International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Classification of Twitter Spam Based on Profile and Message Model Using SVM

Shamstabriz M. Asadullah[1], Dr. S. V. Viraktamath[2],

[1]*M.Tech, Department of Electronics & Communication, SDM College of Engineering & Technology, Karnataka, India*
[2]*Professor Department of Electronics & Communication, SDM College of Engineering & Technology, Karnataka, India*

Abstract - *Social Networking spam has become a significant concern in social media services. These platforms are misused by spammers to spread unwanted messages universally known as spam messages. Users introduce messages in trending topics with promotional messages providing a link. With increasing advancement of the internet technology it has become very difficult in detection fake profiles and spam messages. This paper gives a reasonable method to classify these spam tweets in a Twitter. Using of a machine learning algorithm such SVM (Support Vector Machine) from which the classification is made easier and more prominent in detecting Spam messages. Once the classification is done it will be easier to remove such profiles as well as the tweets.*

***Key Words*: Twitter, Support Vector Machine, SVM, Spam, Social Networking, Features**

1. INTRODUCTION

Twitter is a excellent initial point for social media analysis because people explicitly share their opinions to the general public. This is very different from Facebook, where social relations are often private. Twitter is fastest growing online entity [1]. Twitter had the greatest growth compared to other social network sites. Twitter aims to allow individuals have relation together through tiny message. Unluckily, spammers use twitter as a tool to hurl malicious links and messages to user. The studies show that more than 6% of tweets in twitter are spam [2].

To attack spam filters there are n number of sophisticated tools developed by spammers. This can be seen in Naive Bayes classifiers where haphazard paragraphs and complicated keywords are used to break through it. Unwanted additional information apart from spam is included in the spam to bypass the spam detection techniques. May f the spam content do not effect much to user but click on the URL posted along with the message can infect the account as well as the device [1][3].

The link/URL can infect the profile of the user by downloading malicious software or content and sometimes sends unwanted content or malicious links to the contacts of the user [4]. To avoid this social networking sites have their own spam filters and spam reporting features which are

i. Click on more icon of the tweet which you want to mark as spam
ii. Select "Report" option
iii. Select "Spam" option
iv. Submit your report

The dilemma is that spammers come in to contact with unsuspicious user to spread the spam content [5]. Machine learning algorithms and Data mining can effectively reduce spam content by taking benefit of the gigantic quantity of information on the social media sites. In this paper Support Vector Machine (SVM), a machine learning algorithms is used to categorize similar type of spam in twitter. SVM was developed based on statistical theory by Guyon, Vapnik, in which the training data is mapped into the feature space by using the Kernel functions.

2. METHODOLOGY

This system is mainly divided in three component i.e.,

i. Mapping and Assembly
ii. Pre-Filtering and
iii. Classification

In "mapping and assembly" the frame work defines a standard model for each and every object. In our proposed system, we have used two models: "Message Model and Profile Model". In Pre-Filtering the entering object is checked by equating it with blacklists which are present in database[6].

The system architecture is shown in Fig.1, two models are considered for detecting spam namely profile model and message model. Profile and message model of the object are mapped and assembled. Semi supervised classifier SVM is trained with this data along with blacklist and a

knowledge base is created [7]. In the testing phase, the social network considered is Twitter.

