ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 3 , March 2017

Spam Detection in Social Networks Using Efficient Classifier

Vishalakshi N S, S.Sridevi,Asha Rani Borah

P.G. Student, Department of Computer Science, New Horizon College Of Engineering, Bangalore, India
Assistant Professor, Department of Computer Science, New Horizon College Of Engineering, Bangalore,
Sr.Assistant Professor, Department of Computer Science, New Horizon College Of Engineering, Bangalore, India

ABSTRACT: Social Networks have become more and more popular in the whole world. People share their personal activities, views and opinions among different SNs. At the same time, social spam appears more frequently and in various formats throughout popular SNs. Therefore, efficient detection of spam has become an important and popular problem. The major problem, users face spammer's interaction which leads to misunderstanding and inconvenience for social activities. This work concentrates on detecting the spammer actions using feature relevance analysis and applying efficient classifier. The main objective of the proposed work is to find relationship between features and classifying patterns for detecting spam message from the unwanted sites. The system applies efficient classification algorithms after feature relevance analysis for detecting spam in better way. The outcome of this project will serve for the users participating in social networks for effective purpose like business, marketing and establishing contacts. In this paper we have reviewed the existing techniques for detecting spam users in Twitter social network. Features for the detection of spammers could be user based or content based or both and spam classifier methods.

I. INTRODUCTION

In recent years, internet has become an integral part of our life. With increased use of internet, numbers of email and social media (Facebook, Twitter, Linkln, and Google+) users are increasing day by day. In which user can connect to other users and post messages to each other and on other pages within in the network. Some social networking sites rely on their users not only to generate content, but also fight spam and other inappropriate content. The content of the spam messages include pharmaceuticals, jewellery, electronics, loans, stocks, weight loss, and gambling. Within the past few years, online social network, such as Face-book, Twitter, Weibo, etc., has become one of the major way for internet users to keep communications with their friends. According to Statist report [1], 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 behaviour.

A spam filter is a program that is used to detect unsolicited and unwanted email and prevent those messages from getting to a user's inbox. Like other types of filtering programs, a spam filter looks for certain criteria on which it bases judgments. For example, the simplest and earliest versions (such as the one available with Microsoft's Hotmail) can be set to watch for particular words in the subject line of messages and to exclude these from the user's inbox. This method is not especially effective, too often omitting perfectly legitimate messages (these are called *false positives*) and letting actual spam through. More sophisticated programs, such as Bayesian filters or other heuristic filters, attempt to identify spam through suspicious word patterns or word frequency.

Non-machine learning techniques, such as heuristics, blacklisting and signatures, have been complemented in recent years with new, ML-based technologies. In the last few years, substantial academic research has taken place to evaluate new ML-based approaches to filtering spam. ML filtering techniques can be further categorized into complete and complementary solutions. Complementary solutions are designed to work as a component of a larger filtering system, offering support to the primary filter (whether it be ML or non-ML based). Complete solutions aim to construct a comprehensive knowledge base that allows them to classify all incoming messages independently

.

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

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 3 , March 2017

discussions. For example, trending topics are always abused by spammers to publish comments with URLs, misdirecting all kinds of users to completely unrelated web-sites. Because most social networks provide shorten service on URLs inside messages it is difficult to identify the content without visiting the site.

Types of Spammers:

1) Spammers: are the malicious users who contaminate the information presented by legitimate users and in turn pose a risk to the security and privacy of social networks. Spammers belong to one of the following categories
2) Phishers: are the users who behave like a normal user to acquire personal data of other genuine users.
3) Fake Users: are the users who impersonate the profiles of genuine users to send spam content to the friends' of that user or other users in the network.
4) Promoters: are the ones who send malicious links of advertisements or other promotional links to others so as to obtain their personal information.

