International Journal for Modern Trends in Science and Technology, 9(05): 644-648, 2023
Copyright © 2023International Journal for Modern Trends in Science and Technology
ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0905109
Available online at: http://www.ijmtst.com/vol9issue05.html

A Robust Detection of Bots in Twitter

A.E. Kokila, K. Usha, M. Yamuna, B. Tejaswini Priya, K.V.L. Ramya and Ch. Sailaja

Department of Comuter Sciecne Engineering, Narayana Engineering College, Nellore, Andhra Pradesh, India

To Cite this Article
A.E. Kokila, K. Usha, M. Yamuna, B. Tejaswini Priya, K.V.L. Ramya and Ch. Sailaja. A Robust Detection of Bots in Twitter. International Journal for Modern Trends in Science and Technology 2023, 9(05), pp. 644-648. https://doi.org/10.46501/IJMTST0905109

Article Info
Received: 21 April 2023; Accepted: 18 May 2023; Published: 22 May 2023.

ABSTRACT

During the last decades, the volume of multimedia content posted in social networks has grown exponentially and such information is immediately propagated and consumed by a significant number of users. In this scenario, the disruption of fake news providers and bot accounts for spreading propaganda information as well as sensitive content throughout the network has fostered applied research to automatically measure the reliability of social networks accounts via Artificial Intelligence (AI). In this paper, we present a Deep learning (DL) approaches to support end-users when checking the credibility of a certain Twitter account. To do so, several experiments were conducted using text-based features of the user account that are later on concatenated with the rest of the metadata to build a potential input vector on top of a Dense Network denoted as Bot-DenseNet. Consequently, this paper assesses the language constraint from previous studies where the encoding of the user account only considered either the metadata information or the metadata information together with some basic semantic text features. Moreover, the Bot-DenseNet produces a low-dimensional representation of the user account which can be used for any application within the Information Retrieval (IR) framework.

KEYWORDS: Artificial intelligence, bot detector, deep learning, feature representation, language models, misinformation detection, social media mining.

INTRODUCTION

In recent years, social media platforms such as Twitter or Facebook have gained a large levelof both popularity and influence among millions of users due to the benefits of publishing, propagating and exchanging large volumes of multimedia content along the network. Therefore, these platforms allow users to establish a digital community as remarked in which has made possible not only to discover and embrace new relationships but to maintain and boost existing ones.

On the other hand, due to both the great influence these platforms have on the lifestyle of people and its evolving as a potential communication tool, they have exponentially promoted its attraction for marketing and commercial purposes by analysing the behaviour and opinion of users in different topics or events such as political elections. Consequently, numerous research studies have been fostered in the social media field with different purposes including sentiment analysis traffic control or consumer behaviour mining.

However, the considerable growth of social media platforms has also provoked the desire of altering people's opinion in certain topics by spreading propaganda or bias information. Many of these controlling procedures are carried out by Bots which are widely described in numerous investigations such as automatic systems which are capable of generating and spreading multimedia content throughout the network without the supervision of a human being.

Furthermore, with the disruptive growth of Artificial Intelligence (AI) algorithms, the identification of bots or non-reliable sources has become a crucial challenge to be investigated. It raised many studies and publications with the goal of building robust automatic systems to improve the quality of experience of consumers in such platforms by reducing their privacy risks as well as increasing the trustworthiness on the platform itself atthe same time.

Furthermore, with the disruptive growth of Artificial Intelligence (AI) algorithms, the identification of bots or non-reliable sources has become a crucial challenge to be investigated. It raised many studies and publications with the goal of building robust automatic systems to improve the quality of experience of consumers in such platforms by reducing their privacy risks as well as increasing the trustworthiness on the platform itself at the same time.

Therefore, this paper aims to contribute to the state-of-theart in this field by proposing a novel method for automatically

(i) encoding an input user account as a low-dimensional feature vector independently of its language,

(ii) identifying the input encoding vector as a suspicious bot account with a certain probability throughout a Deep Neural Network (DNN) referred as Bot-DenseNet.

(iii) generating a low-dimensional embedding which represents the original input encoding vector of the user account and which can be used for any other purpose related to Information Retrieval (IR).

