International Journal of Research in Advent Technology, Special Issue, June 2018
E-ISSN: 2321-9637
International Conference on "Multimedia Processing, Communication and Information Technology" (MPCIT -2018)
Available online at www.ijrat.org

An Efficient Way of Filtering Sensitive Messages Using N-Gram Technique over Social Media

Althaf Patel, Aparna V.P, Apoorva C.G, Piyush Kataria
Guide Dr. Sandhya N.
B. *E, Department of Computer Science*
City Engineering College
Bengaluru, India

Abstract - Online Social Networking sites (OSNs) helps to connect people easily. There are various online social sites that are available like Facebook, Twitter etc. which brought world closer. In OSN, user has to create an account on social sites after which they are able to perform various actions like adding friends, sharing videos and images. OSN sites provide some space or area called Wall for users to post their status. But sometimes people post offensive messages on a particulars wall which may cause a serious problem to user's reputation. The state of being unknown via smartphones with pre-paid SIM cards, public Wi-Fi hotspots or over distributed networks like Tor has drastically complicated the task of identifying users of social media during forensic investigations. To overcome problem like these, we can apply Information Filtering (IF) technique where it can be used for formless data in contrast to database application in which data required is in ordered manner. There are various types of Information Filtering methods but here we will be using Content-based filtering method and Policy-based filtering method for text filtering process.
Keywords— Social Media, Bag-of-words, N-gram classification, sensitive messages.

I. INTRODUCTION

As we know online social networking sites are more interactive channel for people to share various information's, ideas and other forms of expressions through Multimedia sources like video, audio etc. with the world.

There are great advantages of such social networking sites excepting a few drawbacks with security which might create problems to active users on such sites. As we know these sites allows users to post comment on another user's wall even when they are unknown to each other, but if that comment is a misleading or harmful then it may cause serious problem to user's reputation where this practice on internet is termed as trolling[3].

To avoid such problems, Information filtering is used to filter the message contents. Forensic authorship attribution isthe process of summarizing something about the characteristics of an author from the form and content of their writing existing collection of evidences that is through the features derived from the style of their writing; also called as Stylometry, capture the diversity of the language deployed therein. In the context of a criminal investigation, this endeavor is termed Forensic Authorship Attribution.

II. NEED FOR AUTHORSHIP ATTRIBUTION

User can post the messages on social media anonymously, sometime causing serious problem to other users on the same platform and can be affected individually by their social images which means instead of all those advantages provided by social sites there are some disadvantages also. Thus, the user has to be restricted [1][4] from posting sensitive messages on these sites by providing filtering techniques to those social media platforms.

A primary problem in this area has been identifying authorship attribution for short messages along with the problem is exacerbated by the unconventional punctuation, abbreviations, and character based signifiers common in Internet culture for which this system uses features like N-Gram technique for content-based filtering and Weight-age concept for policy-based filtering method along with various classification approaches applied to new data set collected. With the help of these concepts the system will detect whether the post contains sensitive data or not.

Identifying sensitive post and block the post, this helps in avoiding unnecessary conflicts in the society also provides Statistic Report to admin

regarding which user is trying to post sensitive post frequently.

III. SYSTEM ARCHITECTURE

System architecture primarily concentrates on the internal interfaces among the system's components or subsystems, and on the interface(s) between the system and its external environment, especially the user.

In this system (Fig1), the admin inputs the sensitive messages into database, the processor tests the data with the help of "word-N- Gram" technique, and output of theprocessor goes to "bag-of- words" where it removes the unwanted words and again the message goes through sensitive classification process. This whole procedure is to filter out the messages based on the sensitivity.

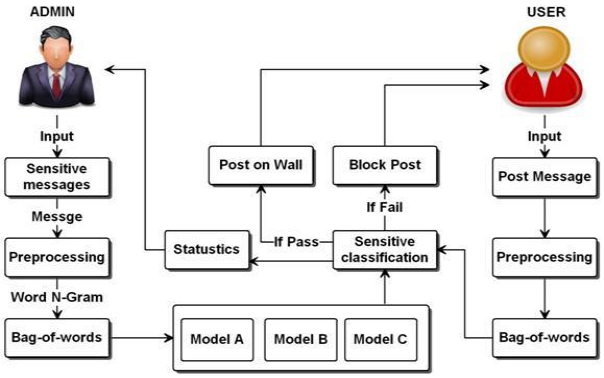


Fig1: System Architecture

The User gives the input in the form of text (post message), processor process the data [2][7] and gives the input data to "bad-of- words" to exclude the unwanted words. Output of "bad-of-words" goes to "sensitive classification" where the processor divides the message in to 3different classification model A (less sensitive), model B (medium sensitive) and model C (more sensitive) if the message is more sensitive (model C) then the processor blocks the post and if the message is less or medium sensitive the processor allows to post the message on wall.

IV. METHODOLOGY

The various methods and techniques used for this system can be seen as follows,

- Bag of 'N' Words Module

This module extracts features from text for use in modeling, such as with machine learning algorithms. It is only concerned with whether known words occur in the document (sentence), not where in the document, this describes the occurrence of words within a document based on the vocabulary of known words and measure of the presence of known words because machine learning algorithms cannot work with raw text directly; the text must be converted into numbers. Specifically, vectors of numbers.

- Classification and Decision Making

The Admin classifies words from bag-of-words model as Less sensitive, Medium sensitive, More sensitive based on which the user input text (comments) will be compared (weightage calculation) and sensitivity result will be given.

- Policy Details Module

In Policy module admin has to give sentence and select the classification id and store it into m_sentences table. Then the unnecessary words from those sentences will beremoved by using preprocessing, after which 'N' gram technique is performed on keydata (result) and based on the classification they will be stored onto the respective sensitive table.