II. INPUT DESIGN And OUTPUT DESIGN

A. INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

B.OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2.It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

C.OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1.Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 3 , March 2017

effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

III. IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods. It consists of different Modules that is shown below:

A. CONTENT NON SPAM MODELING

Content NON SPAM modelling is used to classify content (e.g., Web pages, images, and videos) as spam or legitimate. In this case, the target of NON SPAM is content (resource), and thus a NON SPAM score is given to each content based on its content and/or associated tags. Content NON SPAM models reduce the prominence of content likely to be spam, usually in query-based retrieval results. They try to provide better ordering of the results to reduce the exposure of the spam to users. Koutrika et al. [20] proposed that each incorrect content found in a system could be simply removed by an administrator. The administrator can go a step further and remove all content contributed by the user who posted the incorrect content, on the assumption that this user is a spammer (polluter)..

B. USER NON SPAM MODELING (static)

The aforementioned studies consider users' reliability as static at a specific moment. However, a user's NON SPAM in a social tagging system is dynamic, i.e., it changes over time. The tagging history of a user is better to consider, because a consistent good behaviour of a user in the past can suddenly change by a few mistakes, which consequently ruins his/her NON SPAM in tagging.

C. USER NON SPAM MODELING (Dynamic)

A dynamic NON SPAM score, called SocialNON SPAM, is derived for each user. It depends on the quality of the relationship with his/her neighbours in a social graph and personalized feedback ratings received from neighbours so that NON SPAM scores are updated as the social network evolves.

The dynamics of the system is modelled by including the evolution of the user's NON SPAM score to intent long-term good behaviour and to penalize users who build up a good NON SPAM rating and suddenly "defect." It was shown that SocialNON SPAM is resilient to the increase in number of malicious users, since the highly NON Spammed users manage to keep them under control thanks to the NON SPAMaware feedback scheme introduced in this approach. It was also shown that SocialNON SPAM outperforms NON SPAMRank-based models, because SocialNON SPAM model incorporates relationship quality and feedback ratings into the NON SPAM assessment so that bad behaviour is punished.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 3 , March 2017

D. DATA SET

Data sets used for development and evaluation of NON SPAM modelling techniques have a wide range of diversity in terms of content, numbers of resources, tags and users, and type of spam. Social bookmarking is the most popularly explored domain for NON SPAM modelling, especially user NON SPAM modelling.

E.ALGORITHM

NON SPAM modelling can be formulated as either a classification problem or a ranking problem, depending on the way of treatment. In the classification problem, the results of an algorithm can be summarized by a confusion matrix from ground-truth data and predicted labels, which contains the number of true positives, true negatives, false positives, and false negatives. From these values, classical measures such as a receiver operating characteristic (ROC), the area under the ROC curve (AUC), precision-recall (PR) curves, and F-measure can be derived.

IV. CONCLUSION

In this article, we dealt with one of the key issues in social send systems: combatting noise and spam. We classified existing studies in the literature into two categories, i.e., content and user trust modelling. Representative techniques in each category were analysed and compared. In addition, existing databases and evaluation protocols were re viewed. An example system was presented to demonstrate how trust modelling can be particularly employed in a popular application of image sharing and sending. Finally, open issues and future research trends were prospected. As online social networks and content sharing services evolve rapidly, we believe that the research on enhancing reliability and trustworthiness of such services will become increasingly important.