RELATED WORK

Recently, AI techniques including Deep Learning (DL) and Machine Learning (ML) methods have gained popularity and interest in many applied research and industry services related to social media analysis where, sentiment analysis and text classification have been the central focus of these investigations specially for searching engines or recommender systems.

Furthermore, the continuous growth of the social media platforms such as Twitter or Facebook in the last decade along with the considerable propagation of non-trusted information throughout them, have raised applied research to automatically identify these non-trusted sources which in many cases correspond to non-human or Bot accounts. Most of the aforementioned approaches were limited due to lack of large volumes of annotated data for this specific task by the time their experiments were conducted. Additionally, although many approaches employ both metadata and text-based features from the user accounts, the text-based features are either extracted at a lexical level or they only cover a limited number of languages such as Spanish or English.

Unlike previous studies, our proposed model encodes all the text-based features of an input user account via novel multilingual Language Models (LM) including transformer models such as the so-called BERT or Contextual string embeddings. Thus, by concatenating both the metadata set of features along with the output vector provided by these LMs, an input vector of the user account is obtained. Finally, this paper proposes a Dense-based DL model to produce both the final decision of the account and a low-dimensional embedding of the user based on the aforementioned input vector.

A. DATASET GENERATION

There are several public datasets to address the bot identification problem from a binary classification perspective. Moreover, some of these datasets were already used to train and evaluate the so-called Botometer (formerly BotOrNot) service. However, the generation of bot accounts continuously changes over time and additionally, some of the provided accounts have been already suspended by Twitter. Thus, a preprocessing is needed to improve the usability of this large collection of datasets by removing the identifiers those accounts that were already removed by Twitter. This aspect is also critical since several previous Bot detectors have not been updated with the new tendencies and features that bots accounts may currently have, so that, they are no longer as reliable as they used to be. On the other hand, due to policy restriction terms from Twitter, the datasets only contain the identifier of the Twitter account but not any significant feature. Consequently, an additional data crawling process via the Twitter API is performed to gather further information about the available accounts which is needed for the analysis. the following information is collected:

• Popularity features including total number of both friends and followers,

• Activity features including the following fields: creation date, average tweets per day, tweets & favourites counts, account age.

• Profile information features including: screen name, description, language, location, verified indicator, default profile indicator

B. INPUT USER ENCODING GENERATION

The crucial part of this first stage lies in the generation of an user encoding vector based on the aforementioned collection of features to serve as input of the proposed deep learning model. However, our proposal addresses the aforementioned constraints by combining relevant metadata features along with powerful models capable of transforming text-based features into vectors independently of the language of the input text. More specifically, given an input set of Users U = {u1, u2, . . . , um}, a certain user account is represented as ui = [u t i , u z i] ∀i = 1, . . . , m where u t i indicates its text-based feature vector whereas u z i represents the remaining metadata-based vector. Our proposed solution employs a mapping function f (u) to generate a new set of Users U˙ = {u˙1, u˙2, . . . , u˙m}, where u˙i = f (ui) = g(u t i) | | h(u z i). In this case, we denote to indicate a concatenation operation between this pair of vectors. The reason behind using a concatenation layer at the end of this process lies in the fact that the system only consider information coming from the same target object: the user account. Other alternatives widely used in Collaborative Recommender Systems such as computing the outer product were not consider for this approach since in those scenarios, the information from the embeddings comes from two different sources: Users and Items, and the goal of the outer product is thus, to catch similarities and discrepancies between this two sets.

C. ENCODING TEXT-BASED FEATURES

Regarding the generation of the text-based vector from an input user account ui , a certain function g(u) is needed. Different state-of-the-art sentence-level encoders from several NLP frameworks were explored and investigated. In particular, the so-called Flair framework was employed to combine state-of-the-art Word Embeddings (WE) and Transformers for extracting robust document embeddings from the text-based features. More precisely, the following main families of embeddings have been employed in this study:

(i) Contextual string embeddings which are trained without any explicit notion of words and therefore, words are modelled as sequences of characters. Moreover, words are contextualized by the surrounding text. The employed model was trained using the so-called JW300 Dataset described in . In this study, both multi-forward and multi-backward embeddings are used. The dimension of their outputs is equal to 2048.

