SUPERVISED LEARNING FOR SPAM DETECTION: A ROBUST AND EFFECTIVE METHODOLOGY

[1]S VIJAY KUMAR, [2]NISHRA MAHVEEN
[1]Assistant Professor,[2]Student
Department of CSE
Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT

Short Message Service, or SMS, has become less relevant in this day of widely used instant messaging apps. Instead, service providers, companies, and other organisations have come to rely on this service to target regular users for spam and marketing purposes. The usage of regional language material written in English is a new trend in spam communications, which makes it more difficult to identify and filter such messages. This study uses an expanded version of a conventional SMS corpus that includes labelled text messages printed in English that are written in regional languages like Bengali or Hindi, as well as non-spam communications. The labelled text messages were obtained from local mobile users. Using a collection of characteristics and machine learning algorithms that are often used by academics, the Monte Carlo technique is used for learning and classification in a supervised manner. The results show how various algorithms perform in successfully tackling the given task.

1. INTRODUCTION

Because humans are sociable animals, their capacity for efficient communication is fundamental to their socialising nature. Effective and timely communication has always been essential to human survival, as shown by the earliest cave drawings and the modern, lightning-fast instant messaging apps.

The fundamental elements of a standard communication are shown in Figure 9.1, where sender(s) and receiver(s) exchange messages over a communication channel. With the many decades of human civilisation, this communication medium has undergone various transformations. For example, text messages, letters (pages), and cave walls are examples of the many communication mediums that humans have used.

A new kind of communication known as the Short Message Service, or SMS, took the role of handwritten letters when mobile technology first entered people's lives. Text messaging on mobile devices was first documented in 1992 [1], and the technology has advanced significantly since then. Over the past 20 years, this service has been more popular and essential to how technology has improved human existence. Each mobile device user may create a text message using the SMS that is up to 160 characters long and contains special symbols, numbers, and alphabets [2]. This is the "short message" that may be sent to a receiver (someone else who uses a mobile device). This kind of communication is useful, particularly when brief information has to be sent quickly or when returning calls is not practical.

Nonetheless, the use of internet-based messaging services, which are often quicker and less expensive than SMS, has increased dramatically during the last ten years. In addition, these services are enhanced with features particular to their applications, such stickers, GIFs, and no message length limits, making them the go-to option for mobile communication. Due to this, the once preferred communication method has fallen out of favour and is now seldom utilised by regular mobile users for daily contact. Rather, a variety of service and/or product-based businesses now find this service to be a useful tool for implementing their direct marketing approach.

Several businesses have adopted the SMS-based marketing approach, which gives a unique chance to locate and entice new customers by presenting them with alluring incentives and deals on certain goods or services. According to a recent poll, 42% of Indian participants reported receiving approximately seven unsolicited spam messages each day, with 96% of participants admitting to receiving such messages daily [3]. Only approximately 6% of Indian mobile customers find the Do Not Disturb (DND) service beneficial, despite the regulatory and preventative restrictions put in place by the Telecom Regulatory Authority of India (TRAI) on the transmission of unwelcome messages [4].

To properly avoid, identify, and filter spam at the user end, one must have a broad knowledge of spam as uninvited or unwelcome communications. Unaware smartphone users are much more likely to unintentionally sign up for these annoying SMS while they are using a service or making a purchase. The majority of unsolicited or spam communications that Indian consumers often get are from online marketers, banks, telecom providers, etc. Even more dangerous are the series of fake spam messages (see Figure 9.2) that prey on unsuspecting users in an attempt to trick them into divulging vital information about them, such as banking passwords and personal information.

Conversely, ham messages are the expected electronic texts that a mobile subscriber wants to receive. These SMS might include information about flight tickets or updates about bank accounts, among other things. Therefore, it's critical to differentiate between these two SMS kinds correctly. Generally, SMS-based communication, which includes spam filtering, may be depicted as Figure 9.3. Many spam detection and filtering strategies have been the subject of much study over the years, but not all of them have produced effective and useful end-user applications.The goal of the present study is to determine the resilience of widely used classification methods, which include both modern Deep Neural Network architecture-based models and traditional machine learning classifier models. The training and classification tasks are carried out using the Monte Carlo method on various combinations of spam and ham data for a maximum of 100 times. This makes it possible to determine the final performance statistics for every classification model and designate the best-performing model as the optimal one. The literature review below discusses the current status of research on spam identification.

