IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

Spam Classification Using Deep Learning Technique

A.B.Singh[1*], S.B.Singh[2], Kh.M.Singh[3]

[1*]Dept. of Computer Science and Engineering, National Institute of Technology, Manipur, India
[2] Dept. of Computer Science and Engineering, National Institute of Technology, Manipur, India
[3] Dept. of Computer Science and Engineering, National Institute of Technology, Manipur, India

[*]*Corresponding Author: angom102@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 18/May/2018, Published: 31/May/2018

***Abstract*—** Deep Learning technique which is a new area of Machine Learning is showing huge promise in achieving the original goals of Machine learning: Artificial Intelligence. Deep Learning is being applied in every machine learning problem and has shown great results. In this paper, we evaluate the problem of spam classification using Deep Learning Technique and compare the result with other state-of-art machine learning techniques. The machine learning techniques used in the comparison are: Random Forest, Multinomial Naïve Bayesian and Support Vector Machine. The dataset used in the experiment is the CSDMC_2010 and Enron dataset and the platform used is the WEKA interface. Common features are extracted from the body of the spam and feature vector table is constructed, which is used on all the model. Our experiment shows that Deep Learning model outperform all the other machine learning techniques in terms of true positive & true negative and even in the overall accuracy.

***Keywords*—**Spam, Deep Learning, Machine Learning, Classify, WEKA

I. INTRODUCTION

Spam or unsolicited emails has grown with the growth of Internet technology and with the exponential increase in the volume of spam sent daily and the ever increasing notoriety of spam from being a nuisance to a cyber threat, classification of spam is becoming an important research issue. Advancement in the area of anti-spam techniques has led to many solution to counter spam. However, a complete solution to eradicate spam is still not yet available.

The two primary method of filtering spam emails are content-based and blacklist-based. The first approach considers several factors based on the content of the email such as spammy words, number of words and URLs in page title and body along with their average length, compressibility, n-gram likelihoods etc. during spam detection. On the other hand, in blacklist-based approach, blacklist of well-known spamming host are first constructed and used for filtering purpose. Popular Spam filters uses both the approaches to improve the accuracy of the filtering process.

Spam filtering which is a form of binary classification process, where emails are either classified as legitimate or spam. Some of the common Machine Learning techniques [1][2] used in designing spam filtering are Naïve Bayes [3], Support Vector Machine [4]. Decision Trees [5] etc. These algorithms does not rely on hand coded fixed rules for filtering purpose, instead they have the ability to improve their performance through experience. These Machine Learning Algorithms are capable of extracting knowledge from a set of message supplied and used the same for the classification of newly received messages.

Deep Learning is an area of Machine Learning which is continuously evolved to mimic the function of the Human brain and is broadly classified under artificial neural network. In recent time deep learning techniques has become very popular, primarily because they are delivering on their promise and has shown great results across a suite of very challenging artificial intelligence problem from computer vision and natural language processing.

In this research, we use Deep Learning technique in the classification of spam. The Deep Learning model is constructed using WEKA interface. The evaluation of the model is done using the publicly available CSDMC_2010 and Enron spam dataset. We find that like in other area, Deep Learning technique shows great result in the classification of spam, outperforming other state-of-art machine learning techniques.

The rest of the paper is organized as follows: Section II contain the related work on spam classification, Section III contain the proposed architecture of spam classification,

followed by results & discussion in Section IV, and Section V concludes research work with future directions.

II. RELATED WORK

Many works in the area of spam filtering has been done and various solution to the spam problem has been developed. Some of the solution are formulated based on the information available in the message header while other are based on the body of the email. While some of the solution use information from both the header as well as the body of the message.

Message header contains reliable information such as the sender address, various Date & Time, Subject, MIME information, etc., from which useful features can be extracted that can be used in spam filtering. Zhang et al. [6], proposed spam filter which are based on information extracted from header only. Their spam filter could achieved result which are better or comparable with the results obtained using information extracted from the body of the email.

