International Journal of Engineering

Journal Homepage: www.ije.ir

An Effective Model for SMS Spam Detection Using Content-based Features and Averaged Neural Network

S. Sheikhi*, M. T. Kheirabadi, A. Bazzazi

Department of Computer, Gorgan Branch, Islamic Azad University, Gorgan, Iran

PAPER INFO

Paper history:
Received 04 August 2019
Received in revised form 06 November 2019
Accepted 16 Januray 2020

Keywords:
Spam
Mobile Spam
SMS Spam
Anti-Spam
Machine Learning

ABSTRACT

In recent years, there has been considerable interest among people to use short message service (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.

doi*: 10.5829/ije.2020.33.02b.06*

1. INTRODUCTION

SMS (Short Message Service) is a standard of mobile protocols which allows users to communicate without connecting to the internet. Due to the cheap cost of SMS services in most telecommunications service providers, and its accessibility and efficiency compared to email services, it is one of the most common communication tools worldwide [1]. However, the attention focused at this service attracts criminals to use it as a mean for performing malicious activities and has created trouble for customers and service providers [2, 3]. An SMS spam message is an unwanted or unsolicited text message which is sent to the user's mobile phone with various content types, such as advertisements, awards, free services and promotions [4]. The main goal of spammers is to steal critical user information such as username, password, and credit card details [5]. They may impose various strategies in order to steal information, hence, SMS messages being one the most straightforward tactics [6]. One of the most frequent online attacks is phishing which usually happens through email, but the simplicity and extensive usage of mobile phones have made phishers consider SMS messages as a suitable method. In phishing attacks, the phisher sends a malicious URL using SMS messages and invites users to visit that URL address in order to steal sensitive and personal information from the user's mobile phone [7]. Moreover, SMS phishing has no severe limitation for spammers, and they can easily buy various phone numbers within any area or country code to send malicious SMS messages. This makes it challenging to recognize and distinguish attackers based on their mobile number [8].

Accordingly, a reliable and accurate method to filter spam messages is essential, although different security

*Corresponding Author Email: *s.sheikhi@outlook.com* (S. Sheikhi)

Please cite this article as: S. Sheikhi, M. T. Kheirabadi, A. Bazzazi, An Effective Model for SMS spam Detection using Content-based features and Averaged Neural network, International Journal of Engineering (IJE), IJE TRANSACTIONS B: Applications Vol. 33, No. 2, (February 2020) 221-228

techniques are available to block these messages. Researchers have introduced different solutions and built various mobile apps to filter SMS spam messages. However, these methods are either not well developed or not reliable enough, making the filtering methods focus mainly on the detection of email spam which is an old and non-concerning security issue [9].

Therefore, in the current study we introduced a novel method for classification of SMS spam messages in order to alleviate the challenges and detect spam messages effectively with high detection rate. Our proposed model focuses on the three following issues:

- Studying the behavior and properties of spam and legitimate messages in the selected dataset
- Extracting and selecting most correlated features with high discriminative ability from message content using a .net program to build feature vectors and a new dataset
- Using the extracted features as input for an averaged neural network algorithm to classify messages more accurately with high performance.

The remaining part of the paper proceeds as follows: Second 2 provides a brief overview of existing methods in SMS spam detection, and highlights their limitations and achievements. Section 3 is concerned with the methodology used for this study, including the feature extraction, dataset, and the proposed method. Section 4 presents experiments and evaluation metrics, then discusses results and findings of the research. Finally, the conclusion is provided in Section 5.

2. RELATED WORKS

In recent years, researchers have investigated a variety of approaches to identify SMS spam messages using machine learning techniques. However, many of these published solutions for filtering spam messages have remained at their initial stage of classification, and are not very mature and reliable [10, 11]. In this section, we will overview some of the recently published SMS spam detection approaches and discuss their achievements and weaknesses.