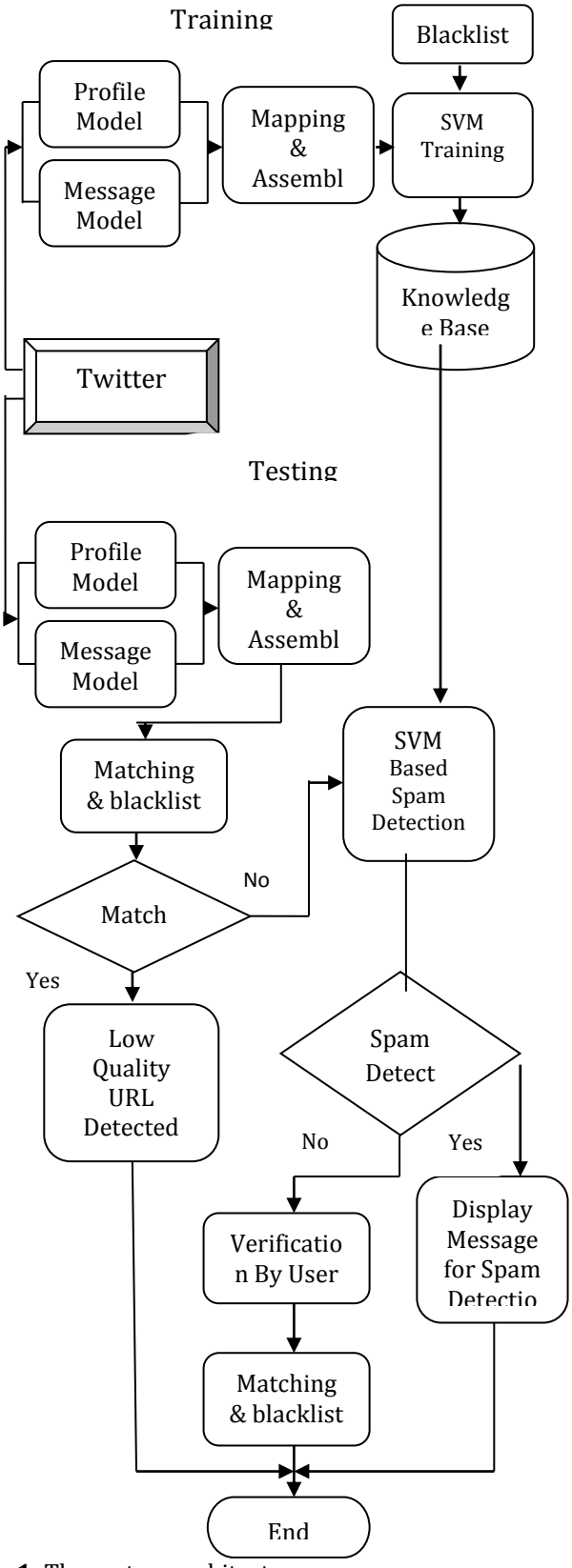

Fig -1: The system architecture

Profile model and message model are formed and the resulting URL is matched with the blacklist. If it matches then that particular URL is reported as spam. If it does not match then the URL is further analyzed by SVM. If according to classification result the URL is classified as spam then the URL is added to the blacklist. In categorization SVM is used for classification of the incoming object.

2.1 Support Vector Machine

Support Vector Machine uses Kernel Functions to map the training data into feature space. Let us consider the mapping of X and Y, were "x ∈ X" is an object and "y ∈ Y" is a label. Thus the classifier is given as y= f(x, α), where α gives us the parameters of the functions. In most of the circumstances data set can be linearly separable. For this we require a simple classifier,

$$S = \{ x \mid < w, x > + b = 0 \} \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Here w and b are taken from 'x' training set.

The decision function is given as

$$f(x_{new}) = sign\ (< w, x_{new} > + b\ ) \ldots\ldots\ldots\ldots (2)$$

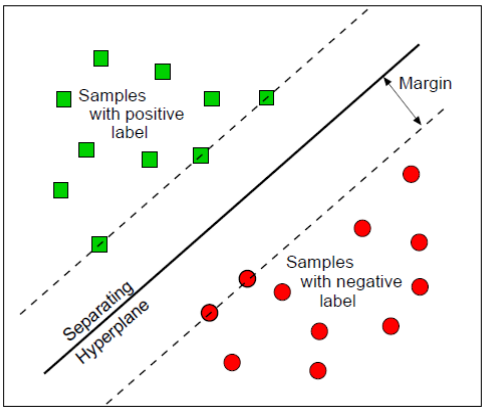

Fig -2: The system architecture

It is advisable to separate the training set with maximum margin as shown in the Fig-2.

3. IMPLEMENTATION

The overall execution is explained by using a flow diagram as shown in Fig-3. It gets splits up into two phase namely, training the data set and selecting the query data. In the training phase, the training data is taken from both

message model and profile model along with the URL linked with it.

