SMS SPAMING DETECTION USING NLP TECHNIQUES

KARUTURI CHANDINI[1], Y.SRINIVASA RAJU [2]

[1]MCA Student, B V Raju College, Kovvada, Andhra Pradesh, India.

[2]Assistant Professor, B V Raju College, Kovvada, Andhra Pradesh, India.

ABSTRACT:

Short Message Service (SMS) Spam is one form of mobile device attack that can affect mobile user's security and privacy. This is because such attack applies social engineering method to trick the user for information gathering. This study proposed an SMS Spam detection framework specifically for Malay language by using Naïve Bayes. There are several solutions to detect SMS Spam, but machine learning is one of the most effective technique to detect spam attack. In addition, the existing detection framework using machine learning technique is not effective for Malay language SMS. This is because the features used are not based on Malay language to detect the SMS content as Spam or not Spam. This framework consists of several processes such as Data Collection, Pre-processing, three types of Features Selection, Classification and Detection. Based on the result, it shows that the classification derives acceptable accuracy which is over 90%.

Keywords: SMS, Naive bayes, classification, data collection.

1. INTRODUCTION

Short Message Services (SMS) is one of the alternatives as communication medium by mobile phone. However, SMS can be used to fraud users [1]. There is anti-Spam available to be installed on mobile devices for protection, but it is still lacking to detect SMS Spam in Malay language. Malay language is the main language in Malaysia and it is used in formal and informal communication throughout Malaysia. The SMS Spam operates by sending SMS to users randomly. The message contains unwanted content such as business promotion or web link. Usually, each SMS sent is charged to the user although unsolicited, and the user needs to prompt a reply to stop the SMS. Even the charge to stop the SMS is borne by the user. This

indicates that the security and the privacy of the user‟s mobile device has been violated. There are several detection techniques that have been applied in SMS Spam detection studies such as Content-Based Filtering, White list or Blacklist, Machine Learning, Matching Pattern and Artificial Immune System. This paper developed a detection framework for SMS Spam for Malay language which consists of Data Collection for Malay language SMS Spam, Pre-processing, Features Selection based on Malay language, Classification and Detection. In this paper, several experiments have been done to analyze the proposed technique and framework. Moreover, the results in each process of the framework had been validated through Naive Bayes classification technique. By developing this framework, it will help to provide a Malay language SMS Spam feature for future work in spam detection since the existing studies are only focused in English SMS Spam features and less on the Malay language SMS features.

Text messages are sent from one mobile device to another through SMS(Short Message Service). Spams are junk text messages or unsolicited messages [1]. Nowadays, people use their mobile phones not only for making and receiving audio calls but also for various other purposes like banking (like transferring money, checking balance etc.), sending and receiving e-mails, accessing Facebook, online shopping etc. which require their confidential information like password, PIN, bank account number, credit card or debit card number etc.. Apart from this people also keep their personal information in their mobile phones like phone number of their friends and relatives, photos, images of their IDs and other important documents. People can be victim of cyber attack through spam SMS and the information stored in their phone can be leaked. Mobile users are disturbed by spam SMS and may be frustrated [2]. Spam SMS wastes network bandwidth and cause loss of productivity [3]. National Customer Preference Registry (NCPR) was set up by Government of India, junk calls have been reduced to some extent by it but spam SMS are not filtered by it [1]. Text classification techniques are widely used for spam filtering [4]. In text classification a category (from a set of predefined categories) is assigned to a

document [5]. SMS are text messages and our goal is to classify a SMS as spam or genuine. In text classification, supervised machine learning approach is used. Therefore, in text classification already labeled data set is required for constructing a classifier. In our data set genuine messages were labeled as ham. Many issues of SMS spam detection are inherited from email spam detection [6]. Since there is similarity in email spam filtration and SMS spam filtration, the techniques used for spam email filtration can be used for spam SMS filtration [7]. Naïve Bayes classifier is a popular method for spam email filtration [8]. In this paper we have used Naïve Bayes algorithm for spam SMS detection.

2. AN OVERVIEW OF PROPOSED SYSTEM

1. Performance Analysis of E-Mail Spam Classification using different Machine Learning Techniques

Authors: V. Sri Vinitha, D. Karthika Renuka.[2020]

Most of the business and general communication is done through email because of its cost effectiveness. This efficiency leads email exposed to various attacks including spamming. Nowadays spam email is the foremost concern for email users. These spams are used for sending fake proposals, advertisements, and harmful contents in the form of executable file to attack user systems or the link to the malicious websites resulting in the unessential consumption of network bandwidth. This paper elucidates the different Machine Learning Techniques such as J48 classifier, Adaboost, K-Nearest Neighbor, Naive Bayes, Artificial Neural Network, Support Vector Machine, and Random Forests algorithm for filtering spam emails using different email dataset. However, here the comparison of different spam email classification technique is presented and summarizes the overall scenario regarding accuracy rate of different existing approaches.