(ii) BERT (Bidirectional Encoder Representations from Transformers) embeddings which were proposed and developed and are based on a bidirectional transformer architecture.

(iii) RoBERTa which is an adaptive version of the BERT embedding where the goal is to improve the performance in longer sequences, or when there are vast volumes of data as

suggests. In this case, we employed the so-called roberta-large-mnli pre-trained model.

METHODOLOGY

Bot Sentinel uses machine learning algorithms to study Twitter accounts tweet messages. It identifies bots and classifies them as either trustworthy or untrustworthy. It then stores those accounts in a database to track them daily.

1. Sevice Provider: In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train and Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Tweet account Predict Type, Find Tweet account Prediction Type Ratio, Download Tweet account Prediction Type, View Tweet account Prediction Type Ratio Results, View All Remote Users.

2. View Users: In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, user name, email, address.

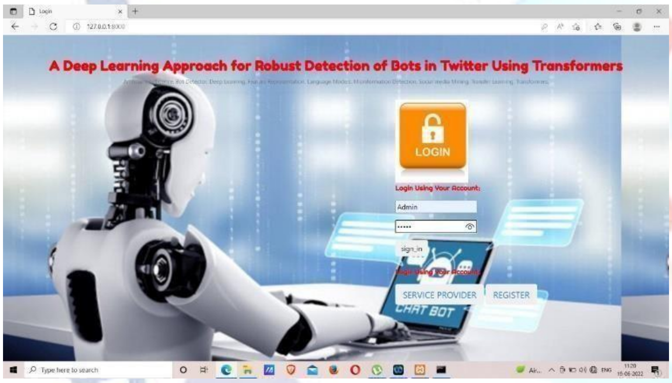
3. Remote Users: In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like, Predict Tweet account Type, View Your Profile.

RESULTS AND DISCUSSIONS

As a Result it provides two main contributions to the scientific community including a DL model for automatically detecting bots as well as a robust manner of representing any Twitter account as a low-dimensional feature vector throughout an intermediate layer of the aforementioned model. Moreover, this compact representation of the Twitter account can be used as a baseline for recommender or search engines, similarity analysis or any other application related with social media mining.

1. **Home page**

This is the home page which consists of login using your Accounts consists of two options namely, Service Provider and Register.



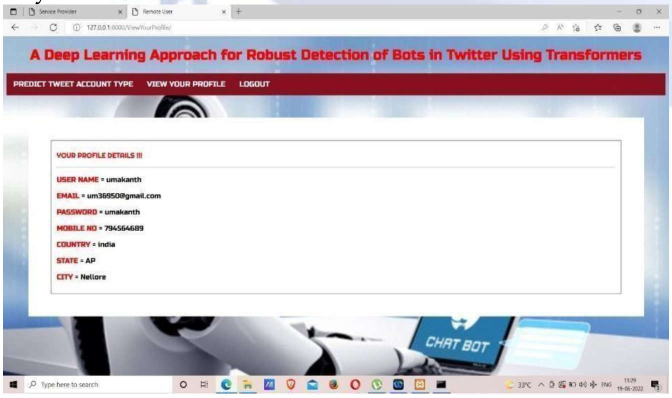
A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers

Login Using Your Account:

Admin

2. User Profile

In this page view the User Profile details that consists of User Name, Email, Password, Mobile No, Country, State, City.



A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers

YOUR PROFILE DETAILS !!!

USER NAME = umakanth

EMAIL = um36950@gmail.com

PASSWORD = umakanth

MOBILE NO = 794564689

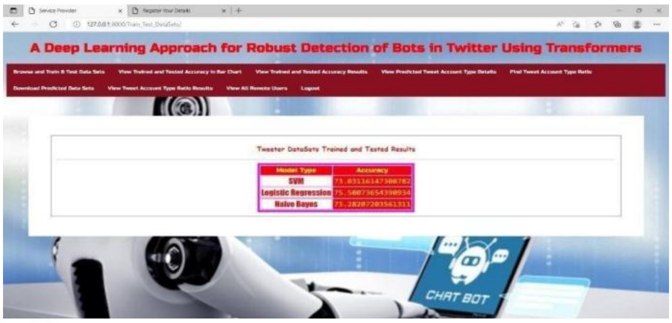
COUNTRY = india

STATE = AP

CITY = Nellore

3. Accuracy Scores

After user login with User Name and Password we access the accuracy scores of the twitter trained and tested results of Model Type and their Accuracy.