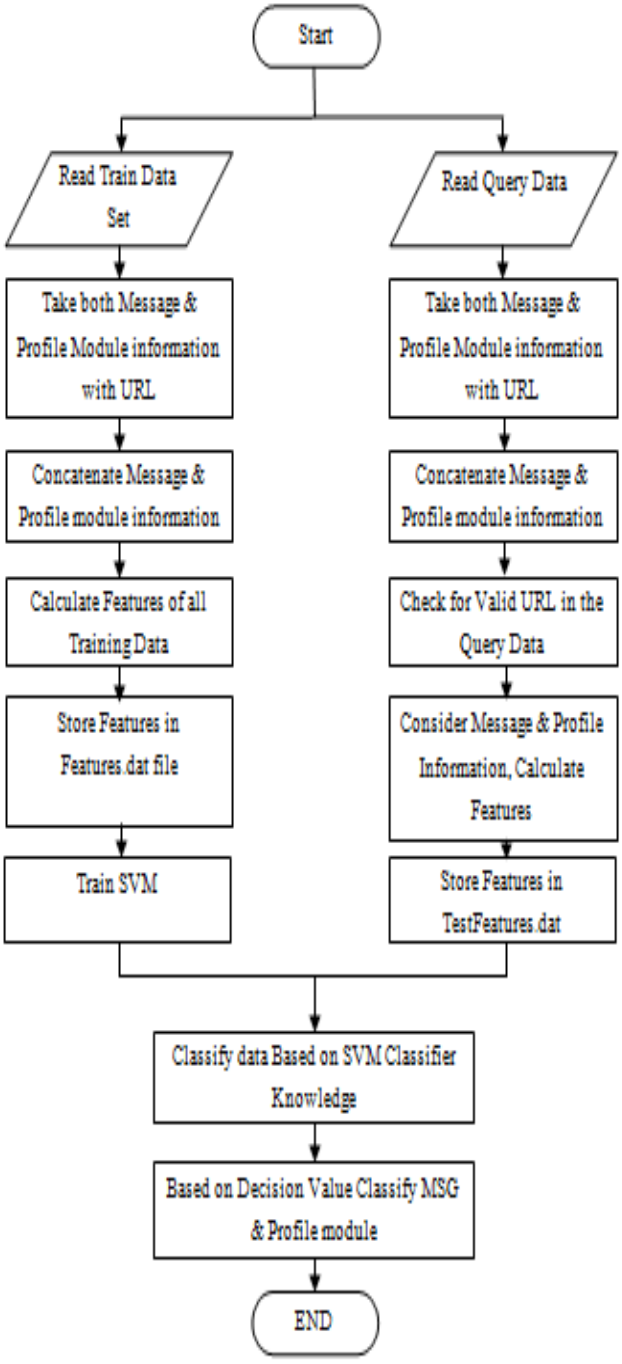

Fig- 3: Overall Flow Chart

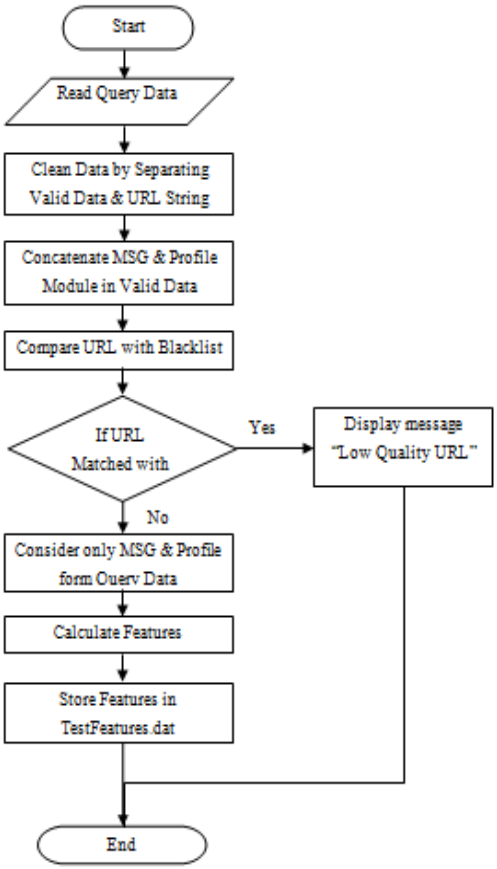

Fig- 4: Flowchart for selecting of Query

The data includes good as well as recognized spam messages. The information containing the profile details and message details are stored in ".dat" file. This data is used for training the Support Vector Machine. In the select query phase, the text file having the profile details and message details are read through the path name. In the select query phase, the text file containing the profile details and message details are read through the path name, which is shown in fig-4.

4. EXPERIMANTAL RESULTS

In this paper, our designed system detects spam, by using a Support Vector Machine model on social networks. To check the feasibility of the proposed system several experiments are performed to check the performance of spam detection. The dataset consists of dataset consisting of five legitimate messages and five spam message.

