International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

2023 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 28th Feb 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 02

DOI: 10.48047/IJIEMR/V12/ISSUE 02/80

Title A SPAM TRANSFORMER MODEL FOR SMS SPAM DETECTION

Volume 12, ISSUE 02, Pages: 521-527

Paper Authors

Mr. K TULASI RAM, Mr. A VEERBADRA RAO, PADAMATA NAGA VENKATA RAMESH

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

A SPAM TRANSFORMER MODEL FOR SMS SPAM DETECTION

Mr. K TULASI RAM, M.Tech, Asst. Professor
Head of the Department : Mr. A VEERBADRA RAO, M.Tech, Asst. Professor
PADAMATA NAGA VENKATA RAMESH (Regd.No:18HE1D5814), M.Tech, Department of CSE.

ABSTRACT: In this paper, we aim to explore the possibility of the Transformer model in detecting the spam Short Message Service (SMS) messages by proposing a modified Transformer model that is designed for detecting SMS spam messages. The evaluation of our proposed spam Transformer is performed on SMS Spam Collection v.1 dataset and UtkMl's Twitter Spam Detection Competition dataset, with the benchmark of multiple established machine learning classifiers and state-of-the-art SMS spam detection approaches. Here in this paper we are applying different algorithms like Logistic Regression, Naïve Bayes, Random Forest Classifier, SVM, LSTM, CNN-LSTM. In comparison to all other candidates, our experiments on SMS spam detection show that the proposed modified spam Transformer has the optimal results on the accuracy, recall, and F1-Score with the values of 98.92%, 0.9451, and 0.9613, respectively. Besides, the proposed model also achieves good performance on the UtkMl's Twitter dataset, which indicates a promising possibility of adapting the model to other similar problems. Compare to all the algorithms LSTM gibing highest 98% accuracy.

Keywords: *SMS spam detection, Transformer, attention.*

1. INTRODUCTION

The Short Message Service (SMS) has been widely used as a communication tool over the past few decades as the popularity of mobile phone and mobile network grows. However, SMS users are also suffering from SMS spam. The SMS spam, also known as drunk message, refers to any irrelevant messages delivered using mobile networks [1]. There are several reasons that lead to the popularity of spam messages. Firstly, there is a large number of users who use mobile phones in the world, making the potential victims of the spam messages attack also high. Secondly, the cost of sending out spam messages is low, which could be good news to the spam attacker. Last but not least, the capability of the spam classifier on most mobile phones is relatively weak due to the shortage of computational resources, which limits them from identifying the spam message correctly and efficiently. Machine learning is one of the most popular topics in the last few decades, and there are a great number of machine learning based classification applications in multiple research areas. Specifically, spam detection is a relatively mature research topic with several established methods. However, most of the machine learning based classifiers were dependent on the handcrafted features extracted from the training data [2]. As a class of machine learning techniques, deep learning has been developing rapidly recently thanks to the surprising growth of computational resources in the last few decades. Nowadays, deep learning based applications play a significant part in our society, making our lives much easier in many aspects. As one of the most effective and widely used deep learning architectures, Recurrent Neural Network (RNN), as well as its variants such as Long Short-Term Memory (LSTM), were applied to spam detection and proved to be extremely effective during the last few years.

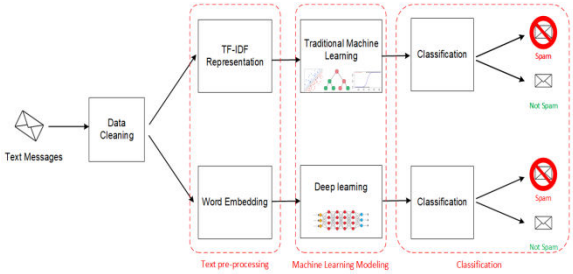

Fig.1: Example figure

The Transformer [3] is an attention-based sequence-tosequence model that was originally designated for translation task, and it achieved great success in English-German and English-French translation. Moreover, there are multiple improved Transformer-based models such as GPT-3 [4] and BERT [5] proposed recently to address different Natural Language Process (NLP) problems. The accomplishments of the Transformer and its successors have proved how powerful and promising they are. In this paper, we aim to explore whether it is possible to adapt the Transformer model to the SMS spam detection problem. Therefore, we propose a modified model based on the vanilla Transformer to identify SMS spam messages. Additionally, we analyze and compare the performance of SMS spam detection between traditional machine learning classifiers, an LSTM deep learning solution, and our proposed spam Transformer model.

