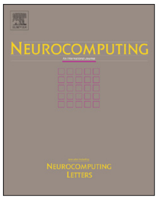
Neurocomputing 201 (2016) 51–65

Contents lists available at ScienceDirect

Neurocomputing

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



Co-detecting social spammers and spam messages in microblogging via exploiting social contexts

Fangzhao Wu [a,*], Jinyun Shu [b], Yongfeng Huang [a], Zhigang Yuan [a]

[a] *Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*

[b] *Beijing University of Posts and Telecommunications, Beijing, China*

ARTICLE INFO

Article history:
Received 21 December 2015
Received in revised form
2 March 2016
Accepted 25 March 2016
Communicated by Y. Chang
Available online 26 April 2016

Keywords:
Spam detection
Social spammer
Microblog

ABSTRACT

Microblogging websites, such as Twitter, have become popular platforms for information dissemination and sharing. However, they are also full of spammers who frequently conduct social spamming on them. Massive social spammers and spam messages heavily hurt the user experience and hinder the healthy development of microblogging systems. Thus, effectively detecting the social spammers and spam messages is of great value to both microblogging users and websites. Existing studies usually treat social spammer detection and spam message detection as two separate tasks. However, social spammers and spam messages have strong inherent connections, since social spammers tend to post more spam messages and spam messages have high probabilities to be posted by social spammers. Thus combining social spammer detection with spam message detection has the potential to boost the performance of both tasks. In this paper, we propose a unified approach for social spammer and spam message co-detection in microblogging. Our approach utilizes the posting relations between users and messages to combine social spammer detection with spam message detection. In addition, we extract the social relations between users and the connections between messages to refine detection results. We regard these social contexts as the graph structure over the detection results and incorporate them into our approach as regularization terms. Besides, we introduce an efficient optimization algorithm to solve the model of our approach and propose an accelerated method to tackle the most time-consuming step. Extensive experiments on a real-world microblog dataset demonstrate that our approach can improve the performance of both social spammer detection and spam message detection effectively and efficiently.

1. Introduction

Microbblogging websites, such as Twitter[1] and Sina Weibo,[2] have become popular platforms for information dissemination and sharing [1]. Hundreds of millions of users frequently post short messages on these websites to release latest news and share their opinions on various topics, such as political events, products, and daily life. However, due to their huge popularity, microbblogging websites are also recognized as ideal places to conduct spamming [2,3,1]. Massive fake microblogging accounts, which are known as social spammers [4], post masses of spam messages for various purposes, including conducting social advertising, collecting users' personal information, promoting affiliate websites and so on [5,6,3]. These spam messages may contain dangerous content and URLs that are related to scams, malware, and phishing [7,6,1]. Spam messages are also used to conduct political astroturfing [5,8]. These massive social spammers and spam messages seriously hurt the user experience and hinder the healthy development of microblogging systems [6]. Thus, it will be beneficial to both microblogging websites and users if we can effectively detect and filter these social spammers and spam messages.

Social spammer detection and spam message detection have been studied for several years, and various methods have been proposed. For example, some researchers proposed to detect social spammers via social network analysis [9,10]. The core assumption behind these methods is that compared with legitimate users social spammers cannot build enough social trust relations [1]. However, different from other online social networking platforms, such as Facebook and Youbute, in microblogging websites a user can follow anyone else without the followee's permission, and it is not too hard for social spammers to gain sufficient social relations

[*] Corresponding author.
E-mail address: wfz12@mails.tsinghua.edu.cn (F. Wu).

[1] https://twitter.com/

[2] http://www.weibo.com/

http://dx.doi.org/10.1016/j.neucom.2016.03.036
0925-2312/© 2016 Elsevier B.V. All rights reserved.

with legitimate users [11]. Thus these social network analysis based methods may not work well on social spammer detection in microblogging platforms [1]. Another line of research for social spammer detection is analyzing their attributes, online behaviours, and textual content of the microblog messages they post [2,12–16]. These methods are based on the assumption that the online behaviours and textual content of social spammers are different from those of legitimate users. However, social spammers may continue to change their posting behaviours and try to behave like normal users [15]. In addition, besides spam messages, social spammers may also post some normal messages. These strategies may lead to these content and behaviour analysis based methods less accurate. As to spam message detection, existing methods mainly focus on extracting effective features and building a classifier using machine learning techniques [17–19]. For example, some researchers found that URLs and timestamps are informative features for identifying spam messages [20].

