ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Detecting Spam Comments Using Spam Word Dictionary and NLP Technique

Rohini D. Warkar[1], Prof.I.R. Shaikh[2]

P.G. Student, Department of Computer Engineering, SND College of Engineering, Yeola, SPPU, Maharashtra, India[1]

HOD, Department of Computer Engineering, SND College of Engineering, Yeola, SPPU Maharashtra, India[2]

ABSTRACT: There are different topics on social media sites that trending towards some popularity on sites. To find such popular or hot topics is becoming one of the challenging task on social media site. Now a day, we are using social media to discuss the hot topics with friends so spamming in comments is increasing day by day. Due to this there is necessity to control spamming carried out on social media sites. To solve this problem of spam detection we used different concepts of data mining. This is the system which is work on detecting the spam comments using self-extensible dictionary and natural language processing.

KEYWORDS: Control Spamming, Information filtering, Natural Language Processing,Social Networking site, Text mining.

I. INTRODUCTION

As we know the popularity of on-line media is increases day byday. Millions of people are connected to social media at a time with their friends, family, colleague etc. People were discuss different issues, news or leading topics with all their circle on on social sites using comments. People use to comments as opinions or their views. there were different comments present on social media sites. some of them are spam comments. So need to detect that spam comments and spammer.

There are several methods that are used for detecting spam comments. Some of them are listed as term frequency-inverse document frequency (tfidf), Naive Bayes classifier and SVM. The disadvantages of these techniques are they do not consider semantic information of the spam. Due to this issue we get incomplete result of spam detection. So there is need to detection take place on the basis of spam word as well as semantic.

Natural Language processing (NLP) is a technique refers to the Artificial intelligence. The NLP Algorithm contains different phases which are given below:1.Lexical analysis 2.Syntactic analysis 3.Semantic analysis 4.Disclosure Integration and 5.Pragmatic analysis. Text mining is nothing but the text data mining. It is the process of extraction or deriving the high quality information from large text data.

The goal of system is to detect spam comments. There are two phases of proposed system first is pre-processing and second is feature extraction. The system will give overall better results than previous techniques. Further, We discuss the details of proposed system.

II. LITERATURE REVIEW

Cristina Radulescu, Mihaela Dinsoreanu,and RodicaPotolea proposes identification and detection of spam using NLP technique. It gives tokenization method for implementation efficiently [2].

Hurst, Maykov and Sayyadi proposes clustering of the word approaches for extracting the words. In this paper the author used twitter comments as a input data for formalizing and frequency count is done after mining the words. This frequency is counted for specific time interval and clustered data[4].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Backstorm and Kleinberg gives the information of scalable algorithm in clustering which helps to checking the phases or words. It is the algorithm which is used for analysing matching the contents. For this purpose it analyse the large data of social media sites.[5]

Andrew Y.Ng and David M.Blei gives the information and use of model named as Latent Dirichlet Allocation(LDA)model. It is model which is works on the basis of word frequency of given document. this is used to find the accuracy of data according to reasonable data set.[6]

III. PROPOSED SYSTEM

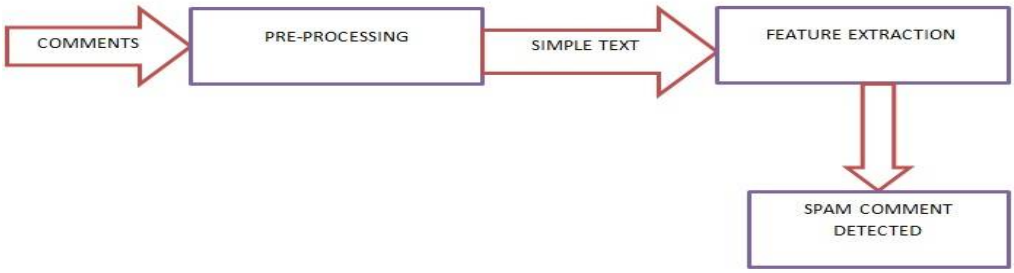

Figure1: System Overview

The system takes the input as comments from social media sites. These comments are pre-processed in pre-processing module or phase. The output of pre-processing is simple plain text. This plane text is give as an input to feature extraction module. In this phase actual spam is detected by comparing text with self-extensible spam word dictionary. If spam word detected then it is replace with star (***).Finally we get the output as spam has been detected or not.

