IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

Mobile SMS Spam Recognition Using Machine Learning Techniques with the help of Biasian and Spam Filters

Ashok Koujalagi

Dept. of Computer Science, Basaveshwar Science College, Bagalkot Rani Channamma University, Karnataka, INDIA

Corresponding: koujalagi.ashok@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i4.540542 | Available online at: www.ijcseonline.org

Accepted: 19/Apr/2019, Published: 30/Apr/2019

***Abstract*:** Summary Spam SMS is unwanted messages to users who are worried and harmful from time to time. Currently, group survey papers are available on SMS detection techniques. Study and review their used techniques, approaches and algorithms, their advantages and disadvantages, evaluation measures, discussion of data sets, as well as the end result of the studies. Although the SMS spam detection techniques are additionally demanding as sms spam detection techniques, as the local content, the use of abbreviated words, unfortunately does not meet any of the existing research on these challenges. There is an enormous amount of emerging research in this region and this survey can serve as a point of reference for the upcoming direction of research.

***Keywords*:** Mobile SMS spam detection

I. INTRODUCTION

Small message once (SMS) is the majority, often and extensively used message medium. The term "SMS" is used to compile user activity and all kinds of text messages in many parts of the world. It has evolved into a medium of announcement as well as endorsement of products, bank updates, agricultural information, flight updates and internet offers. SMS also works in direct marketing that is recognized as SMS marketing. SMS marketing is occasionally a problem for users. This kind of SMS Called spam sms. Spam is one or more unwanted messages that are unwanted for users, sent or posted as part of a better collection of messages. Everyone has considerable content. The purpose of SMS spam is the announcement and marketing of a variety of products, the sending of political issues, the distribution of inappropriate adult content and internet offers. For the reason, spam floods are a serious problem. All over the world.SMS spamming gets reputation over additional spamming approaches like electronic mail with twitter due to the increasing popularity of SMS Communication.

II. BACKGROUND AND RELATED WORK

SMS spam detection is moderately a fresh research area after the SMS electronic mail social tags, and twitter and web spam detection. a number of spam detection investigations include [1] [2] etc. These researchers typically run after 2011. There are some well-known SMS spam detection techniques, a little more challenging than SMS spam detection, such as limited message size using local and shortcut words and incomplete slogan information. These challenges must be solved. There is scope of research in this field and some research works have been included Guided by this present be different category of SMS spam filter such as pallid record and black record. Content-based non-content basics are two-way approaches and challenge-response techniques. [4], [5], [8] use the techniques Client surface server in these various machine learning algorithms such as

Naïve Bayes.
Bayesian is a probabilistic move towards what began under a previous belief, observe some data and then updated that faith is the probabilistic life form spam and not spam from a word can be meant by the appearance of that word in ham and spam messages with the Bayesian algorithm [12].

Support vector machine. (SVM)
Supporting vector machines are taught under supervision using linked algorithms that use data intended for categorization and regression analysis. As a teaching example that includes spam and legitimate SMS, it is known after the SVM teaching algorithm builds a model that can assign a new example to spam and legitimate groups. An SVM model is a demonstration of the example, because a point in space, such an example of the division category, is separated by a clear gap so that it is widely available. [9]

Decision Trees
A decision hierarchy is a decision support tool that uses a hierarchy similar to or model of decisions and their likely

punishment, counting the possibility of event outcomes. A decision tree can be used to decide whether a fresh message is spam or spam. [11]

Logistic Regression

A logistic regression is a prognostic analysis. Logistic regression is used to explain data and explain the association that is flanked by single dependent binary variable and single or additional putative ordinal, interval or percentage level independent variables. From time to time logistic regression is difficult to understand, the intellectual statistics tool without Difficulty enables you to do the analysis, and then interprets the production in simple English.

Random Forest

Random Forest is a trademark term for an ensemble of decision trees. In random forests our collection of decision trees. To classify a new object based on properties, each tree agrees for that class. The forest chooses the classic

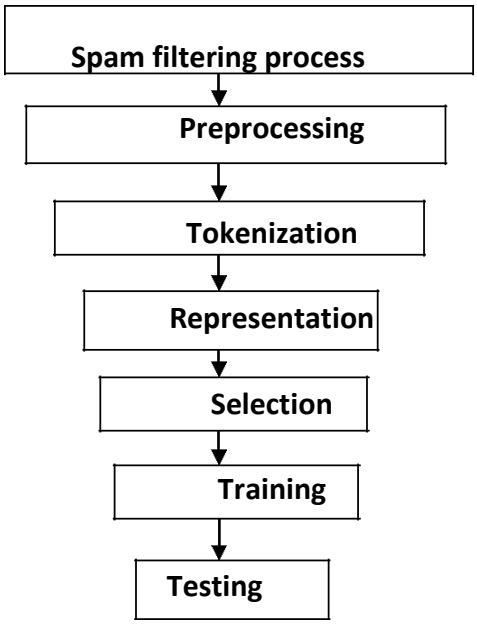

Fig.1

Spam filters process. A physical confidential spam and ham message is the import or teaching position for a spam filtering algorithm. The algorithm consists of the following steps.

Preprocessing. Removing irrelevant content such as stop word is part of the data preprocessing

Tokenization. Segmentation of the message according to words character or symbols called tokens. There are different tokenization approaches like word Tokenization, sentence, word or character N-gram and orthogonal sparse bigrams.

