



7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India

HashMiner: Feature Characterisation and analysis of #Hashtag Hijacking using real-time neural network

Deepali Virmani^a, Nikita Jain^a, Ketan Parikh^{a,*}, Abhishek Srivastava^{a,1}

^aDepartment of Computer Science, Bhagwan Parshuram Institute of Technology, New Delhi, India

Abstract

Online social media has become a vital platform to discuss common topics which are being categorised under a single name: Hashtag where people put their views, opinions and data. Thus hashtags have become a victim for spam, fake and un-related advertising content dissemination. In this paper we propose a novel approach designed on 9 distinctive parameters which extends to 4 other derived statistic from Twitter Streaming API, to detect Hashtag hijacking using Neural network analysis which shows a mean hijacking percentage of 28.5 over 10, 240 test tweets collected whereas, manual based annotation performed results in 17.14 %hijacking. Our method over collected dataset results in 94.025% accuracy.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications.

Keywords: Hijacking; Information; Hashtag; Neural network; Feature; Activation feature.

1. Introduction

Misusing various hashtags for personal and unwanted information distribution has become a common practice on social media platform. Such mal-practices are being are being conducted by advertisers, certain attention seekers or marketers. The methods adopted are discussed as follows

- Advertisers are using certain popular hashtags trending in demand on social media platform to popularize their own brands or websites. Such method of popularizing their content assures them a fast and high publicity for their brands or content as there is a high probability for a user following a particular hashtag of clicking any content added along [1]

* Corresponding author. Tel.: +91-8826682954;

E-mail address: ketanparikh1211@gmail.com

- Certain users are also misusing trending hashtags to distribute fake, unwanted, spam and unrelated content. These users are clearly exploiting the popularity of that trend to spread his/her data online to gain maximum views or popularity.[2].
- Several hashtags are also being used for political hijacks as discussed by authors in [3] where the exploitation. These methods of hashtag have also been categorized as hijacking or misuse by reports published in [4]. As the above mentioned hijacking hacks can neither be generalized nor are scalable, we hereby propose a novel approach to detect hashtag hijacking on live tweets based on eight diverse features formulated for every incoming tweet used to tag it as hijacked or non hijacked.
- Every trending hashtag has different story to discuss, it's nearly infeasible to know these stories and can only be studied and then defined. This detection analysis explores a vast and in depth features tagged with a hash-tagged tweet including the user account information mined using Twitter Streaming API, associated with that tweet. The methodology adopted has been present in fig 2. The major contributions of our paper is as follows :
- Data Characterisation: The data was collected using streaming API for 10 trending hashtags over a period of 4 months from September to December 2017. This data accounts to 10,240 tweets in total
- Feature Identification: Eight distinguished features derived into 4 more features have been formulated for every hash-tagged tweet which includes the user account information associated with the tweet.
- Classifier: We propose a novel approach on the above categorised features using neural network analysis and thus tag the hijacked tweets along with the identification of most correlated and distinguishing feature among the above mentioned.

2. Literature Review

The requirement of the hashtag came to light when it was realized to handle unordered data. Hashtags classify the content of the tweet as related to a certain topic. The hashtags mentioned in the tweet relate the tweet to certain topic and facilitate the searching of tweet.

2.1 The Impact of Hashtag

Hashtags are a powerful source when it comes to information flow. As per [5], social movements have long used media and technology to disseminate, escalate, and enlarge the scope of their struggles. Hashtags allowed a message to get out, called global attention to a small corner of the world, and attempted to bring visibility and accountability to repressive forces. The study by [6] concluded that the Twitter hashtags depend on the usage patterns of their network neighbours. We have found that hashtags of different types and topics spread differently. These differences can be analyzed in terms of the probabilities that users utilize a hashtag after repeated exposure to it, with variations occurring not just in the absolute magnitudes of these probabilities but also in their rate of decay.

2.2 Twitter Hashtags as a platform for advertisement and Ad hoc Publics

Although Twitter is a platform for information dissemination and facts propagation, this platform is exploited by companies for advertisement of their products and schemes. The advertising agencies exploit the excitement of the people about the products by promoting certain hashtags as much as possible. As per [7] Hashtags enable "ambient affiliation," which means users can engage in a conversation about a celebrity, or election using hashtags without having to follow the other users involved.

2.3 Semantic sentiment analysis of twitter data using Hashtags

The tweet can be classified as to convey a sentiment to the reader whoever reads it therefore there is an attached emotion with a tweet. Sentiment extraction systems usually require an extensive dataset consisting of words associated with a certain sentiment. With the recent popularity of article tagging, some social media types like blogs allow users to add sentiment tags to articles. This allows using blogs as a large user-labelled dataset for sentiment learning and identification.

2.4 Hashtag spamming and hijacking

This section highlights the potential misuse of the hashtags, its hijacking and spamming. [8] broadly identifies fake hashtags as a tweet that follows the trend or is synchronous with other similar tweets of a particular hashtag can be thus considered as non hijacked tweet. The idea behind tweeting such text is misusing the hashtag for acquiring the user attention connected to that hashtag. [9] Classified tweets as spam and ham tweets by analyzing two aspects, namely, (i) the usage of uppercase letters, numbers and (iii) the position of hashtags in tweets.

Download English Version:

<https://daneshyari.com/en/article/6902440>

Download Persian Version:

<https://daneshyari.com/article/6902440>

[Daneshyari.com](https://daneshyari.com)