Zainal et al. [12] introduced a Bayesian method which was developed using the RapidMiner and Weka tools; they have used two freeware tools to perform the experiment. They have run the experiments on the UCI repository dataset. The outcomes of their research show that both tools produce almost a similar result on the same dataset with the same clustering and classification methods. El-Alfy and AlHasan [13] introduced a method to identify spam messages on both email and SMS platforms. They investigated and tested many techniques and features to achieve the best set of features with low complexity. They used Support Vector Machine (SVM) and Naïve Bayes techniques with 11 different features. Finally, they evaluated the performance of their methods on five SMS and email datasets.

Nuruzzaman et al. [14] introduced a system for SMS spam filtering. They evaluated performance of their SMS spam detection method with the help of text classification methods on independent mobile phones. They performed all of their processes such as filtering, training, and updating using independent mobile phones. The result of their method showed their model was able to filter SMS spam messages while using less storage space and achieving an acceptable classification accuracy. In another research, Chan et al. [15] proposed two approaches for detecting and filtering SMS spam. The focuss of their method was on the weight and length of each message; they also evaluated the experiment on SMS messages and comments datasets. Uysal et al. [16] introduced a new approach for filtering SMS spam messages. They used a hybrid method of chi-square and information gain algorithms for feature selection purposes. Moreover, they introduced an android SMS spam filtering application, which is based on two different Bayesian classification methods. According to the author's result, their method is advantageous and can classify ham and spam messages with high classification accuracy.

Serrano et al. [17] presented a new method for SMS spam filtering based on writing style using extrinsic information. In their research, they saved spam and ham writing styles using sequential labeling and term clustering extraction techniques. All experiments were performed in Weka environment using 10-fold cross-validation. The result shows their method achieved good classification accuracy in tested plans and also produced an efficient low dimensional feature space. Junaid and Farooq [18] proposed a system to identify SMS spam messages on a mobile phone using an evolutionary learning classifier. Their system used hexadecimal format SMS messages, where they extracted two features. They estimated the possibility and performance of many evolutionary and non-evolutionary classifiers in their research. Finally, the result of the experiments recommends using supervised classifiers, and their introduced model obtained over 89% percent classification accuracy. Hidalgo et al. [19] have studied the role of the Bayesian detection approaches on two datasets, one of the datasets was in English, and another one in Spanish. The outcomes of their study showed the Bayesian techniques were useful in filtering spam emails and also could be effective in SMS spam classification.

Choudhary and Jain [20] presented a system for detection of SMS spam messages. The main aim of their introduced method was to identify spam messages efficiently while maintaining high performance. They extracted and selected ten different features and applied them in the evaluation of their approach. The basis of their investigation in their experiments were True Positive (TP) rate, False Positive (FP) rate, precision, and F-measure. In their research, they compared various classification algorithms, and among them, the Random Forest algorithm achieved the best results with 96.1% TP rate. In similar research, Suleiman and Al-Naymats [21] introduced a new method for filtering SMS spam messages using the HBO framework. They applied the HBO framework for feature selection, and extracted ten features for SMS spam messages identification. They then compared selected features on various machine learning algorithms. Finally, the random forest algorithm achieved the highest classification accuracy results in their system and classified 96% of messages accurately.

This section reviewed the advantages of recent developed approaches in detecting and filtering SMS spam messages while also noting their weakness and limitations. According to the literature, many different SMS spam detection techniques have been introduced in the past years, but there is still no comprehensive approach towards solving the problem, which means many of the existing solutions are not accurate enough. Therefore, in the current study, we propose a machine learning technique to identify SMS spam messages with high performance and acceptable classification accuracy.

3. PROPOSED METHODOLOGY

In this section, the used dataset and feature extraction process are described in the beginning, and then technical details of the proposed method are presented.

3. 1. Data Collection In the current research, we have used a dataset consists of 5,574 text messages classified as spam and ham (legitimate), which is publicly available in the UCI machine learning repository [22]. 747 of the messages are spam while 4,827 of them were labelled as ham. The format of the dataset is text where every individual line describes a message that includes two parts, the message string, and its category.

