

Survey on Spam SMS filtering using Data mining Techniques

Nikunj Chaudhari¹, Prof. Jayvala², Prof. Vinitashah³

PG Scholar, Department of IT, G.H Patel College of Engineering and Technology, V.V Nagar, Anand¹

Assistant Professor, Department of IT, G.H Patel College of Engineering and Technology, V.V Nagar, Anand^{2,3}

Abstract: Data Mining is the process of “Extracting” Use full information from large amount of Data. It is called the Data mining. Data mining is also called the knowledge discovery in Database (KDD). Now days Short Message Service(SMS) is the most popular way to communication for Short message Service for mobile phone user because it is the cheapest mode of Communication compare to other. The advantages to this Fact some company or Spammer used this service for marketing and send the unwanted Spam message to cell phone user and create the disturbance in communication and also consume the networked bandwidth so decrease the network efficiency. To Avoid this Problem Spam SMS filtering Techniques are used in this paper provide the Detail Summary of Spam SMS filtering techniques and Algorithms which will help to overcome this Problem.

Keywords: Data mining; Spam SMS, filtering techniques, Evolutionary Algorithms.

I. INTRODUCTION

In simple word Data Mining is the process of “Extracting” Use full information from large amount of Data. It is called the Data mining. Data mining is also called the knowledge discovery in Database (KDD) [1] In Data mining various techniques are including like classification, Association, Clustering etc. Now modern world the Short Message Service (SMS) is the most popular and widely used by variety of Service for Communication like banking updates, agriculture information Flight updates etc. In developing countries like Indian SMS is the Cheapest mode for Communication [6] since the mobile phone lunch we have Seen the an evolution of Device and Services provided by the mobile phone network Variety of Service is offered. Sort message is Simple and cheapest way to communication [4] Million of people Send the SMS for communication in daily life but the main Problem of user is Spam SMS. [9] The Definition of Spam SMS does not very much in Case of Emails or SMS Spam in simple word the it can be Describe as “Unsolicited Bulk Message” these are unwanted for the user Sent by Samper [8] due to low price the company and Spammer used this service for marketing and promotion. This message is not use full for user and its message is consume the networked bandwidth so it reduced networked Efficiency. So the main objective of the Spam SMS filtering to reduce or blocks the unwanted message Send by the Spammer. This paper is organized as follow Section II introduction to the Spam SMS filtering. And Section III Literature Survey, IV conclusion.

II. SPAM SMS FILTERING

The Defination of Spam SMS does not very much in Case of Emails or SMS Spam in simple word it can be Describe as “Unsolicited Bulk Message” these are unwanted for the

user Sent by Samper [8] is Send to the user for the marketing purpose, for promoting and Advertisement. The Disadvantage of the Spam SMS is it is unwanted and meaningless for the user and it is not Delete without open. Various Techniques are used in Sam SMS filtering are used for the Spam SMS filtering some the most popular algorithms are K-Nearest Neighbor (KNN),Support vector machine,(SVM) Naïve Bayes (NB),Black list/ white list [12]

III. LITERATURE SURVEY

There are many approaches have been used to Spam SMS filtering using data mining techniques. Some of this are presented here.[8]

In [2] Authors have used novel frame worked for SMS Spam Filtering make used of the two different Future selection approach based on the information gain and Chi-square metrics find out Discriminative feature representing SMS messages. The Discriminative feature subsets are then employed in two different Bayesian-based classifiers, so that SMS messages are classified as Spam or ham. Considering all experimental results, the highest overall accuracy is obtained as 90.17% by the binary classification model using top-10 CHI2 based features.

In [3] Authors have used an algorithm is based on the Naïve Bayes and Word Occurrences table they objective to does not depend another computer filtering for Spam SMS filtering. Develop the personal and independent Spam SMS filtering to reduced the communication cost and hardware cost and also they research objective to Develop Spam filtering System on mobile phone to provide user to independent, private, secure, personal, Simple, updatable, filtering system.