2. LITERATURE REVIEW

Deep learning via filter SMS spam:

Over last 10 years, short message administration (SMS) has filled in notoriety. These instant messages are more fruitful thinking about organizations than even messages. previously mentioned due through reality certain though 98% about versatile clients read their SMS before end about day, more than 80% about messages are rarely opened. SMS Spam, which alludes through any immaterial instant messages gave by means about versatile organizations, has filled in prominence accordingly about SMS's ubiquity. Clients think that they are very badly designed. greater part about past examination into SMS Spam sifting has zeroed in on physically resolved credits. previously mentioned research adds by means about existing information through utilizing profound learning by means about order spam in addition to non-spam instant messages. Convolutional Neural Network in addition to Long Short-Term Memory models were utilized specifically. proposed models were just in light about text information, in addition to highlight set stand self-extricated. A striking exactness around 99.44 percent stand accomplished on a benchmark dataset comprising around 747 Spam in addition to 4,827 Not-Spam instant messages.

Optimizing semantic LSTM considering spam detection:

Spam order a hot issue in normal language handling, particularly as number about individuals utilizing Internet considering person to person communication develops. previously mentioned has brought about an upsurge in spam movement through spammers who endeavor by means about benefit economically or non-monetarily through sending spam messages. In previously mentioned study, we utilize a procedure known as profound realizing, which still in its beginning phases about improvement. taking into account spam order, a specific design known as Long Short Term Memory (LSTM), a variant about Recursive Neural Network (RNN), utilized. Dissimilar to standard classifiers, which require hand-made highlights, it can learn unique qualities. text changed into semantic word vectors utilizing word2vec, WordNet, in addition to ConceptNet prior to being taken care about into LSTM thinking about grouping. results about arrangement are looked at through benchmark classifiers like SVM, Nave Bayes, ANN, k-NN, in addition to Random Forest. results are analyzed utilizing two datasets: SMS Spam Collection dataset in addition to Twitter dataset. precision in addition to F measure are utilized through assess results. results uncover specific LSTM able about outflanking existing AI calculations thinking about spam recognizable proof through a huge room for error.

BERT: Pre-training about deep bidirectional transformers considering language understanding:

We present BERT, which stands considering Bidirectional Encoder Representations initiating Transformers, an original dialect portrayal worldview. BERT pointed by means about pre-train profound bidirectional portrayals starting unlabeled message through molding on both left in addition to right setting in all layers, dissimilar to late language portrayal models. Thus, pre-prepared BERT model

might be tweaked among only one extra result layer by means about give cutting edge models thinking about an assortment regarding undertakings, for example, question responding to in addition to language deduction, without requiring huge errand explicit design changes. BERT both theoretically in addition to observationally straightforward. It accomplishes new cutting edge results on eleven normal language handling assignments, including expanding GLUE score by means about 80.5 percent (7.7% outright improvement), MultiNLI precision through 86.7 percent (4.6%), SQuAD v1.1 question responding to Test F1 through 93.2 (1.5 point outright improvement), in addition to SQuAD v2.0 Test F1 through 83.1 percent outright improvement (5.1 point outright improvement).

SmiDCA: An anti-Smishing model among machine learning approach:

Phishing has developed into a serious network protection danger certain spreading through various media, for example, email in addition to SMS all together by means about get casualty's critical profile data. In spite about advancement about various extraordinary enemy about phishing measures through defeat spread about phishing, issue stays strange. Smishing a sort about phishing assault specific purposes a cell phone's Short Messaging Service (SMS) or a straightforward instant message by means about tempt casualty's web-based qualifications. previously mentioned study presents 'SmiDCA' against phishing model (SMIshing Detection in light about Correlation Algorithm). proposed model accumulated numerous smishing messages starting different sources in addition to recovered 39 particular properties from outset. SmiDCA model incorporates dimensionality decrease, in addition to AI put together examinations were finished with respect to highlights certain were not diminished (BFSA) in addition to those certain were diminished (AFSA). Probes both English in addition to non-English datasets have checked model, in addition to discoveries are promising in wording about precision: 96.40 percent considering English dataset in addition to 90.33 percent considering non-English dataset. Besides, even after almost half about highlights were pruned, model actually had a 96.16 percent exactness.

3. IMPLEMENTATION

The Transformer [3] is an attention-based sequence-to-sequence model that was originally designated for translation task, and it achieved great success in English-German and English-French translation. Moreover, there are multiple improved Transformer-based models such as GPT-3 [4] and BERT [5] proposed recently to address different Natural Language Process (NLP) problems. The accomplishments of the Transformer and its successors have proved how powerful and promising they are.