3. 2. Feature Extraction and Selection The features extraction phase is a critical task in the detection of spam messages since the choice of features can significantly affect the performance of machine learning techniques. Therefore, in most cases, it is a challenging task to discover the most useful features that can efficiently classify SMS spam messages. Hence, features with the best correlation should be selected to improve the detection rate and produce a shorter process time [23]. In the feature extraction phase, the C# .net framework was used to read lines from the dataset and to extract features from text messages and save them in a new structure. Furthermore, in order to find the most efficient features for detection of SMS spam messages, we have thoroughly investigated different characteristics of spam messages and selected features which are essential and helpful for identification of these messages in our dataset. The extracted features that were utilized in the experiments of our study are illustrated in Table 1 as follows:

In Table 2, we have described how each attribute value was extracted from the dataset of spam and ham messages.

3. 3. Proposed Model Artificical neural networks are somewhat inspired by neurons in the human brain,

TABLE 1. Description of selected features

Index	Feature	Feature Description
1	URLs	The existence of URL in the message
2	Punctuation marks	The existence of dot and comma symbols in a message seems to be a good sign for legitimate messages since people use dots for separate sentences and chatting.
3	Mathematical Symbols	Spammers usually use mathematical symbols in their spam messages. Symbols like: +, −, <, >, / and ^.
4	Special symbols	The existence of special symbols in a message will usually signify that a message is spam since many spammers use these symbols for different reasons. Special symbols like: “$”, “!”, &, #, ~, and *.
5	Emoji symbols	Many normal people use emoji symbols in their messages, and it seems like a good sign for detecting legitimate messages. Symbols like: :), :*, :p, :-), :’(, etc.
6	Uppercased words	Many spammers usually use uppercased words to gain the user’s attention. For example, UNLIMITED, AWARD, URGENT, WINNER, FREE, DATE, etc.
7	Phone Number	Many spammers usually send a phone number in a message, requesting users to call that given number, and eventually steal their personal information in the process.
8	Special Keywords	Many spam messages contain some suspicious keywords such as cash, ringtone, bonus, congrats, prize, voucher, etc. These keywords could reflect spam messages.
9	Message Length	Defined as the total number of characters in the message.
10	Number of Words	Defined as the total number of words in the message. Usually, spam messages contain a large number of words.

TABLE 2. List of extracted features for spam and ham messages

Feature Name	**Text Messages**	
	"I like you peoples very much:) but am very shy pa." *(Legitimate message)*	"SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day,6days, 16+ TsandCs apply Reply HL 4 info" *(Spam message)*
URLs	No	No
Punctuation marks	Yes	Yes
Mathematical Symbols	No	Yes
Special symbols	No	Yes
Emoji symbols	Yes	No
Uppercased words	No	Yes
Phone Number	No	No
Special Keywords	No	Yes
Message Length	50	137
Number of Words	11	26

and they have a wide range of application in different domains. In artificial intelligence, neural networks have made numerous breakthroughs and help many researchers to introduce solutions for many real-world problems [24]. In this research, the primary goal of the proposed approach is to identify the SMS spam messages with high detection accuracy using the model-averaged neural network (avNNet). Hence, in the first step, we collected the dataset and finalized the features for the experiment. Next, the features extracted from all of the SMS messages in the dataset were employed to create feature vectors which were used for training and testing the proposed model. Finally, we applied the extracted features to the proposed model to classify SMS messages. The architecture of the proposed model is presented in Figure 1.

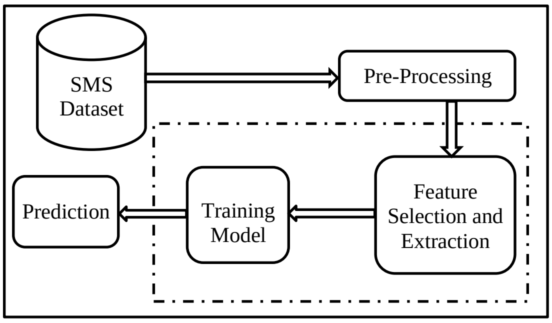

Figure 1. The structure of the Proposed Method

4. EXPERIMENTS AND RESULT ANALYSIS

This section includes the experimental setup, evaluation metrics, and outcomes of our research. Subsection 4.1 describes the preparation process and method of experimentation, subsection 4.2 describes performance evaluation measures to evaluate performance of the proposed method, and subsection 4.3 describes analysis on the obtained results to determine its effectiveness in detection of SMS spam messages. Moreover, achieved results were compared with the results of other techniques.