REFERENCES

[1] Wikimedia Foundation Inc. (2011, Dec.). Flickr.[Online]. Available: http://en.wikipedia.org/wiki/Flickr
[2] Pingdom Blog. (2011, Jan.).Internet 2010 in numbers.[Online]. Available:http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers
[3] C. Marlow, M. Naaman, D. Boyd, and M. Davis, "HT06, tagging paper, taxonomy,Flickr, academic article, to read," in *Proc. ACM HT*, Aug. 2006, pp. 31–40.
[4] K. Liu, B. Fang, and Y. Zhang, "Detecting tag spam in social tagging systemswith collaborative knowledge," in *Proc. IEEE FSKD*, Aug. 2009, pp. 427–431.
[5] L. S. Kennedy, S.-F. Chang, and I. V. Kozintsev, "To search or to label?: Predictingthe performance of search-based automatic image classifiers," in *Proc.ACM MIR*, Oct. 2006, pp. 249–258.
[6] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Fighting spam on social Websites: A survey of approaches and future challenges," *IEEE Internet Comput*., vol.11, no. 6, pp. 36–45, Nov. 2007.
[7] L. von Ahn, M. Blum, N. J. Hopper, and J. Langford, "CAPTCHA: Using hardAI problems for security," in *Proc. Eurocrypt*, May 2003, pp. 294–311.
[8] L. von Ahn, B. Maurer, C. Mcmillen, D. Abraham, and M. Blum, "reCAPTCHA: Human-based character recognition via Web security measures," *Science*, vol. 321,no. 5895, pp. 1465–1468, Aug. 2008.
[9] Yahoo, Inc. (2011, Dec.). Flickr—Tags. [Online]. Available: http://www.flickr.com/help/tags
[10] G. Mori and J. Malik, "Recognizing objects in adversarial clutter: Breaking avisual CAPTCHA," in *Proc. IEEE CVPR*, June 2003, pp. I-134–I-141.
[11] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian approachto filtering junk e-mail," AAAI Workshop on Learning for Text Categorization,Madison: WI, *Tech. Rep*. *WS-98-05*, July 1998.
[12] D. Fetterly, M. Manasse, and M. Najork, "Spam, damn spam, and statistics:Using statistical analysis to locate spam Web pages," in *Proc. ACM WebDB*, June2004, pp. 1–6.
[13] S. Marti and H. Garcia-Molina, "Taxonomy of trust: Categorizing P2P reputationsystems," *Comput. Netw*., vol. 50, no. 4, pp. 472–484, Mar. 2006.
[14] A. Thomason, "Blog spam: A review," in *Proc. CEAS*, Aug. 2007.
[15] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systemsfor online service provision," *Decision Support Syst*., vol. 43, no. 2, pp. 618–644,Mar. 2007.
[16] A. Whitby, A. Jøsang, and J. Indulska, "Filtering out unfair ratings in Bayesianreputation systems," in *Proc. IEEE AAMAS*, July 2004, pp. 106–117
[17] Y. Yang, Y. L. Sun, S. Kay, and Q. Yang, "Defending online reputation systems against collaborative unfair raters through signal modeling and trust," in *Proc. ACM SAC*, Mar. 2009, pp. 1308–1315.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 3 , March 2017

[18] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," in *Proc. ACM AIRWeb,* Apr. 2009, pp. 41–48.
[19] B. Krause, C . Schmitz, A. Hotho, and G. Stum, "The anti-social tagger: Detecting spam in social bookmarking systems," in *Proc. ACM AIRWeb*, Apr. 2008, pp. 61–68.
[20] G. Koutrika, F. A. Effendi, Z. Gyöngyi, P. Heymann, and H. Garcia-Molina, "Combating spam in tagging systems: An evaluation," *ACM TWEB*, vol. 2, no. 4, pp. 22:1–22:34, Oct. 2008.
[21] Z. Gyongyi, H. Garcia-Molina, and J. Pedersen, "Combating Web spam with TrustRank," in *Proc. VLDB*, Aug. 2004, pp. 576–587.
[22] C.-T. Wu, K.-T.Cheng, Q. Zhu, and Y.-L. Wu, "Using visual features for antispamfiltering," in *Proc. IEEE ICIP*, Sept. 2005, vol. 3, pp. 509–512.
[23] T. Bogers and A. Va n den Bosch, "Using language models for spam detection in social bookmarking," in *Proc. ECML PKDD*, Sept. 2008, pp. 1–12.
[24] I. Ivanov, P. Vajda , J.-S. Lee, L. Goldmann, and T. Ebrahimi, "Geotag propagation in social networks based on user trust model," *Multimedia Tools Applicat*., pp. 1–23, July 2010.
[25] Z. Xu, Y. Fu, J. Ma o, and D. Su, "Towards the semantic Web: Collaborative tag suggestions," in *Proc. ACM WWW*, May 2006, pp. 1–8.