IV. IMPLEMENTATION DETAILS

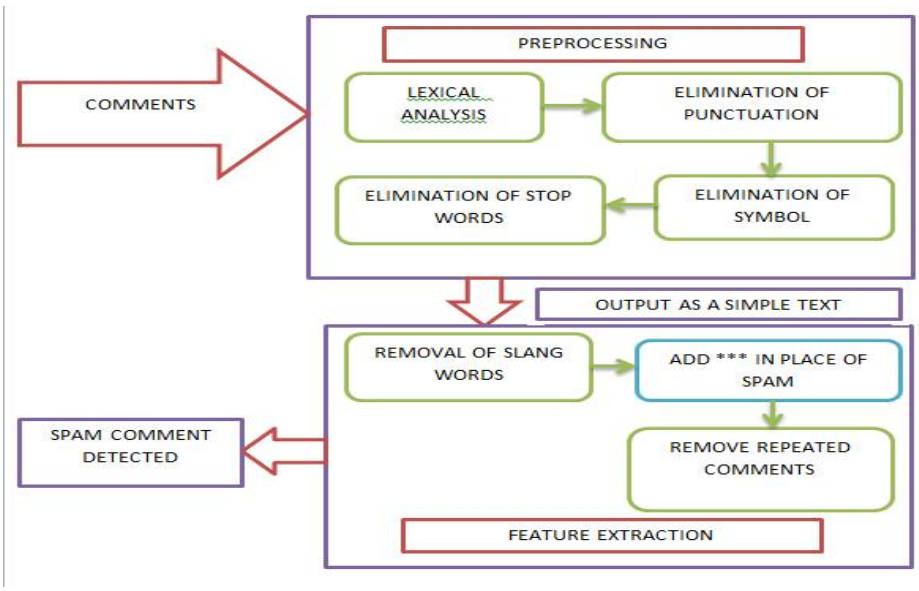

Figure2: Block Diagram of Proposed System

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

The system is divided as two phases or modules Pre-processing and Feature extraction.The figure gives the detailflow of system.
1) Pre-processing
2) Feature Extraction

A. **Pre-processing**: Pre-processing is the initial phase or first phase of proposed system. In this phase we filter comments and convert comments into simple plane text. In this phase we use the concept of natural language processing (NLP) algorithm. We mainly use the lexical analysis phase in NLP.

- Lexical analysis: Lexical analysis is used as analyzer or as a parser. It converts sentence into words and also words into character.
- Elimination of Punctuations: In this sub module we are removing punctuations. For example comma, full stops etc.
- Elimination of Symbols: In this sub module we are eliminating symbols like $,#, % etc.
- Elimination of stop words : In this sub module we remove words that break the sentence. For example for, and, is, the, of, in etc.

B**. Feature Extraction**: Feature extraction is second phase of proposed system. In this system we take input as simple plain text from pre-processing. then we detect the spam word by comparing it with spam word dictionary. In this phase we use iteration algorithm for creating the spam word dictionary.

- Remove of slang word : In this sub module we control the spamming in comment. for this task we create a dictionary of slang words and comparing word of comments with this dictionary if word is available then this word detected as spam word and then it is replaces by star(***).

- Remove Repeated comments: In extracting the data from document remove ambiguity in result.

Algorithm for Spam Word Dictionary

1. Procedure : Construct the Spam words Dictionary.
2. Input: spam words of AD dictionary and Basic vulgar or added in the result of the previous iteration.
3. Compare semantic of the most similar words from dictionary with our spam word selected;
4. And spam word is added into the candidate spam word dictionary;
5. Delete words of candidate spam dictionary if they exist in basic vulgar dictionary;
6. Calculate the average weight of same words in candidate spam word dictionary.
7. For each spam word in candidate spam word dictionary do Acquire 4 most similar words for spam word by comparing the semantic similarity between them;
8. If there are more than 4 words among exist in Basic vulgar and AD dictionary then Add given spam word into final dictionary otherwise Drop it;
9. Empty the candidate dictionary
10. Output: the newly added spam words in this iteration.

V. RESULT ANALYSIS

A. Data Set And Expected Output

The expected input of proposed system is comments. The constraint is comment must be in textual format only. The expected output is the given comment is spam or not. We take twitter comments as a input data sets. From these comments spam word are detected and spam word is replaced by *. The overall result of proposed system is spam detected means comment is spam or not.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

B. Experimental Result

We use classification algorithm to measure a result of our proposed system. The confusion matrix of system is define as,

	Spam	Not Spam
True Spam	A	B
Not Spam(Normal comments)	C	D

There are three threshold values such as Precision Rate, Recall Rate and F1
Precision Rate (P)=a/(a+b);
Recall Rate (R)=a/(a+c);
F1 is the balanced value of P and R used to evaluate the overall result of classification.
F=2*P*R/(P+R);
Consider as we compile 10,000 comments taken from twitter as input file.

	Spam	Not Spam
True Spam	4480	1048
Not Spam(Normal comments)	1254	3218

By calculating value of a Precision rate ,Recall rate and F1 we get,

	Precision rate	Recall rate	F1
True Spam	81.04	78.13	79.55
Not Spam(Normal comments)	88.01	89.23	88.61