4. 1. Experimental Setup

The proposed model is implemented using the R programming language and the Caret package. The avNNet method employed for this research consistes of one hidden layer, and this layer forms the output using input information it receives from the predictors. Finally, the output consists of the averaged output of five individual models.

In general, performance of a neural networks is inspired by hyperparameters such as the number of hidden neurons in the hidden layer or values of decay variable which restricts neuronal connections weights.

Therefore, we have applied grid searches in order to find the optimal set of decay variable value and hidden neurons. The achieved result of this process was then used to tune the model. The details of the additional parameters used for model tuning is exhibited in Table 3.

To have reliable results, we splitted 80 percent of data for training and 20 percent for validation purpose. Moreover, to improve performance of the model in predicting unknown instances, we used the well-known k-fold cross-validation technique with k=10 for model training.

4. 2. Evaluation Metrics

The proposed model's performance was evaluated using standard evaluation metrics, namely, accuracy (ACC), precision, recall value, F-measure, and the area under the receiver-operating characteristic curve (AUC). These measurements are calculated with the following equations:

TABLE 3. Neural Network – detailed parameters.

Parameter	**Setting**
Size	6
Decay	0.1
Bagging	False
Regularization	L2
Maximum number of iterations	100

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F-measure = \frac{2\times Precision\times Recall}{Precision+Recall}, \tag{4}$$

where TP is an abbreviation for true-positive and expresses cases where the prediction and sample lable are both positive, FN is false-negative and denotes cases when the prediction is negative while the sample label is positive. TN is true-negative and expresses conditions where the prediction and sample label are both negative, and FP is false-positive and denotes cases where the sample label is negative, while the prediction is positive.

4. 3. Result Analysis We have performed a series of experiments to measure the performance of the proposed model in identifying spam messages. Our main objective was to investigate the detection capabilities of the proposed model on actual SMS corpus, then compare it with several benchmark classification algorithms. Initially, we selected best-correlated features based on ham and spam message behaviors. Next, all selected features were extracted from SMS messages dataset in order to create the feature vector. Then, the extracted features are applied using the proposed approach to get the performance metrics. Finally, the proposed model results are compared with several classification algorithms, and some recently published works which were benchmarked on the same dataset. The achieved results of the proposed model and other classification techniques are presented in Table 4.

As presented in Table 4, after comparing the performance metrics for different researches, we analyzed that our neural network model achieved the highest results in comparison with several classification techniques. The proposed method achieved 0.988% accuracy and 0.9929% F-measure rate, respectively, and outperformed other classification methods in terms of classification accuracy and F-measure rate. The efficiency of the model is measured based on the percentage of correctly classified messages.

The receiver operating characteristic curve (Roc Curve) is a graphical tool used for visualizing the diagnostic ability of a model in order to discover whether a model is suitable with regards to cost sensitivity. In Figure 2, the curve's x-axis represents the true-positive rate (sensitivity) while the y-axis represents the false positive rate (FPR). The ROC curve result indicates that it is a suitable model for classifying spam and legitimate messages.

To have a better evaluation of the achieved results for the introduced model, we have done a comparative study to compare the obtained results of the proposed model with previous researches. The results of this comparative study are shown in Table 5.

Next, we show in Figure 3 the distributions of features' importance in the proposed model.

The data presented in Figure 3 shows that the Message_Length features have more influence on class prediction than other considered features. Moreover, the role of Special_keywords and Special_symbols features was significant in the identification of spam messages. It also represents that these keywords are one of the best differentiators between the spam and legitimate messages as many spammers usually use these keywords in their messages to get the attention of users.

