IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SPAM DETECTION APPROACH FOR SECURE SMS USING MACHINE LEARNING ALGORITHMS

[1]Dr.Nancy Jasmine Goldena, [2]T.Ancy Selciya

ABSTRACT

Short Message Service (SMS) has become one of the most important ways of communication in today's world, where digitization is ubiquitous. SMS does not require active participation, unlike other chatting-based messaging systems such as Facebook, WhatsApp and others. As we all know, hackers and spammers try to break into mobile computing devices, and SMS support for mobile devices has become vulnerable, as attackers try to break into the system by sending unwanted links into , the attacker can gain remote control of the mobile computing device by clicking those links. So, in order to identify those communications, this work created a system that will detect malicious messages and determine whether they are SPAM or HAM (malicious or not malicious) using deep learning VGG16 algorithm.

Keywords--Short Message Service, Spam, Machine Learning, Deep Learning.

INTRODUCTION

The spam detection is a big issue in mobile message communication due to which mobile message communication is insecure. In order to tackle this problem, an accurate and precise method is needed to detect the spam in mobile message communication.The growth of the mobile phone users has led to a dramatic increase in SMS spam messages. Though in most parts of the world, mobile messaging channel is currently regarded as “clean” and trusted, on the contrast recent reports clearly indicate that the volume of mobile phone spam is dramatically increasing year by year. It is an evolving setback especially in the Middle East and Asia. SMS spam filtering is a comparatively recent errand to deal such a problem. It inherits many concerns and quick fixes from Email spam filtering. However it fronts its own certain issues and problems. This paper inspires to work on the task of filtering mobile messages as Ham or Spam for the Indian Users by adding Indian messages to the worldwide available SMS dataset. The paper analyses different machine learning classifiers on large corpus of SMS messages for Indian people.

Literature Survey:

The daily traffic of SMS keeps increasing. As a result, it leads to dramatic increase in mobile attacks such as spammers who plague the service with spam messages sent to the groups of recipients. Mobile spams are a growing problem as the number of spams keep increasing day by day even with the filtering systems. Spam's are defined as unsolicited bulk messages in various forms such as unwanted advertisements, credit opportunities or fake lottery winner notifications. Spam classification has become more challenging due to complexities of the messages imposed by spammers. Hence, various methods have been developed in order to filter spam's. In this study, methods of Term Frequency-Inverse Document Frequency (TF-IDF) and Random Forest Algorithm will be applied on SMS spam message data collection. Based on the experiment, Random Forest algorithm outperforms other algorithms with an accuracy of 97.50%.

The spam detection is a big issue in mobile message communication due to which mobile message communication is insecure. In order to tackle this problem, an accurate and precise method is needed to detect the spam in mobile message communication. We proposed the applications of the machine learning-based spam detection method for accurate detection. In this technique, machine learning classifiers such as Logistic Regression (LR), K-Nearest Neighbor (K-NN) and Decision Tree (DT) are used for classification of ham and spam messages in mobile device communication. The SMS spam collection data set is used for testing the method. The dataset is split into two categories for training and testing the research. The results of the experiments demonstrated that the classification performance of LR is high as compared with K-NN and DT, and the LR achieved a high accuracy of 99%. Additionally, the proposed method performance is good as compared with the existing state-of-the-art methods.

Many machine learning methods have been applied for Short Messaging Service (SMS) Spam Detection, including traditional methods such as Naïve Bayes (NB), Vector Space Model (VSM), and Support Vector Machine (SVM), and novel methods such as long Short-Term Memory (LSTM) and the Convolutional Neural Network (CNN). These methods are based on the well-known bag of words (BoW) model, which assumes documents are unordered collection of words. This assumption overlooks an important piece of information, i.e., word order. Moreover, the term frequency, which counts the number of occurrences of each word in SMS, is unable to distinguish the importance of words, due to the length limitation of SMS. This paper proposes a new method based on the discrete Hidden Markov Model (HMM) to use the word order information and to solve the low term frequency issue in SMS spam detection. The popularly adopted SMS spam dataset from the (University of California, Irvine) UCI machine learning repository is used for performance analysis of the proposed HMM method. The overall performance is compatible with deep learning by employing CNN and LSTM models. A Chinese SMS spam dataset with 2000 messages is used for further performance evaluation. Experiments show that the proposed HMM method is not language-sensitive and can identify spam with high accuracy on both datasets.

SMS spam is a contemporary issue fundamentally because of the accessibility of very modest mass SMS bundles and the way that SMS induces higher reaction rates as it's far a depended on and personal service. In this paper, we will be differentiating the messages into two categories: Ham and Spam. Ham is described as the dataset that includes the textual content of SMS messages at the side of the label indicating whether the records is legitimate message or now not. Spam is defined as the dataset that includes the textual content of SMS messages along with the label indicating the junk messages. In SMS Spam messages, the advertisers utilize the SMS text messages to target the customers with unwanted advertising. But it is troublesome, because the users pay a fee per SMS received. To overcome this, we perform a comparison between the machine learning algorithms to predict the messages and calculate the accuracy criterion by using the SMS spam dataset.

