

Detection of Malicious Social Bots Using Instagram hashtags

V. Pradeep¹, V. Vaidehi²

1. PG Student, 2. Professor

Department of Computer Application

Dr. MGR Educational and Research institute, Chennai-600095

Date of Submission: 25-03-2024

Date of Acceptance: 05-04-2024

ABSTRACT

In this paper, we have presented a new 2FA (including both a user secret key and a lightweight security device) access control system for web-based cloud computing services. Based on the attribute-based access control mechanism, the proposed 2FA access control system has been identified to not only enable the cloud server to restrict access to those users with the same set of attributes but also preserve user privacy. A detailed security analysis shows that the proposed 2FA access control system achieves the desired security requirements. Through performance evaluation, we demonstrated that the construction is "feasible." We leave it as future work to further improve the efficiency while keeping all the nice features of the system. Online social networks have attempted to design new ways of collecting and analyzing such big data. For example, social bots have been used to perform automated analytical services and provide users with improved quality of service. However, malicious Social bots have also been used to disseminate false information (e.g., fake news), and this can result in real-world consequences. Therefore, detecting and removing malicious social bots from online social networks is crucial. Most existing detection methods for malicious social bots analyze the quantitative features of their behavior. These features are easily imitated by social bots, thereby resulting in low accuracy of analysis. A novel method of detecting malicious social bots, including both feature selections based on the transition probability of clickstream sequences and semi-supervised clustering, is presented in this paper. This method not only analyzes the transition probability of user behavior clickstreams but also considers the time feature of behavior. Findings from our experiments on a real online social network Platforms demonstrate detection accuracy for different types of malicious social bots by using the detection method of malicious social bots based on the

transition probability of user behavior. Clickstreams increase by an average of 12.8% in comparison to the detection method-based on quantitative analysis of user behavior.

KEYWORDS: Messenger, SVM, JDBC, Text Mining.

I. INTRODUCTION

Social networking is going to have a significant impact on people's lives today. Social media platforms such as Facebook, Google+, Twitter, and others have millions of global users. Social media now makes up a significant part of the online environment. It has altered how people interact with one another and live. The primary purpose of online social networks (OSNs) is to facilitate the sharing and exchange of personal information. OSN is primarily used for the sharing of various kinds of content, such as text, links, images, audio, and video files. An online social network is a location where users meet new people and form social bonds through text, pictures, and real-time communication with those who share interests.[1].

A social network service manages profiles of each user, his interests, social links, and a wide range of additional services like finding new people with the same interests, hobbies, and locations. OSN is a web-based service that allows individual users to create their profile and a list of users, such as friends, family, or colleagues, with whom to connect and whom to allow connection within the system. [2].

Information filtering using web content mining can be used for many other purposes, as per the requirements in OSN. This is because there is the possibility of posting a message or commenting on other posts in the user's message area called Wall. Unwanted messages are filtered using information filtering, which is mainly used to give

users the ability to control the messages written on their own walls. [3].

The Internet is the basis of modern life as we know it. People today find it impossible to envision life without the Internet. People have been using social networking sites to exchange opinions, ideas, and information with one another for the past few years. These exchanges may involve a variety of content types, including text, images, audio, and video. The typical Instagram user generates 90 pieces of content each month, according to #tag statistics, while over 30 billion pieces of material (weblinks, news articles, blog entries, notes, photo albums, etc.) are shared monthly.[4].

Information filtering has been widely used and employed for textual documents and web contents. However, the goal of this proposal is mainly to provide categorization techniques to protect user walls from useless and meaningless data. This is especially true for OSNs; users can comment on posts in public or private areas of other user walls. These comments can be useless, meaningless, or unwanted. So, here, information filtering plays a vital role in protecting the user walls in OSNs from undesired messages and giving the user the authority to automatically control the undesired data on their walls. [5].

II. LITERATURE SURVEY

L. Fang (2010), et al. says ,Privacy is an enormous problem in online social networking sites. While sites such as Facebook allow users fine-grained control over who can see their profiles, it is difficult for average users to specify this kind of detailed policy. The intuition for the design comes from the observation that real users conceive their privacy preferences (which friends should be able to see which information) based on an implicit set of rules. Thus, with a limited amount of user input, it is usually possible to build a machine learning model that concisely describes a particular user's preferences, and then use this model to configure the user's privacy settings automatically.[6].

