A SURVEY ON THE ANALYSIS OF EMOTIONS IN SPAMS IN SOCIAL NETWORK

Meira Yendrembam
Computer Science and Engineering
New Horizon College of Engineering
Bangalore, India

Deepika N
Computer Science and Engineering
New Horizon College of Engineering
Bangalore, India

Abstract: Social media, such as Twitter, is a very popular way of expressing opinions and interacting with other people in the online world. When taken in aggregation, tweets can provide a reflection of public sentiment towards events. Also, Twitter spam has become a critical problem worldwide. In this paper, we provide a positive or negative sentiment on Twitter posts using a well-known machine learning method for text categorization. In addition, we use manually labelled (positive/negative) tweets to build a trained method to accomplish a task. The trained model is based on the Naive Bayes classification method. Our aim is to detect the spam tweets and the sentiment of the tweets.

Keywords: Social Network, Sentiment, Spam, Machine Learning, Text Categorization, Naive Bayes Classification Method

I. INTRODUCTION

Social media, such as Twitter, is a very popular way of expressing opinions and interacting with other people in the online world. When taken in aggregation, tweets can provide a reflection of public sentiment towards events. Also, Twitter spam has become a critical problem. In this paper, we provide a positive or negative sentiment on Twitter posts using a well-known machine learning method for text categorization. In addition, we use manually labelled (positive/negative) tweets to build a trained method to accomplish a task. The task is looking for a correlation between twitter sentiment and events that have occurred. The trained model is based on the Naive Bayes classification method. Our aim is to detect the spam tweets and the sentiment of the tweets.

Ideas and information are spreading in social networks like a pathogen, each infected person effects friends around him/her. Moreover, Emotions such as happiness and depression have been studied and found to be contagious both by media and by personal contact with strangers or friends [1]. “Arab Spring” and “Occupy movement” have recently showed the powerful influence of social media [2]. Negative messages spreads within a short period of time. Because of these messages, there is a huge impact on the users and economics. Twitter is one of the most common online social media and micro-blogging services. It is used for expressing opinions and interacting with other people. Twitter messages (or Tweets) are of short texts that express opinions, ideas and events captured in the moment. Tweets are well-suited sources of streaming data for opinion mining and sentiment polarity detection. Opinions, evaluations, emotions and speculations often reflect the states of individuals. There also has been an increase of spamming activities. Twitter spam contains malicious link that directs victims to external sites containing malware downloads, phishing, drug sales, or scams, *etc.* [3]. This not only interferes user experiences but also damages the whole Internet. Twitter and other Security companies are trying to make Twitter a spam-free platform. For example, Trend Micro uses a blacklisting service called Web Reputation Technology system to filter spam URLs for users who have its products installed [4].

The main motive of any spammer is to lure the legitimate user towards a malicious spam. A number of Machine Learning based detection schemes have been proposed by researchers [3], [5], [6], and [7].

According to Twitter Policy, there are various tactics that are considered as Spam:

- Post harmful and malicious links
- Abusive replies to various users
- Posting duplicate and unrelated data for tweets
- Posting about current topics for seeking attention

II. LITERATURE SURVEY

Several analysis have been carried out in order to detect the spam and sentiments on online social networks, like Twitter. Some of these are given below.

Reference [3] dealt with the problem of detecting spammers on Twitter. Based on the dataset and using manual inspection, a labelled collection have been created with users classified as spammers or non-spammers. Using a classification technique, it was able to correctly identify a significant fraction of the spammers while incurring in a negligible fraction of misclassification of legitimate users.

Advantages:

- Increase and improve the labelled collection in a collaborative manner.
- Aim at investigating other kinds of attacks on Twitter.
- Correctly identify a significant fraction of the spammers.
- Detect spammers with high accuracy.
- Also investigates the feasibility of detecting spam instead of spammers.

Disadvantages:

- Problem of detecting spammers on Twitter.
- Pollute real time search.
- Interfere on statistics presented by tweet mining tools and consume extra resources from users and systems.

Reference [8] shows that for messages about time-sensitive topics, automatically newsworthy topics can be separated from other types of conversations.

Advantages:

- Assess automatically the level of social media credibility of newsworthy topics.
- Facilitates real-time propagation of information to a large group of users.
- Enable users to assess information credibility.

Disadvantages:

- Easily mislead by unreliable information.
- Lack the clues that assess the credibility of the information to which they are exposed.
- Critical to provide tools to validate the credibility of online information.

Reference [9] demonstrated the potential of using linguistic features as a means of classifying automated activity on Twitter by using a flexible and transparent classification scheme.

Advantages:

- Flexible and transparent.
- More deeply analysed to identify organic versus SPAM related hyperlinks.

Disadvantage:

- Lack of interoperability.
- Lack of standardization and use of data dictionaries which results in a lack of precision concerning our ability to collate signs, symptoms, and diagnoses.

Reference [10] presented a novel approach to detect compromised accounts in social networks by using COMPA, which characterizes the behavior of social network users, and anomaly detection techniques to identify sudden changes in their behavior.

Advantages:

- Detect compromised accounts in social networks.
- Use of anomaly detection techniques to identify sudden changes in their behaviour.
- Effectively detect compromised accounts with very low false positives.

Disadvantages:

- Do not have full visibility of every message exchanged on Facebook and Twitter.

Reference [11] provided online spam filtering for social networks by using text shingling and URL comparison to incrementally reconstruct spam messages into campaigns, which are then identified by a trained classifier.

Advantages:

- Provide online spam filtering for social networks.
- Achieves high accuracy, low latency and high thorough output.
- System is able to remain accurate for more than 9 months.
- Also shows very low maintenance cost after deployment.