The body of an email also provide various features that can be used in spam filtering. Common features extracted from the body are the bag-of-word which is nothing but the collection of words in the message, URLs information, structure and layout of the message etc. a simple content-based heuristic spam filter analyzes the features extracted from the body of the message and uses the pattern observed in the extracted information to see the occurrence of 'Spammy' words like 'Win', 'Viagra', or 'Free, and use this information to either classify the message as spam or ham. Almedia et al [7], shown in their experiment that SVM acquired the best average performance among the different statistical classifier based on Naïve Bayes Probability and Linear Support Vector Machine filter used.

M. Sahami et al [3], proposed the use of probabilistic learning methods based on Bayes Theorem in conjunction with a notion of differential misclassification cost to produce effective spam filters. Lin Li et al [8], proposed to improve the Naïve Bayesian spam filter by using improved IDF weighing algorithm of the TF-IDF feature selection. These reduces the feature dimension and also increase the weight of the high frequency words corresponding to its class.

Druker et al [4], proposed to use Support Vector Machine for spam categorization which was found to poses high speed in the classification process and also remarkably intolerant of the relative size of the number of training example of the two classes. Amayri et al [9], proposed to use various string kernels for spam filtering and feature mapping variants in text classification. This increases the overall performance of the standard SVM in the spam filtering task.

Koprinska et al [10], proposed the use of random forests, to e-mail categorization and spam filtering. It was found that Random Forest produce the best overall results, with naïve Bayes performing the worst.

III. METHODOLOGY

We tested the classification using publicly available ENRON dataset and CSDMC_2010 Spam dataset. The proposed architecture of the classification process is given below in Figure-1.

Initially, we perform the pre-processing of the dataset to remove the less informative and noisy terms in the messages. Pre-processing process, apart from transforming the dataset in to a uniform format, the feature space is also reduce thereby enhancing the overall classification performance.

Next, features which are likely to improve the classification process are extracted. These consist of Lexical analysis (Tokenization), Stop-word removal, Stemming and final representation as feature vector. The feature vector table thus constructed are divided into training set and test set.

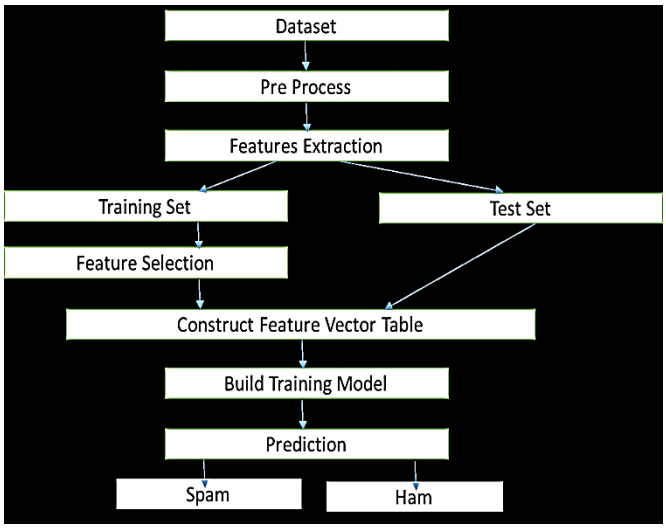

Figure 1: Spam Classification Process

One of the parameter commonly used in the evaluation of classification algorithm are the False Positive Rate (FPR) and False Negative Rate (FNR). The False Positive Rate is defined as the rate of the legitimate emails that are wrongly classified as spam and is given as:

$$FPR = \frac{\#\text{of FalsePositives}}{\#\text{of FalsePositives} + \#\text{of TrueNegative}} \quad (1)$$

On the other hand, False Negative Rate is defined as the rate of spam messages that were classified as legitimate and is given as:

$$FNR = \frac{\#\text{of FalseNegatives}}{\#\text{of TruePositives} + \#\text{of FalseNegative}} \quad (2)$$

Both the values of FPR and FNR should be low for the classifier to be effective. Another way of representing the effectiveness of the classifier is to plot a curve based on the values of FPR and FNR. The resulting curve is also known as an ROC (Receiver Operating Characteristics) curve. For measuring the performance metric of the various Machine Learning Algorithm used in our experiment, we adopt the ROC based analysis and concluded that a spam filter whose ROC curve strictly lies above that of another or in other word the filter with the largest Area under the Curve (AUC) is the better filter in all deployment scenarios [10].