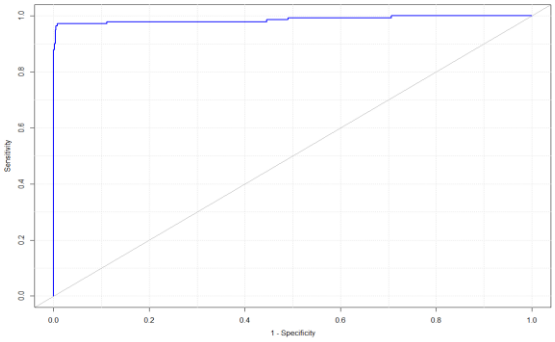

Figure 2. ROC curve for the ham (legitimate) class. (AUC=0.9872)

TABLE 4. Comparisons between the proposed method and several classification techniques

Algorithm	Accuracy	Precision	F1	Recall
Random Tree	0.9671	0.967	0.967	0.967
Naive Bayes	0.9634	0.964	0.963	0.963
J48	0.9726	0.972	0.972	0.972
Decision Stump	0.9332	0.933	0.935	0.933
HoeffdingTree	0.9675	0.966	0.966	0.967
Proposed Model	**0.988**	**0. 9892**	**0. 9929**	**0. 9967**

TABLE 5. Comparisons between the proposed model and previous researches

Research Paper	Method	Accuracy
Content based spam detection [13]	SVM	95.5
Filtering of SMS Spam Messages [20]	Random Forest	96.5
SMS Spam Detection using H2O Framework [21]	H2O+Random Forest	96.1
Proposed Model	Hybrid Model	98.8

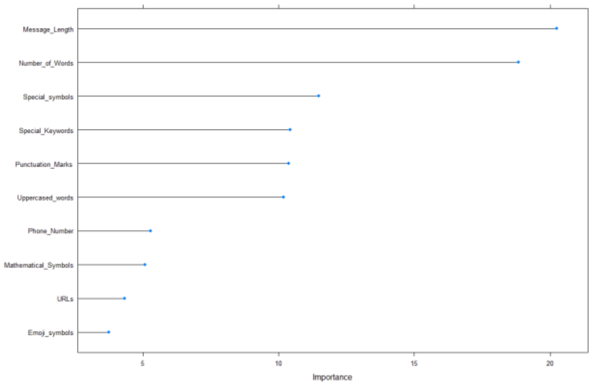

Figure 3. variable importance information

Moreover, we used different SMS datasets with the rising number of messages in each of them to identify the best feature type and evaluate the performance of the model in different sizes data set. For this evaluation, we considered accuracy and F-measure metrics as the basis of our investigation. F-measure is one of the essential metrics used in many classification studies; which is an overall estimation of the efficiency of a classification technique because it is a combination of both recall and precision. When the algorithm is measured using this metric, it shows its capability in detecting legitimate and spam messages. Hence, in Table 6, we summarize the F-measure, accuracy, and the best feature type for each set of the SMS datasets.

As shown in Table 6, the "Message_Length" feature has the highest results in most sets of data. It has the most impacts on the obtained results in terms of F-measure and accuracy. The closest result in the evaluation was obtained by "Special_Keywords" and "Number_of_Words" features. These results also indicate that there is a progressive improvement in detection rates as the dataset size increases. This constant increase in detection rate shows that the extracted features performed efficiently on different size of data sets. In this research, we have worked with a dataset of 5574 SMS messages, and the present evaluation explains that increment in dataset size could lead the model to achieve a higher detection rate, which means more records can help us to produce a better classification accuracy.

TABLE 6. Results of the proposed model on datasets with various sizes

NO. of Items	Best Feature	F-measure	Accuracy
1116	Number_of_Words	0.9891	98.1
2230	Message_Length	0.9833	97.1
3346	Message_Length	0.9872	97.8
4460	Number_of_Words	0.9870	97.8
5574	Message_Length	0.9929	98.8

2. CONCLUSION

In this paper, we aimed to develop an accurate model to identify SMS spam messages based on content-based features using averaged neural networks. In the proposed model, we extracted the most relevant features from messages on the UCI SMS spam dataset, which contains over 5000 messages. Then we applied averaged neural networks method on extracted features in order to classify them into two categories of ham and spam messages. The evaluation results show that the extracted features have a high correlation with the class of messages, and averaged neural network algorithm can effectively distinguish the class of messages with good classification accuracy and high F-measure rate.