2. Email Spam Detection Using Machine Learning Algorithms

Authors: Nikhil Kumar, Sanket Sonowal, Nishant [2020]

Email Spam has become a major problem nowadays, with Rapid growth of internet users, Email spams is also increasing. People are using them for illegal and unethical conducts, phishing and fraud. Sending malicious link through spam emails which can harm our system and can also seek in into your system. Creating a fake profile and email account is much easy for the spammers, they pretend like a genuine person in their spam emails, these spammers target those peoples who are not aware about these frauds. S o, it is needed to Identify those spam mails which

are fraud, this project will identify those spam by using techniques of machine learning, this paper will discuss the machine learning algorithms and apply all these algorithm on our data sets and best algorithm is selected for the email spam detection having best precision and accuracy .

3. Feature Extraction and Classification of Spam Emails

Authors: Muhammad Ali Hassan, Nhamo Mtetwa [2018]

Emails are a popular and preferred way of written communication in our daily life. The problem with emails is spam. These spam emails are sent with different intentions, but advertisement and fraud are the main reasons. As being inexpensive to send, it causes many problems to the internet society. This paper discusses the use of different feature extraction methods coupled with two different supervised machine learning classifiers evaluated using four performance metrics on two publicly available spam email datasets for spam filtering. We highlight the importance of the correct coupling of feature extraction and classifier, and the merits of using two independent datasets.

4. A Proposed Data Science Approach for Email Spam Classification using Machine Learning Techniques

Authors: Aakash Atul Alurkar, Sourabh Bharat Ranade, Shreeya Vijay Joshi, Siddhesh Sanjay Ranade [2017)

With the facility of email being accessible to any individual with an internet connection, the proliferation of spam emails is one of the biggest problems which plagues our globally integrated communication systems. The various solutions to filter and hide spam previously included the manual detection of specific keywords and the blacklisting of certain domains created to send spam. However, these methods have certain shortcomings in classifying whether emails are spam or ham. This proposed system attempts to use machine learning techniques to detect a pattern of repetitive keywords which are classified as spam. The system also proposes the classification of emails based on other various parameters contained in their structure such as Cc/Bcc, domain and header. Each parameter would be considered as a feature when applying it to the machine learning algorithm. The machine learning model will be a pre-trained model with a feedback mechanism to distinguish between a proper output and an ambiguous output. This method provides an alternative architecture by which a spam filter can be implemented.

5. A Study of Machine Learning Classifiers for Spam Detection

Authors: Shrawan Kumar Trivedi

In the present world, there is a need of emails communication but unsolicited emails hamper such communications. The present research emphasises to build a spam

classification model with/without the use of ensemble of classifiers methods have been incorporated. Through this study, the aim is to distinguish between ham emails and spam emails by making an efficient and sensitive classification model that gives good accuracy with low false positive rate. Greedy Stepwise feature search method has been incorporated for searching informative feature of the Enron email dataset. The comparison has been done among different machine learning classifiers (such as Bayesian, Naïve Bayes, SVM (support vector machine), J48 (decision tree), Bayesian with Adaboost, Naïve Bayes with Adaboost). The concerned classifiers are tested and evaluated on metric (such as F-measure (accuracy), False Positive Rate, and training time). By analysing all these aspects in their entirety, it has been found that SVM is the best classifier to be used. It has the high accuracy and the low false positive rate. However, training time of SVM to build the model is high, but as the results on other parameters are positive, the time does not pose such an issue.

Methodology:

We cleaned the data and then we split the dataset into training dataset and test dataset. Training data set was used to train Naïve Bayes classifier. Performance of trained classifier was tested on test dataset.

A. Dataset Description

We used SMS Spam Collection v.1 dataset [9]. We downloaded this dataset from [10]. This dataset has 5572 text messages which were classified as ham or spam. It has two columns two labeled as v1 and v2. First column v1 has only two values spam or ham describing whether the text message in second column v2 is spam or genuine. The dataset is available as CSV (comma-separated values) file. The messages in this dataset are collected from these sources: Grumbletext Web site, NUS SMS Corpus (NSC), Caroline Tag's PhD Thesis, SMS Spam Corpus v.0.1 Big. In this dataset 4825 text messages were labeled as ham and 747 text messages were labeled as spam.

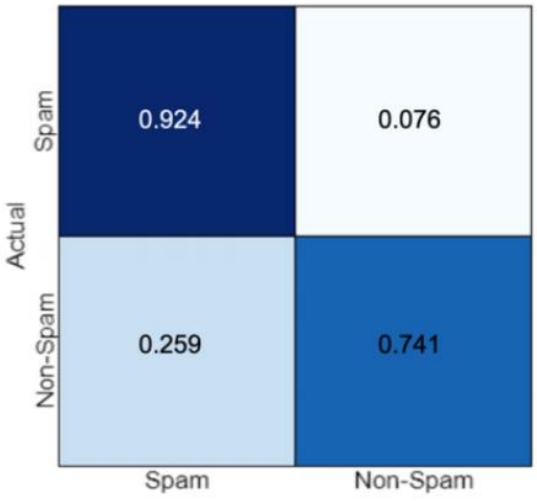
B. Data Preprocessing

We renamed the column v1 as class and v2 as text. After renaming the columns we shuffled the dataset to reduce overfitting. After shuffling, dataset was cleaned. To clean the dataset all text messages were converted to lowercase, and punctuations, numbers, stopwords and URLs were removed.