Representation. Conversion to applying value parks

Selection. Select key feature values that contain crash on top of categorization rather than choosing each pair of feature value.

Training. Learn the algorithm using the selected quality values.

Test. Experiment with the recent data through the teaching model.

III. STUDY SELECTION PROCESS

Nearly related studies, we mostly searched for Google Academies. Together we have a number of papers present, a number of additional conferences and journals such as when IEEExplore, IJCSI ITJ ACM, is made from side to side. Google academic tool the journals file and Conference from which we also contain selected paper contains many references. We also searched for the referral papers and used a number of them as our appropriate paper. We have connected the Google scientist Articles and quoted feature celebrate our search process.

Table 1. SMS Spam Detection dataset description

Study ID	Available At	Total No of Messages
[1]	[13]	5574

IV. VALIDATION OF THE STUDY

Our SLR is performed to each investigate the used approaches Techniques within SMS spam detection. The intimidation to the strength of our review is that the presence may be selection bias and short enough resources. We try to get every possible and relevant information resources. Some resources do not contain a straight line published. Another danger is that a number of resources are not available for community use.

V. RESULT ANALYSIS

An initial that we physically used on Google's search, used the topic of spam detection to increase the speed detection field. This led to many SMS spam detection headers. Then we changed our search using only SMS spam detection through our study selection procedure we chose Paper published in various conferences and journals linking only on the road to mobile SMS spam detection in mid-13 studies.

VI. DATASET DESCRIPTION

A data set preparation should be envisaged for a number of machine learning classification algorithms. The result of the

machine knowledge algorithms depends on the data set. Because an effect can run spam detection algorithms without a data set. Within us there are several different openly available data sets available in different studies. Data set link and a number of metrics such as total spam number and spam messages are shown in table [13].

VII. CONCLUSION

This paper results from the systematic literature review on SMS detection. We selected a total of 13 research questions in this field and reviewed their proposed techniques. Advantages and disadvantages. And challenges they addressed. We also investigated them Evaluation Procedures. We have shown the publicly available data set information that is a previous need for a spam filtering algorithm.

REFERENCES

[1] K. Yadav, S. K. Saha, P. Kumaraguru, and R. Kumra,'' take control of your smses: designing an usable spam sms filtering system," in 2012 IEEE 13th International Conference on Mobile Data Management. IEEE, 2012, pp. 352–355.

[2] S. J. Warade, P. A. Tijare, and S. N. Sawalkar, "An approach for sms spam detection."

[3] A. Narayan and P. Saxena, "The curse of 140 characters: evaluating the efficacy of sms spam detection on android," in Proceedings of the Third ACM workshop on Security and privacy in smartphones & mobile devices ACM, 2013, pp. 33–42.

[4] A. S. Onashoga, O. O. Abayomi-Alli, A. S. Sodiya, and D. A.Ojo "An adaptive and collaborative server side sms spam filtering scheme using artificial immune system," Information Security Journal: A Global Perspective, vol. 24, no. 4-6, pp. 133–145, 2015.

[5] J. W. Yoon, H. Kim, and J. H. Huh, "Hybrid spam filtering for mobile communication," computers & security, vol. 29, no. 4, pp. 446–459, 2010.

[6] Q. Xu, E. W. Xiang, Q. Yang, J. Du, and J. Zhong, "Sms spam detection using noncontent features," IEEE Intelligent Systems, vol. 27, no. 6, pp. 44–51, 2012.

[7] I. Ahmed, D. Guan, and T. C. Chung, "Sms classification based on naïve bayes classifier and apriori algorithm frequent itemset," International Journal of machineLearning and computing, vol. 4, no. 2, p. 183, 2014.

[8] J. M. G´omez Hidalgo, G. C. Bringas, E. P. S´anz, and F. C. Garc´ıa, "Content based sms spam filtering," in Proceedings of the 2006 ACM symposium on Document engineering. ACM, 2006, pp. 107–114.

[9] https://en.wikipedia.org/wiki/Support_vector machine [last Accessed: 05-11-2016]

[10] https://en.wikipedia.org/wiki/Decision tree[Last Accessed:05-11-2016]

[11] http://en.wikipedia.org/wiki/K-

[12] http://fastml.com/bayesian-machine- learning/ [Last Accessed: 05-11-2016

[13] SMS Spam Collection data set from UCI Machine learning Repository,"http://archive.ics,uci.edu/ml/dataset/SMS+Spam+Collection

AUTHORS PROFILE

Dr. Ashok Koujalagi is an Author, Professor and Postdoctoral Researcher. He received his M.Sc degree from Bangalore Central University. And Ph.D from Central University of Allahabad. He authored four books among one is International, he also has more than 25 research publications & Proceedings indexed in ICI, Scopus.

Presently serving as an Assistant Professor & Postdoctoral Researcher in the Post Graduation Department of Computer Science, Basaveshwar Science College - Bagalkot, affiliated to Rani Channamma University - Belgaum, Karnataka, INDIA.

He is also Professional Board member of 21 International Organizations / Association, Editorial Board Member in 18 International Journals, Peer-Reviewal Board Member in 15 International Journals and Advisory Board Member in 3 International Journals. He has also been invited by 3 international conferences as a "Conference Organizing Committee Member, also invited as a speaker in 7 international Conferences. And his area of research covers, Computer Networking, Cloud Computing, Data Mining, and Mobile Computing.