2. LITERATURE SURVEY

SMS spam detection using H2O framework

One of the issues is SMS spam, which many people find bothersome and dislike receiving. Numerous SMS spam detection techniques are now in use, and several classifiers, including Support Vector Machine, Naïve Bays, and numerous more machine learning algorithms, were used. This research proposes a novel classifier that primarily relies on utilising H2O as a platform for comparing various machine learning techniques. Furthermore, naïve bays, deep learning, and random forests are machine learning methods that are used in comparisons. They are used to identify the most significant features that may be supplied as input to naïve Bayes, random forest, and deep learning classifiers, in addition to being utilised as classifiers themselves. The amount of digits and the presence of a URL in the SMS text are the two most important characteristics that might influence the identification of SMS spam, according to experimental data. The dataset provided by UCI Machine Learning Repositories is the one utilised in the experiment. As a result, testing reveal that naïve Bayes, with a runtime of 0.6 seconds, is the quicker approach that produces good performance. However, when compared to deep learning and random forest, it has the lowest precision, recall, f-measure, and accuracy. However, random forest, which has 50 trees and 20 maximum depths, has the highest accuracy of any algorithm. Its precision, recall, f-measure, and accuracy are 96%, 86%, 91%, and 0.977%, respectively; however, its runtime is lengthy—30.28 seconds.

Spam detection on social media using semantic convolutional neural network

This article explains how the exponential rise in spam volume throughout the network has made spam detection in social media text more crucial. It is difficult, particularly when writing in a restricted character count. Learning more efficient characteristics is necessary for effective spam detection. The present paper suggests using a convolutional neural network (CNN), a deep learning technique, in conjunction with an additional semantic layer for spam identification. Semantic convolutional neural network, or SCNN, is the name given to the resulting model. To get a semantically enhanced word embedding, a semantic layer consists of training random word vectors with the aid of Word2vec. If a word is absent from the word2vec, WordNet and ConceptNet are used to locate a word that is comparable to the provided word. Two corpora are used to assess the architecture: the Twitter dataset and the SMS Spam dataset from the UCI repository. tweets that were taken from live public tweets. With 98.65% accuracy on the SMS spam dataset and 94.40% accuracy on the Twitter dataset, the authors' method beats the state-of-the-art findings.

Convolutional neural network based SMS spam detection

Unwanted text messages are referred to as SMS spam. The ability of machine learning techniques to classify spam communications has proven notably useful in anti-spam filters. Tiago's dataset is the one that was utilised for this study. Data preparation, which included tokenisation, stopword removal, and text reduction to lower case, was an essential part of the experiment. A Convolutional Neural Network was suggested as the classification technique. The accuracy of the whole model was 98.4%. The obtained model is a useful tool for a variety of situations.

Spam detection using ensemble learning

We connect with each other often in our everyday lives via email and SMS, but with the rise in spam, these methods have become more problematic for both senders and recipients. To identify spam, we need a spam detection tool. While there are a lot of spam detection solutions on the market, their effectiveness is limited since they focus just on a single classifier or a small number of classifier combinations. Four distinct classifiers—the "Gaussian Naive Bayes," "Multinomial Naive Bayes," "Bernoulli Naive Bayes," and "Decision Tree"—are presented in various combinations in our study. Voting classifiers are a kind of ensemble learning that we have used to determine the accuracy of various classifier combinations. The results demonstrate that using a voting classifier yields predictions that are more accurate than using a single classifier. For the same reason, we had also developed an Android application. The mobile application uses the client-server model of operation. In essence, the mobile application serves as a client, sending the user-selected data from the mobile device to the server. A machine learning script on the server categorises the incoming data and provides the client with a forecast.

3. EXISTING SYSTEM

Back in 2015, Agarwal et al. [5] utilized the comprehensive data corpus consolidated by [6] and extended it by adding a set of spam and ham SMS collected from Indian mobile users. They demonstrated how different learning algorithms like Support Vector Machine (SVM) and Multinomial Naïve Bayes (MNB) performed on the Term Frequency–Inverse Document Frequency (TF-IDF)–based features extracted from the corpora. Starting at around this time, a plethora of research works have used the same corpus and similar set of features and learning algorithms for designing spam detection systems. In the following set of similar works, it is observed that a set of learning and classification algorithms are used for a performance comparison study. Also, there is a paradigm shift toward neural network-based learning algorithms in more recent times.

In such a work in 2017, Suleiman et al. [7] demonstrated a comparative study of the performance of MNB, Random Forest, and Deep Learning algorithm–based models by using the H2O framework and a self-determined set of novel features on the same SMS corpus. Using word embedding features, Jain et al. [8] showed in 2018 how Convolutional Neural Network (CNN) can be utilized to achieve a better performance than a number of other baseline machine learning models in determining the spam messages from the corpus of [6].