The detection results are evaluated by calculating the True positive rate (i.e., true positive rate, a real spam is classified as spam) and false positive rate (i.e., false positive rate, a good message is misclassified as a spam)

[6]. The true rate and false rate for spam and good messages for the proposed system is calculated below and comparison is shown in Fig-5

Spam Messages:

- True Rate = No of spam messages truly classified / total no of messages
  (4/5) X 100% = 80%
- False Rate= No of spam messages Falsely classified – True rate
  (1/5) X 100% = 20%

Good Messages:

- True Rate = No of good messages truly classified/ total no of messages
  (3/5) X 100% = 60%
- False Rate= No of good messages Falsely Classified / total no of messages.
  (2/5) X 100% = 40%.

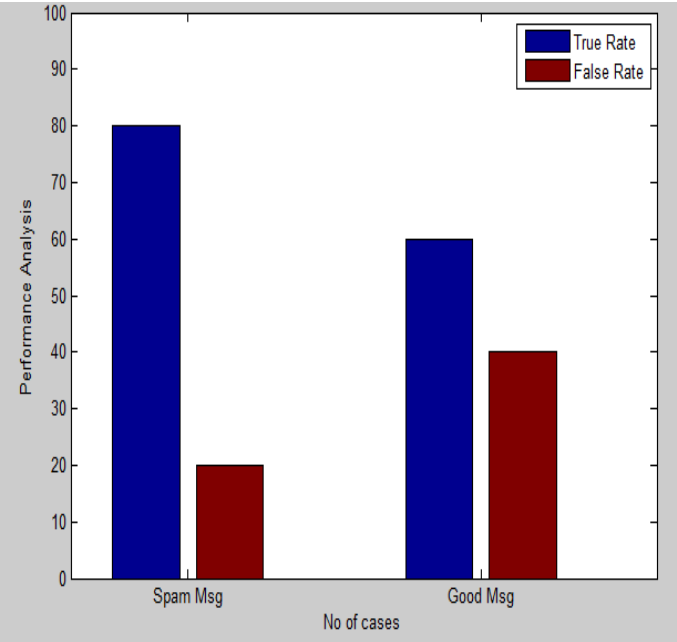

Fig- 5: Comparison between true rate and false rate

5. CONCLUSION

In order to detect and prevent spammers in social networks several methods have been proposed and developed by many researchers. During our survey it is seen that spam detection in social networks using Decision Tree, Support Vector Machine, Random Forest and Naïve Bayesian approaches is highly effective and a combination of spam prevention filters will give higher accuracy. Spammers are involved in posting multiple messages by creating fake profiles. Spammers also try to hack different user profiles. URL in Twitter is intensively analyzed to find out helpful methods to avoid URL spam effluence. This study includes how spammer use different spamming techniques in spreading URL spam. Tit also shows our detection problem and explained the trained SVM, will classify the testing data considering both the profile model and message model into spam message and good message.

REFERENCES

[1] M Robertson, Y. Pan, B. Yuan, "A Social Approach to Security: Using Social N1etworks to Help Detect Malicious Web Content," in *2010 International Conference on Intelligent Systems and Knowledge*

[2] Wang, A.H. *Don't follow me: Spam detection in twitter*. in *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*. 2010. IEEE.

[3] K. Beck, "Analyzing Tweets to Identify Malicious Messages," in *Proceedings of the 2011 IEEE International Conference onElectro/Information Technology*, Mankato, MN, pp. 1-5, 2011.

[4] X. Jin, J. Luo, C. Xide Lin, and J. Han. A Data Mining-based Spam Detection System for Social Media Networks. 2011.

[5] D. Wang, D. Irani, and C. Pu, "A Social-Spam Detection Framework,"in *Proceedings of the 8th Annual Collaboration, Electronic messaging,Anti-Abuse and Spam Conference*, New York, NY, pp. 46-54, 2011.

[6] D. Wang. Analysis and Detection of Low Quality Information in Social Networks. "*30th IEEE International Conference on Data Engineering Workshops"*, 2014

[7] D. Wang and C. Pu. "BEAN: a BEhavior Analysis approach of URL Spam filtering in Twitter" in *16th International Conference on Information Reuse and Integration* on 2015.IEEE

[8] N. Chinchor, "Muc-4 evaluation metrics," in *The Fourth Message Understanding Conference*, 1992.