C. Naive Bayes Classifier

After data preprocessing, dataset was split into training dataset and test dataset. There are 5572 text messages in dataset in which 747 text messages are labeled as spam and 4825 text messages are labeled as ham. The data were split into two datasets. Training dataset had 4000 text messages in which 3461 were labeled as ham and 539

were labeled as spam. Test dataset had remaining 1572 text messages in which 1364 were labeled as ham and 208 were labeled as spam. For classification, a model or classifier is constructed then this model or classifier is further used for predicting the class labels [11]. First we converted text messages of training dataset into document term matrix and the terms having frequency less than five were removed. The entries 0 of document term matrix were replaced by "No" and other non-zero entries were replaced by "Yes". So, this document term matrix had only two values: "Yes" and "No". The Naïve Bayes classifier was trained by using this document term matrix and class labels of text messages of training dataset. In same manner, document term matrix for text messages of test dataset was also created and used for predicting the class labels of text messages by Naïve Bayes classifier.



4. CONCLUSION

The existing SMS attack detection framework can only detect specific features attack. By detecting SMS Spam in the Malay language, spam features in Malay language has been introduced which contributes in detecting SMS Spam in Malaysia. There are five (5) text mining techniques that can be applied to detect these attacks using the proposed framework. The experiments from data mining tool showed the acceptable result by using Naive Bayes. The basis to the selection of this technique is because it is commonly used by other researchers in SMS Spam attack detection and its availability in existing machine learning tools. As a conclusion, according to the study that has been done, it shows that it is significant to detect SMS Spam in Malay language using machine learning technique because most studies focus on detecting SMS Spam in English and that creates a limitation in detecting Malay language SMS Spam. The increasing number of SMS Spam on mobile device violates mobile device user"s security and privacy. Although there are detection and filtering mechanisms to prevent SMS Spam, it is still lacking for SMS Spam in Malay language and needs more suitable

features to detect Malay language SMS Spam attack.

REFERENCES

1. S. Chhabra, "Fighting spam, phishing and email fraud," UNIVERSITY OF CALIFORNIA RIVERSIDE, 2005.

2. K. Yadav, P. Kumaraguru, A. Goyal, A. Gupta, and V. Naik, "SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering," Proceedings of the 12th Workshop on Mobile Computing Systems and Applications. ACM, Phoenix, Arizona, pp. 1– 6, 2011.

3. H. Shirani-Mehr, "SMS Spam Detection using Machine Learning Approach," 2014.

4. Cloudmark, "Annual Security Threat Report 2014," Cloudmark, San Francisco, USA, 2014.

5. E. Vall, #233, and P. Rosso, "Detection of near-duplicate user generated contents: the SMS spam collection," Proceedings of the 3rd international workshop on Search and mining user-generated contents. ACM, Glasgow, Scotland, UK, pp. 27–34, 2011.

6. H. Najadat, N. Abdulla, R. Abooraig, and S. Nawasrah, "Mobile SMS Spam Filtering based on Mixing Classifiers," 2014.

7. K. Yadav, S. K. Saha, P. Kumaraguru, and R. Kumra, "Take Control of Your SMSes: Designing an Usable Spam SMS Filtering System," in Mobile Data Management (MDM), 2012 IEEE 13th International Conference on, 2012, pp. 352–355.

8. T. M. M. and A. M. Mahfouz, "SMS Spam Filtering Technique Based on Artificial Immune System," IJCSI Int. J. Comput. Sci. Issues, vol. 9, no. 2, 2012.

9. Q. Xu, E. Xiang, J. Du, J. Zhong, and Q. Yang, "SMS Spam Detection using Content-less Features," Intell. Syst. IEEE, vol. PP, no. 99, p. 1, 2012.

10. M. Taufiq Nuruzzaman, C. Lee, M. F. A. bin Abdullah, and D. Choi, "Simple SMS spam filtering on independent mobile phone," Secur. Commun. Networks, vol. 5, no. 10, pp. 1209–1220, 2012.

11. M. Z. R. que and M. Farooq, "SMS Spam Detection By Operating On Byte-Level Distributions Using Hidden Markov Models (HMMS)," Virus Bulletin Conference September 2010. 2010.

12. S. J. Delany, M. Buckley, and D. Greene, "SMS spam filtering: Methods and data," Expert Syst. Appl., vol. 39, no. 10, pp. 9899–9908, 2012.

13. Tiago A. Almeida and J. M. G. Hidalgo, "SMS Spam Collection Data Set," 2012. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection.

14. I. Androulidakis, V. Vlachos, and A. Papanikolaou, "FIMESS: filtering mobile external SMS spam," in BCI, 2013, pp. 221–227.

15. T. Charninda, T. T. Dayaratne, H. K. N. Amarasinghe, and J. Jayakody, "Content based hybrid sms spam filtering system," 2014.