In the same year, Popovac et al. [9] illustrated how CNN algorithm performs on the
same SMS corpus using TD-IDF features.

In 2019, Gupta et al. [10] proposed a voting ensemble technique on different learning algorithms, namely, MNB, Gaussian Naïve Bayes (GNB), Bernoulli Naïve Bayes (BNB), and Decision Tree (DT) for spam identification using the same corpus.

The trend of classifier performance comparison continues till recent times in 2020,
where the work by Hlouli et al. [11], illustrated how Multi-Layer Perceptron (MLP), SVM, k-Nearest Neighbors (kNN), and Random Forest algorithms perform on the same SMS corpus for detecting spam and ham using Bag of Words and TF-IDF–based features. In a similar contemporary work, GuangJun et al. [12] highlighted the performance of kNN, DT, and Logistic Regression (LR) models on SMS spam corpus, though the feature extraction techniques were not discussed.

A recent but different type of work by Roy et al. [13] shows how the same SMS corpus by Hidalgo et al. [6] is classified using Long Short Term Memory (LSTM) and CNN-based machine learning models with a high accuracy. The authors also noted that dependence on manual feature selection and extraction results often influences the efficacy of the spam detection system and consequently utilized the inherent features determined by the LSTM and CNN algorithms.

Disadvantages

The system is not implemented Inverse Document Frequency (IDF).

SMS data is to be finally used by the mathematical model–based supervised learning algorithms. These algorithms fail to deal with textual content in the data and are more comfortable with numeric values.

4. PROPOSED SYSTEM

It is observed that in spite of the comparative study of classification performance undertaken by the aforementioned state-of-the-art works, none of them have attempted to determine and establish the robustness of the classification techniques in spam identification. Also, the abundance of spam messages in regional language is largely ignored in such works.

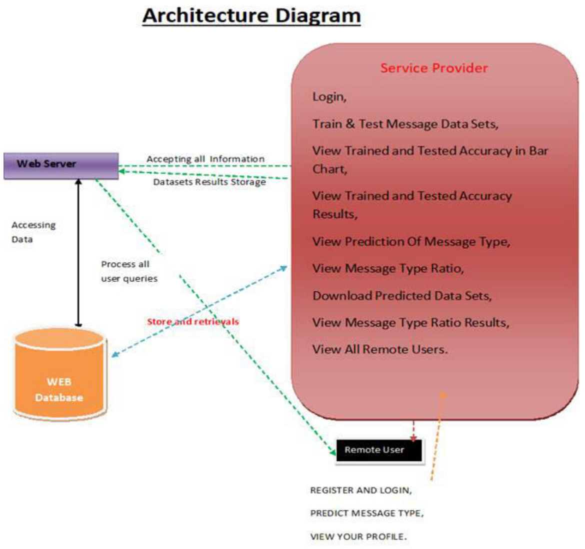
1. The system introduces the novel context of identifying spam and ham SMS in regional languages that are typed in English, along with the general English corpus of
spam and ham by extending it.
2. The system employs a Monte Carlo approach and ML Classifiers to repeatedly perform classification using different machine learning algorithms on different combinations of spam and ham text from the extended corpus (with k-fold cross-validation for a large value of k = 100) in order to determine the efficiency of baseline learning algorithms in comparison to the CNN-based model.

Advantages

The proposed system is more effective due to presence of many ml classifiers.

The proposed system implemented with an accurate prediction for the corresponding dataset.

5. SYSTEM ARCHITECTURE



6. IMPLEMENTATION

Modules

Service Provider

The Service Provider must provide a valid user name and password to log in to this module. Following a successful login, one may do a number of tasks, including Train and Examine Message Sets, Examine the Bar Chart for Trained and Tested Accuracy. View the results of trained and tested accuracy, view the message type prediction, view the message type ratio, download the predicted data sets, View Results for Message Type Ratio, See Every Remote User.

View and Authorize Users

The administrator may see a list of all enrolled users in this module. The administrator may see user information here, including name, email address, and address, and they can also approve people.

Remote User

There are n numbers of users present in this module. Prior to beginning any actions, the user must register. The user's information is saved in the database when they register. Upon successful registration, he must use his authorised user name and password to log in. Upon successful login, the user may do several tasks such as registering and logging in, predicting the message type, and seeing their profile.