Moreover, the proposed model was investigated and compared against several recently proposed algorithms concerning classification accuracy and F-measure on the same dataset. The results demonstrated that the proposed model outperformed other considered research in terms of accuracy and F-measure, and had a high performance in identifying SMS spam messages with good classification accuracy. Furthermore, the paper findings also show that classification based on content-based features is a more useful metric for filtering spam messages since most spammers use suspicious content in their messages. The accuracy of the proposed model can be improved further on evaluation using bigger datasets and finding a standard sizable dataset which was one of limitations in this research.

2. REFERENCES

1. Cormack, G.V., "Email spam filtering: A systematic review", ***Foundations and Trends® in Information Retrieval***, Vol. 1, No. 4, (2008), 335-455.
2. Almeida, T.A., Hidalgo, J.M.G. and Yamakami, A., "Contributions to the study of sms spam filtering: New collection and results", in Proceedings of the 11th ACM symposium on Document engineering., (2011), 259-262.
3. Parandeh Motlagh, F. and Khatibi Bardsiri, A., "Detecting fake websites using swarm intelligence mechanism in human learning", ***International Journal of Engineering***, ***Transactions A: Basics,*** Vol. 31, No. 10, (2018), 1642-1650.
4. Mohammadi, A. and Hamidi, H., "Analysis and evaluation of privacy protection behavior and information disclosure concerns in online social networks", ***International Journal of Engineering***, ***Transactions B: Applications,*** Vol. 31, No. 8, (2018), 1234-1239.
5. Jain, A.K. and Gupta, B.B., "Phishing detection: Analysis of visual similarity based approaches", ***Security and Communication Networks***, Vol. 2017, No., (2017).
6. Yamakami, T., "Impact from mobile spam mail on mobile internet services", in International Symposium on Parallel and

Distributed Processing and Applications, Springer., (2003), 179-184.

7. Gupta, B.B., Tewari, A., Jain, A.K. and Agrawal, D.P., "Fighting against phishing attacks: State of the art and future challenges", ***Neural Computing and Applications***, Vol. 28, No. 12, (2017), 3629-3654.

8. Choudhary, N. and Jain, A.K., "Comparative analysis of mobile phishing detection and prevention approaches", in International Conference on Information and Communication Technology for Intelligent Systems, Springer., (2017), 349-356.

9. Puniškis, D., Laurutis, R. and Dirmeikis, R., "An artificial neural nets for spam e-mail recognition", ***Elektronika ir Elektrotechnika***, Vol. 69, No. 5, (2006), 73-76.

10. Ji, H. and Zhang, H., "Analysis on the content features and their correlation of web pages for spam detection", ***China Communications***, Vol. 12, No. 3, (2015), 84-94.

11. Kim, S.-E., Jo, J.-T. and Choi, S.-H., "Sms spam filterinig using keyword frequency ratio", ***International Journal of Security and Its Applications***, Vol. 9, No. 1, (2015), 329-336.

12. Zainal, K., Sulaiman, N. and Jali, M., "An analysis of various algorithms for text spam classification and clustering using rapidminer and weka", ***International Journal of Computer Science and Information Security***, Vol. 13, No. 3, (2015), 66.

13. El-Alfy, E.-S.M. and AlHasan, A.A., "Spam filtering framework for multimodal mobile communication based on dendritic cell algorithm", ***Future Generation Computer Systems***, Vol. 64, (2016), 98-107.

14. Taufiq Nuruzzaman, M., Lee, C., Abdullah, M.F.A.b. and Choi, D., "Simple sms spam filtering on independent mobile phone", ***Security and Communication Networks***, Vol. 5, No. 10, (2012), 1209-1220.

15. Chan, P.P., Yang, C., Yeung, D.S. and Ng, W.W., "Spam filtering for short messages in adversarial environment", ***Neurocomputing***, Vol. 155, (2015), 167-176.

16. Uysal, A.K., Gunal, S., Ergin, S. and Gunal, E.S., "A novel framework for sms spam filtering", in 2012 International Symposium on Innovations in Intelligent Systems and Applications, IEEE., (2012), 1-4.

17. Serrano, J.M.B., Palancar, J.H. and Cumplido, R., "The evaluation of ordered features for sms spam filtering", in Iberoamerican Congress on Pattern Recognition, Springer., (2014), 383-390.

