Detection of Spam Messages in Social Networks based on SVM

Sumaiya Pathan
Computer Network Engineering
Visvesvaraya Technological
University. Belgavi- 590018,
Karnataka

R. H. Goudar
Associate Professor
Computer Network Engineering
Visvesvaraya Technological
University,Belgavi- 590018,
Karnataka

ABSTRACT
Social networks are platforms through which people communicate and share information. Some users commonly known as spammers are misusing these platforms for spreading unsolicited messages commonly known as spam messages. Due to the advancement of internet, it is very difficult to detect spam messages and fake profiles. This research article presents the use of a machine learning algorithm such SVM (Support Vector Machine), which is based on statistical learning methods to detect spam in social networks. This paper also evaluates the classification efficiency of Non Linear SVM using RBS (Radial Basis Function) Kernel.

Keywords
Spam, SVM, RBS, Kernel, Machine Learning.

1. INTRODUCTION
Social networks provide efficient channel for the growth of spam. Spam is unsolicited posting in social networks. Spammers are involved in posting multiple messages which have the same URL. There are several definitions of spam. The most widely excepted definition is "unsolicited bulk messages". Another commonly known definition is "spam in internet includes unsolicited messages all having similar content, which are posted or sent to large numbers of legitimate users". The spam track defines spam as "unwanted, unsolicited bulk email which is sent indirectly, directly and indiscriminately to large number of users by an unknown sender". Spammers create fake profiles or hack different user profiles. Hence we have trained our SVM, which will classify the testing data considering both the profile and message model. Spam is present in all types of social networks such as twitter, facebook, email, LinkedIn etc. In this framework, machine learning algorithm such as SVM (Support Vector Machine) is used to classify similar types of spam in all types off social networks. 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. Thus a person who's developing a new social networking site can prevent its users from social spam.

Several algorithms have been proposed by various researchers for detecting spam in the area of email, image, facebook, twitter, VoIP (Voice over Internet Telephony) and SMS. We have reviewed the existing hybrid techniques. Several authors have proposed various hybrid techniques to overcome the limitations of spam filtering algorithms. These hybrid filters greatly improve the classification efficiency. From the above survey, we can presume that machine learning algorithms are more efficient for detection of spam messages. In [1], the author proposes a hybrid technique using artificial immunity spam filter and Bayesian Spam Filter. Bayesian Filter is effective in creating an anti-spam filter because of accuracy. Artificial Immune System is efficient in the field of computing because of its self adaptability, robustness and self learning. These two techniques can be combined to reduce the processing time. The filter takes the whitelist and compares it with the message header and if spam is detected it sends it to the spam folder. In [2]-[4], to provide accurate classification result in email the author combines rough set theory and TF-IDF (Term Frequency-Inverse Document Frequency). In [5]-[8], the technique used is K-means Filter based on local concentration. Filtering occurs in four stages. In the first stage, to generate the terms from the message received we apply string token. In the second stage, information gain is applied to the gained information. In the third stage, for feature selection artificial immune system based on local concentration is used. In the fourth stage, the algorithm used is K-means for feature vectors. In [9]-[10], to optimize the performance of spam filter the two methods were hybridized: PSO (Particle Swam Optimization) and SAIS (Simple Artificial Immune System). Mutation was used by PSO to improve the filtering of artificial immune system. The efficiency gained by hybridizing PSO with SIAS gave higher accuracy than SAIS. In [10]-[12], the author compares four machine learning algorithms for categorizing spam. The four machine learning algorithms used are: TF-IDF (Term Frequency- Inverse Document Frequency), Naives Bayes, SVM (Support Vector Machine) and K-nearest Neighbor. These machine learning algorithms were applied to various parts of an email. It was also observed that the compared to other parts of email, the classification done using header was more accurate. Classification efficiency can be increased by combing the two methods: Naïve Bayes and TF-IDF. From the above survey we can presume that machine learning algorithms are more efficient for detection of spam messages.

In this paper the dataset is processed using a Non-Linear classifier by collecting known spam and good messages from newsgroup which is easily assessable and few profile details are taken and converted into Hex form.

The rest of the paper is organized as follows: Section 2 explains the methodology to detect the spam messages. Section 3 covers the experimental results of SVM. Conclusion and future scope has been discussed in section 4.

