are attracting new users in exponential rates. LBSNs like Foursquare and Yelp allow users to share their geographic location with friends through smartphones equipped with GPS, search for interesting places as well as posting tips about existing locations. By allowing users to comment on locations, LBSNs increasingly have to deal with new forms of spammers, which aim at advertising unsolicited messages on tips about locations. Spammers may jeopardize the trust of users on the system, thus, compromising its success in promoting location-based social interactions. In spite of that, the available literature is very limited in providing a deep understanding of this problem.

In this paper, we investigated the task of identifying different types of tip spam on a popular Brazilian LBSN system, namely Apontador. Based on a labeled collection of tips provided by Apontador as well as crawled information about users and locations, we identified three types of irregular tips, namely local marketing, pollution and, bad-mouthing. We leveraged our characterization study towards a classification approach able to differentiate these tips with high accuracy.

1. Introduction

Location Based Social Networks (LBSNs) are new Web 2.0 systems that are attracting new users in exponential rates. Nearly one in five smartphone owners access this kind of service via their mobile devices [1]. LBSNs like Foursquare and Yelp allow users to share their geographic location with friends through smartphones equipped with GPS, search for interesting places as well as posting tips about existing locations.

In Brazil, a popular LBSN system is named Apontador,[1] and it includes the main features of systems like Foursquare and Yelp. It allows users to search for places, register new locations and check-in in locations using smartphones. Additionally, Apontador contains one of the most interesting features of LBSNs, which is allowing users to post-tips to existing places. Due to these tips, with a smartphone and access to a LBSN, a user might not only find nearby places to visit, but also read

* Corresponding author. Tel.: +55 31 3409 5860.
E-mail addresses: helen.c.s.costa@gmail.com (H. Costa), luizhenrique@iceb.ufop.br (L.H.C. Merschmann), fabricio.barth@vagas.com.br (F. Barth), fabricio@dcc.ufmg.br (F. Benevenuto).

[1] http://www.apontador.com.br/.

http://dx.doi.org/10.1016/j.ins.2014.03.108
0020-0255/© 2014 Elsevier Inc. All rights reserved.

suggestions about what to order, what to buy or even what to avoid on specific places. Thus, tips in LBSN work as crowdsourcing recommendations about specific locations.

Although appealing as a mechanism to enrich the user experience on the system, tips open opportunities for users to disseminate unsolicited messages. LBSNs increasingly have to deal with different forms of attacks, which aim at advertising unsolicited messages instead of true tips about locations. Most of studies in this field have focused in opinion spam [13,14,16], a fake opinion that deliberately mislead readers by giving positive reviews to a place in order to promote it and/or by giving negative reviews in order to damage the place's reputation. However, little is known about the activities of other types of opportunistic behavior in the context of LBSNs. As example, consider tips containing local advertisement or even tips containing irrelevant content that is unrelated or inappropriate to the place (e.g. links to pornography websites posted at schools). This kind of tips may also jeopardize the trust of users on the existing tips available in the system, thus compromising its success in promoting location-based social interactions. Furthermore, tip spam may compromise user patience and satisfaction with the system since users need to filter out spam from what is worth reading. In spite of that, the available literature is very limited in providing a deep understanding of this problem in an environment where places are the central object.

In this paper, we address the issue of identifying different forms of tip spam in LBSNs adopting a 3-step approach. First, we categorized tip spam into three different classes based on a labelled dataset of spam and non-spam tips. Second, we analyzed a number of attributes extracted from the tips' content and from the user behavior on the system aiming at understanding their relative discriminative power to distinguish among the different classes of tip spam. Lastly, we investigated the feasibility of applying supervised machine learning methods to identify these tip spam classes. Our approach was not only able to correctly identify a significant part of the tips as spam or non-spam, but it was also able to differentiate tips from different spam classes. In this study, we identify three types of spam tips, namely (i) *local marketing*: tips containing local advertisement, sometimes about the current place or about a business related to it, (ii) *pollution*: tips with content unrelated to the place, and (iii) *bad-mouthing*: tips containing very aggressive comments about the places, their owners or other users who posted tips in the system.

The rest of the paper is organized as follows. Next section presents related efforts in this theme. Section 3 describes our strategy to categorize tip spam classes. Section 4 investigates a number of attributes and their ability to distinguish tips from different classes. Section 5 describes and evaluates our strategy to detect the three types of tip spam. Finally, Section 6 offers conclusions and directions for future work.

2. Related work

Spam detection has been observed in various social network systems, including YouTube [6], Twitter [5,11], Facebook [10], and MySpace [15]. Particularly, Benevenuto et al. [5] approached the problem of detecting spammers on Twitter. By using a labeled collection of users manually classified, they applied a classification machine learning approach to differentiate spammer users from legitimate ones. Similarly, Lee et al. [15] created social honeypots to identify a set of spammers on MySpace and Twitter. In addition to showing that social honeypots are accurate in identifying spammers, they propose a machine learning method to detect spammers in these two systems. Although these methods inspired the approach we used here, our work is complementary to them as we investigate spam in a different environment, identifying the specific features that allow us to accurately differentiate classes of tip spam. In a recent effort, Vasconcelos et al. [18] crawled Foursquare to characterize the user behavior based on information of *tips*, *dones* and *toDos*. Using an expectation maximization clustering algorithm, they clustered users into four groups, out of which one contained a large number of tip spammers. Thus, they presented the first evidence of spam in LBSNs.

In the context of reviews about products, Jindal and Liu [13] investigated the detection of opinion spam on product reviews, based on the analysis of reviews from amazon.com. Opinions spam are untruthful opinions that deliberately mislead readers by giving undeserving positive reviews to some target objects in order to promote the objects and/or by giving unjust or malicious negative reviews in order to damage the objects' reputation. Thus, they proposed a model to detect harmful opinions, based on duplicate reviews (copies), which inspired a few metrics proposed in our work. Recently, Kakhki et al. [14] approached the problem where users create multiple identities and use these identities to provide positive ratings on their own content or negative ratings on others' content. Then, they developed a system named Iolaus to mitigate the effect of rating manipulation in online content ratings services like tips in LBSNs. Different from these efforts, we explore other kinds of spam in LBSNs. Thus, our work is complementary to theirs.

In a previous study, we preliminarily approached the problem of detecting tip spam by creating a small test collection composed of spam and non-spam tips, and applying a binary classification strategy to detect tip spam [9]. The present work builds on this preliminary effort by providing a much more thorough, richer and solid investigation of the feasibility and tradeoffs in detecting tip spam in LBSN, considering a much larger test collection, a richer set of user attributes, as well as different classes of malicious and opportunistic behaviors. Similarly, Aggarwal et al. [2] presented a method to identify different types of spammers in Foursquare. The authors found that Foursquare users with irregular tipping activity can be classified into four categories – advertising/marketing, self-promotion, abusive or malicious. Using machine learning techniques, they were able to distinguish between legitimate and spam users on Foursquare. Our work here is complementary to [2] as we investigate a different set of features, related to content of tips, to places where the tips were posted, to users

Please cite this article in press as: H. Costa et al., Pollution, bad-mouthing, and local marketing: The underground of location-based social networks, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.03.108