Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 10062 - 10068
Received 25 April 2021; Accepted 08 May 2021.

Detecting Spam Bots on Social Networks using Supervised Learning

Gnanasekar A[1*], Srividhya Lakshmi R[2] Afraah Mariam[3], Deepika K[4], DhivyaShree[5]

[1]Associate Professor, Department of Computer Science and Engineering, R.M.D. Engineering College, India
[2]Assistant Professor, Department of Computer Science and Engineering, RMK College of Engineering and Technology, India
[3, 4, 5] Student, Department of Computer Science and Engineering, R.M.D. Engineering College, India
*ags.cse@rmd.ac.in,srividhya.lakshmi89@gmail.comucs17103@rmd.ac.in,ucs17125@rmd.ac.in,
ucs17131@rmd.ac.in

ABSTRACT

With the advancement of the Internet, social bots are progressively spreading on friendly stages. Hence, a successful discovery calculation is requested to recognize these social bot accounts that jeopardize informal communities. In this paper, a social bots detection model dependent on deep learning algorithm is proposed. The model primarily incorporates three layers. The primary layer is the joint substance highlight extraction layer, which centers around the element extraction of the tweets content and the connection between them. The subsequent layer is the tweet metadata fleeting component extraction layer, which views the tweet metadata as worldly data and utilizations this transient data as the contribution of the LSTM to extricate the client social action transient element. The third layer is the component intertwining layer, which combines the removed joint substance highlights with the worldly highlights to identify social bots. To assess the viability of the social bots detection model (DMbSLM), we led probes three unique sorts of new friendly bot informational indexes from this present reality and the investigation results likewise exhibit the adequacy of our proposed model.

Keywords: Social bots, supervised learning, combined content aspects, user behavior

Introduction

The expansiveness and convenience of online informal organizations have made them ideal climate for the multiplication of bogus and malignant bots [3]. Social bots (otherwise called spam bots) are robots who exist on friendly stages that consequently create content, associate with others, endeavor to impersonate and change the conduct of others [2].The bots in informal communities are practically unique. Some of them are basic in plan that can just send tweets, however a few bots are intricate which can even develop into more perplexing adaptations to avoid location approaches [4]. To better recognition the noxious conduct of progressively complex social bots, this paper proposes social bots location model which utilizes CNN-LSTM calculation to distinguish social bots. Initially, social bots recognition model uses CNN to extricate the joint highlights of the tweet content and the connection between them. Also, it utilizes LSTM to remove the likely worldly highlights of the tweet metadata. At long last, the fleeting highlights are combined with the joint substance highlights to accomplish the reason for distinguishing social bots. To show the adequacy of the model (DMbSLM), we directed examinations on this present reality new friendly bot dataset. And all accomplished almost wonderful distinguishing precision is over 98%.The research work carried out in this paper is presented in four headings and its organization is as follows: In heading 2, discusses about the literature review in this area. Heading 3, discusses about the algorithm model. Heading 4, discusses about the experimental result and evaluation finally conclusion and feature works.

Literature Review

Yang et al. [8] first conducted a comprehensive empirical analysis of the evasive strategies used by Twitter spammers. Then, multiple detection features were further designed to detect more Twitter bots and analyze the robustness of the proposed detection features. Yang et al. [8] only

Received 25 April 2021; Accepted 08 May 2021.

considered the tweet text aspects, but Miller et al. [6] viewed social bot detection as an anomaly detection problem rather than a classification problem. The tweet text aspect is fused with the user information feature, then use this fusing feature as the input of the stream clustering algorithm StreamKM++ and DenStream to achieve the purpose of spam identification. Davis et al. [10] extracted more features to detect social bots. These features include six main aspects: network features, user features, friend features, temporal features, content features and sentiment features, then using clustering algorithms to identify several social bot subsets. While Twitter bot detection is a specific use case on a particular social media, Cresci's[9] approach is platform technology-independent online user behavior modeling: the digital DNA sequences are extracted and analyzed from the user's online actions, by comparing the digital DNA sequences to each other, and then the accounts sharing the suspect long DNA substrings are labeled as spam bots. With the deepening of deep learning algorithms in recent years, Cai et al. [7] proposed a deep bots detection model to detect social bots from another angle. They use user tweets as temporal text data rather than plain text information, and learn the potential temporal patterns of tweet information through CNN-LSTM. Then the social bots is detected by jointly modeling social behavior and content information

Algorithm model

As shown in Figure. 1., DMbSLM generally consist of three layers: combined content aspects extraction layer, temporal aspects extraction layer and fusing aspects layer. The combined content aspects extraction layer mostly extracts the content and the relationship aspects between tweets. The temporal aspects extraction layer mainly extracts the tweet metadata temporal aspects. The specific details of these two layers will be discussed in Sections B, C, and D. In the fusing aspects layer, we combine combined content aspects with the metadata temporal aspects among definite rules. Following the fusing aspects layer is a combined content aspects extraction layer, a temporal aspects extraction layer, and finally, the fusing aspects layer is output.