K.Strater (2007), et al. says that the popularity of social networking websites such as Facebook and the subsequent levels and depth of online disclosures have raised several concerns for user privacy. Previous research into these sites has indicated the importance of disclosures between users as well as an under-utilization of extensive

privacy options. This study qualitatively examines college students' disclosure and privacy behaviors and attitudes on Facebook.com. Results support current research into social networking and privacy and provide user-generated explanations for observed disclosure and privacy trends. Implications for future research into privacy software are discussed.[7].

B. Sriram (2010), et al. says that in micro blogging services such as Twitter, the users may become overwhelmed by the raw data. One solution to this problem is the classification of short text messages. As short texts do not provide sufficient word occurrences, traditional classification methods such as "Bag-Of-Words" have limitations. To address this problem, we propose to use a small set of domain-specific features extracted from the author's profile and text. The proposed approach effectively classifies the text to a predefined set of generic classes such as News, Events, Opinions, Deals, and Private Messages.[8].

V. Bobicev (2008), et al. says that the Classification of texts potentially containing a complex and specific terminology requires the use of learning methods that do not rely on extensive feature engineering. In this work we use prediction by partial matching (PPM), a method that compresses texts to capture text features and creates a language model adapted to a particular text. We show that the method achieves a high accuracy of text classification and can be used as an alternative to state-of-art learning algorithms.[9].

III. PROPOSED SYSTEM

An automated system called filtering wall that is able to filter unwanted messages from OSN user walls. We exploit machine learning text categorization techniques to automatically assign with each short text message a set of categories based on its content. Our contribution is that we are going to implement real time system using facebook app. The project is to develop a system that is going to block the unwanted messages from OSN user's wall. Now we are implementing the software which is going to work for filtering messages/comments in the form of a text, so in future we can extend our project scope to filter images, audio, video format or filtering. Paragraphs must be justified, i.e. both left-justified and right-justified.