Spam is well defined as the unsolicited bulk messages or junk mail will send to email address or phone number that are generally marketable in nature and also carry malicious documents. The main issue of spam is that it can download malicious files which can attack the computers, smart phones and networks, utilize network bandwidth and storage space, degrades email servers and can cause attacks in our devices like spyware, phishing and ransomware. In the existing approach, an exploratory analysis of supervised machine learning algorithms has done and the performance has been evaluated. The drawback of existing approach is that the performance of supervised machine learning algorithms decreases as we increase the size of the dataset. In order to overcome such drawbacks, an efficient spam detection using recurrent neural networks using the BiGRU model has been proposed. By implementing this, it has been achieved with better accuracy of 99.07%. From this, it is concluded that BiGRU model has better performance than existing approaches.

In recent years, there has been considerable interest among people to use SMS as one of the essential and straightforward communications services on mobile devices. The increased popularity of this service also increased the number of mobile devices attacks such as SMS spam messages. SMS spam messages constitute a real problem to mobile subscribers; this worries telecommunication service providers as it disturbs their customers and causes them to lose business. Therefore, in this paper, we proposed a novel machine learning method for detection of SMS spam messages. The proposed model contains two main stages: feature extraction and decision making. In the first stage, we have extracted relevant features from the dataset based on the characteristics of spam and legitimate messages to reduce the complexity and improve performance of the model. Then, an averaged neural network model was applied on extracted features to classify messages into either spam or legitimate classes. The method is evaluated in terms of accuracy and F-measure metrics on a real-world SMS dataset with over 5000 messages. Moreover, the achieved results were compared against three recently published works. Our results show that the proposed approach achieved successfully high detection rates in terms of F-measure and classification accuracy, compared with other considered researches.

Short Message Service (SMS) is the most important communication tool in recent decades. With the increased popularity of mobile devices, the usage rate of SMS will increase more and more in years. SMS is a practical method used to reach individuals directly. But this practical and easy method can cause SMS to be misused. The advertising or promotional SMS of the companies are an examples of this misuse. In this study, a spam SMS detection technique is proposed using SVM. SMSSpamCollection dataset, which is contain 747 spam SMS and 4827 ham SMS, is used. 10 fold cross-validation technique is used to evaluate prediction of Spam SMS in the dataset. Therefore, proposed approach achieved 98.33 % true positive rate and 0,087 false positive rate for SVM classification algorithm.

METHODOLOGY

SMS spam detection phases

The phases of detection of spam involve preprocessing, feature extraction and selection and classification.

Preprocessing

Pre-processing is the first stage in which the unstructured data is converted into more structured data. Since keywords in SMS text messages are prone to be replaced by symbols. In this study, the stop word list remover for English language have been applied to eliminate the stop words in the SMS text messages. Fig. 2(a) shows the frequencies of words in SMS messages, while Fig. 2(b) shows the most frequent words used in spam text messages are from pronoun (e.g. to) and proposition (e.g. your) groups. Similarly, the top words in ham text messages are occupied by either pronoun, proposition among many other types of stop words.

Feature extraction and selection

Feature extraction and selection is important for the discrimination of ham and spam in SMS text messages. For this phases TFIDF will be used. TFIDF is the often-weighting method used to in the Vector Space Model, particularly in IR domain including text mining. It is a statistical method to measure the important of a word in the document to the whole corpus. The term frequency is simply calculated in proportion to the number of occurrences a word appears in the document and usually normalized in positive quadrant between 0 and 1 to eliminate bias towards lengthy documents [10]. To construct the index of terms in TFIDF, punctuation is removed, and all text are lowercase during tokenization. The first two letter TF or term frequency refers to how important if it occurs more frequently in a document. Therefore, the higher TF reflects to the more estimated that the term is significant in respective documents. Additionally, IDF or Inverse Document Frequency calculated on how infrequent a word or term is in the documents [10]. The weighted value is estimated using the whole training dataset. The idea of IDF is that a word is not considered to be good candidate to represent the document if it is occurring frequently in the whole dataset as it might be the stop words or common words that is generic. Hence only infrequent words in contrast of the entire dataset is

relevant for that documents. TF-IDF does not only assess the importance of words in the documents but it also evaluates the importance of words in document database or corpus. In this sense, the word frequency in the document will increase the weight of words proportionally but will then be offset by corpus's word frequency [21]. This key characteristic of TF-IDF assumes that there are several words that appear more often compared to others in the document in general. Hence, the relevancy of a word to a document is shown in Eq. 1.