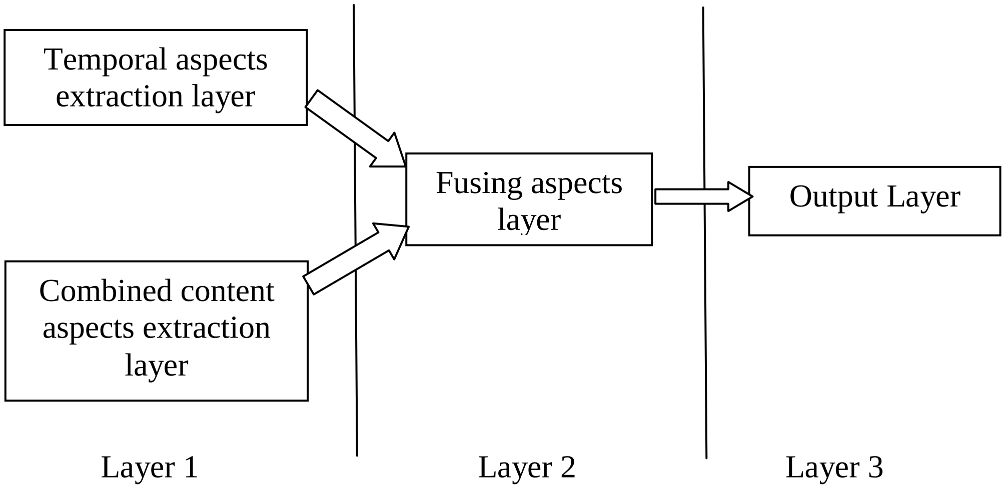
Temporal aspects extraction layer
Combined content aspects extraction layer
Fusing aspects layer
Output Layer
Layer 1
Layer 2
Layer 3

Figure 1. DMbSLM Structure

Experiment and Evaluation

A. Dataset

The dataset utilized in this testing was from [4]. We principally utilize genuine records of this group of informational collections datasets and three social bots record informational collections datasets, these three gatherings of records address the essential realities about the most recent social bots [4]. The three social bot records and one genuine user record are separately partitioned into three test informational collections data sets which individually named test1, test2, and test3. The quantity of certified records and bot represents each test informational data set is appeared in Table 1. To assess the adequacy of the most recent social bot location calculations, Cresci et al. [4], which gives informational collections, executed the most recent clump of social bot recognition calculations on this dataset. Trials have shown that these bot identification data sets don't have a well location impact on the current new friendly bots. To confirm the adequacy of social bots detection model, we played out the investigation of social bot location on the joint substance highlights of tweet, the highlights of tweet metadata and their combining.

Table 1. Comparision of test data set data

Data Set	Record Type	Record Address
Test 1	Genuine Records	1030
	Bot Records	950
Test 2	Genuine Records	456
	Bot Records	460
Test 3	Genuine Records	1046
	Bot Records	1095

B. Social Bots Detection Based on Combined Content Aspects

In this segment, we originally cleaned the informational data set. The examination tracked down that in every informational data set, tweets containing English clients, yet additionally tweets from other language clients. Our proposed social bot detection algorithm predominantly centers on social bots that utilization English to convey, so we eliminated the non-English clients and their tweets to stay away from the effect of the exactness of the model. To keep all client tweets of similar length to make following convolution tasks more helpful, we fill the cleaned social client tweets to a similar length. At that point, the filled client tweet is linked, and the tweet is changed over into a word inserting as a CNN contribution for the connected tweet utilizing word2vec. Since the principle highlight of the client tweet content combined aspects isn't combined with the worldly highlights, the combined aspects of removing the client tweet content through CNN are straightforwardly utilized as the contribution of the full association layer. At long last, the client personality (bot or not) is yield through the soft max layer. In this segment, we select the precision, recall, and F1 score to assess the impact of DMbSLM. The test results have appeared in Table 2. It tends to be seen from the test results that the tweet content is the principle highlight of social client representation. This shows that the proposed DMbSLM is a powerful strategy for distinguishing social bots.