A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers

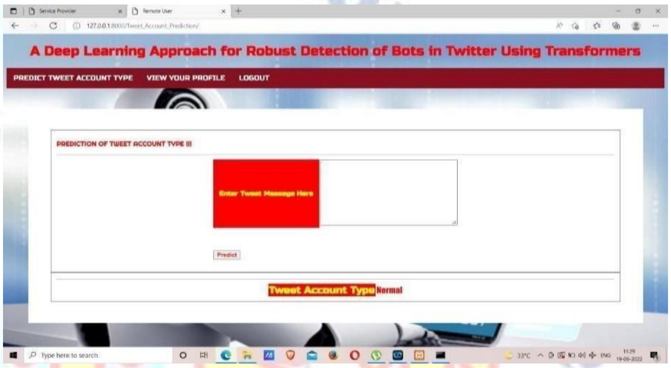
Tweeter DataSets Trained and Tested Results

Model Type	Accuracy
SVM	73.83116147308782
Logistic Regression	75.58073654390934
Naive Bayes	73.28287203561311

Service Provider
Register Your Details
127.0.0.1:8000/Train_Test_DataSets/
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Predicted Tweet Account Type Details
Find Tweet Account Type Ratio
Download Predicted Data Sets
View Tweet Account Type Ratio Results
View All Remote Users
Logout
CHAT BOT

4. Result As Normal

After enter the Tweet Message here it shows that the Result is Normal.



A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers

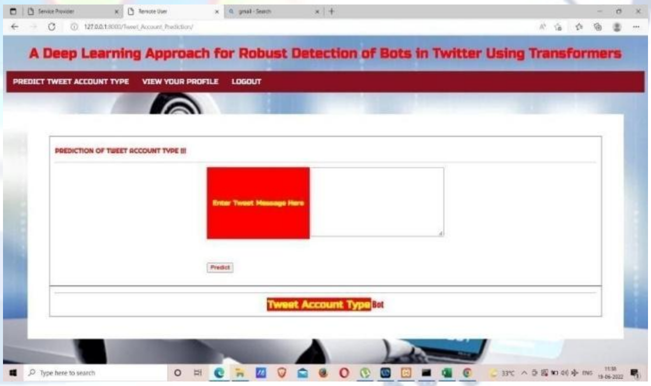
PREDICTION OF TWEET ACCOUNT TYPE !!!

Enter Tweet Message Here

Tweet Account Type Normal

5. Result as Bot

After enter the Tweet Message here it shows that the Result is Bot.



A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers

PREDICTION OF TWEET ACCOUNT TYPE !!!

Enter Tweet Message Here

Tweet Account Type Bot

FUTURE ENHANCEMENT

As Future work, the latest Transformers such as the GPT-3 andT5 will be considered for generating the input vector of the proposed DL model in order to compare the performance with the work described on this paper. Moreover novel approaches such the one described by authors into automatically generated non-parametric Two-Sample tests based on the so-called Maximum Mean Discrepancy (MMD), will be considered once all the user embeddings are generated, to find discrepancies and similarities between the distributions of both bots andnon-bots embeddings.

CONCLUSION

In this paper, a robust solution for detecting Bots in Twitter accounts has been described. In particular, this study has taken advantage of Transfer learning techniques via powerful state-of-the-art NLP models such as Transformers to extract compact multilingual representations of the text-based features associated with user accounts. By doing so, several constraints presented in previous studies related to process text-based features to improve the input feature vector from multiple languages were mitigated. Furthermore, by employing the text encodings along with additional metadata on top of a dense-based neural network ,a final classifier named as Bot-DenseNet has been trained and validated using a large set of samples collected via the Twitter API. More specifically, several experiments were conducted using different combinations of Word Embeddings, document embeddings (Pooling and LSTMs) and Transformers to obtain a single vector regarding the text-based features of the user account. Subsequently, a detailed comparison of the performance of the proposed classifier when using these approaches of Language Models as part of the input has been presented to investigate which input vector provides the best result in terms of performance simplicity in the generation of decision boundaries and feasibility.