$$F - IDF = \frac{Ter \quad Frequency}{Document\ Frequency}$$

SMS message spam classification

Deep Learning VGG16 algorithm will used for classification of ham or spam during this phase. RF is averaging ensemble learning method that can be used for classification problem. This algorithm combines various decision tree models in order to eliminate the overfitting problem in decision trees. As a result, each tree gives different performances, in which the average of their performances will be generalized and calculated. During the training phase, a set of decision trees will be constructed before they can operate on randomly selected features. Regardless, RF can work well with a large dataset with a variety of feature types, similar to binary, categorical and numerical. The algorithm works as follows (see Fig. 3): for each tree in the forest, a bootstrap sample is selected from S where S (i) represents the ith bootstrap. A decision-tree is then learn using a modified decision-tree learning algorithm. The algorithm is modified as follows: at each node of the tree, instead of examining all possible feature-splits, some subset of the features text $f \subseteq F$ is selected randomly. where F is the set of Spam features. The node then splits on the best feature in f rather than F. In practice f is much, much smaller than F. Deciding on which feature to split is oftentimes the most computationally expensive aspect of decision tree learning. By narrowing the set of features, the speed up the learning of the tree is increase drastically.

RESULT ANALYSIS

Dataset

The public dataset of SMS labelled messages is obtained from UCI Machine Learning Repository. The data is originally collected by Almeida and Hidalgo [20]. It contains 5,574 English raw text messages with tag labels either as legitimate (ham) or spam. The text messages have been collected and derived from various sources such as UK forum from Grumbletext website, NUS SMS Corpus (NSC), SMS Spam Corpus v.0.1 Big. Table 1 shows the source of the dataset

Spam Dataset	Total
Grumbletext website	425(spam)
NUS SMS Corpus (NSC)	3,375 (ham)
Caroline Tag's PhD Thesis	450 (ham)
Corpus v.0.1 Big	1,002 (ham) 322 (spam)

From Table 1 above, this study finds that there are only 5,574 labelled messages in the dataset, with 4827 of messages belong to ham messages while the other 747 messages belong to spam messages. Nonetheless, this dataset consists of two named columns starting with the message labels (ham or spam) followed by strings of text messages and three unnamed columns.

CONCLUSION

The SMS spam message problem is plaguing almost every country and keeps increasing without a sign of slowing down as the number of mobile users increase in addition to cheap rates of SMS services. Therefore, this paper presents the spam filtering technique using various machine learning algorithms. Based on the experiment, TF-IDF with Random Forest classification algorithm outperforms good compare to other algorithms in terms of accuracy percentage. However, it is not enough to evaluate the performance based on the accuracy alone since the dataset is imbalanced; therefore, the precision, recall and f-measure of the algorithms must also be observed. After some examinations, RF algorithm still manages to provide good precision and f-measure with 0.98 of precision while 0.97 for f-measure. Different algorithms will provide different performances and results based on the features used. For future works, adding more features such as message lengths might help the classifiers to train data better and give better performance.

REFERENCES

1) Nilam Nur Amir Sjarif, Nurulhuda Firdaus Mohd Azmi , Suriayati Chuprat ,Suriayati Chuprat, "SMS Spam Message Detection using Term Frequency-Inverse Document Frequency and Random Forest Algorithm" January 2019Procedia Computer Science 161:509-515 DOI:10.1016/j.procs.2019.11.150

2) Luo GuangJun, Shah Nazir, Habib Ullah Khan, Amin Ul Haq, "Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms" July 2020Security and Communication Networks 2020(11):1-6 DOI:10.1155/2020/8873639

3) Tian Xia, Xuemin Chen, "A Discrete Hidden Markov Model for SMS Spam Detection" July 2020Applied Sciences 10(14):5011 DOI:10.3390/app10145011

4) Gomatham Sai Sravya, G. Pradeepini, Vaddeswaram, Guntur, "Mobile Sms Spam Filter Techniques Using Machine Learning Techniques" Published 25 March 2020 Computer Science International Journal of Scientific & Technology Research.

5) Sridevi Gadde, A. Lakshmanarao, S. Satyanarayana, "SMS Spam Detection using Machine Learning and Deep Learning Techniques March 2021 DOI:10.1109/ICACCS51430.2021.9441783 Conference: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).

6) Karishma Regina, Akila Gopu, V. Govindasamy,"SMS Spam Detection using RNN" May 2020 DOI:10.35291/2454-9150.2020.0305.

7) Saeid Sheikhi, Mohammad Taghi Kheirabadi, Amin Bazzazi, "An Effective Model for SMS Spam Detection Using Content-based Features and Averaged Neural Network" February 2020International Journal of Engineering, Transactions B: Applications 33(2):221-228 DOI:10.5829/ije.2020.33.02b.06.

8) Adem Tekerek, "Support Vector Machine Based Spam SMS Detection" July 2019Journal of Polytechnic 22(2):779-784 DOI:10.2339/politeknik.429707.