In general, in existing studies social spammer detection and spam message detection are mainly regarded as two separate tasks. However, in microblogging websites, a common phenomenon is that social spammers tend to post more spam messages, and spam messages have high probabilities to be posted by social spammers. Thus there are strong connections between social spammers and spam messages. Detecting social spammers and spam messages simultaneously may achieve better performance than conducting each task in isolation. In addition, the social connections between users and those between messages may also be helpful for social spammer detection and spam message detection. For example, social spammers are often followed by other spammers, because they frequently collaborate in this way to build more social relations and pretend to be normal users. Many spam messages contain the same URLs because they belong to the same promoting campaign [3]. Some spam messages contain the same hashtags for the purpose of advertising for the same product, brand or service. An illustrative example of these social contexts is shown in Fig. 1.

Motivated by the above-mentioned observations, in this paper we propose a unified approach for Social Spammer and S pam M essage Co-D etection (*S3MCD*) in microblogging, which can exploit various kinds of social contexts. In our approach, social spammer detection and spam message detection are bridged by the posting relations between users and messages. The results of social spammer detection can help refine the results of spam message detection, and vice versa. Both the performance of social spammer detection and spam message detection can be boosted. In addition, our approach can incorporate the social contexts of user–user relations and message–message relations to refine the social spammer and spam message detection results by modelling them as the graph structure over the detection results.

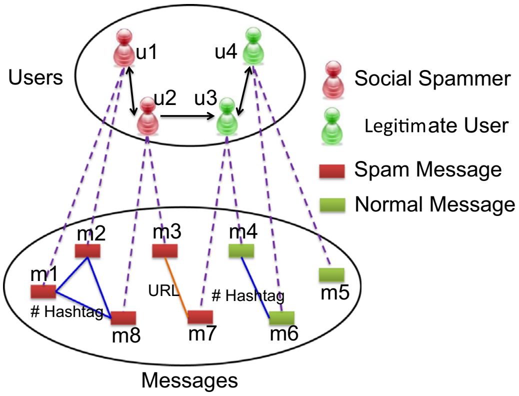


Fig. 1. An illustrative example of the social contexts used in our approach. The red and green user figures represent social spammers and legitimate users respectively. The red and green rectangles stand for spam messages and normal messages respectively. The solid black lines with arrows connecting users represent the relations between users. The dashed purple line connecting a message with a user represents the posting relation between the user and the message. The solid blue lines and orange lines represent the connections between messages introduced by hashtags and URLs respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

The main contributions of this paper are summarized as follows:

- We propose a unified approach to co-detect social spammers and spam messages via exploiting the social contexts of microblogging users and messages.
- We propose to extract three kinds of social contexts, i.e., the user-message relations, user–user relations and message–message relations, for social spammer and spam message co-detection in microblogging.
- We introduce an efficient optimization algorithm based on ADMM [21] to solve the model of our approach, and propose an accelerated method based on FISTA [22] to tackle the most time-consuming step.
- We evaluate our social spammer and spam message co-detection approach (*S3MCD*) via extensive experiments on a real-world microblog dataset. The experimental results validate the effectiveness and efficiency of our approach.

This paper is an extended and improved version of our previous work [23]. In comparison, in this paper we introduce an accelerated method to solve the most time-consuming step in the optimization algorithm for our approach and improve its efficiency significantly. In addition, we add a section in this paper to analyze the time complexity of our approach. Besides, we propose an alternative model for our approach based on label propagation, and compare it with our previous model using experiments to explore the effectiveness of different kinds of models for our approach, We also explore the influence of the type and quality of classifiers on the performance of our approach in this paper. In addition, we validate the effectiveness of our accelerated algorithm empirically.

The rest of this paper is organized as follows. In Section 2, we briefly introduce several related works on social spammer detection and spam text detection. In Section 3, we discuss how to extract various kinds of social contexts for detecting social spammers and spam messages in microblogging. In Section 4, we describe our social spammer and spam message co-detection approach, and the optimization algorithm to solve the model of our approach. In Section 5, we report the experimental results on a real-world microblog dataset. In Section 6 we conclude this paper.

2. Related work

In this section, we briefly introduce several related works on social spammer detection and spam detection.

2.1. *Social spammer detection*

Social spammer detection is a hot research topic and has been studied in various social networking websites, such as Twitter [12,24,25], Facebook [26], Youtube [27] and Sina Weibo [14,28]. Existing studies on social spammer detection can be roughly divided into two categories. The first category is based on social network analysis [9,10,29]. The assumption behind these methods is that social spammers cannot build a large number of social relations with legitimate users [1]. However, due to the special characteristics of microblogging websites, this assumption may