Table 2. Detection effect based on Tweet combined content aspects model

Data set	Precision	Recall	F1 Score
Test 1	0.986	0.987	0.986
Test 2	0.964	0.944	0.953
Test 3	0.974	0.970	0.972

C. Social Bots Detection Based on Tweet Metadata

In this work, we innovatively utilize social client tweet metadata as transient data. Utilizing LSTM as a preparation model for these transient data, the transient data for a while is a contribution to the LSTM, and the output of the LSTM is extricated as the contribution of the completely connected layer. Like the Combined Content Aspects highlight detection part, the full connection layer is trailed by the soft max layer to straightforwardly yield whether the client is an authentic client or bot. The fundamental test boundaries in this part are the quantity of LSTM neurons and the quantity of LSTM layers. To get the impact of the number of various neurons and the quantity of LSTM layers on the last final results, we additionally select the precision, recall, and F1 score as the assessment measurements of the model in this part. At last, the ideal exploratory outcomes got on the three informational data sets test1, test2 and test3 appear in Table 3. It very well may be unmistakably found in Table 3 that the precision of the model is about 98% when the quantity of neurons is 128 or 256 and the LSTM is 1 or 3 layers. This shows that the proposed Combined Content Aspects is an effective method for recognizing bots.

Table 3. Detection effect based on Tweet temporal features

Data set	LSTM layer number	Number of neurons	Precision	Recall	F1 Score
Test 1	3	128	0.985	0.976	0.980
Test 2	1	256	0.978	0.988	0.984
Test 3	3	128	0.964	0.801	0.874

D. Social Bots Detection Based on the Fusing of Tweet Combined Content Aspects and Temporal Aspects

The test informational data set in this segment chooses test1, test2. To be steady with the test experimental measurements of Cresci et al. [9], we additionally chose the precision, recall, F1 score, explicitness, and MCC as the assessment measurements of a social bots detection model. To enhance the present status of the workmanship exploration and social bots detection tests, we recreated the experiment of utilizing deep learning out how to detect social bots in [10] in this informational collection data set. The boundaries chose in the test are reliable with the original text, yet they chose assessment measurements that are equivalent to the above assessment measurements. At long last, the consequences of all near tests have separately appeared in Table 4 and Table 5. The experimental results show that a social bots detection model has preferable test results over different models in precision, recall, and MCC show in figure 2 and figure 3.

Table 4. DMbSLM Detection effect on test 1

Technique	Type	Detection Results					
		Precision	**Recall**	**Specificity**	**Accuracy**	**F-Measure**	**MCC**
Ahmed et al.[5]	Unsupervised	0.944	0.943	0.944	0.942	0.943	0.885
Miller et al. [6]	Unsupervised	0.554	0.357	0.699	0.525	0.434	0.058
Cai et al[7]	Supervised	0.925	0.967	0.908	0.940	0.945	0.881
C. Yang et al[8]	Supervised	0.562	0.169	0.859	0.505	0.260	0.042
Cresci et al.[9]	Unsupervised	0.981	0.971	0.980	0.975	0.976	0.951
BotOrNot[10]	Supervised	0.470	0.207	0.917	0.733	0.287	0.173
DMbSLM	Supervised	0.998	1.000	0.999	1.000	0.999	0.997

Table 5. DMbSLM Detection effect on test 2

Technique	Type	Detection Results					
		Precision	**Recall**	**Specificity**	**Accuracy**	**F-Measure**	**MCC**
Ahmed et al.[5]	Unsupervised	0.912	0.934	0.911	0.922	0.922	0.846
Miller et al. [6]	Unsupervised	0.466	0.305	0.653	0.480	0.369	0.042
Cai et al[7]	Supervised	0.914	0.986	0.895	0.944	0.949	0.893
C. Yang et al[8]	Supervised	0.726	0.408	0.847	0.628	0.523	0.286
Cresci et al.[9]	Unsupervised	0.999	0.857	0.999	0.928	0.922	0.866
BotOrNot[10]	Supervised	0.634	0.949	0.980	0.921	0.760	0.737
DMbSLM	Supervised	0.998	1.000	0.998	1.000	0.999	0.997