To measure the effectiveness and quality of the spam filter, we also use two measures known as 'Recall' and 'Precision'. Spam Recall (R_s) and Spam Precision (P_s) are defined by the equations:

$$R_s = \frac{|S \rightarrow S|}{|S \rightarrow S| + |S \rightarrow L|} \quad (3)$$

And,

$$P_s = \frac{|S \rightarrow S|}{|S \rightarrow S| + |L \rightarrow S|} \quad (4)$$

Where, $|S \rightarrow S|$ signifies the number of spam classified as spam (True Positive) and $|S \rightarrow L|$ signifies the number of spam misclassified as ham (False Negative) and $|L \rightarrow S|$ signifies the number of legitimate messages misclassified as spam (False Positive).

Another combining measure known as F-measure or F-score combines both Precision (P_s) and Recall (R_s) metric in one equation to give the weighted harmonic mean of both.

$$F - measure = \frac{2.\text{Precision.Recall}}{\text{Precision+Recall}} \quad (5)$$

IV. RESULTS AND DISCUSSION

The Machine Learning models were tested on two public spam corpus, CSDMC_2010 and Enron spam dataset. Feature vector is constructed from the header of the message and also from the body. The experiment is conducted using machine learning platform WEKA 3.0. The classification algorithm selected were Random Forest, Multinomial Naïve Bayesian and Support Vector Machine. This Machine Learning algorithms were compared with the Deep Learning Model built using WEKA interface. The testing of the result was done using 10-fold cross validation on the test dataset.

The result of the analysis on the CSDMC_2010 spam data set is given in Table 1 while the result of the analysis of the Enron Spam dataset is given in Table 2. The ROC curve of analysis of CSDMC_2010 and Enron Dataset is given in Figure 1 and Figure 2 respectively.

Table 1: Result of analysis (CSDMC_2010 Spam Dataset)

Sl No	Model	Precision	Recall	F1_score
0	Deep Learning	0.986895	0.979464	0.983039
1	Random Forest	0.905645	0.872994	0.886297
2	Multinomial NB	0.962239	0.927225	0.941818
3	SVM	0.907457	0.856245	0.875000

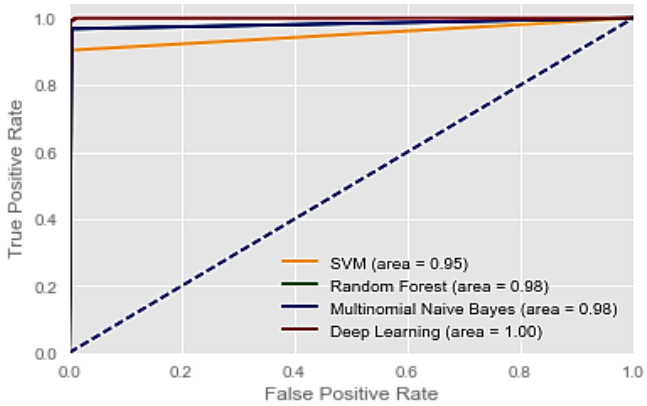

Figure 2: ROC curve (CSDMC_2010 Spam Dataset)

Table 2: Result of Analysis (Enron Spam Dataset)

Sl No	Model	Precision	Recall	F1_score
0	Deep Learning	0.995690	0.994118	0.994877
1	Random Forest	0.886202	0.882992	0.884449
2	Multinomial NB	0.962121	0.968031	0.964444
3	SVM	0.951965	0.960102	0.954449

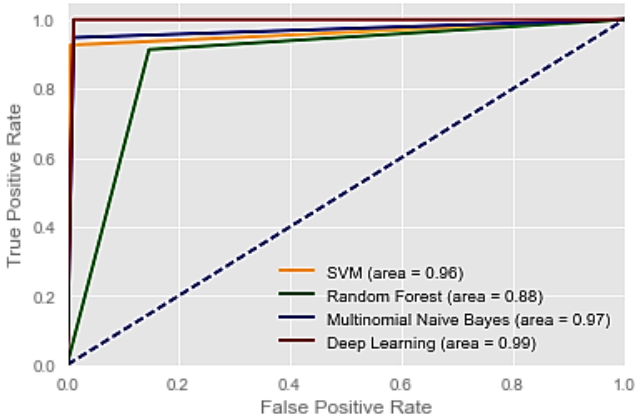

Figure 3: ROC curve (Enron Spam Dataset)

The area under the curve (AUC) for Deep Learning Model is the largest when compare to other Machine Learning model.