ACKNOWLEDMENT

We express our gratitude to our seniors for providing mechanism and the base paper which are used to successfully complete an efficient " A Robust Detection of Bots in Twitter" and we also express our sincere gratitude to Narayana Engineering College for providing infrastructure and services which help us to complete the project.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Ž. Agić and I. Vulić, ''JW300: A wide-coverage parallel corpus for lowresource languages,'' in Proc. 57th Annu. Meeting Assoc. Comput. Linguistics. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 3204–3210.

[2] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and A. Vollgraf, ''FLAIR: An easy-to-use framework for state-of-the-art NLP,'' in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Demonstrations, 2019, pp. 54–59.

[3] A. Akbik, D. Blythe, and R. Vollgraf, ''Contextual string embeddings for sequence labeling,'' in Proc. 27th Int. Conf. Comput. Linguistics, 2018, pp. 1638–1649.

[4] A. S. M. Alharbi and E. de Doncker, ''Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information,'' Cogn. Syst. Res., vol. 54, pp. 50–61, May 2019.

[5] N. R. Aljohani, A. Fayoumi, and S.-U. Hassan, ''Bot prediction on social networks of Twitter in altmetrics using deep graph convolutional networks,'' Soft Comput., vol. 24, pp. 11109–11120, Jan. 2020.

[6] M. Arora and V. Kansal, ''Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis,'' Social Netw. Anal. Mining, vol. 9, no. 1, p. 12, Dec. 2019.

[7] A. Balestrucci, R. De Nicola, O. Inverso, and C. Trubiani, ''Identification of credulous users on Twitter,'' in Proc. 34th ACM/SIGAPP Symp. Appl. Comput., Apr. 2019, pp. 2096–2103.

[8] A. Bhoi and S. Joshi, ''Various approaches to aspect-based sentiment analysis,'' 2018, arXiv:1805.01984. [Online]. Available: http://arxiv. org/abs/1805.01984

[9] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, ''Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?'' IEEE Trans. Depend. Sec. Comput., vol. 9, no. 6, pp. 811–824, Nov./Dec. 2012.

[10] T. Cooijmans, N. Ballas, C. Laurent, Ç. Gülçehre, and A. Courville, ''Recurrent batch normalization,'' 2016, arXiv:1603.09025.[Online].Available: http://arxiv.org/abs/1603.09025

[11] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, ''The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race,'' in Proc. 26th Int. Conf. World Wide Web Companion, 2017, pp. 963–972..

[12] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, ''BotOrNot: A system to evaluate social bots,'' in Proc. 25th Int. Conf. Companion World Wide, 2016, pp. 273–274.

[13] A. Davoudi, A. Z. Klein, A. Sarker, and G. Gonzalez-Hernandez, ''Towards automatic bot detection in twitter for health-related tasks,'' AMIA Summits Transl. Sci. Proc., vol. 2020, p. 136, May 2020.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ''BERT: Pre-training of deep bidirectional transformers for language understanding,'' 2018, arXiv:1810.04805. [Online]. Available: http://arxiv.org/abs/1810.04805

[15] J. Diesner, E. Ferrari, and G. Xu, in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining. Sydney, NSW, Australia: ACM, Aug. 2017. [Online]. Available: https://dblp.org/rec/bib/conf/asunam/2017,doi: 10.1145/3110025.

[16] C. D. Santos and M. Gatti, ''Deep convolutional neural networks for sentiment analysis of short texts,'' in Proc. 25th Int. Conf. Comput. Linguistics (COLING), 2014, pp. 69–78. [17] L. Floridi and M. Chiriatti, ''GPT-3: Its nature, scope, limits, and consequences,'' Minds Mach., vol. 30, pp. 681–694, Nov. 2020.