2. METHODOLOGY
There are three main components of the proposed system as described.

1. First is Mapping and Assembly,
2. Pre-Filtering and
3. Classification.

In mapping and assembly a standard model is specified for each object, which is defined by the framework. For example in our proposed system we have used two models: message model or profile model. In Pre-Filtering the incoming object is checked by comparing it with a blacklist. In classification a machine learning algorithm such as SVM (Support Vector Machine) for classifying the incoming object is used.

2.1 SVM (Support Vector Machine)

SVM uses statistical theory. It has many advantages. SVM is different from other methods which aim to minimize empirical risk such as neural network, nearest neighbors etc. SVM is based on structure risk; it overcomes the problem of local minimum and over fitting. SVM uses Kernel functions which reduce the complexity and computation.SVM shows good result in the classification of statistical data. SVM gives efficient classification for problems based on Linear Separation and Non- Linear Separation.

2.1.1 Linear SVM

In most of the situations data set can be linearly separable. For this we require a simple classifier,

$$s = \left\{\frac{x}{< w}, x > a + b = 0\right\}$$

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

The decision function is given as,

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

It is advisable to separate the training set with maximum margin as shown in the Fig 1(a). There are cases where training errors occur because of non-separable training set as illustrated in Fig 1(b). This problem is solved by Lagrange method. Let x_i be each training point, which is described by a multiplier of Lagrange $\alpha_{i.}$, If $\alpha_{i.} = 0$, then there is no influence on hyper plane by $x_{i.}$ If $\alpha_{i.} > 0$, then the points are near to hyper plane and are called as support vectors.

2.1.2 Non-Linear SVM

The data is mapped into a large feature space which contains even non-linear features. A hyper plane is then constructed in the space.

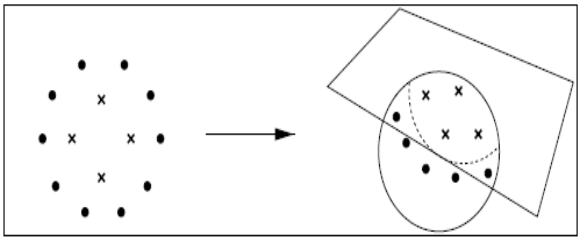
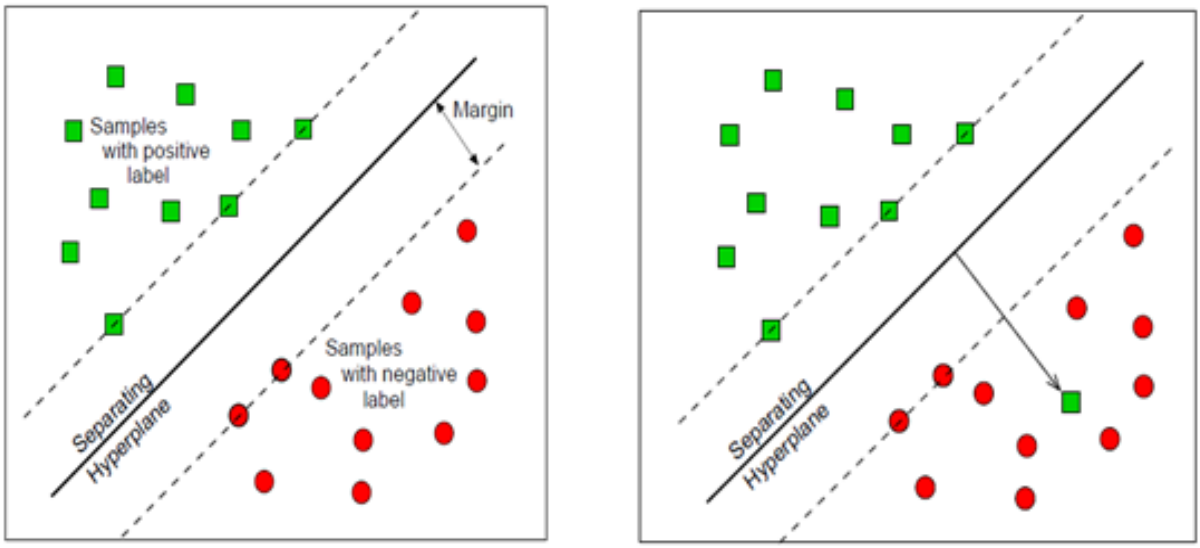


Fig 2: Non-Linear case of SVM



(a) **(b)**

Fig 1: (a) Linear Case of SVM and (b) Non-Separable Case.

