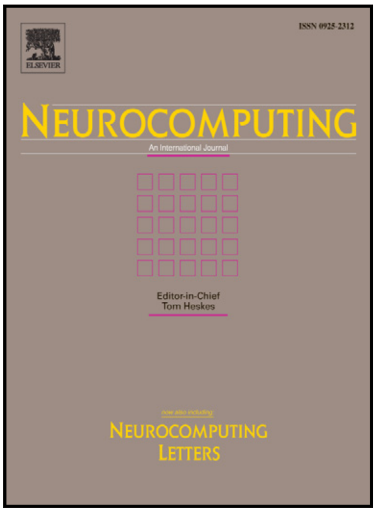
Author's Accepted Manuscript



www.elsevier.com/locate/neucom

Detecting spammers on social Networks

Xianghan Zheng, Zhipeng Zeng, Zheyi Chen, Yuanlong Yu, Chunming Rong

PII: S0925-2312(15)00210-6
DOI: http://dx.doi.org/10.1016/j.neucom.2015.02.047
Reference: NEUCOM15177

To appear in: *Neurocomputing*

Received date: 10 September 2014
Revised date: 26 November 2014
Accepted date: 8 February 2015

Cite this article as: Xianghan Zheng, Zhipeng Zeng, Zheyi Chen, Yuanlong Yu, Chunming Rong, Detecting spammers on social Networks, *Neurocomputing,* http://dx.doi.org/10.1016/j.neucom.2015.02.047

Detecting Spammers on Social Networks

Xianghan Zheng[1,2], Zhipeng Zeng[1,2], Zheyi Chen[3], Yuanlong Yu[1,2]*, Chunming Rong[4]
[1]College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China
[2]Fujian Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou, China
[3]Department of Computer Science, QingHua University, Beijing, China
[4]Department of Computer Science and Electronic Engineering, University of Stavanger, Stavanger, Norway

ABSTRACT

Social network has become a very popular way for internet users to communicate and interact online. Users spend plenty of time on famous social networks (e.g., Facebook, Twitter, Sina Weibo, etc.), reading news, discussing events and posting messages. Unfortunately, this popularity also attracts a significant amount of spammers who continuously expose malicious behavior (e.g., post messages containing commercial URLs, following a larger amount of users, etc.), leading to great misunderstanding and inconvenience on users' social activities. In this paper, a supervised machine learning based solution is proposed for an effective spammer detection. The main procedure of the work is: first, collect a dataset from Sina Weibo including 30,116 users and more than 16 million messages. Then, construct a labeled dataset of users and manually classify users into spammers and non-spammers. Afterwards, extract a set of feature from message content and users' social behavior, and apply into SVM (Support Vector Machines) based spammer detection algorithm. The experiment shows that the proposed solution is capable to provide excellent performance with true positive rate of spammers and non-spammers reaching 99.1% and 99.9% respectively.

Keywords: social network, spammer, machine learning, support vector machine

1. INTRODUCTION

Within the past few years, online social network, such as Facebook, Twitter, Weibo, etc., has become one of the major way for internet users to keep communications with their friends [1-3]. According to Statista report [4], the number of social network users has reached 1.61 billion until late 2013, and is estimated to be around 2.33 billion users globe, until the end of 2017.

However, along with great technical and commercial success, social network platform also provides a large amount of opportunities for broadcasting spammers, which spreads malicious messages and behavior. According to Nexgate's report [5], during the first half of 2013, the growth of social spam has been 355%, much faster than the growth rate of accounts and messages on most branded social networks.

The impact of social spam is already significant. A social spam message is potentially seen by all the followers and recipients' friends. Even worse, it might cause misdirection and misunderstanding in public and trending topic discussions. For example, trending topics are always abused by spammers to publish comments with URLs, misdirecting all kinds of users to completely unrelated websites. Because most social networks provide shorten service on URLs inside message, it is difficult to identify the content without visiting the site.

There has been a few proposals from industry and academia, discussing possible solutions for spam detection and filtering (described in Section 2). However, they are either ineffective or based on too much considered conditions (e.g., a lot of content and behavior feature, etc.). This paper investigates social spammer content and behavior issues, and proposes an effective machine learning model for spammer detection. The paper contains the following four main contributions:

- The paper adopts the spammer feature to detect spammer and test the results over Sina Weibo, the biggest social network site in China. Under the Weibo API, a specific dataset crawler is developed to extract any unauthorized users' public messages inside the Weibo platform. This is the first step for data analysis.
- The major novelty of the paper is to study a set of most important features related to message content and user behavior and apply them on the SVM based classification algorithm for spammer detection. The experiment and comparison work shows that the proposed solution enables to provide higher accuracy.
- Through feature selection algorithms and experiment testing, ten most important feature and the weight of these feature are identified. The experiment results further validate the selected spammer feature (manually classified) and also explain why the proposed solution could achieve excellent performance.
- The paper also develops a prototype software that is capable to distinguish any Weibo user (spammer or non-spammer). With friendly user interface, efficient and accurate classification result, ordinary users are capable to distinguish any Weibo users with simple operation. The software has been published in Sourceforge [6].

It should be mentioned that although the proposed approach is currently tested specifically in the Sina Weibo social network, it is applicable to all other existing social sites (e.g., Twitter, Facebook, etc.) with few revisions. The rest of the paper is organized as follows. Section 2 presents the background of the Weibo social network and displays some related works about spammer detection. Section 3 introduces the method how we collect the dataset and extract feature. Section 4 describes the spammer detection model, experiments and corresponding evaluation. Finally, the conclusion and future works are given in Section 5.