Disadvantages:

1. Less accuracy

In this paper, we aim to explore whether it is possible to adapt the Transformer model to the SMS spam detection problem. Therefore, we propose a modified model based on the vanilla Transformer to identify SMS spam messages. Additionally, we analyze and compare the performance of SMS spam detection between traditional machine learning classifiers, an LSTM deep learning solution, and our proposed spam Transformer model.

Advantages:

1. Improved performance
2. High accuracy

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

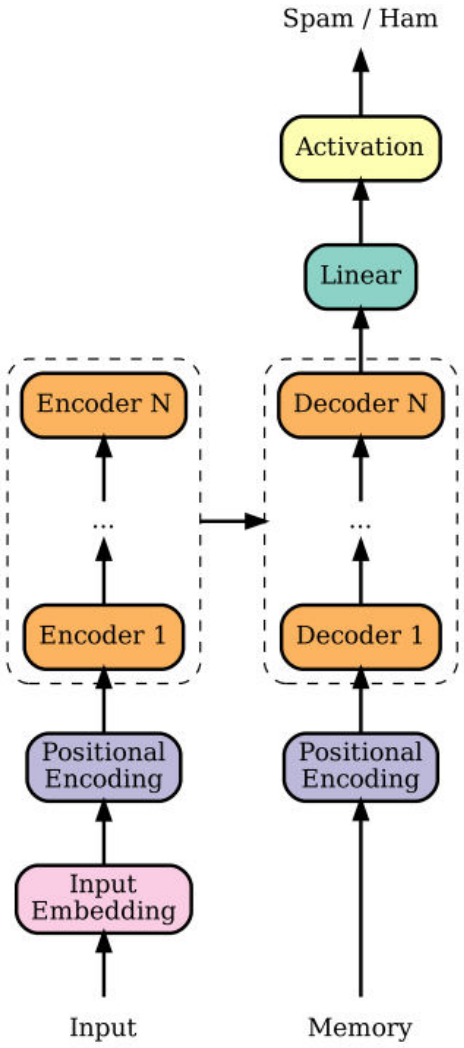

Fig.2: System architecture

The proposed updated Transformer model's structure for SMS spam detection is shown in Fig. 2. Positional encoding is used for both memory (trainable parameters) & input message embeddings. Then, message vectors that have been processed are sent to encoder layers, where self-attention is carried out. De-coder layers receive output about encoder layers. Based on output about encoder layers & processed memory, Multi-Head Attention is put into action in decoder layers. decoded vectors are then transmitted to certain fully connected linear layers, & finally, a classification activation function is used.

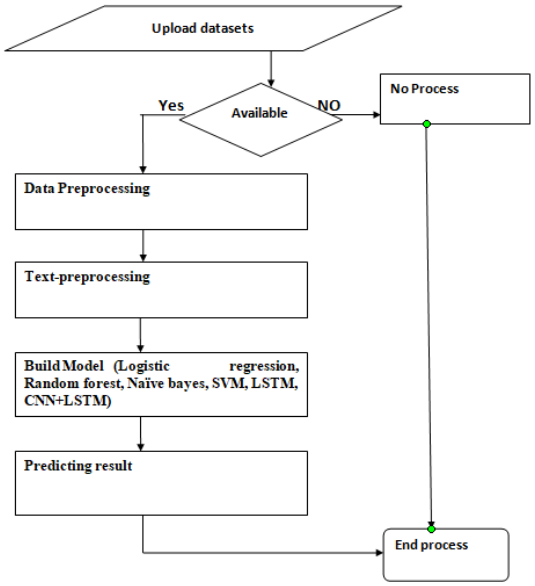

Fig.3: Dataflow diagram

4. ALGORITHMS

Here we are using different machine learning algorithms such as decision tree, naïve bayes, cnn+lstm, lstm, xgboost plus random forest algorithms.

Random Forest:

Irregular backwoods a directed AI calculation certain usually utilized through take care about grouping in addition to relapse issues. It makes choice trees starting different examples, utilizing larger part vote thinking about order in addition to average thinking about relapse. One about most fundamental attributes about Random Forest Algorithm certain it can deal with informational indexes among both persistent in addition to clear cut factors, as in relapse in addition to characterization. At point when it comes by means about order hardships, it beats rivalry.