2.1.3 SVM-Kernel

Kernel methods are used for pattern analysis and are enabled to work in high-dimensional feature space. In higher dimension, classification is simple where as in low dimensions classification is complex. To construct a maximum hyper plane margin, we have

$$< \emptyset(x_1), \emptyset(x_2) > H$$

Where φ is Low to High dimension feature map.

There are 3 types of kernels in SVM,

1. Linear Kernel: $<x_{1,} x_2> = K\,(x_{1,} x_2)$
2. Polynomial Kernel: $(\sqrt{} < x_{1,} x_2 > + C)^d = K\,(x1, x_2)$
3. Radial Basic Function (RBF): $\exp\,(- \sqrt{} \| x_{1,} x_2 \|^2) = K\,(x_{1,} x_2)$

The Kernel used in our proposed model is the most popular kernel function RBF. Around each data point a curve is added as shown in the figure 3.

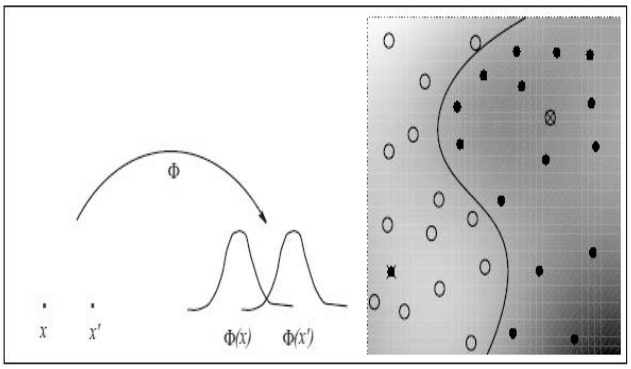


Fig 3: RBF Kernel

2.1.4 System Architecture

As seen in Fig 4 architecture diagram, 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. In the testing phase, the social network considered is Twitter. 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. According to classification result if the URL is classified as spam then the URL is added to the blacklist.