ARCHITECTURE DIAGRAM

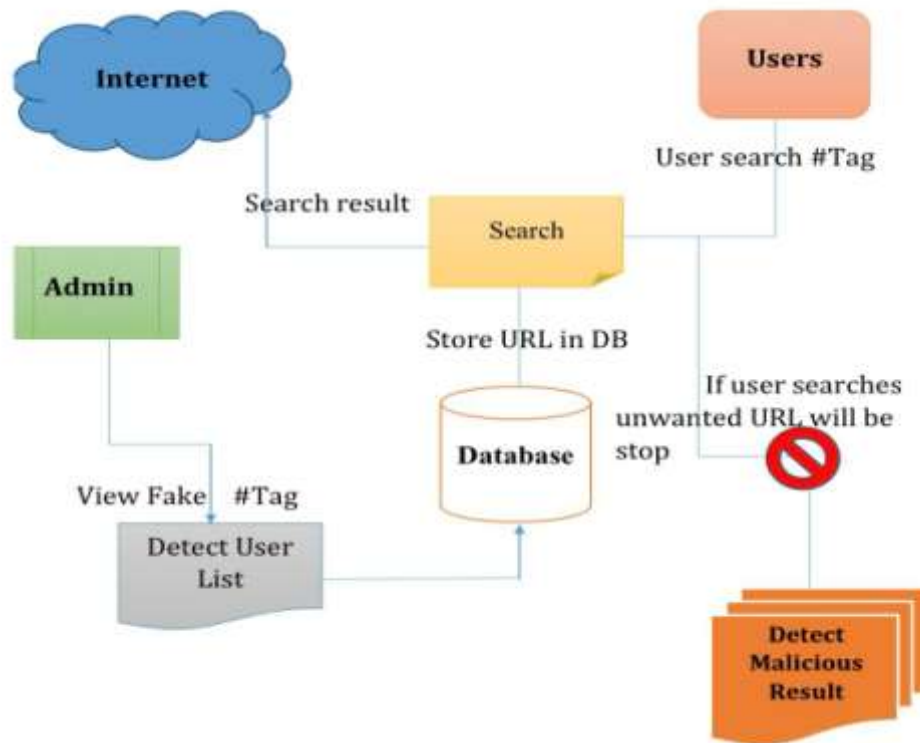


Fig 1: Architecture of proposed system

EXPLANATION

Fig 1, demonstrates the architecture diagram for the proposed system. To identify potential malicious social bots in online social networks in real-time, we analyze the social situation behavior of users in online social networks. We also evaluate user behavior features and select the transition probability of user behavior on the basis of general behavior characteristics. We then analyze and classify situation aware user behaviors in social networks using our proposed semisupervised clustering detection method. We then analyze and classify situation aware user behaviors in social networks using our proposed semisupervised clustering detection method. This allows us to promptly detect malicious social bots using only a small number of tagged users.

MODULE DESCRIPTION

1. Registration
2. User Login
3. Compose Mail
4. Inbox
5. Spam

6. Sent Mail
7. Classification

Registration

As like Gmail, we have created the application for sending and receiving mail. By using this application, spam mails are filtered. In this module, User needs to register first by giving all the details. Then the given details are get stored in the database. So, all the users should register before login.

User Login

After Registration process done, User can login with username and password. If both get matches, then the user will be considered as a valid user otherwise invalid user. Valid user can send mail to other users.

Compose Mail

In this module, user can compose mail after login. The User can enter the message what he wants to send to other user. After composing the message, user enters the To mail and send the mail to the other user. User can compose both the spam

and ham mails.

Inbox

In this module, Inbox which contains all the received mails sent by the user. It accepts only the ham mails and it will not allow the spam mail in inbox. All the received mails can be read by the users.

Spam

If any user sends spam mail means, automatically it will come to Spam module. It will not show in Inbox. So that the user can get to know about the spam mails. It avoids the user to click on

the spam mails.

Sent Mail

In this module, User can able to see the sent messages. All the sent mails are show here.

Classification

We have collected the dataset which contains both spam and ham mails. By using the LDA and SVM algorithm, we have classified the Spam and Ham mails. The Result shows that LDA has the higher accuracy than the SVM.

IV. RESULT AND DISCUSSION



Figure 2: Home Page

In figure 2, it includes the admin module, user module and the registration page. After logging in, they access respective functionalities through the home page.

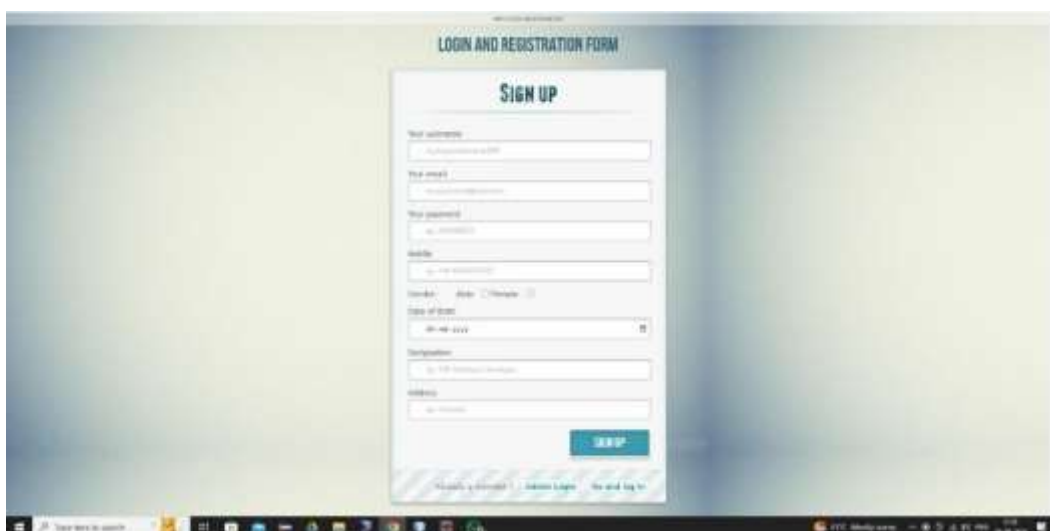


Figure 3: Register Page

Figure 3, demonstrates the user register form. The Registration Module holds all the information related to registration. It is responsible for tasks such as user authentication, storing user

preferences. It generally contains the basic information about the user such as name, E-mail, password, address and mobile number.

