Chapter 73
Detecting Pharmaceutical Spam in Microblog Messages

Kathy J. Liszka
University of Akron, USA

Chien-Chung Chan
University of Akron, USA

Chandra Shekar
University of Akron, USA

ABSTRACT

Microblogs are one of a growing group of social network tools. Twitter is, at present, one of the most popular forums for microblogging in online social networks, and the fastest growing. Fifty million messages flow through servers, computers, and cell phones on a wide variety of topics exchanged daily. With this considerable volume, Twitter is a natural and obvious target for spreading spam via the messages, called tweets. The challenge is how to determine if a tweet is a spam or not, and more specifically a special category advertising pharmaceutical products. The authors look at the essential characteristics of spam tweets and what makes microblogging spam unique from email or other types of spam. They review methods and tools currently available to identify general spam tweets. Finally, this work introduces a new methodology of applying text mining and data mining techniques to generate classifiers that can be used for pharmaceutical spam detection in the context of microblogging.

INTRODUCTION

Social networking, in its many forms, is overtaking email in popularity for communication. Facebook[1] and Twitter[2] are leading the race, but many other notable sites exist and are very popular. Twitter, identi.ca[3] and bentio.com[4] are known as open microblogging services. They operate under free software licenses, providing a public application interface (API) for accessing and mining the messages posted by registered users. The parallel to text messaging is short message content but that's where the similarity ends. In general, the audience and subject content are of a totally different nature. People subscribe to sites like Twitter so they can post about topics such as "I'm waiting for a flight and I'm bored" to "the new Android minifigs rock."

DOI: 10.4018/978-1-4666-2455-9.ch073

Given the public, open-nature of sites like Twitter, it's not surprising that spammers have found new, fertile ground for distributing their unsolicited messages, catching users unaware. Twitter lets users post messages, called tweets. These can be saved in their personal profile and forwarded to others in their circle of friends. The information may be kept private among the list, or by default, remain public and unrestricted. Spammers use a variety of techniques to exploit Twitter for nefarious purposes.

In this chapter, we discuss how Twitter is used, the open API, and present some statistics about tweets in general. From there, we focus on a specific type of spam, those that are classified as pharmaceutical spam. It turns out that while it accounts for a significant and increasing percentage of traditional email spam, it only accounts for a small percentage of Twitter spam. Nonetheless, there are many dangers associated with these illegal scams and feel that identifying tweets engaging in this type of spam should be taken seriously and addressed in these early stages. Our contribution is a new classification scheme specifically targeting pharmaceutical spam that appears specifically in microblog messages. The decision strategy and data mining techniques used in this work take into account the unbalanced nature of the data set. A broader decision set is used, classifying a post as strongly identified as pharmaceutical spam, yes, maybe, and no.

BACKGROUND

When email became a popular form of communication, junk mail flowing through the United States Postal Service morphed into spam messages flowing through Internet Service Providers. It is well known that spam is more than an annoyance dealt with over morning coffee. It consumes massive amounts of bandwidth, spreads malware, entices users to phishing sites, and offers products for sale that are either illegal or fake. We start by discussing Twitter as a medium for this undesirable activity and specific aspects that spammers use to mount successful campaigns. Then we focus on the pharmaceutical industry and the difference between traditional spam techniques and adaptations for intruding into the microblog world.

The Nature of Spam

Unwanted solicitations, better known as spam, are not new. Fighting spam is a never ending battle as spammers are relentless. Original anti-spam methods included keyword analysis, honeypots, and black lists (mxLogic, 2004). Honeypots are a reactive approach analyzing spam email and blocking identical messages for clients. Alone, this is not good enough to combat spam, but it does increase effectiveness when combined with other approaches. Challenge-response is a variation of white listing where a new sender not on a recipient's personal white list must answer a question that presumably only a human could answer. This is meant to deter spammers who would not find it feasible to do this manually. The way around this today is via botnets that infiltrate machines and hijack accounts that are already white listed. Email header analysis is another technique for identifying spam emails by checking the IP address against the domain in an attempt to identify IP spoofing attempts. Other analysis includes tracking the amount of email sent from an IP address. Bayesian statistical filtering, pioneered by (Sahimi et al., 1998) is still probably the most effective technique in the email spam fighting toolbox. This is still a popular and widely implemented tool.

On the horizon of the social networking revolution, spammers lay in wait, devising new ways to score a hit on unsuspecting victims. Their intent is malicious including monetary theft, identity theft, fake merchandise and malware propagation. (Stringhini et al., 2010) set up a honeypot to study the new landscape of spam in social networks. What they did not find is a clear pattern of behavior. Different bot campaigns operate completely