18. Junaid, M.B. and Farooq, M., "Using evolutionary learning classifiers to do mobilespam (SMS) filtering", in Proceedings of the 13th annual conference on Genetic and evolutionary computation., (2011), 1795-1802.

19. Gómez Hidalgo, J.M., Bringas, G.C., Sánz, E.P. and García, F.C., "Content based sms spam filtering", in Proceedings of the 2006 ACM symposium on Document engineering., (2006), 107-114.

20. Choudhary, N. and Jain, A.K., "Towards filtering of sms spam messages using machine learning based technique", in International Conference on Advanced Informatics for Computing Research, Springer., (2017), 18-30.

21. Suleiman, D. and Al-Naymat, G., "Sms spam detection using h2o framework", ***Procedia Computer Science***, Vol. 113, (2017), 154-161.

22. Dua, D. and Graff, C., *Uci machine learning repository*. 2017.

23. Gholami, M., "Islanding detection method of distributed generation based on wavenet", ***International Journal of Engineering***, ***Transactions B: Applications***, Vol. 32, No. 2, (2019), 242-248.

24. Gharvirian, F. and Bohloli, A., "Neural network based protection of software defined network controller against distributed denial of service attacks", ***International Journal of Engineering***, ***Transactions B: Applications,*** Vol. 30, No. 11, (2017), 1714-1722.

An Effective Model for SMS Spam Detection Using Content-based Features and Averaged Neural Network

S. Sheikhi, M. T. Kheirabadi, A. Bazzazi

Department of Computer, Gorgan Branch, Islamic Azad University, Gorgan, Iran

PAPER INFO

Paper history:
Received 04 August 2019
Received in revised form 06 November 2019
Accepted 16 Januray 2020

Keywords:
Spam
Mobile Spam
SMS Spam
Anti-Spam
Machine Learning

چکیده

در سال های اخیر، علاقه بسیار زیادی در بین مردم برای استفاده از خدمات پیام کوتاه (SMS) به عنوان یکی از خدمات ارتباط اساسی در دستگاه های تلفن همراه به وجود آمده است. این محبوبیت باعث افزایش تعداد حملات به دستگاه های تلفن همراه مانند هرزنامه ها شده است. هرزنامه ها به یک مشکل واقعی برای مشترکین تلفن همراه بدل شده است. این موضوع حتی باعت نگرانی ارائه دهندگان خدمات مخابراتی نیز شده است، زیرا باعث آزرده شدن مشترکان آنها می شود و یا حتی می تواند باعث از دست رفتن تجارت آنها شود. بنابراین، در این مقاله، ما یک روش یادگیری ماشین جدید برای تشخیص پیام های اسپم ارائه داده ایم. مدل ارائه شده شامل دو مرحله اصلی است: استخراج ویژگی و تصمیم گیری. در مرحله اول، ما ویژگی های و اطلاعات مربوطه را از مجموعه داده ها بر اساس ویژگی های شخصیتی پیام های اسپم و مشروع به منظور کاهش پیچیدگی و بهبود عملکرد مدل استخراج کردیم. سپس از یک مدل شبکه عصبی به منظور طبقه بندی پیام ها به کلاس های اسپم و مشروع با استفاده از ویژگیهای استخراج شده استفاده کردیم . این روش بر یک مجموعه داده اس ام اس واقعی با بیش از ۵۰۰۰ پیام از نظر دقت و معیارهای اندازه گیری F1 مورد ارزیابی قرار گرفت. علاوه بر این، نتایج به دست آمده در این مقاله با سه تحقیق اخیر در این زمینه مقایسه و ارزیابی شده است. مبنای تحقیق در این مقاله دقت طبقه بندی و معیار ارزیابی F1 است. نتایج بدست آمده نشان می دهد که رویکرد ارائه شده نرخ تشخیص بالایی به دست آورده و از نظر معیارهای ارزیابی F1 و دقت طبقه بندی در مقایسه با دیگر تحقیقات در نظر گرفته شده بسیار موفق عمل کرده است.

***doi**: 10.5829/ije.2020.33.02b.06*