- User Registration Module

In this module, the user will get registered with basic information of his onto any social media sites so that he can have access to view, post any images or comments.

- Posting Management Module

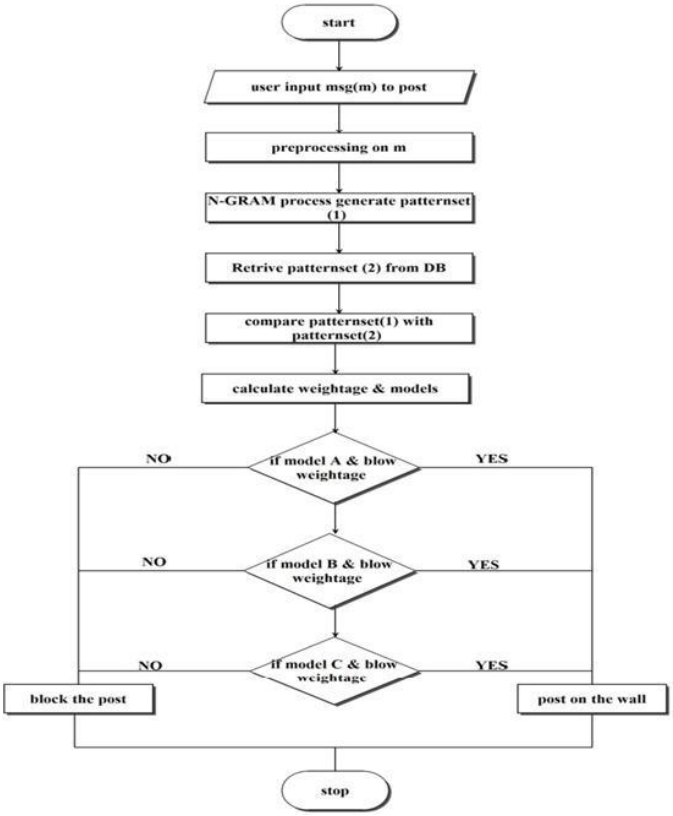
In this module, users are allowed to post images and comment. To know the level of sensitivity of these comments [5], our system first removes the unnecessary words with the help of preprocessing algorithm as we discussed in policy details. The system now uses 'N' gram technique for the keydata and compares with all the values in sensitive tables. The system allows user to post the comments in user wall if there is a match for less sensitive values, else it restricts them from posting it on the site. The system will also show sensitive result to users.

- View Postings Module

In this module, the individual user can

provide privacy in their account by using friend group management where only those people can view and comment on the posts.

• Flowchart



V. RESULT AND DISCUSSION

In this section, we present a comprehensive and performance evaluation of the system by describing the data sets and the comparative experiments. The Admin of the system has the authority for accessing various operations like has authority over user details, manages classifications of sensitive messages where they are classified to their sensitive level (less, medium, more) with ids and policies which are added by the admin who also gives the classification id for them, can change his/her profile password.

The user first registers to the corresponding social media sites where he/ she is asked to give their personal details like name, password, birth date, gender etc for further activities to be carried out. The user can login to their respective accounts on social sites and can perform general activities such as connecting with other people on the identical site, can post images, can write comments, can view his profile .

The comments [7] that have more sensitive words then those messages will be restricted from posting based on the filtering techniques/ methods used.

VI. CONCLUSION

Today various social networking sites are available which make people remain in constant touch with each other. Sharing any type of data has become easy. There are great advantages of such social networking sites excepting a few minor drawbacks like poor security which create huge problems to people when they were active on such sites. As we have seen Facebook allows users to post comment on another users wall even when they were unknown to each other. But if that comment is a vulgar one then it may cause serious problem to user reputation. To avoid such a problem Information filtering is used to filter the content of the message. So we have analyzed various Information filtering methods like content based filtering, policy-based filtering, and collaborative filtering in this paper. Content-based filtering method is best filtering method than any other methods, because it has filtered out bad or non-neural words from the input message and allows posting only pleasant comment on a user's wall. This will help us to avoid unwanted messages

from ever spoiling reputation which carries the at most importance in the world of socialization.

REFERENCES

[1] P. Juola, "Authorship attribution", *Found*, vol. 1, no.3, pp. 233-334, Dec. 2006.

[3] K. Gimpel et al., "Part-of-speech tagging for Twitter: Annotation features and experiments", *Proc. Annu. MeetingAssoc. Comput. Linguistics Human Lang. Technol.*, vol. 2,pp. 42-47, 2011.

[4] R. Layton, S. McCombie, P. Watters, "Authorship attribution on *Cybercrime Trustworthy Compute".Workshop*, pp. 7-13, Oct. 2012.

[5] D. V. Khmelev, F. Tweedie, "Using Markov chains for identification of writer", *Literary Linguistic Comput.*

[5] W. J. Scheirer, L. P. Jain, T. E. Boult, "Probability models for open set recognition", *IEEE Trans. PatternAnal. Mach. Intell.*, vol. 36, no. 11, pp. 2317-2324, Nov.2014.

[6]S. Okuno, H. Asai, H. Yamana, "A challenge of authorship identification for ten-thousand-scale microblog users", *Proc. IEEE Int. Conf. Big Data*, pp. 52-54, Oct. 2014.

[7] M. Eder, M. Kestemont, J. Rybicki, "Stylometry: A suite of tools", *Proc. Digit. Humanities*, pp. 1-4, 2013.

[8]Mohamad Hassoun, *Fundamentals of ArtificialNeural Networks*, A Bradford Book.

[9]Sandhya Samarasinghe, *Neural Networks for AppliedSciences and Engineering*, Auerbach Publications.[10]Martin T Hagan, Howard B Demuth, Mark H Beale, *Neural NetworkDesign*, Martin H