In [4] Proposed the implementation of content filter for SMS System based on the Bayesian classifier and word Grouping to evaluate the performance of this filter 120,000 message Sent from a content provider that service mobile operator were tested .the new future ware Added in Calculation the Result Show high Accuracy to classify message is ham or Spam. The Result for test done with all added attribute ware 99.1% accurate.

In [5] Authors have used Method is Simulation based method (SB), Real time Database method (RDB), and real System based method (RSB) Gobble trusted management(GTM) are used for the Design and implementation of Spam control System for the based on the trusted management.

In [6] Authors is used Bayesian filtering for developing the SMSAssassin system for filtering Spam SMS. They develop the SMSAssassin application Developed for the Android and Symbian Phones both. The limited flexibility offered to the current Messaging inbox designs and growing unwanted Content on SMS channel and user is disturbed. So the author present the solution to this types of problems design the an application SMSAssassin which replace the current messaging inbox and can give the Spam filtering capability giving the Control to user to received this type of contented based filtering method provide the limited Capabilities in the Space due to short Content of SMS here developed application Smartly automatic filter based on the Bayesian and user generate preference.

In[7] Proposed System give the approach for detecting and Restrting the Spam message Sent through Spammer cell phone .The System Detect the Spam Message by checking the mutual relation between the Sender and Receiver and content of the message. If the system not found the relation between sender and Receiver and if the message content is found Spamming then System will treat as a Spam and forward Message with either Spam tag or Reject it.

In [8] Authors various algorithms are used to Spam Filtering for the Mobile Text Message. In this paper compare some filtering techniques on a publically available SMS Spam Croups. The Bayesian Method was very effective just they expected giving very high Success up to 98% so the indicate the Bayesian method is the best approach for the SMS Spam filtering for the Mobile Text Message.

In [9] Authors used an algorithm Bayesian filtering Techniques can easily transfer to the SMS spam. Author used Combine mechanism of Naïve Bye's and dynamic nature in to single Algorithms for Spam SMS filtering. Most up Spam Detection techniques are unable to find to this Spam's because regular training of these Classifiers is not done yet, database of Spam be should be updatable Existing Spam filter are statics in nature because of that

these Spam filter show false Positive and false negative Result. So Dynamic training improved the Spam filtering techniques.

In [10] this paper various classifier are Multinomial Naïve Bayes (MNB),Support vector(SVM) Random Forest(RF), Adaboost are used for the Indian Spam SMS filtering . Multinomial Naïve Bayes (MNB)and support vector machine (SVM) are the best classifier compare the other But the time Required to SMS Spam Detection is Support vector machine is less than the Multinomial Naïve Bayes and Accuracy ,precision, Recall-measure ,Matthews Correlation Coefficient, Spam Caught , Blocked ham, Area Under the curve is high. For Indian Spam SMS filtering.

In [11] authors have an algorithm Support vector machine (SVM) and Naïve Bayesian (NB)used for the Thai – English Spam filtering. Two method are used Spam SMS filtering the First method Simply used Current Spam English Message Filtering filter the filtering and upgrade for Thai language support and Second method applies text normalization, word segmentation and analyzing the Thai Semantic word. Method #2 using SVM the accuracy is high fir Thai- English but processing time is more. And method #2 using the NB the filtering is interesting because its processing time is faster than SVM and Provide the Acceptable Accuracy.

In [12] Authors Suggest an algorithm Gentle Boost algorithm for Spam SMS Detection because they have unbalance Data and found that Gentle Boost performed that better than other. Boosting Classifier and lead to the batter Performance the Reason for that this might be a batter method is that it works batter for unbalanced and binary classification.

In [13] in this paper author performed a detailed analysis of potential near duplicate in Collection by using the Standard “String to text method “on three sub collections: the original one(INIT) ,the Added message (ADD) and the final collection . the near- duplicate method consist of the Finding near N-gram match between message ,for N=5,6 and 10 with each collection in order to verify that there is no a Significant duplicates in the Final sub-collection, apart from those previously Exiting in the INT and ADD sub collections.

In [14] Authors have used an algorithm is based on Bayesian filtering and Spam filtering technology is based on the black list, white list and keyword has certain defect, at present filtering of Spam SMS turns to Study based on the Content Filtering.