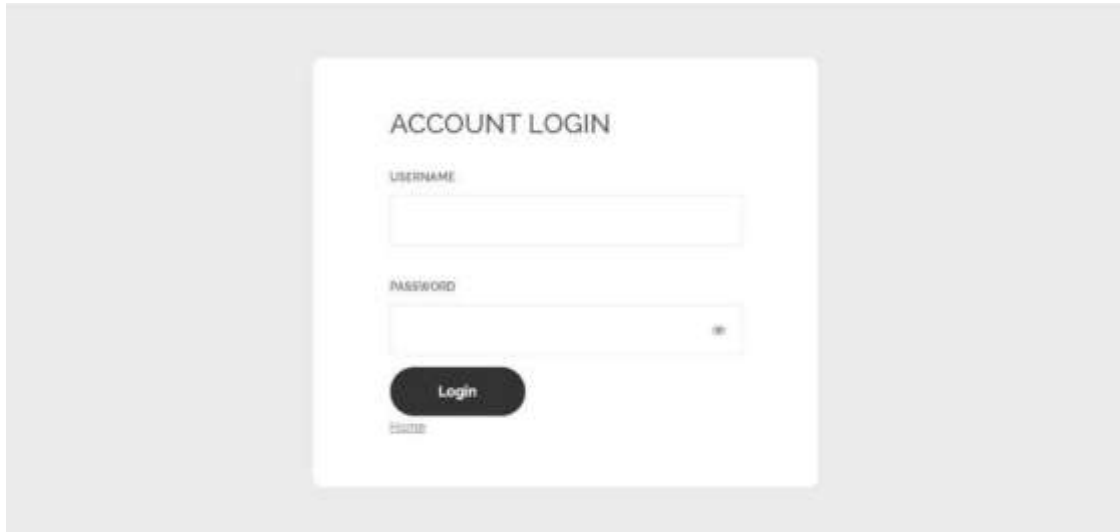


Figure4: Admin login page

Figure 4, demonstrates the admin login form which manages the authentication process. This involves validating the provided credentials, such as a username and password against stored

user data. The portal's admin login page serves as a secure door for authorized persons to get to and oversee the data.



Figure5: Admin add url and hash tags page

Figure 5, demonstrates the admin add url and hastages page. This page allows the admin to update profile, search user, send messages.






S.No	First Name	Last Name	Password	Email	Gender	DOB	Profile Pic
1	Rejany	R	123456	priyavimala.2002@gmail.com	female	2002-03-01	
2	Sanjuka	R	123456	sanjuka@gmail.com	female	2002-06-15	
3	deep	s	1234567	devidisha07@gmail.com	female	2022-04-03	

Figure6: Viewall userpage

Figure 6, demonstrates the view all user page. This page allows the admin to view all the information about the registered user.



Figure7: Chat page

Figure 7, demonstrates the chat page. In this module, user can compose mail after login. The User can enter the message what he wants to send

to other user. After composing the message, user enters the To mail and send the mail to the other user.