Disadvantages:
- Identifying spamming accounts is not sufficient to fight OSN spam.
- Long lag-time or limited efficiency.

Reference [12] finds that a relatively simple sentiment detector based on Twitter data replicates consumer confidence and presidential job approval polls.

Advantages:
- Simple sentiment detector.

Disadvantages:
- Is a noisy measurement instrument.

Reference [13] presented a tool, Taratweet, to analyse conversations in Twitter for predicting elections results.

Advantages:
- Is used for political discussion, and that the references to the different political parties correlate, significantly, with the votes of the electors.
- Manage the tweets between different political parties.

Disadvantages:
- Managing the live political event is little bit challenging.

Reference [14] analyzed a large social network in a new form of social media known as microblogging to determine an individual user's intention.

Advantages:
- Is used to describe the current status in short posts distributed by instant messages, mobile phones, email or the Web.
- Instead of spending hours together to narrate a blog post, micro blogging platform can be used effectively to share the things we wish to do.

Disadvantages:
- The character is limited to just 140.
- Also prone to getting off topic quite easily.

Reference [15] provided an efficient way to observe public opinion on temporal dimension. The method can identify break points, and find related events that caused these opinion changes.

Advantages:
- Efficient.
- Provides accuracy.

Disadvantages:
- Tracking is not possible.
- Limited amount of data.

Reference [16] provided an introduction to the WEKA workbench, reviews the history of the project, and, in light of the recent 3.6 stable release, briefly discusses what has been added since the last stable version (Weka 3.4) released in 2003.

Advantages:

- Provide a comprehensive collection of machine learning algorithms and data pre-processing tools to researchers and practitioners alike.
- It allows users to quickly try out and compare different machine learning methods on new data sets.

Disadvantages:

- Lack of proper and adequate documentations.

III. CONCLUSION

It has been observed that different types of analysis have been carried out in order to detect spams and sentiments on Online Social Network (OSN). But there are still few drawbacks in the analysis. Relating machine learning skills for detecting spams and sentiments on OSN, like Twitter, can repeatedly shape the model based on the training data set, which holds data instances that can be labelled by means of a usual set of attributes and associated labels. We will try to detect spams based on how many times a user tweets in 1 or 2 minutes, whether 2 or more users tweet the same sentence or URL at a particular time and whether a user(s) promote a blacklisted URL in his/her tweet.

IV. REFERENCES

[1] Emotions as infectious diseases in a large social network: the SISa model, Alison L. Hill, David G. Rand, Martin A. Nowak and Nicholas A. Christakis, Proc Biol Sci. 2010 Dec 22; 277(1701): 3827–3835. Published online 2010 Jul 7. doi: 10.1098/rspb.2010.1217

[2] Bastos MT, Mercea D, Charpentier A. Tents, tweets, and events: The interplay between ongoing protests and social media. Journal of Communication. 2015;65:320–350

[3] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, “Detecting spammer on twitter,” in Proc. 7th Annu. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf., Jul. 2010, p. 12.

[4] J. Oliver, P. Pajares, C. Ke, C. Chen, and Y. Xiang, “An in-depth analysis of abuse on twitter,” Trend Micro, Irving, TX, USA, Tech. Rep., Sep. 2014.

[5] G. Stringhini, C. Kruegel, and G. Vigna, “Detecting spammers on social networks,” in Proc. 26th Annu. Comput. Security Appl. Conf., 2010, pp. 1–9.

[6] C. Yang, R. Harkreader, and G. Gu, “Empirical evaluation and new design for fighting evolving twitter spammers,” IEEE Trans. Inf. Forensics Security, vol. 8, no. 8, pp. 1280–1293, Aug. 2013.

[7] X. Zhang, S. Zhu, and W. Liang, “Detecting spam and promoting campaigns in the twitter social network,” in Proc. Data Mining IEEE 12th Int. Conf. (ICDM), 2012, pp. 1194–1199.

[8] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on twitter,” in Proc. 20th Int. Conf. World Wide Web, 2011, pp. 675–684.

[9] E. M. Clark, J. R. Williams, C. A. Jones, R. A. Galbraith, C. M. Danforth, and P. S. Dodds, “Shifting robotic from organic text: A natural language approach for detecting automation on twitter,” J. Comput. Sci., vol. 16, p. 1–7, Sep. 2016.

[10] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, “Compa: Detecting compromised accounts on social networks,” in Proc. Annu. Netw. Distrib. Syst. Security Symp., 2013.

[11] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary, “Towards Online spam filtering in social networks,” in Proc. NDSS, 2012, pp. 1–16.

[12] B. O'Connor, R. Balasubramanyan, B. Routledge, and N. Smith, “From tweets to polls: Linking text sentiment to public opinion time series”, in Proc. Int. AAAI Conf. Weblogs Social Media, May 2010, pp. 2633.

[13] J. M. Soler, F. Cuartero, and M. Roblizo, ``Twitter as a tool for predicting elections results," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2012, pp. 11941200.

[14] A. Java, X. Song, T. Finin, and B. Tseng, ``Why we Twitter: Understanding microblogging usage and communities," in Proc. 9th WebKDD 1st SNA- KDD Workshop Web Mining Social Netw. Anal., Aug. 2007, pp. 5665.

[15] C. G. Akcora, M. A. Bayir, M. Demirbas, and H. Ferhatosmanoglu, ``Identifying breakpoints in public opinion," in Proc. 1st Workshop Social Media Anal., Jul. 2010, pp. 6266.

[16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, ``The WEKA data mining software: An update," SIGKDD Explorations Newslett., vol. 11, no. 1, pp. 1018, Jun. 2009.