SPAM DETECTION USING LINGUISTICS AND SPAMMER BEHAVIORAL

Dr. V. Geetha[1], Mr. N. Venkat Aniket Reddy[2], Mr. P. Anurag Kashyap[3]

[1]Assistant Professor, Dept. of CSE, SCSVMV University, Kanchipuram, TN, India
[2]Student, Dept. of CSE, SCSVMV University, Kanchipuram, TN, India
[3]Student, Dept. of CSE, SCSVMV University, Kanchipuram, TN, India

---***---

Abstract - *Digital review systems plays a vital role in affecting consumers' behaviors and decision making. Attracting many spammers to upload fake reviews to manipulate the review content and ratings. To develop the utility and improve the user experience. Some digital review systems allow one to form social relationships between each other and encourage their interactions. We try to provide an effective method or technique to detect review spammers by integrating social relations based on the assumptions that people will likely to consider. Reviews from the users who are likely to be genuine do not maintain huge relationship network with spammers. The contributions of this paper are:*

(1) We expand how social relationships can be implemented into review rating prediction and propose a trust-based rating prediction model using proximity as trust weight.

(2) We design a trust-aware detection model based on rating variance which calculates user-specific and overall trustworthiness scores as the indicator.

Experiments on the dataset collected from Kaggle show that the proposed trust-based prediction achieves a higher accuracy than standard CF method, and there exists a strong correlation.

***Key Words*: Machine Learning; Spam Detection; Scalability; Twitter**

1. INTRODUCTION

What Is a Social Network?

Internet defines it as a service which "focuses on verifying of online social networks for people who share the interests and activities, or who are interested in finding the interests and activities of others, and which requires the use of a software."

A report published by OCLC gives the following definition of social networking sites. "Websites are mainly designed to facilitate interaction between the users who share interests, and activities, such as "Facebook.""

What Can Social Networks Be Used For?

Social networks can give a wide variety of benefits to members of the organization.

Support for learning: Social networks can enhance the informal learning and support the social relations within group of learners and those involved in the support of learning.

Support for members: Social networks can potentially be used by all members of an organization, and not just those involved in working with students. Social networks can help the improvement of communities of practice.

Engaging with others: Passive use of social networks can provide valuable business intelligence and feedback on institutional services (although this may give rise to ethical concerns).

Ease of access to information and applications: The ease of use of many social services can provide benefits to users by easy access to other tools and applications. The company Facebook is an example of how a social networking service can be used as an environment for other tools.

Common interface: A benefit of social networks may be the common interface which spans work / social boundaries. Since such services are mostly used in a personal capacity, the interface and the way the service works may be familiar, thus minimizing training and support needed to exploit the services in a professional context. This can also be an obstacle to those who want to have strict boundaries between work and social activities.

1.1 Opportunities and Challenges

The popularity and ease of use of social networking services have excited institutions with their potential in a variety of areas. However effective use of social networking services poses a number of challenges for institutions including long-term sustainability of the services; user concerns over use of social tools in a work or study context; a variety of technical issues and legal issues such as copyright, privacy, accessibility; etc.

Institutions would be advised to consider carefully the implications before promoting significant use of such services.

1.2 Scope of the project:

Social media can also be used for spreading fake news and malicious links (like Ransomware and malware). So, we need

to be careful while using social media. Since social media is widely used now-a-days, these spams are increasing rapidly. In this project we are mainly focusing on spam detection in social media (specifically in Twitter) using machine learning techniques.

2. Implementation:

- We will be using Flask API, a web framework written in Python.

- Tweepy an open-source python package to access the Twitter API.

- Importing packages like Pandas, NumPy, CSV, matplotlib for graphical representation and importing algorithms from Sklearn as we are using python Library.

- Fetching the dataset and training a model to detect fake users by reading the CSV file.

- Extracting the features.

- Checking and display the user info from the dataset.

- Building and evaluating the results.

- Displaying whether the user is fake or legitimate with the help of matplotlib.

3. Result:

- Algorithm's accuracy depends on the type and size of your dataset.

- More the data, more chances of getting correct accuracy.

- Machine learning depends on the variations and relations

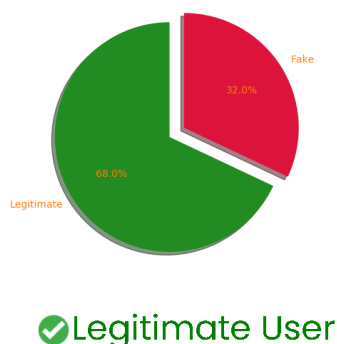- Understanding what is predictable is as important as trying to predict it.

Figure 1

4. CONCLUSIONS

- We performed a review of techniques used for detecting fake users on Twitter.
- We also presented a typology of Twitter spam detection approaches and sorted them as fake content detection, URL based spam detection, spam detection in trending topics, and fake user detection techniques.
- We also related the presented techniques based on several features, such as user features, content features, graph features, structure features, and time features.
- The techniques were also compared by the goals and datasets used.
- It is certain that the presented paper will be useful for researchers to find the information on Twitter spam detection techniques.
- Despite the development of efficient and effective approaches for the spam detection and fake user identification. Twitter, there are still certain open areas that require considerable attention by the researchers.
- The issues are briefly highlighted as under: False news identification on social media networks is an issue that needs to be explored because of the serious repercussions of such news at individual as well as collective level.
- Another associated topic that is worth investigating is the identification of rumor sources on social media.
- Although a few studies based on statistical methods have already been conducted to detect the sources of rumors, more sophisticated approaches, e.g., social network-based approaches, can be applied because of their proven effectiveness.

REFERENCES

[1] Guanjan Lin, Junyang Qui & Paul Rimba, "Near real-time twitter spam detection with machine learning techniques", International Journal of computers and applications, 16 April 2020

[2] I. David, O. S. Siordia, and D. Moctezuma, "Features combination for the detection of malicious Twitter accounts," in Proc. IEEE Int. Autumn Meeting Power, Electron. Comput. (ROPEC), Nov. 2016, pp. 1–6.

[3] M. Babcock, R. A. V. Cox, and S. Kumar, "Diffusion of pro- and anti-false information tweets: The Black Panther movie case," Comput. Math. Org. Theory, vol. 25, no. 1, pp. 72–84, Mar. 2019.

[4] S. Keretna, A. Hossny, and D. Creighton, "Recognizing user identity in Twitter social networks via text mining," in

Proc. IEEE Int. Conf. Syst., Man, Cybern. Oct. 2013, pp. 3079–3082.

BIOGRAPHIES

1. Dr .V. Geetha is an assistant professor in Computer science and engineering at Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya deemed to be a university.

2. Mr. N. Venkat Aniket Reddy, Student, B.E Computer science and engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya deemed to be a university.

3. Mr. P. Anurag Kashyap, Student, B.E Computer science and engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya deemed to be a university.