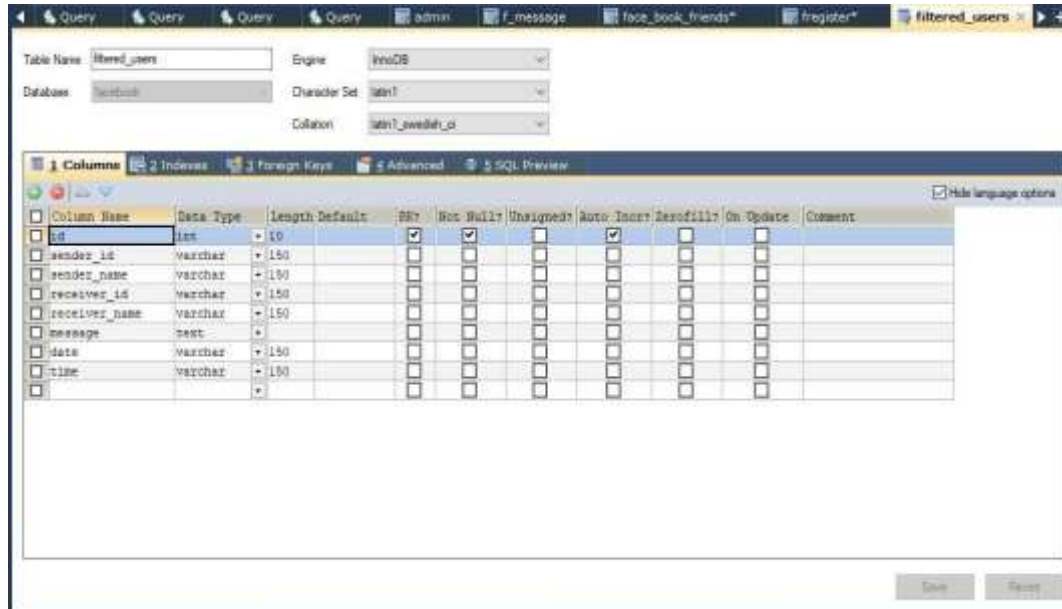


Figure 8: Filter user db page

Figure 8, demonstrates the Filter User Database page. This allows us to promptly detect malicious social bots using only a small number of tagged users.

V. CONCLUSION

Because spam issues are massive and continuous, spam filtering approaches with higher performance are still required to be developed urgently. We proposed a Chinese spam filtering approach with a text classification method based on semantic information. The extraction of semantic information from text was achieved by attaching semantic annotations to the words and sentences of it. The results of the experiment conducted on the corpus of TREC06c showed a satisfactory classifying performance on Chinese texts and indicate enormous potentiality in spam filtering with multiple classes and fewer feature terms.

The character of spam issues, massive and continuous, and spam filtering approaches with higher performance are still required to be developed urgently. A spam filtering approach with a text classification method based on the semantic information. The extraction of semantic information from text was achieved by attaching semantic actions to the words and sentences of it. The results of the experiment conducted on the corpus showed a satisfactory classifying performance on text, indicating enormous potential

in spam filtering with multiple classes and fewer feature terms.

FUTURE ENHANCEMENT

There is plenty of further work that can proceed based on semantics-based text classification because of the highly effective and flexible method of feature selection. For example, the application of online SVM is limited because of the low training speed, but the semantics-based feature term selection by attaching annotations would highly promote the efficiency of online SVM, which is also promising in online spam filtering. Moreover, the method of attaching annotations based on semantics has great potential in text classification in situations where multiple languages are involved, since this method focuses mainly on the semantic meanings other than the words. It can also play an efficient role in the classification of SMS messages, news, scientific literature, and messages delivered through social networks. Besides, the convenience of realizing personalized spam filtering is significant for a user-friendly and approachable commercial spam filtering system.

REFERENCES

- [1] M.Tsikerdekis, "Identity deception prevention using common contribution network data," IEEE Transactions on Information Forensics and Security, vol. 12,

- no. 1, pp.188–199, 2017.
- [2] T.Anwarand M.Abulaish,“Ranking radically influential web forumusers,” IEEE Transactions on Information Forensics and Security, vol.10, no. 6,pp. 1289–1298, 2015.
- [3] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, “Design and analysis of socialbotnet,”Computer Networks, vol. 57, no. 2, pp. 556– 578, 2013.
- [4] D.Fletcher,“Abriefhistoryofspam,”TIME, Tech.Rep.,2009.
- [5] Y.Boshmaf,M.Ripeanu,K.Beznosov,andE.Santos-Neto,“Thwartingfakeosnaccountsby predicting their victims,”in Proc. AISec., Denver,2015, pp. 81-89.
- [6] L. FangandK.LeFevre “Privacy wizards for social networking sites”, 2010, Vol. 6, pp. 1–10.
- [7] K.StraterandH.Richter “Examining Privacy and Disclosure in a Social Networking Community”, 2007,pp. 435– 42.
- [8] B. Sriram, D. Fuhry, E. Demir, “Short Text Classification in Twitter to Improve Information Filtering”, 2010, pp. 1–9.
- [9] V. Bobicev andM. Sokolova(2008), “An Effective and Robust Method for Short Text Classification” Vol. 16 pp. 576–589.
- [10] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, “Detecting and characterizing socialspam campaigns,”inProc. IMC, Melbourne, 2001,pp. 35–47.