7. CONCLUSION

Numerous workable ideas have been put forward in the very popular topic of study on effective spam detection and filtering. Reviewing pertinent, current state-of-the-art literature makes it clear that the most notable advancements have been made in the employment of more sophisticated, newer algorithms that are able to identify more subtle patterns in the fundamental characteristics of various spam and ham messages in a text corpus. The majority of these algorithms are based on neural networks and Deep Neural Network variations, such CNN and LSTM. In the present study, a spam detection system has been created and assessed using an extensive and well validated SMS corpus as input. The corpus has been expanded to include the context of regional messages written in English. In order to identify whether supervised classification algorithm—cluster neural network (CNN) or more traditional machine learning algorithms, such as support vector machines (SVM, kNN, and DT—is the most reliable at successfully identifying spam messages, the system uses a Monte Carlo method. K-fold cross-validation has been used for this, with intervals of 10 folds and a high value of k = 100. Experimental research has shown that the suggested strategy consistently outperforms the other classifiers, with CNN emerging as the most reliable method with an accuracy and F1 score of almost 99.5%. SVM is also the most reliable of the traditional learning algorithms, with typical evaluation measure values greater than 98%. As a result, the system that was created has successfully categorised the provided unique text corpus, and CNN may be used as a powerful learning and classification method. Additionally included is a cloud-based framework for using the suggested classifier. This study may serve as a guide in the future for developing reliable, real-time spam identification and filtering systems that must operate on difficult SMS datasets with unique circumstances.

REFERENCES

1. Hppy bthdy txt!, BBC, BBC News World Edition, UK, 3 December 2002, [Online]. Available:http://news.bbc.co.uk/2/hi/uk_news/2538083.stm. [Accessed October 2020].

2. Short Message Service (SMS) Message Format, Sustainability of Digital Formats, United Statesof America, September 2002, [Online]. Available: https://www.loc.gov/preservation/digital/formats/fdd/fdd000431.shtml. [Accessed, October 2020].

3. India's Spam SMS Problem: Are These Smart SMS Blocking Apps the Solution?, Dazeinfo, India, August 2020, [Online]. Available: https://dazeinfo.com/2020/08/24/indias-spam-sms-problemare-these-smart-sms-blocking-apps-the-solution/. [Accessed October 2020].

4. The SMS inbox on Indian smartphones is now just a spam bin, Quartz India, India, March 2019, [Online]. Available: https://qz.com/india/1573148/telecom-realty-firms-banks-sendmost-sms-spam-in-india/. [Accessed October 2020].

5. Agarwal, S., Kaur, S., Garhwal, S., SMS spam detection for Indian messages, in: 1st International Conference on Next Generation Computing Technologies (NGCT) 2015, UCI Machine Learning Repository, United States of America, IEEE, pp. 634–638, 2015.

6. Almeida, T.A. and Gómez, J.M., SMS Spam Collection v. 1, UCI Machine Learning Repository, United States of America, 2012. [Online]. Available: http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/,[Accessed October 2020].

7. Suleiman, D. and Al-Naymat, G., SMS spam detection using H2O framework. Proc. Comput.Sci., 113, 154–161, 2017.

8. Jain, G., Sharma, M., Agarwal, B., Spam detection on social media using semantic convolutional neural network. Int. J. Knowl. Discovery Bioinf. (IJKDB), IGI Global, 8, 12–26, 2018.

9. Popovac, M., Karanovic, M., Sladojevic, S., Arsenovic, M., Anderla, A., Convolutional neural network based SMS spam detection, in: 2018 26th Telecommunications Forum (TELFOR), Serbia, 2018.

10. Gupta, V., Mehta, A., Goel, A., Dixit, U., Pandey, A.C., Spam detection using

ensemble learning, in: Harmony Search and Nature Inspired Optimization Algorithms, pp. 661–668, 2019.

11. El Hlouli, F.Z., Riffi, J., Mahraz, M.A., El Yahyaouy, A., Tairi, H., Detection of SMS Spam Using Machine-Learning Algorithms, Embedded Systems and Artificial Intelligence: Proceedings of ESAI 2019, Fez, Morocco, 1076, 429, Springer Nature, Singapore, 2020.

12. GuangJun, L., Nazir, S., Khan, H.U., Haq, A.U., Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms. Secur. Commun. Netw., Hindawi, 2020, 1–6, 2020.

13. Roy, P.K., Singh, J.P., Banerjee, S., Deep learning to filter SMS spam. Future Gener. Comput. Syst., 102, 524–533, 2020.

14. Ghourabi, A., Mahmood, M.A., Alzubi, Q.M., A Hybrid CNN-LSTM Model for SMS Spam Detection in Arabic and English Messages. Future Internet, 12, 156, 2020.

15. Sammut, C. and Webb, G.I., TF-IDF, in: Encyclopedia of Machine Learning, pp. 986–987, Springer, US, 2010.