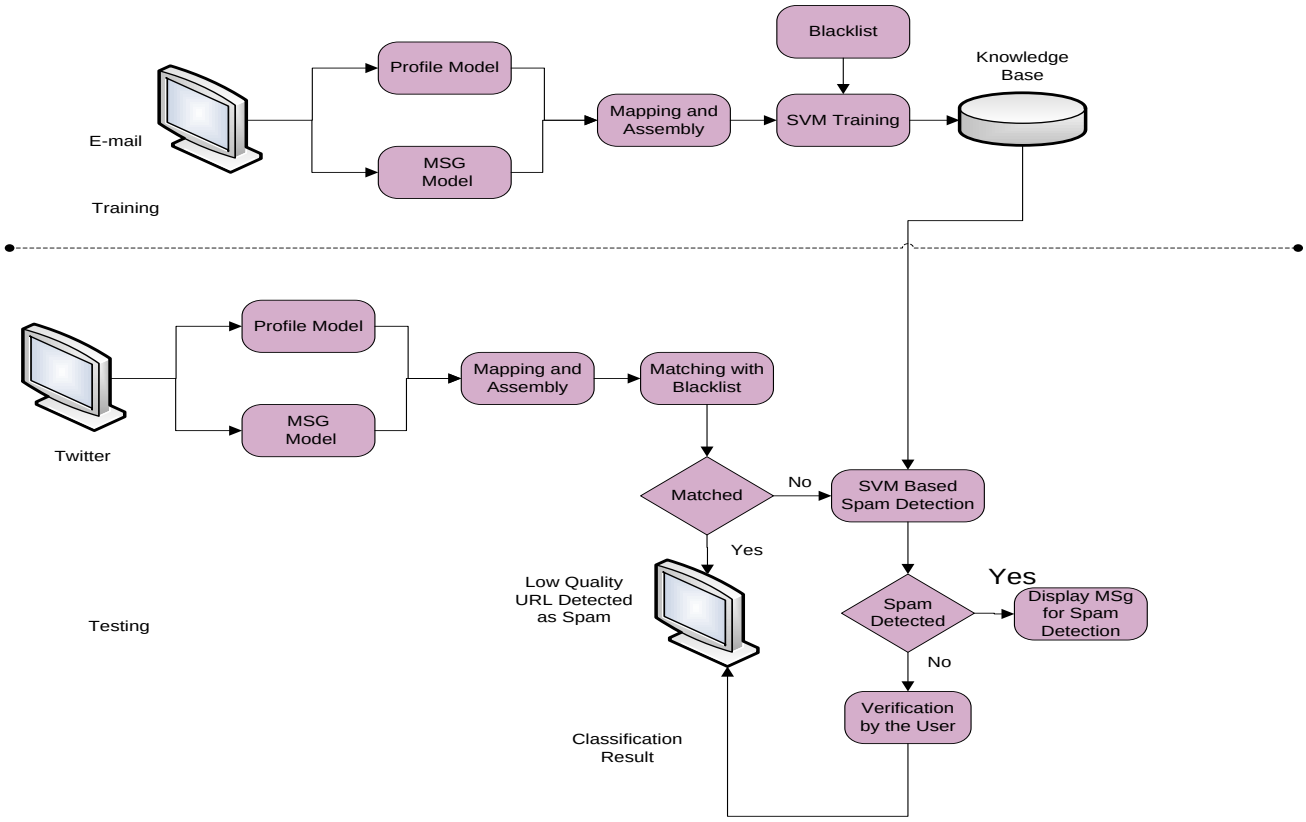


Fig 4: System Architecture

2.1.5 Implementation

The overall implementation is explained by using a flow diagram as depicted in Fig 5. It gets divided into two phases 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 associated with it. The data contains good as well as known spam messages. The information containing the profile details and message details is stored '.dat' file. This data is used for training the SVM. In the select query phase, the text file containing the profile details and message details are read through the path name. The flow diagram is shown in Fig 5.

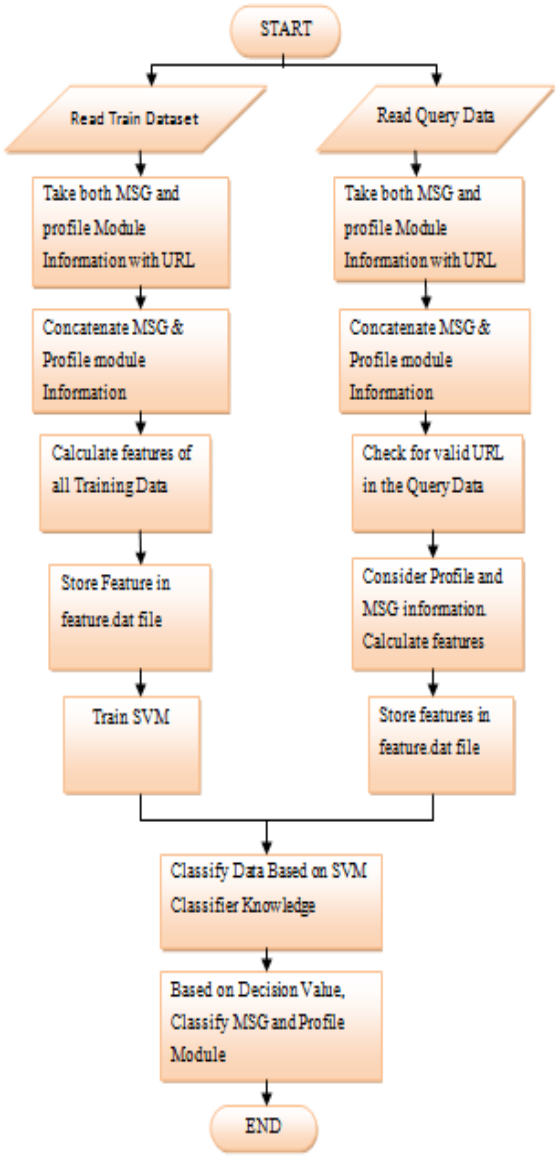


Fig 5: Overall Flow Diagram.

3. EXPERIMENTAL RESULTS

The above explained section 2 is implemented on a training dataset consisting of five legitimate messages and five spam message. The true rate and false rate for spam and good messages for the proposed system is calculated from equation (1) to equation (4). True rate is number of messages truly classified as spam message and good message. False rate is number of messages falsely classified as spam message and good message.

Spam Messages:

- True Rate = (No of spam messages truly classified / total no of messages) *100% (1) (4/5) *100% = 80%
- False Rate= (No of spam messages Falsely classified – True rate) *100% (2)(80-60) * 100% = 20%

Good Messages:

- True Rate = (No of good messages truly classified/ total no of messages) *100% (3)

(3/5) * 100% = 60%

- False Rate= (No of good messages Falsely Classified / total no of messages) * 100% (4)

(100- 60) = 40%.