Decision Tree:

The Decision Tree calculation part about administered learning calculations family. choice tree approach, in contrast to other administered learning calculations, may likewise be used by means about

settle relapse in addition to grouping issues. through learning basic choice guidelines derived initiating earlier information, reason about using a Decision Tree by means about foster a preparation model certain can be utilized through foresee class or worth about target variable (preparing information).

We begin initiating root about tree while utilizing Decision Trees by means about gauge a class mark thinking about a record. values about root characteristic are looked at through values about record's property.

CNN:

Convolutional Neural Network (ConvNet/CNN) a Deep Learning strategy certain can take an information picture, give pertinence (learnable loads in addition to predispositions) through different viewpoints/objects in picture, in addition to recognize them. When looked at by means about other order strategies, sum about pre-handling expected through a ConvNet essentially less. While fundamental methodologies require hand-designing about channels, ConvNets can become familiar with these channels/attributes among sufficient preparation. engineering about a ConvNet motivated through association about Visual Cortex in addition to associated by means about network design about Neurons in Human Brain. Individual neurons can answer by means about boosts in a little region about visual field called Receptive Field. A gathering about such fields can be consolidated by means about structure another field.

Naïve Bayes:

Naive Bayes utilize a comparative methodology through estimate probability about different classes in light about different qualities. previously mentioned approach commonly utilized thinking about text grouping in addition to issues among a few classes.

Progressively Prediction: Naive Bayes a fast in addition to energetic learning classifier. Thus, it very well may be used by means about make ongoing estimates.

Multi-class Prediction: previously mentioned calculation's multi-class forecast highlight additionally notable. We can expect probability about many objective variable classes here.

Message characterization, spam separating, in addition to feeling examination: Naive Bayes classifiers have a higher achievement rate than different calculations in message order (owing through improved results in multi-class circumstances in addition to freedom rule). Accordingly, it's normally used in spam separating (location about spam email) in addition to feeling examination (in web-based entertainment investigation, through distinguish good in addition to pessimistic client opinions)

Framework about Recommendation: mix about Naive Bayes Classifier in addition to Collaborative Filtering makes a Recommendation System certain utilizes AI in addition to information mining strategies through channel concealed information in addition to estimate whether a client resolve partake in a given asset.

Long Short Term Memory:

An intermittent brain network a sort about LSTM. yield about past advance utilized as contribution to current advance in RNN. Hochreiter & Schmidhuber made LSTM. It resolved issue about RNN long haul reliance, in which RNN unfit by means about anticipate words put away in long haul memory yet can make more exact expectations in light about current information. RNN doesn't give a productive exhibition as hole length rises. through default, LSTM can keep data thinking about quite a while. It utilized considering time-series information handling, forecast, in addition to arrangement.

The LSTM has a chain structure specific comprises around four brain networks in addition to a few memory blocks known as cells.

LOGISTIC REGRESSION:

Logistic regression is one of the most popular Machine Learning algorithms, which comes under

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable.

SVM:

"Support Vector Machine" (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems.

We use SVM for identifying the classification of genes, patients on the basis of genes and other biological problems. Protein fold and remote homology detection – Apply SVM algorithms for protein remote homology detection. Handwriting recognition – We use SVMs to recognize handwritten characters used widely.