In [15] Authors have used an Algorithm based on the traditional balanced window algorithm the System has been improved in many aspect that help to get better performance for Spam filtering design to process row text materials without the time consuming word Segmentation

for Chinese and most importantly to be easily extended to multicast pattern by extending the Original window to multiple classifier.

In [16] this paper Vector Space model based on Spam SMS filtering it addresses the particularly of Short message Service ,such as short ,vocal , domain related etc this technology Considers much about the particular and Apply much modification on the traditional VSM model. This technology has been deployed in Production environment of Dahan Tricom Corporation and results in Production Department turn out be Applied in SMS Commercial Companies.

IV. CONCLUSION AND FUTURE SCOPE

In this Paper we have seen the various Spam SMS filtering techniques for Mobile Short message service. Most of the Naïve byes, Bayesian classifier and Support vector machine (SVM) techniques are more Accuracy for Spam SMS filtering compare to other. We can also use the hybrid Spam Filtering techniques using combine two or more different techniques and increased the Efficiency and Accuracy of the Existing Spam SMS filtering techniques.

REFERENCES

- [1] J.Han and M. Kamber, jian pei “Data mining: Concepts and techniques (3rd edition)”,Morgan Kaufman Publishes, 2012.
- [2] Alper Kursat Uysal, Serkan Gunal, Semih Ergin, Efnan Sora Gunal. “A Novel Framework for SMS Spam Filtering” 2012 IEEE
- [3] M.Taufiq Nuruzzaman,ChangmooLee, “independent and Personal SMS Spam filtering 11th international Conference on Computer and information technology 2011 IEEE
- [4] Dirceu Belem, Fatima Duarte-Figueiredo, “Content Filtering for SMS Systems Based on Bayesian Classifier and Word Grouping”2011 IEEE
- [5] Liang Chen ,Zheng Yang, weidong zhang ,Raimo Kantola, “Implementation of an SMS Spam Control System on trusted management”2013 IEEE
- [6] Kuldeep Yadav, Swetank K. Saha, Ponnuranga Kumaraguru”Take Control of Your SMSes : Designing an Usable Spam SMS Filtering System”13th International Conference on Mobile Data Management 13th International Conference on Mobile Data Management2012 IEEE
- [7] Suraj J.Warade,Pritish A. Tijare,Swapnil N.Sawalkar”An Approach for SMS Spam Detection “International Journal of Research in Advent Technology, Vol.2, No.12, December2014(E-ISSN: 2321-9637)
- [8] Kuruville Mathew,Biju Issac “Intelligent Spam Classification for Mobile Text Message” International conference on Computer Science and Network Technology,2011 IEEE
- [9] GuravSheti, "SMS Spam Filtering Application Using Android “International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4624-4626
- [10] Sakshi Agarwal,Sanmeet Kaur,Sunita Garhwal”SMS Spam Detection for Indian Messages”2015 1st International Conference on Next Generation Computing Technologies (NGCT-2015)Dehradun, India, 4-5 September 2015 IEEE
- [11] Chaiyaporn Khemapatapan”Thai-English Spam SMS Filtering” 16th Asian –pacific Conference on Communication (APCC) 201010 IEEE
- [12] Fatemeh Akbari, Hedieh Sajedi “SMS Spam Detection Using Selected Text Futures and Boosting Classifiers” IKT2015 7th international Conference on information and Knowledge Technology. 2015 IEEE
- [13] Jose Maria Gomez Hidalgo, Tiago A. Almeida, akebo Yamakami “on validity of New SMS Spam Collection” 11th International Conference on Machine learning and Applications 2012 IEEE
- [14] Hong-yan Zhang ,wei Wang, “Application of Bayesian Method Spam SMS filtering” 2009 IEEE
- [15] Jie CAI,Yuezhong TANG,Rile HU “Spam filter for Short Message Using Window” International conference Advance Language Processing And Web information Technology 2008 IEEE
- [16] Wei Li, Sisheng Zeng “ A vector Space Model based on Spam SMS filtering” 11th international conference on computer Science & Education (ICCC2016) 2016 IEEE