Based on the true rate and false rate values of spam and good message, the following graph is generated.

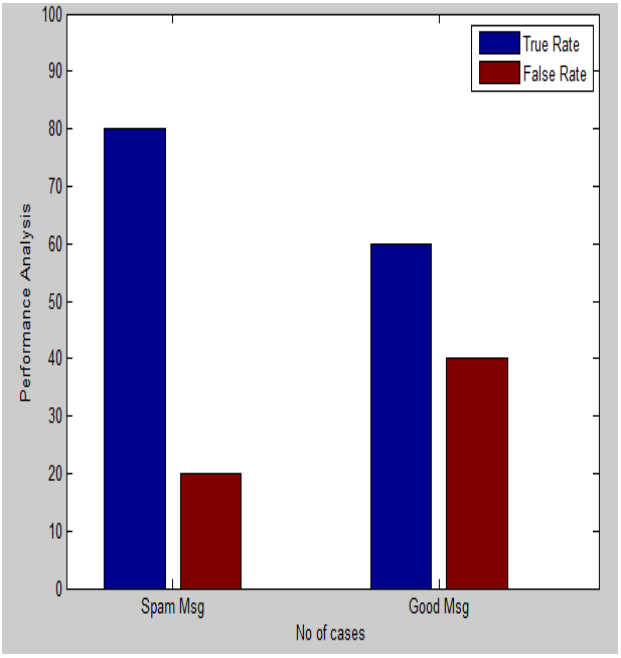


Fig 6: Graph showing True Rate and False Rate for spam and good messages.

4. CONCLUSION AND FUTURE WORK

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, SVM, 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. Hence SVM is trained in such a manner in this research work, that it will classify the testing data considering both the profile model and message model. Future work involves to implement a new SVM Kernel which has enlarged dataset for classifying messages which have non-english words and spam messages which are encrypted.

5. REFERENCES

[1] Agarwal S, Jain. K "Hybrid Approach For Spam Detection using Support Vector Machine and Artificial Immune System", First International Conference on Network and Soft Computing", Aug 2014, pg no: 05-09.

[2] Selamat, Mohammed .M, " An Evaluation on Efficiency of Hybrid Features for Spam Email Classification",2015 International Conference on Computer Communication and Control Technology ,April 2015, pg no : 227-231

[3] "A Hybrid Approach for Spam Filtering using Local Concentration and K- means Clustering", 2014, 5th International Conference, pg no: 194- 199.

[4] Salehi, Solmat. A "Hybrid Simple Artificial Immune System and Particle Swam Detection", "5th Malaysian

Conference In Software Engineering", Aug 2011, pg no: 124-129.

[5] Chin-Lai, Ming-Chin Tsai, "An Emperical Performance Comparision for Spam Categorization", 4th International Conference on Hybrid Intelligent Systems, Dec 2014, pg no : 44-48.

[6] Fabrico Benevenuto, Gabriel Magno, Tiago Rodrigues and Virgilis Almeida, " Detecting Spammers on Twitter", CEAS Seventh Annual Collaboration, Electronic Messaging, Anti Abuse and Spam Conference, July 2010.

[7] M. Mc Cord, M. Chuah, " Spam Detection On Twitter Using Traditional Classifiers", ATC,Banff Canada, Sept 2011, pg no: 2-4 IEEE

[8] Ayon Chakraborty, Jyotinmoy Sundi, Som Satapathy, "SPAM: A Framework For Social Profile Abuse Monitoring".

[9] Saber Salehi, Ali Selmat, "Enhanced Genetic Algorithm for Spam Detection in Email", IEEE 2nd International Conference on Software Engineering and Service Science, July 2011, pg no: 594-597

[10] Congfu Xu, Baojun Su and Yunbiao Cheng, "An Adaptive Fusion Algorithm For Spam Detection", IEEE Intelligent System, vol : 29, no: 4, July-Aug 2014, pg no:2-8

[11] M. Andreolini, A. Bulgarelli, M. Colajanni and F. Mazzoni, " HoneySpam: Honeypots fighting spam at the source", in Proc. USENIX Step to Redoducing Unwanted Traffic on the Internet Workshop, Cambridge, March 2005.

[12] C. Tseng, J.Hvang and M.Chen, " ProMail: Using Progressive Email Social Network for Spam Detection", Advances in Knowledge Discovery And Data Mining, LNCS vol: 4426, pg no: 833-840, 2007