This indicates that Deep Learning Model performs better in the spam classification both in terms of accuracy and quality.

V. CONCLUSION and Future Scope

In this paper, we built a Deep Learning model for spam classification and compared the result with other Machine Learning Algorithms. Our empirical performance shows that Deep Learning outperformed all the other Machine Learning Model and an accuracy of more than 99% was achieved.

An important aspect of the experiment was the use of multiple features in the classification process and these set of features were selected at the beginning of the process. For updating the features which are no longer informative or when new features are required to be added, the whole model need to be re-built. An area of future research would be to design method that could allow incremental addition or removal of features, without re-building the entire model.

ACKNOWLEDGMENT

This work was supported by grants from MeitY, Govt. of India. (AA No.12(3) 2016 ESD).

REFERENCES

[1] Thiago S. Guzella and Walmir M. Caminhas, "*A review of machine learning approaches to spam filtering*", Expert System with Applications", Elsevier, Vol-36, pp 10206-10222, 2009.

[2] G. Cormack, "*Email spam filtering: A systematic Review*", Foundations and Trends in Information Retrieval , Vol-1, no. 4, pp. 335–455, 2008.

[3] M. Sahami, S. Dumais, D. Heckerman and E. Horvitz, "*A Bayesian Approach to Filtering Junk Email*," AAAI Technical Report WS-98-05, AAAI Workshop on Learning for Text Categorization, 1998.

[4] Drucker H, Wu D, Vapnik VN. "*Support Vector Machines for Spam Categorization*", IEEE Transactions on Neural Networks Vol-10, Issue-5, pp 1048-1054, 1999.

[5] Yudong Zhang, Shuihua Wang, Preetha Phillips, Genlin Ji, "*Binary PSO with mutation operator for feature selection using decision tree applied to spam detection*", knowledge-Based Systems, Elsevier, Vol-64, pp 22-31, 2014.

[6] Zhang L, Zhu J, Yao T, "*An Evaluation of Statistical Spam Filtering Techniques Spam Filtering as Text Categorization*", ACM Transactions on Asian Language Information Processing (TALIP), Vol-3, Issue 4, pp 243-269, 2004.

[7] Almeida TA, Yamakami A, "*Content-Based Spam Filtering*", The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, pp 1-7, 2010.

[8] Lin Li and Chi Li, "*Research and Improvement of a Spam Filter based on Naïve Bayes*", Proceedings of the 2015 Seventh International Conference on Intelligent Human-Machine Systems and Cybernetics, 2015

[9] Amayri O, Bouguila N, "*A study of Spam Filtering using Support Vector Machines*", Artificial Intelligence Review, Vol-34, Isuue 1, pp 73-108, 2010.

[10] Koprinska I, Poon J, Clark J, Chan J, "*Learning to Classify e-mail*", Information Sciences, Vol-177, issue 10, pp 2167-2187, 2007.

Authors Profile

Mr.A.B.Singh is a research scholar in the Department of Computer Science and Engineering at the National Institute of Technology, Manipur. He received his B.E in Computer Science and Engineering from Annamalai university and M.Tech in Computer Science and Engineering from National Institute of Technology, Agartala. His research area include IT security, Cyber Forensics and Digital Investigation.

Dr. S.Birendra Singh, Professor, Department of Computer Science & Engineering, National Institute of Technology, Manipur. His research area include IT security and Image Processing.

Dr. Kh. M. Singh is an Associate Professor in the Department of Computer Science and Engineering at National Institute of Technology, Manipur. His research area include Cryptography, IT security, Image Processing and Digital Investigation.