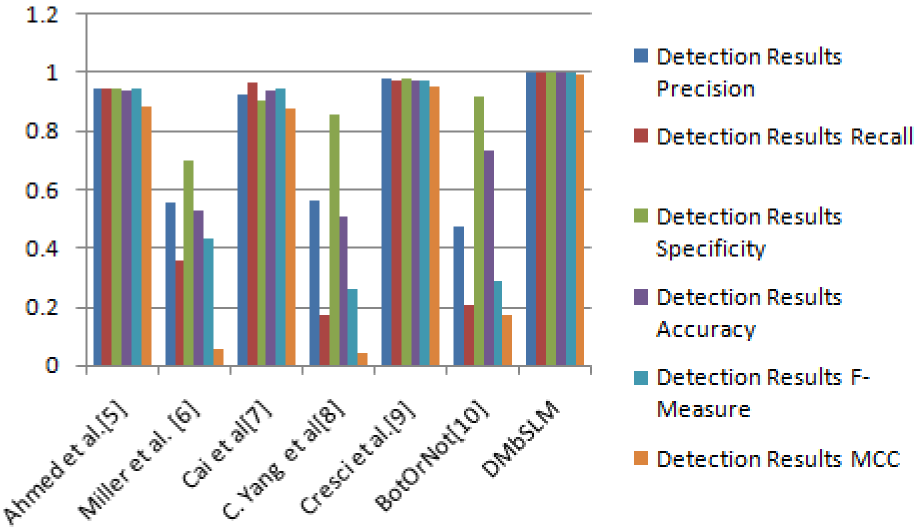

Figure 2. DMbSLM Detection effect on test 1

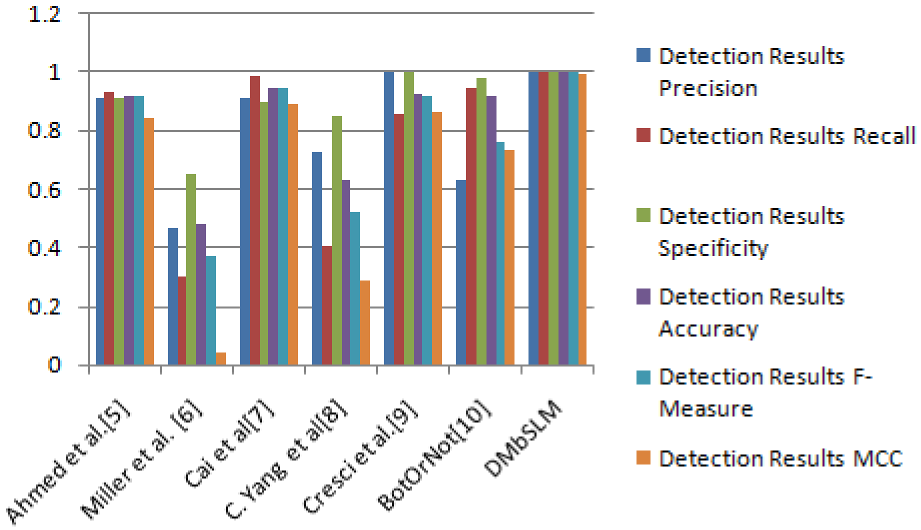

Figure 3. DMbSLM Detection effect on test 2

Conclusion and Future Work

The social bots detection method consists of three parts: social bot detection based on tweet joint features, social bot detection based on tweet metadata temporal aspects and features aspects In the first part, the user tweet is transformed into a word embedding and concatenates them. Then CNN is used to extract the aspects of the tweet content and the relationship between them. In the second part, we treat the metadata of the user tweets as temporal information represented by social users rather than purely digital features. Counting the user's temporal information for a period of time and using it as an input to the LSTM neural network. The experimental results of these two parts show that the proposed social bots detection model is effective for detecting bots. CNN and LSTM focus on different aspects of tweet features and fuse the extracted features of CNN and LSTM. Although the social bot detection based on deep learning achieves nearly perfect accuracy on different data sets, it requires a large amount of tweet information from the user. In future work, we can use the user tweets and information to detect social bots as little as possible while ensuring a high detection rate.

References

[1] Laszlo A and Castro K. “Technology and values: Interactive learning environments for future generations. Educational Technology,” vol.35, pp.7-13, 2013.

[2] Blunkett, Jiang D M, Cui P, and Faloutsos C, “Suspicious Behavior Detection:Current Trends and Future Directions,” IEEE Intelligent Systems, vol. 31, pp. 31-39, 2016.

[3] Zhou Y et al., , “ProGuard: Detecting malicious accounts in socialnetwork- based online promotions,” IEEE Access, vol. 5, pp. 1990_1999, 2017.

[4] Cresci S, Petrocchi M, et al. “The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race,” Proceedings of the 26th International Conference on World

Wide Web Companion. International World Wide Web Conferences Steering Committee, pp. 963-972, 2017.

[5] Ahmed F and Abulaish M, “A generic statistical approach for spam detection in Online Social Networks,” Computer Communications, vol. 36, pp. 1120-1129, 2013.

[6] Miller Z, Deitrick W, et al. “Twitter spammer detection using data stream clustering,” Information Sciences, vol. 260, pp. 64-73,2014.

[7] Cai C, Li L and Zeng D, “Detecting Social Bots by Jointly Modeling Deep Behavior and Content Information,” ACM, pp. 1995-1998, 2017.

[8] Yang C and Gu G, “Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers”. IEEE Transactions on Information Forensics & Security, vol. 8, pp. 1280-1293, 2013.

[9] Cresci S, Pietro R D, et al. “DNA-Inspired Online Behavioral Modeling and Its Application to Spambot Detection,” IEEE Intelligent Systems, vol. 31, pp. 58-64, 2017.

[10] Davis C A, Varol O, et al. “BotOrNot: A System to Evaluate Social Bots,” pp. 273-274, 2016.