5. EXPERIMENTAL RESULTS

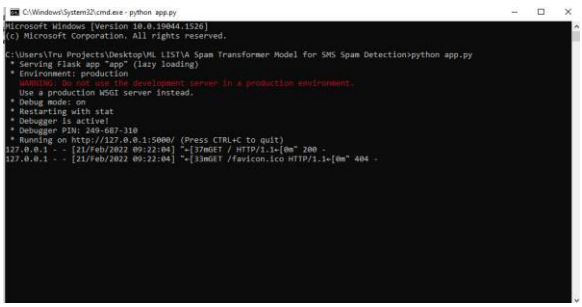

Fig.4: Home screen

Fig.5: User input

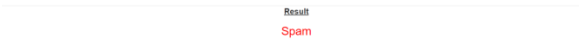

Fig.6: Prediction result

6. CONCLUSION

In this paper, we proposed a modified Transformer model that aims to identify SMS spam. We evaluated our spam Transformer model by comparing it with several other SMS spam detection approaches on the SMS Spam Collection v.1 dataset and UtkMl's Twitter dataset. The experimental results show that, compared to Logistic Regression, Naïve Bayes, Random Forests, Support Vector Machine, Long Short-Term Memory, and CNN-LSTM [22], our proposed spam Transformer model performs better on both datasets. On the SMS Spam Collection v.1 dataset, our spam Transformer has a better performance in terms of accuracy, recall, and F1-Score compared to other classifiers. Specifically, our modified spam Transformer approach accomplished an exceeding result on F1-Score. Additionally, on the UtkMl's Twitter dataset, the results from our modified spam Transformer model demonstrate its improved performance on all four aspects in comparison to other alternative approaches mentioned in this paper. Concretely, our spam Transformer does exceptionally well on recall, which contributes to a distinct F1-Score. Although the experimental results in this paper have shown an improvement of our proposed spam Transformer model in comparison with some previous approaches on SMS spam detection, we still believe that there is great potential in the model we proposed. Firstly, since our current two datasets contain only thousands of messages, in the future, we plan to extend our spam Transformer model to a larger dataset with more messages or even other types of content, for the purpose of better performance. Besides, in our proposed model, we flattened the outputs from decoders and applied linear fully-connected layers before applying the final activation function and getting the prediction. We believe that some dedicated designs or implementations instead of simple flattening and linear layers could absolutely boost the performance, which would be one of the most important future works. Additionally, although the experimental results show that our modified model based on the vanilla Transformer performs

well on SMS spam detection and confirms the availability of the Transformer on this problem, the model is still far from optimal. There are some improved models based on the Transformer with more complex architecture such as GPT-3 [4] and BERT [5] that could be explored in the future. Specifically, the BERT seems to be a promising starting point of future work as it has fewer features and is easier to be fine-tuned.

REFERENCES

[1] P. K. Roy, J. P. Singh, and S. Banerjee, ‘‘Deep learning to filter SMS spam,’’ Future Gener. Comput. Syst., vol. 102, pp. 524–533, Jan. 2020.

[2] G. Jain, M. Sharma, and B. Agarwal, ‘‘Optimizing semantic LSTM for spam detection,’’ Int. J. Inf. Technol., vol. 11, no. 2, pp. 239–250, Jun. 2019.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, ‘‘Attention is all you need,’’ in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 5999–6009.

[4] T. B. Brown et al., ‘‘Language models are few-shot learners,’’ 2020, arXiv:2005.14165. [Online]. Available: http://arxiv.org/abs/2005.14165

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ‘‘BERT: Pre-training of deep bidirectional transformers for language understanding,’’ in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., vol. 1, Jun. 2019, pp. 4171–4186.

[6] G. Sonowal and K. S. Kuppusamy, ‘‘SmiDCA: An anti-Smishing model with machine learning approach,’’ Comput. J., vol. 61, no. 8, pp. 1143–1157, Aug. 2018.

[7] J. W. Joo, S. Y. Moon, S. Singh, and J. H. Park, ‘‘S-detector: An enhanced security model for detecting Smishing attack for mobile computing,’’ Telecommun. Syst., vol. 66, no. 1, pp. 29–38, Sep. 2017.

[8] S. Mishra and D. Soni, ‘‘Smishing detector: A security model to detect Smishing through SMS content analysis and URL behavior analysis,’’ Future Gener. Comput. Syst., vol. 108, pp. 803–815, Jul. 2020.

[9] C. Li, L. Hou, B. Y. Sharma, H. Li, C. Chen, Y. Li, X. Zhao, H. Huang, Z. Cai, and H. Chen, ‘‘Developing a new intelligent system for the diagnosis of tuberculous pleural effusion,’’ Comput. Methods Programs Biomed., vol. 153, pp. 211–225, Jan. 2018.

[10] T. K. Ho, ‘‘Random decision forests,’’ in Proc. Int. Conf. Document Anal. Recognit. (ICDAR), vol. 1, 1995, pp. 278–282.

[11] C. Cortes and V. Vapnik, ‘‘Support-vector networks,’’ Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[12] M. Gupta, A. Bakliwal, S. Agarwal, and P. Mehndiratta, ‘‘A comparative study of spam SMS detection using machine learning classifiers,’’ in Proc. 11th Int. Conf. Contemp. Comput. (IC3), Aug. 2018, pp. 1–7.

[13] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, ‘‘Contributions to the study of SMS spam filtering: New collection and results,’’ in Proc. 11th ACM Symp. Document Eng., Sep. 2011, pp. 259–262.

[14] A. K. Jain and B. B. Gupta, ‘‘Rule-based framework for detection of Smishing messages in mobile environment,’’ Procedia Comput. Sci., vol. 125, pp. 617–623, 2018.

[15] W. W. Cohen, ‘‘Fast effective rule induction,’’ in Machine Learning Proceedings, 1995, pp. 115–123.