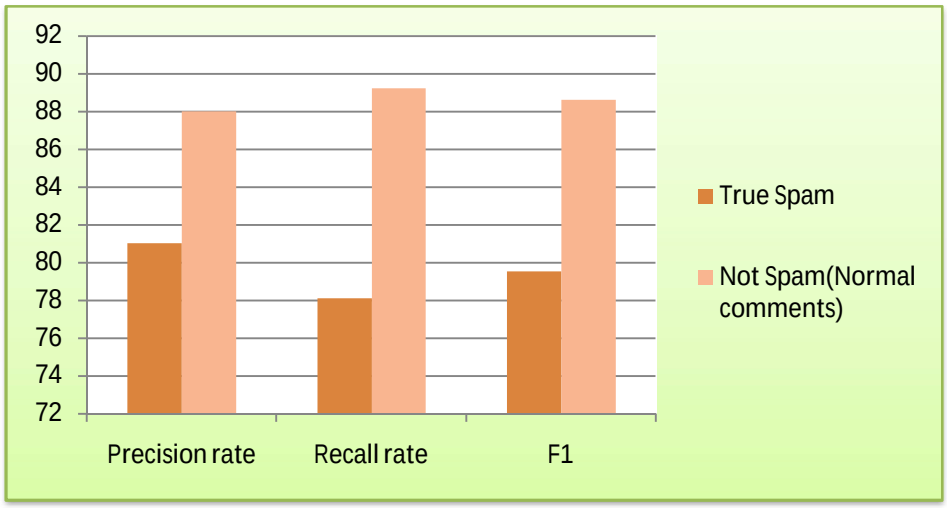

Figure3: Graphical representation of Precision rate &Recall rate

It is observed that the proposed approach distinguish the normal comments from the spam comments. With respect to the accuracy and efficiency threshold, We observed that it is really difficult to confirm such a threshold

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

adapted to any comments from any social media site(as we used twitter), especially considering the diversity of online language used. As we know language understanding of human and machine is differs a lot and this is constraints to get maximum accuracy and efficiency Thus accordingly in our approach, we use the relatively accurate threshold values to get confirm accuracy.

VI. CONCLUSION AND FUTURE WORK

The main aspects of proposed work are detection of spamming occurred in social media sites. This work is divided into two approaches pre-processing and feature extraction. Pre-processing works on filtering data and convert it into simple plain text and feature extraction extract the spam word by comparing it with self-extensible spam word dictionary. Natural Language processing algorithms improves the semantic analysis results in improved spam detection. Self -extensible spam word dictionary minimize the complexity of spam detection and improve performance.

The Next work will be research of spam detection based on order and format of spam words also combination of spam words. And also to implement on-line security and current event detection.

REFERENCES

1. Chenwei Liu, Jiawei Wang, Kai Lei, "Detecting Spam Comments Posted in Micro-BlogsUsing the Self-Extensible Spam Dictionary", IEEE ICC 2016 SAC Social Networking
2. Cristina Radulescu,Mihaela Dinsoreanu,and Rodica Potolea,"Identification of spam comments using natural languageprocessing techniques", In Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on, pages 2935.IEEE, 2014.
3. H. Otori and S. Kuriyama,"Emerging topic detection on Twitter based on temporal and social terms evaluation",in Proc. MDMKDD: 10th Int. Workshop Multimedia Data Mining, New York, NY, USA, 2010, pp. 4:14:10, ACM.
4. Sayyadi, M. Hurst and A.Maykov,"Event detection and tracking in social streams", in ICWSM, E. Adar, M. Hurst, T. Finin, N. S. Glance, N. Nicolov, and B. L. Tseng, Eds. Palo Alto, CA, USA:AAAI Press, 2009.
5. J. Leskovec, L. Backstrom, and J. Kleinberg,"Meme-tracking and the dynamics of the news cycle", in Proc. KDD: 15th ACM Int. Conf. Knowledge Discovery and Data Mining, New York, NY, USA, 2009, pp. 497506.
6. D. M. Blei, A. Y. Ng, and M. I. Jordan,"Latent dirichlet allocation", J. Mach. Learn. Res., vol. 3, pp. 9931022, Mar. 2003
7. Roshani M. Shete, Prof. S. W. Mohod,"Using Natural Language Processing for Detection of Events and Spam Control from user Data Stream in Social Sites", International Journal of Engineering Research and Technology (IJERT) ISSN: 2278-0181 Vol. 4 Issue 04, April-2015
8. Bai Xue, Chen Fu, and Zhan Shaobin,"A study on sentiment computing and classification of sina weibo with word2vec",In Big Data (BigData Congress), IEEE International Congress on, pages 358363 IEEE, 2014.
9. Huiyu Wang, Kai Lei, and Kuai Xu,"Profiling the followers of the most influential and verified users on sina weibo", In Communications (ICC), IEEE International Conference on, pages 11581163. IEEE, 2015.
10. Ala M. Al-Zoub, Jafar Alqatawna, Hossam Faris,"Spam Profile Detection in Social Networks Based on Public Features",8th International Conference on Information and Communication Systems (ICICS),2017.
11. Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon,"What is Twitter, a Social Network or a News Media?", In Proceedings of the 19th international conference on World wide web, pages 591600. ACM,2010
12. Rohit Giyanani, Mukti Desai,"Spam Detection using Natural Language Processing",IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. IV (Sep Oct. 2014), PP 116-119.

BIOGRAPHY

Rohini D. Warkar is a PG Student **I.R.Shaikh** is Professor and H.O.D. in the Computer Engineering Department, College of Engineering, Yeola, SPPU, Pune. Rohini D. Warkar pursuing Master's Degree in S.N.D COE,Yeola, SPPU, and Pune India. Her research interests are Data mining etc